STAT/MATH 521, 522, 523

Advanced Probability

Instructor 2012-2013: Jon Wellner

Text

Probability for Statisticians

by

Galen R. Shorack

Department of Statistics
University of Washington
Seattle, WA 98195

This dedication to my parents is offered as a permanent gift to my family,
so that my parents' basic history may not be lost.

To my Father—who loved me

- Theodore James Shorack (August 20, 1904–July 31, 1983) Charleston, WV
- With only a third-grade education, he taught me that mathematics is fun.
- Effie, Minnesota; the Aleutian Islands; Eugene, Oregon. Homesteader, boxer, carpenter and contractor. He loved the mathematics of his carpenter's square.
- He loved his children with all his being.

To my Mother—who praised me

- Marcella (Blaha) Shorack (November 4, 1902–April 25, 1987) St. Paul, MN
- "What you don't have in your head, son, you'll have to have in your feet."
- Effie, Minnesota; Battle Creek, Michigan; Flagstaff, Arizona; Eugene, Oregon. Homesteader and dedicated teacher. She had a heart for her troubled students.
- She cared dearly about who and what her children would become.

Theodore J. Shorack and Marcella (Blaha) Shorack, wed 6/12/1929

We, their descendents, are entrusted with their memory.

1. Theodore James Shorack Jr. (1929–1966; Vietnam pilot, my boyhood hero)
and Elva (Buehler) Shorack (1925)
Candace (1953).
Helen (1989)–Vietnam
Kathleen (1955), and Walter Petty (1953)
Elizabeth (1987), Angela (1990)
Theodore III (or Todd) (1957), and Karie (Lott) Shorack (1960)
Theodore IV (1985), Wesley (1988)
John (1960), and Birgit (Funck) Shorack (1958)
Johanna (1990), Marna (1993), John Mark (1995)
2. Charlene (or Chari) Rose Boehnke (1931)
and George Boehnke (1931)
Michael (1956), and Betsy Foxman (1955)
David (1985), Kevin (1987), Richard (1989)
Richard (1958–1988)
Barbara (1961), and Terrance Aalund (1952)
Katherine (1989), Daniel (1991), Gary (1995)
3. Roger Allen Shorack (1934) daughter Stefani (1968)
and Heather (Cho) Shorack (1949)
4. Galen Richard Shorack (1939; the author)
and Marianne (Crabtree) Shorack (1938); Sandra Ney Wood (1943)
Galen (or GR) (1964), and Lanet (Benson) Shorack (1967)
Nikolai (1999), Luca (2002), Nadia (2005)
Bart (1966), and Kerri (Winkenweder) Shorack (1968)
Landon (1995), Kyle (1998), Titus (2001)
Matthew (1969), and Julie (Mitchell) Shorack (1969)
(Isaac (1999), Thomas (2000), Alexia (2001))–Uganda, Tessa (2005)

My grandparents' generation

- Peter Shorack and Anna Miliči (immigrants from Miliči, near Karlovatz)
  (They married in 1890, and came to the US in 1898.)
  Nicholas died by age 10
  Annie to Ivan Harrington (Archie (wed Ellen; 5) + 9)
  William to Kate (William Jr. (wed Virginia; 2))
  Amelia to Charlie Lord
  Theodore James Shorack (my father)
  Jenny to Godfrey Knight (5)
- Frank Blaha Jr. (Chicago 3/24/1893, cooper) and Marcella Nekola (immigrant from the Prague area as a child)
  Marcella Barbara Blaha (my mother)
  Marie (1904) to Carlos Halstead (Carlos, Gilbert, Christine)
  George (1906) (my father's dear friend) to Carmen Jirik
  Julia (1908) to Lyle Dinnell
  Nan (1911) (first child born on the Effie, Minnesota homestead)
  Rose (1913)
  Helen (1916)
  Carol (twin) (1918) to Ashley Morse (Leigh (wed Kent; 2), Laurel)
  Don (twin) (1918) to Jean Dora

Frank J. Blaha Sr. (1850, lumber mill and railroad) wed Rose Hřda (1852) 1/30/1872. They had immigrated separately to the US in their twenties; he from the Prague area. (Frank Jr. (my mother's father, homesteader), James (married Aunt Anna Nekola, printer), Joseph, Agnes.) Frank Jr. was an inept farmer, but he enjoyed his books and raised educated daughters.

Thomas Nekola (wagon maker in the Prague area) married Mary Tomăsek. (Barbara, Anna (married Uncle James), Marcella (my mother's mother), Albert, Pete, Frances)

Peter Shorack (an only child) seems to have been "on the move" when he arrived in Miliči, but his origin is unclear. Was he fleeing a purge in the east (he said) or Austro–Hungarian conscription (his wife said)? His parentage is unknown. He died from alcohol when my father was nine.

Anna Miliči was the fourth of five children of Maximo and Martha Miliči. Maximo (appropriately 6'9") came to the US, but fled home when two attackers did not survive. Later, his neighbors there banded together and killed him with pitchforks. Anna visited Miliči with her children when my father was five, but had to leave her children there for one year. She had hidden enough to get herself home, thus thwarting Peter's efforts to strand his family. A hard woman in most ways, she used her gun to run off robbers (when linguistic Peter was running a railroad gang) and poachers (after she was alone on the homestead).

My father worked in logging camps as a young kid; his mother and older brother would not allow him to go to school. He hopped a freight when he was bigger, but "hammer toes" allowed him to negate his mistake of an army enlistment. Back home, he trained religiously as a boxer and a fighter. He thus "survived" his older brother, boxed the county fairs with George, defended his interests in my mother, won two professional fights (but lost two teeth), fought (with some success) for his full winter logging earnings (each spring the same companies would go bankrupt, leaving their debts unpaid). WW II construction work on the Al-Can Highway and in the Aleutian Islands, gave him the nestegg to get us out of that country. With boxing and the gym for entertainment, he lived entirely off a whiskey allotment sold late in each month, every full paycheck came home—and he learned carpentry. On to Oregon! After ten years he was building his own houses on speculation, in spite of his financial raw fear common to so many of that depression generation. But that gave his sons jobs to go to college, and he sent his daughter. He took incredible pride in even the smallest of the accomplishments of any of us. Part of him died with my brother, flying cover on a pilot pickup. My mother provided the stability in our family, not an easy task. She provided the planning, tried to challenge us, watched for opportunities to expand our horizons. A shy woman, she defended the value of her son's life by following the anti-Vietnam circuit and challenging all speakers. Her's was the quiet consistency that I beter appreciated after having a family of my own.

Preface

Chapters 1–5 and Appendix B provide the mathematical foundation for the rest of the text. Then Chapters 6–7 hone some tools geared to probability theory. Appendix A provides a brief introduction to elementary probability theory, that could be useful for some mathematics students. (The appendices begin on page 425.)

The classical weak law of large numbers (WLLN) and strong law of large numbers (SLLN) as presented in Sections 8.2–8.4 are particularly complete, and they also emphasize the important role played by the behavior of the maximal summand. Presentation of good inequalities is emphasized in the entire text, and this chapter is a good example. Also, there is an (optional) extension of the WLLN in Appendix C that focuses on the behavior of the sample variance, even in very general situations. It will be appealed to in the optional Section 10.5 and Chapter 11.

The classical central limit theorem (CLT) and its Lindeberg and Liapunov and Berry–Esseen generalizations are presented in Chapter 10 using the characteristic function (chf) methods introduced in Chapter 9. Conditions for both the weak bootstrap and the strong bootstrap are also developed in Chapter 10, as is a universal bootstrap CLT based on light trimming of the sample. This approach emphasizes a statistical perspective. Gamma and Edgeworth approximations appear at the end of Chapter 11.

Both distribution functions (dfs $F(\cdot)$) and quantile functions (qfs $K(\cdot) \equiv F^{-1}(\cdot)$) are emphasized throughout (quantile functions are important to statisticians). In Chapter 6 much general information about both dfs and qfs and the Winsorized variance is developed. The text includes presentations showing how to exploit the inverse transformation $X \equiv K(\xi)$ with $\xi \cong \text{Uniform}(0,1)$. In particular, Appendix C inequalities relating the qf and the Winsorized variance to some empirical process results of Chapter 12 were used in the first edition to treat trimmed means and L-statistics, rank and permutation tests, sampling from finite populations.

I have learned much through my association with David Mason, and I would like to acknowledge that here. Especially (in the context of this text), Theorem 12.4.3 is a beautiful improvement on Theorem 12.10.3, in that it still has the potential for necessary and sufficient results. I really admire the work of Mason and his colleagues. It was while working with David that some of my present interests developed. In particular, a useful companion to Theorem 12.10.3 is knowledge of quantile functions. Section 7.6 and Sections C.2–C.X present what I have compiled and produced on that topic while working on various applications, partially with David.

Jon Wellner has taught from several versions of this text. In particular, he typed an earlier version and thus gave me a major critical boost. That head start is what turned my thoughts to writing a text for publication. Sections 14.2, and the Hoffman–Jorgensen inequalities came from him. He has also formulated a number of exercises, suggested various improvements, offered good suggestions and references regarding predictable processes, and pointed out some difficulties. My thanks to Jon for all of these contributions. (Obviously, whatever problems may remain lie with me.)

My thanks go to John Kimmel for his interest in this text, and for his help and guidance through the various steps and decisions. Thanks also to Lesley Poliner, David Kramer, and the rest at Springer-Verlag. It was a very pleasant experience.

This is intended as a textbook, not as a research manuscript. Accordingly, the main body is lightly referenced. There is a section at the end that contains some discussion of the literature.

Galen R. Shorack
Seattle, Washington
April 14, 2000

Contents

Chapter 6. Distribution and Quantile Functions

Chapter 7. Independence and Conditional Distributions

Chapter 8. WLLN, SLLN, LIL, and Series

Chapter 9. Characteristic Functions

Chapter 10. CLTs via Characteristic Functions

Chapter 11. Infinitely Divisible and Stable Distributions

Chapter 12. Brownian Motion and Empirical Processes

Chapter 13. Martingales

Use of This Text

The University of Washington is on the quarter system, so my description will reflect this fact. My thoughts are offered as a potential guide to an instructor. They certainly do not comprise an essential recipe.

The reader will note that the exercises are interspersed with the text. It is important to read all of the exercises as they are encountered, as most of them contain some worthwhile contribution to the story.

Chapters 1–5 provide the measure-theoretic background that is necessary for the rest of the text. Many of our students have had at least some kind of an undergraduate exposure to part of this subject. Still, it is important that I present the key parts of this material rather carefully. I feel it is useful for all of them.

Chapter 1 (measures; 5 lectures)
Emphasized in my presentation are generators, the monotone property of measures, the Carathéodory extension theorem, completions, the approximation lemma, and the correspondence theorem. Presenting the correspondence theorem carefully is important, as this allows one the luxury of merely highlighting some proofs in Chapter 5. [The minimal monotone class theorem of Section 1.1, claim 8 of the Carathéodory extension theorem proof, and most of what follows the approximation lemma in Section 1.2 would never be presented in my lectures.] {I always assign Exercises 1.1.1 (generators), 1.2.1 (completions), and 1.2.3 (the approximation lemma). Other exercises are assigned, but they vary each time.}

Chapter 2 (measurable functions and convergence; 4 lectures)
I present most of Sections 2.1, 2.2, and 2.3. Highlights are preservation of σ-fields, measurability of both common functions and limits of simple functions, induced measures, convergence and divergence sets (especially), and relating $\to_\mu$ to $\to_{a.s}$ (especially, reducing the first to the second by going to subsequences). I then assign Section 2.4 as outside reading and Section 2.5 for exploring. [I never lecture on either Section 2.4 or 2.5.] {I always assign Exercises 2.2.1 (specific σ-fields), 2.3.1 (concerning $\to_{a.e.}$), 2.3.3 (a substantial proof), and 2.4.1 (Slutsky's theorem).}

Chapter 3 (integration; 7 lectures)
This is an important chapter. I present all of Sections 3.1 and 3.2 carefully, but Section 3.3 is left as reading, and some of the Section 3.4 inequalities (C_r, Hölder, Liapunov, Markov, and Jensen) are done carefully. I do Section 3.5 carefully as far as Vitali's theorem, and then assign the rest as outside reading. {I always assign Exercises 3.2.1–3.2.2 (only the zero function), 3.3.3 (differentiating under the integral sign), 3.5.1 (substantial theory), and 3.5.7 (the Scheffé theorem).}

Chapter 4 (Radon–Nikodym; 2 lectures)
I present ideas from Section 4.1, sketch the Jordan–Hahn decomposition proof, and then give the proofs of the Lebesgue decomposition, the Radon–Nikodym theorem, and the change of variable theorem. These final two topics are highlighted. The fundamental theorem of calculus of Section 4.4 is briefly discussed. [I would never present any of Section 4.3.] {I always assign Exercises 4.2.1 (manipulating Radon–Nikodym derivatives), 4.2.7 (mathematically substantial), and 4.4.1, 4.4.2, and 4.4.4 (so that the students must do some outside reading in Section 4.4 on their own).}

Chapter 5 (Fubini; 2 lectures)
The first lecture covers Sections 5.1 and 5.2. Proving Proposition 5.2.1 is a must, and I discuss/prove Theorems 5.1.2 (product measure) and 5.1.3 (Fubini). The remaining time is spent on Section 5.3. [I rarely lecture from Section 5.4, but I do assign it as outside reading.] {I always assign Exercises 5.3.1 (measurability in a countable number of dimensions) and 5.4.1 (the finite-dimensional field).}

Appendix B (topology and Hilbert space, 0 lectures)
This appendix is presented only for reference. I do not lecture from it.

The mathematical tools have now been developed. In the next three chapters we learn about some specialized probabilistic tools and then get a brief review of elementary probability. The presentation on the classic topics of probability theory then commences in Chapter 8.

Chapter 6 and Appendix B (distribution functions (dfs) and quantile functions (qfs); 4 lectures) This chapter is quite important to this text. Skorokhod's theorem in Section 6.3 must be done carefully, and the rest of Sections 6.1–6.4 should be covered. Section 6.5 should be left as outside reading. [Lecturing from Sections 6.6 and Sections B.2–B.X is purely optional, and I would not exceed one lecture.] {I always assign Exercises 6.1.1 (on continuity of dfs), 6.3.3 ($F^{-1}(\cdot)$ is left continuous), 6.3.3 (change of variable), and 6.4.2 (for practice working with $X \equiv K(\xi)$). Consider lecturing on Theorem 6.6.1 (the infinite variance case).}

Chapter 7 (conditional expectation; 2 lectures)
Lecture one on Sections 7.1– 7.2 highlights Proposition 7.1.1 (on preserving independence), Theorem 7.1.2 (on extending independence from π-systems), and Kolmogorov's 0-1 law. The other lecture provides some discussion of the definition of conditional probability in Section 7.4, includes proofs of several parts of Theorem 7.4.1 (properties of conditional expectation), and discusses Definition 7.5.1 of regular conditional probability. [I never lecture on Sections 7.3, or 7.5.] {I always assign Exercises 7.1.2 and 7.1.3 (they provide routine practice with the concepts), Exercise 7.4.1 (discrete conditional probability), Exercise 7.4.3 (repeated stepwise smoothing in a particular example), and part of Exercise 7.4.4 (proving additional parts of Theorem 7.4.1).}

Appendix A (elementary probability; 0 lectures)
Sections A.1 and A.2 were written to provide background reading for those graduate students in mathematics who lack an elementary probability background. Sections A.3 and A.4 allow graduate students in statistics to read some of the basic multivariate results in appropriate matrix notation. [I do not lecture from this material.] {But I do assign Exercises A.1.8 (the Poisson process exists) and A.2.1(ii) (so that the convolution formula is refreshed).}

Chapter 8 (laws of large numbers (LLNs) and inequalities; 3 lectures for now)
Since we are on the quarter system at the University of Washington, this leaves me 3 lectures to spend on the law of large numbers in Chapter 8 before the Christmas break at the end of the autumn quarter. In the first 3 lectures I do Sections 8.1 and 8.2 with Khinchine's weak law of large numbers (WLLN), Kolmogorov's inequality only from Section 8.3, and at this time I present Kolmogorov's strong law of large numbers (SLLN) only from Section 8.4. {I always assign Exercises 8.1.1 (Cesàro summability), 8.2.1 (it generates good ideas related to the proofs), 8.2.3 (as it

practices the important $O_p(\cdot)$ and $o_p(\cdot)$ notation), 8.4.4 (the substantial result of Marcinkiewicz and Zygmund), 8.4.7 (random sample size), and at least one of the alternative SLLN proofs contained in 8.4.8, 8.4.9, and 8.4.10.}

At this point at the beginning of the winter quarter the instructor will have his/her own opinions about what to cover. I devote the winter quarter to the weak law of large numbers (WLLN), an introduction to the law of the iterated logarithm (LIL), and various central limit theorems (CLTs). That is, the second term treats material from Chapters 8-10, with others optional. I will now outline my choices.

Chapter 8 (LLNs, inequalities, LIL, and series; 6 lectures)
My lectures cover Section 8.3 (symmetrization inequalities and Lévy's inequality for the WLLN, and the Ottovani–Skorokhod inequality for series), Feller's WLLN from Section 8.4, the Glivenko–Cantelli theorem from Section 8.5, the LIL for normal rvs in Proposition 8.6.1, the strong Markov property of Theorem 8.7.1, and the two series Theorem 8.8.2. [I do not lecture from any of Sections 8.9, 9.10, or 8.11 at this time.] {I always assign Exercise 8.6.1 (Mills' ratio).}

Chapter 9 (characteristic functions (chfs); 8 lectures) Sections 9.1 and 9.2 contain classic results that relate to deriving convergence in distribution from convergence of various classes of integrals. I also cover sections 9.3–9.8. {I always assign Exercises 9.3.1 and 9.3.3(a) (deriving specific chfs) and 9.6.1 (Taylor series expansions of the chf).}

Chapter 10 (CLTs via chfs; 6 lectures)
The classical CLT, the Poisson limit theorem, and the multivariate CLT make a nice lecture. The chisquare goodness of fit example and/or the median example (of Section 10.3) make a lecture of illustrations. Chf proofs of the usual CLTs are given in Section 10.2 (Section 9.5 on Esseen's lemma could have been left until now). Other examples from Section 10.2 or 10.3 could now be chosen, and Example 10.3.4 (weighted sums of iid rvs) is my first choice. [The chisquare goodness of fit example could motivate a student to read from Sections A.3 and A.4.]

At this stage I still have at least 7 optional lectures at the end of the winter quarter and about 12 more at the start of the spring quarter. In my final 16 lectures of the spring quarter I feel it appropriate to consider Brownian motion in Chapter 12 and then martingales in Chapter 13 (in a fashion to be described below). Let me first describe some possibilities for the optional lectures, assuming that the above core was covered.

Chapter 10 (bootstrap)
Both Sections 10.8 and 10.9 on the bootstrap require only a discussion of section 10.**??**.

Chapter 19 (convergence in distribution)
Convergence in distribution on the line is presented in Chapter 10. [This is extended to metric spaces in Chapter 14, but I do not lecture from it.]

Chapter 10 (domain of normal attraction of the normal df)
The converse of the CLT in Theorem 10.6.1 requires the Giné–Zinn symmetrization inequality and the Khinchine inequality of Section 8.3 and the Paley–Zygmund inequality of Section 3.4.

Chapters 7, 10 and 11 (domain of attraction of the Normal df)
Combining Sections 6.6, C.1-C.4, Section 8.3 subsection on maximal inequalities of another ilk, and Sections 10.5–10.6 makes a nice unit. Lévy's asymptotic normality

condition (ANC) of (10.6.3) for a rv X has some prominence. In Section B.2 purely geometric methods plus Cauchy–Schwarz are used to derive a multitude of equivalent conditions. In the process, quantile functions are carefully studied. In Section 10.1 the ANC is seen to be equivalent to a result akin to a WLLN for the rv X^2, and so in this context many additional equivalent conditions are again derived. Thus when one comes to the general CLT in Sections 10.5 and 10.6, one already knows a great deal about the ANC.

Chapter 11 (infinitely divisible and stable laws)
First, Section 11.1 (infinitely divisible laws) is independent of the rest, including Section 11.2 (stable laws). The theorem stated in Section 11.4 (domain of attraction of stable laws) would require methods of Section B.4 to prove, but the interesting exercises are accessible without this.

Chapter 11 (higher-order approximations)
The local limit theorem in Section 10.4 can be done immediately for continuous dfs, but it also requires Section 9.8 for discrete dfs. The expansions given in Sections 11.5 (Gamma approximation) and 11.6 (Edgeworth approximation) also require Exercise 9.6.7.

Assorted topics suitable for individual reading
Possibilities include Section 13.8 (counting process martingales), and Section 13.9 (martingale CLTs). Section 15.1 on trimmed means and Section 15.2 on R-statistics (including a finite sampling CLT) are both fun; both require some discussion of Section C.6.

The primary topics for the spring quarter are Chapter 12 (Brownian motion and elementary empirical processes) and Chapter 13 (martingales).

Chapter 12 (Brownian motion; 6 lectures)
I discuss Section 12.1, sketch the proof of Section 12.2 and carefully apply that result in Section 12.3, and treat Section 12.4 carefully (as I believe that at some point a lecture should be devoted to a few of the more subtle difficulties regarding measurability). I am a bit cavalier regarding Section 12.5 (strong Markov property), but I apply it carefully in Sections 12.6, 12.7, and 12.8. I assign Section 12.9 as outside reading. [I do not lecture on Theorem 12.8.2.] {I always assign Exercises 12.1.2 (on $(C, \mathcal{C})$), 12.3.1 (various transforms of Brownian motion), 12.3.3 (integrals of normal processes), 12.4.1 (properties of stopping times), 12.7.3(a) (related to embedding a rv in Brownian motion), and 12.8.2 (the LIL via embedding).}

At this point let me describe three additional optional topics that could now be pursued, based on the previous lectures from Chapter 12.

Chapter 12 (elementary empirical processes)
Uniform empirical and quantile processes are considered in Section 12.10. Straightforward applications to linear rank statistics and two-sample test of fit are included. One could either lecture from Section 12.12 (directly) or 12.11 (with a preliminary lecture from Sections 10.10–10.11, or leave these for assigned reading.)

Chapter 11 (martingales; 10 lectures)
I cover most of the first seven sections. {I always assign Exercises 11.1.4 (a counting process martingale), 11.3.2 (a proof for continuous time mgs), 11.3.7, and 11.3.9 (on $\mathcal{L}_r$-convergence).}

Definition of Symbols

$\cong$ means "is distributed as"

$\equiv$ means "is defined to be"

$a = b \oplus c$ means that $|a - b| \leq c$

$U_n =_a V_n$ means "asymptotically equal" in the sense that $U_n - V_n \to_p 0$

$X \cong (\mu, \sigma^2)$ means that X has mean μ and variance σ^2

$X \cong F(\mu, \sigma^2)$ means that X has df F with mean μ and variance σ^2

$\bar{X}_n$ is the "sample mean" and $\ddot{X}_n$ is the "sample median"

$(\Omega, \mathcal{A}, \mu)$ and $(\Omega, \mathcal{A}, P)$ denote a measure space and a probability space

$\sigma[\mathcal{C}]$ denotes the σ-field generated by the class of sets $\mathcal{C}$

$\mathcal{F}(X)$ denotes $X^{-1}(\bar{\mathcal{B}})$, for the Borel sets $\mathcal{B}$ and $\bar{\mathcal{B}} \equiv \sigma[\mathcal{B}, \{+\infty\}, \{-\infty\}]$

ξ will always refer to a Uniform$(0, 1)$ rv

$\nearrow$ means "nondecreasing" and $\uparrow$ means "strictly increasing"

$1_A(\cdot)$ denotes the indicator function of the set A

"df" refers to a distribution function $F(\cdot)$

"qf" refers to the left continuous quantile function $K(\cdot) \equiv F^{-1}(\cdot)$

The "tilde" symbol denotes $\tilde{\text{W}}$insorization

The "háček" symbol denotes $\check{\text{T}}$runcation

$\lambda(\cdot)$ and $\lambda_n(\cdot)$ will refer to Lebesgue measure on the line R and on R_n

See page 115 for "dom(a, a')"

Brownian motion $\mathbb{S}$, Brownian bridge $\mathbb{U}$, and the Poisson process $\mathbb{N}$

The empirical df $\mathbb{F}_n$ and the empirical df $\mathbb{G}_n$ of Uniform$(0, 1)$ rvs

$\rightsquigarrow$ is associated with convergence in the LIL (see page 175)

"mg" refers to a martingale

"smg" refers to a submartingale

$\overset{\geq}{=}$ means "$\geq$" for a submartingale and "$=$" for a martingale

The symbol "$\overset{\geq}{=}$" is paired with "s-mg" in this context.

Prominence

Important equations are labeled with numbers to give them prominence. Thus, equations within a proof that are also important outside the context of that proof are numbered. Though the typical equation in a proof is unworthy of a number, it may be labeled with a letter to help with the "bookkeeping." Likewise, digressions or examples in the main body of the text may contain equations labeled with letters that decrease the prominence given to them.

Integral signs and summation signs in important equations (or sufficiently complicated equations) are large, while those in less important equations are small. It is a matter of assigned prominence. The most important theorems, definitions, and examples have been given titles in boldface type to assign prominence to them. The titles of somewhat less important results are not in boldface type. Routine references to theorem 10.4.1 or definition 7.3.1 do not contain capitalized initial letters. The author very specifically wishes to downgrade the prominence given to this routine use of these words. Starting new sections on new pages allowed me to carefully control the field of vision as the most important results were presented.

Chapter 1

Measures

1 Basic Properties of Measures

Motivation 1.1 (The Lebesgue integral) The Riemann integral of a continuous function f (we will restrict attention to $f(x) \geq 0$ on $a \leq x \leq b$ for convenience) is formed by subdividing the domain of f, forming approximating sums, and passing to the limit. Thus the mth Riemann sum for $\int_a^b f(x)\,dx$ is defined as

$$(1) \qquad RS_m \equiv \sum_{i=1}^{m} f(x_{mi}^*)\,[x_{mi} - x_{m,i-1}],$$

where $a \equiv x_{m0} < x_{m1} < \cdots < x_{mm} \equiv b$ (with $x_{m,i-1} \leq x_{mi}^* \leq x_{mi}$ for all i) satisfy $\text{mesh}_m \equiv \max[x_{mi} - x_{m,i-1}] \to 0$. Note that $x_{mi} - x_{m,i-1}$ is the measure (or length) of the interval $[x_{m,i-1}, x_{mi}]$, while $f(x_{mi}^*)$ approximates the values of $f(x)$ for all $x_{m,i-1} \leq x \leq x_{mi}$ (at least it does if f is continuous on $[a,b]$). Within the class $\mathcal{C}^+$ of all nonnegative continuous functions, this definition works reasonably well. But it has one major shortcoming. The conclusion $\int_a^b f_n(x)\,dx \to \int_a^b f(x)\,dx$ is one we often wish to make if f_n "converges" to f. However, even when all f_n are in $\mathcal{C}^+$ and $f(x) \equiv \lim f_n(x)$ actually exists, it need not be that f is in $\mathcal{C}^+$ (and thus $\int_a^b f(x)\,dx$ may not even be well-defined) or that $\int_a^b f_n(x)\,dx \to \int_a^b f(x)\,dx$ (even when it is well defined).

A different approach is needed. (Note figure 1.1.)

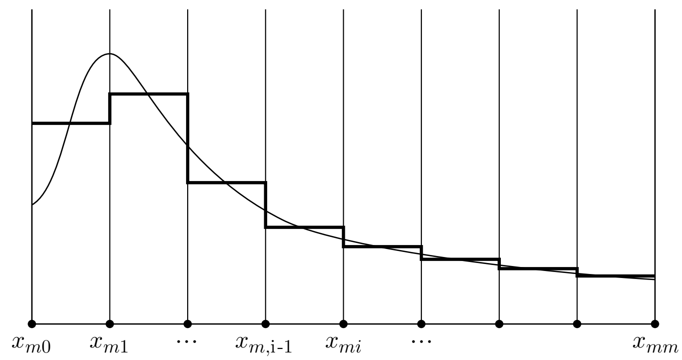
The Lebesgue integral of a nonnegative function is formed by subdividing the range. Thus the mth Lebesgue sum for $\int_a^b f(x)\,dx$ is defined as

$$(2) \qquad LS_m \equiv \sum_{k=1}^{m2^m} \frac{k-1}{2^m} \times \text{measure}\left(\left\{x : \frac{k-1}{2^m} \leq f(x) < \frac{k}{2^m}\right\}\right),$$

and $\int_a^b f(x)\,dx$ is defined to be the limit of the LS_m sums as $m \to \infty$. For what class $\mathcal{M}$ of functions f can this approach succeed? The members f of the class $\mathcal{M}$ will need to be such that the measure (or length) of all sets of the form

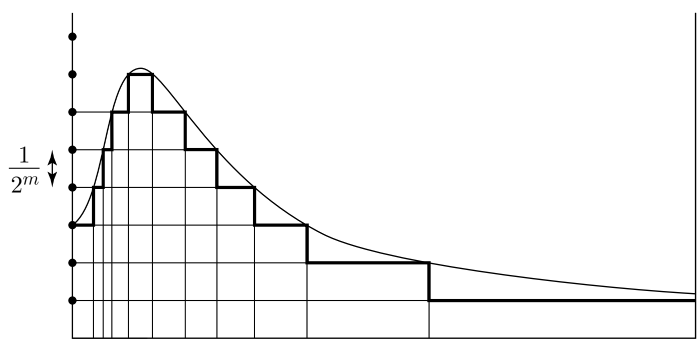
$$(3) \qquad \left\{x : \frac{k-1}{2^m} \leq f(x) < \frac{k}{2^m}\right\}$$

can be specified. This approach leads to the concept of a σ-field $\mathcal{A}$ of subsets of $[a, b]$ that are measurable (that is, we must be able to assign to these sets a number called their "length"), and this leads to the concept of the class $\mathcal{M}$ of measurable functions. This class $\mathcal{M}$ of measurable functions will be seen to be closed under passage to the limit and all the other operations that we are accustomed to performing on functions. Moreover, the desirable property $\int_a^b f_n(x)\,dx \to \int_a^b f(x)\,dx$ for functions f_n "converging" to f will be broadly true. □



Riemann sums

The domain of $f(\cdot)$ is equally divided.



Lebesgue sums

The range of $f(\cdot)$ is equally divided.

Figure 1.1 Riemann sums and Lebesgue sums.

Definition 1.1 (Set theory) Consider a nonvoid class $\mathcal{A}$ of subsets A of a nonvoid set Ω. (For us, Ω will be the sample space of an experiment.)
(a) Let A^c denote the *complement* of A, let $A \cup B$ denote the *union* of A and B, let $A \cap B$ and AB both denote the *intersection*, let $A \setminus B \equiv AB^c$ denote the *set difference*, let $A \triangle B \equiv (A^c B \cup AB^c)$ denote the *symmetric difference*, and let $\emptyset$ denote the *empty set*. The class of all subsets of Ω will be denoted by 2^Ω. Sets A and B are called *disjoint* if $AB = \emptyset$, and sequences of sets A_n or classes of sets $\mathcal{A}_t$ are called disjoint if all pairs are disjoint. Writing $A + B$ or $\sum_1^\infty A_n$ will also denote a union, but will imply the disjointness of the sets in the union. As usual, $A \subset B$ denotes that A is a *subset* of B. We call a sequence A_n *increasing* (and we will nearly always denote this fact by writing $A_n \nearrow$) when $A_n \subset A_{n+1}$ for all $n \geq 1$. We call the sequence *decreasing* (denoted by $A_n \searrow$) when $A_n \supset A_{n+1}$ for all $n \geq 1$. We call the sequence *monotone* if it is either increasing or decreasing. Let ω denote a generic element of Ω. We will use $1_A(\cdot)$ to denote the *indicator function* of A, which equals 1 or 0 at ω according as $\omega \in A$ or $\omega \notin A$.
(b) $\mathcal{A}$ will be called a *field* if it is closed under complements and unions. (That is, A and B in $\mathcal{A}$ requires that A^c and $A \cup B$ be in $\mathcal{A}$.) [Note that both Ω and $\emptyset$ are necessarily in $\mathcal{A}$, as $\mathcal{A}$ was assumed to be nonvoid, with $\Omega = A \cup A^c$ and $\emptyset = \Omega^c$.]
(c) $\mathcal{A}$ will be called a *σ-field* if it is closed under complements and countable unions. (That is, $A, A_1, A_2, \ldots$ in $\mathcal{A}$ requires that A^c and $\cup_1^\infty A_n$ be in $\mathcal{A}$.)
(d) $\mathcal{A}$ will be called a *monotone class* provided it contains $\cup_1^\infty A_n$ for all increasing sequences A_n in $\mathcal{A}$ and contains $\cap_1^\infty A_n$ for all decreasing sequences A_n in $\mathcal{A}$.
(e) $(\Omega, \mathcal{A})$ will be called a *measurable space* provided $\mathcal{A}$ is a σ-field of subsets of Ω.
(f) $\mathcal{A}$ will be called a *π-system* provided AB is in $\mathcal{A}$ for all A and B in $\mathcal{A}$; and $\mathcal{A}$ will be called a *$\bar{\pi}$-system* when Ω in $\mathcal{A}$ is also guaranteed.

If $\mathcal{A}$ is a field (or a σ-field), then it is closed under intersections (under countable intersections); since $AB = (A^c \cup B^c)^c$ (since $\cap_1^\infty A_n = (\cup_1^\infty A_n^c)^c$). Likewise, we could have used "intersection" instead of "union" in our definitions by making use of $A \cup B = (A^c \cap B^c)^c$ and $\cup_1^\infty A_n = (\cap_1^\infty A_n^c)^c$. (This used *De Morgan's laws.*)

Proposition 1.1 (Closure under intersections)
(a) Arbitrary intersections of fields, σ-fields, or monotone classes are fields, σ-fields, or monotone classes, respectively.
[For example, $\mathcal{F} \equiv \cap\{\mathcal{F}_\alpha : \mathcal{F}_\alpha$ is a field under consideration$\}$ is a field.]
(b) There is a minimal field, σ-field, or monotone class *generated by* (or, containing) any specified class $\mathcal{C}$ of subsets of Ω. Call $\mathcal{C}$ the *generators*. For example,

$$(4) \qquad \sigma[\mathcal{C}] \equiv \bigcap\{\mathcal{F}_\alpha : \mathcal{F}_\alpha \text{ is a } \sigma\text{-field of subsets of } \Omega \text{ for which } \mathcal{C} \subset \mathcal{F}_\alpha\}$$

is the *minimal σ-field generated by $\mathcal{C}$* (that is, containing $\mathcal{C}$) .
(c) A collection $\mathcal{A}$ of subsets of Ω is a σ-field if and only if it is both a field and a monotone class.

Proof. (c) ($\Leftarrow$) $\cup_1^\infty A_n = \cup_1^\infty(\cup_1^n A_k)) \equiv \cup_1^\infty B_n \in \mathcal{A}$ since the B_n are in $\mathcal{A}$ and are $\nearrow$. Everything else is even more trivial. □

Exercise 1.1 (Generators) Let $\mathcal{C}_1$ and $\mathcal{C}_2$ denote two collections of subsets of the set Ω. If $\mathcal{C}_2 \subset \sigma[\mathcal{C}_1]$ and $\mathcal{C}_1 \subset \sigma[\mathcal{C}_2]$, then $\sigma[\mathcal{C}_1] = \sigma[\mathcal{C}_2]$. Prove this fact.

Definition 1.2 (Measures and events) Consider a measurable space $(\Omega, \mathcal{A})$ and a set function $\mu : \mathcal{A} \to [0, \infty]$ (that is, $\mu(A) \geq 0$ for each $A \in \mathcal{A}$) having $\mu(\emptyset) = 0$.
(a) Now $\mathcal{A}$ is a σ-field and if μ is *countably additive* (abbreviated *c.a.*) in that

(5) $$\mu\Big(\sum_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} \mu(A_n) \qquad \text{for all disjoint sequences } A_n \text{ in } \mathcal{A},$$

then μ is called a *measure* (or, equivalently, a *countably additive measure*) on $(\Omega, \mathcal{A})$. The triple $(\Omega, \mathcal{A}, \mu)$ is then called a *measure space*. We call μ *finite* if $\mu(\Omega) < \infty$. We call μ *σ-finite* if there exists a *measurable decomposition* of Ω as $\Omega = \sum_1^\infty \Omega_n$ with $\Omega_n \in \mathcal{A}$ and $\mu(\Omega_n) < \infty$ for all n. The sets A in the σ-field $\mathcal{A}$ are called *events*.

[Even if $\mathcal{A}$ is not a σ-field, we will still call μ a measure on $(\Omega, \mathcal{A})$, when (5) holds for all sequences $A_n \in \mathcal{A}$ for which $\sum_1^\infty A_n$ is in $\mathcal{A}$. We will not, however, use the term "measure space" to describe such a triple. We will consider below measures on fields, on certain $\bar{\pi}$-systems, and on some other collections of sets. A useful property of a collection of sets is that along with any sets $A_1, \ldots, A_k$ it also includes all sets of the type $B_k \equiv A_k A_{k-1}^c \cdots A_2^c A_1^c$; then $\bigcup_1^n A_k = \sum_1^n B_k$ is easier to work with.]
(b) Of less interest, call μ a *finitely additive measure* (abbreviated *f.a.*) on $(\Omega, \mathcal{A})$ if

(6) $$\mu(\,\textstyle\sum_1^n A_k) = \sum_1^n \mu(A_k)$$

for all disjoint sequences A_k in $\mathcal{A}$ for which $\sum_1^n A_k$ is also in $\mathcal{A}$.

Definition 1.3 (Outer measures) Consider a set function $\mu^* : 2^\Omega \to [0, \infty]$.
(a) Suppose that μ^* also satisfies the following three properties.
Null: $\mu^*(\emptyset) = 0$.
Monotone: $\mu^*(A) \leq \mu^*(B)$ for all $A \subset B$.
Countable subadditivity: $\mu^*(\bigcup_1^\infty A_n) \leq \sum_1^\infty \mu^*(A_n)$ for all sequences A_n.
Then μ^* is called an *outer measure*.
(b) An arbitrary subset A of Ω is called *μ^*-measurable* if

(7) $$\mu^*(T) = \mu^*(TA) + \mu^*(TA^c) \qquad \text{for all subsets } T \subset \Omega.$$

Sets T used in this capacity are called *test sets*.
(c) We let $\mathcal{A}^*$ denote the class of all *μ^*-measurable sets*, that is,

(8) $$\mathcal{A}^* \equiv \{A \in 2^\Omega : A \text{ is } \mu^*\text{-measurable}\}.$$

[Note that $A \in \mathcal{A}^*$ if and only if $\mu^*(T) \geq \mu^*(TA) + \mu^*(TA^c)$ for all $T \subset \Omega$, since the other inequality is trivial by the subadditivity of μ^*.]

Motivation 1.2 (Measure) In this paragraph we will consider only one possible measure μ, namely the Lebesgue-measure generalization of length. Let $\mathcal{C}_I$ denote the set of all intervals of the types $(a, b]$, $(-\infty, b]$, and $(a, +\infty)$ on the real line R, and for each of these intervals I we assign a measure value $\mu(I)$ equal to its length, thus $\infty, b - a, \infty$ in the three special cases. All is well until we manipulate the sets in $\mathcal{C}_I$, as even the union of two elements in $\mathcal{C}_I$ need not be in $\mathcal{C}_I$. Thus, $\mathcal{C}_I$ is not a very rich collection of sets. A natural extension is to let $\mathcal{C}_F$ denote the collection of all finite disjoint unions of sets in $\mathcal{C}_I$, where the measure $\mu(A)$ we assign to each such set A is just the sum of the measures (lengths) of all its disjoint pieces. Now $\mathcal{C}_F$ is a field, and is thus closed under the elementary operations of union, intersection, and complementation. Much can be done using only $\mathcal{C}_F$ and letting "measure" be the "exact length." But $\mathcal{C}_F$ is not closed under passage to the limit, and it is thus insufficient for many of our needs. For this reason the concept of the smallest σ-field containing $\mathcal{C}_F$, labeled $\mathcal{B} \equiv \sigma[\mathcal{C}_F]$, is introduced. We call $\mathcal{B}$ the Borel sets. But let us work backwards. Let us assign an outer measure value $\mu^*(A)$ to every subset A in the class 2^R of all subsets of the real line R. In particular, to any subset A we assign the value $\mu^*(A)$ that is the infimum of all possible numbers $\sum_{n=1}^\infty \mu(A_n)$, in which each A_n is in the field $\mathcal{C}_F$ (so that we know its measure) and in which the A_n's form a *cover* of A (in that $A \subset \cup_1^\infty A_n$). Thus each number $\sum_1^\infty \mu(A_n)$ is a natural upper bound to the measure (or generalized length) of the set A, and we will specify the infimum of such upper bounds to be the outer measure of A. Thus to each subset A of the real line we assign a value $\mu^*(A)$ of generalized length. This value seems "reasonable," but does it "perform correctly"? Let us say that a particular set A is μ^*-measurable (that is, it "performs correctly") if $\mu^*(T) = \mu^*(TA) + \mu^*(TA^c)$ for all subsets T of the real line R, that is, if the A versus A^c division of the line divides every subset T of the line into two pieces in a fashion that is μ^*-additive. This is undoubtedly a combination of reasonableness and fine technicality that took some time to evolve in the mind of its creator, Carathéodory, while he searched for a condition that "worked." In what sense does it "work"? The collection $\mathcal{A}^*$ of all μ^*-measurable sets turns out to be a σ-field. Thus the collection $\mathcal{A}^*$ is closed under all operations that we are likely to perform; and it is big enough, in that it is a σ-field that contains $\mathcal{C}_F$. Thus we will work with the restriction $\mu^*|\mathcal{A}^*$ of μ^* to the sets of $\mathcal{A}^*$ (here, the vertical line means "restricted to"). This is enough to meet our needs.

There are many measures other than length. For an $\nearrow$ and right-continuous function F on the real line (called a generalized df) we define the Stieltjes measure of an arbitrary interval $(a, b]$ (with $-\infty \le a < b \le \infty$) in C_I by $\mu_F((a, b]) = F(b) - F(a)$, and we extend it to sets in $\mathcal{C}_F$ by adding up the measure of the pieces. Reapplying the previous paragraph, we can extend μ_F to the μ_F^*-measurable sets. It is the important Carathéodory extension theorem that will establish that all Stieltjes measures (including the case of ordinary length, where $F(x) = x$, as considered in the first paragraph) can be extended from $\mathcal{C}_F$ to the Borel sets $\mathcal{B}$. That is, all Borel sets are μ^*-measurable for every Stieltjes measure. One further extension is possible, in that every measure can be "completed" (see the end of section 1.2). We note here only that when the Stieltjes measure μ_F associated with the generalized df F is "completed," its domain of definition is extended from the Borel sets $\mathcal{B}$ (which all Stieltjes measures have in common) to a larger collection $\hat{\mathcal{B}}_{\mu_F}$ that depends on the particular F. It is left to section 1.2 to simply state that this is as far as we can go. That is, except in rather trivial special cases (especially, mass at only countably many points), we find that $\hat{\mathcal{B}}_{\mu_F}$ is a proper subset of 2^R. (That is, it is typically impossible to try to define the measure of all subsets of Ω in a suitable fashion.) $\square$

Example 1.1 (Some examples of measures, informally)
(a) *Lebesgue measure*:
Let $\lambda(A)$ denote the length of A.
(b) *Counting measure*:
Let $\#(A)$ denote the number of "points" in A (or the *cardinality* of A).
(c) *Unit point mass*:
Let $\delta_{\omega_0}(A) \equiv 1_A(\omega_o)$, assigning measure 1 or 0 to A as $\omega_0 \in A$ or not. □

Example 1.2 (Borel sets)
(a) Let $\Omega = R$ and let $\mathcal{C}$ consist of all finite disjoint unions of intervals of the types $(a, b]$, $(-\infty, b]$, and $(a, +\infty)$. Clearly, $\mathcal{C}$ is a field. Then $\mathcal{B} \equiv \sigma[\mathcal{C}]$ will be called the *Borel sets* (or the *Borel subsets* of R). Let $\mu(A)$ be defined to be the sum of the lengths of the intervals composing A, for each $A \in \mathcal{C}$. Then μ is a (c.a.) measure on the field $\mathcal{C}$, as will be seen in the proof of theorem 1.3.1 below.
(b) If (Ω, d) is a metric space and $\mathcal{U} \equiv \{$all d-open subsets of $\Omega\}$, then $\mathcal{B} \equiv \sigma[\mathcal{U}]$ will be called the *Borel sets* or the *Borel σ-field* .
(c) If (Ω, d) is $(R, |\cdot|)$ for absolute value $|\cdot|$, then $\sigma[\mathcal{C}] = \sigma[\mathcal{U}]$ even though $\mathcal{C} \neq \mathcal{U}$. [This claim is true, since $\mathcal{C} \subset \sigma[\mathcal{U}]$ and $\mathcal{U} \subset \sigma[\mathcal{C}]$ are clear. Then, just make a trivial appeal to exercise 1.1.]
(d) Let $\bar{R} \equiv [-\infty, = \infty]$ denote the *extended real line*; let $\bar{\mathcal{B}} \equiv \sigma[\mathcal{B}, \{-\infty\}, \{+\infty\}]$. □

Proposition 1.2 (Monotone properties of measures) Let $(\Omega, \mathcal{A}, \mu)$ denote a measure space. (Of course, $\mu(A) \leq \mu(B)$ for $A \subset B$ in $\mathcal{A}$.) Let $A_1, A_2, \ldots$ be in $\mathcal{A}$.
(a) If $A_n \subset A_{n+1}$ for all n, then

(9) $$\mu(\bigcup_1^\infty A_n) = \lim_{n\to\infty} \mu(A_n).$$

(b) If $\mu(A_{n_0}) < \infty$ for some n_0, and $A_n \supset A_{n+1}$ for all n, then

(10) $$\mu(\bigcap_{n=1}^\infty A_n) = \lim_{n\to\infty} \mu(A_n)\,.$$

[Letting Ω denote the real line R, letting $A_n = [n, \infty)$, and letting μ denote either Lebesgue measure or counting measure, we see the need for some requirement.]
(c) (Countable subadditivity) Whenever $A_1, A_2, \ldots$ and $\cup_1^\infty A_n$ are all in $\mathcal{A}$, then

$$\mu(\bigcup_1^\infty A_k) \leq \sum_1^\infty \mu(A_k)\,;$$

(d) All this also holds true for a measure on a field (via the same proofs).

Proof. (a) Now,

$$\mu(\cup_1^\infty A_n) = \mu(\textstyle\sum_1^\infty (A_n \setminus A_{n-1})) \qquad \text{with } A_0 \equiv \emptyset$$

(p) $$= \textstyle\sum_1^\infty \mu(A_n \setminus A_{n-1}) \qquad \text{by c.a.}$$

$$= \lim_n \textstyle\sum_{k=1}^n \mu(A_k \setminus A_{k-1})$$

$$= \lim_n \mu(\textstyle\sum_{k=1}^n (A_k \setminus A_{k-1})) \qquad \text{by f.a.}$$

(q) $$= \lim_n \mu(A_n).$$

(b) Without loss of generality, redefine $A_1 = A_2 = \cdots = A_{n_0}$. Let $B_n \equiv A_1 \setminus A_n$, so that $B_n \nearrow$. Thus, on the one hand,

$$\lim_n \mu(B_n) = \mu(\cup_1^\infty B_n) \qquad \text{by (a)}$$
$$= \mu(\cup_1^\infty (A_1 \cap A_n^c))$$
$$= \mu(A_1 \cap \cup_1^\infty A_n^c)$$
$$= \mu(A_1 \cap (\cap_1^\infty A_n)^c)$$
$$\text{(r)} \qquad = \mu(A_1) - \mu(\cap_1^\infty A_n).$$

On the other hand,

$$\lim_n \mu(B_n) = \lim_n \{\mu(A_1) - \mu(A_n)\} \qquad \text{by f.a.}$$
$$\text{(s)} \qquad = \mu(A_1) - \lim_n \mu(A_n).$$

Equate (r) and (s); since $\mu(A_1) < \infty$, we can cancel it to obtain the equality $\mu(\cap_1^\infty A_n) = \lim_n \mu(A_n)$.

(c) Let $B_1 \equiv A_1, B_2 \equiv A_2 A_1^c, \ldots, B_k \equiv A_k A_{k-1}^c \cdots A_1^c$. Then these newly defined sets B_k are disjoint, and $\cup_{k=1}^n A_k = \sum_{k=1}^n B_k$. Hence [a technique worth remembering]

$$\text{(11)} \qquad \mu(\bigcup_{k=1}^n A_k) = \mu(\sum_{k=1}^n B_k) = \sum_{k=1}^n \mu(B_k)$$
$$\text{where } \bigcup_{k=1}^n A_k = \sum_{k=1}^n B_k \text{ is } \nearrow \text{ for } B_k \equiv A_k A_{k-1}^c \cdots A_1^c$$
$$\leq \sum_1^n \mu(A_k)$$
$$\text{(t)} \qquad \leq \sum_1^\infty \mu(A_k) \qquad \text{by monotonicity.}$$

Let $n \to \infty$ in (11), and use part (a) to get the result. □

Definition 1.4 (liminf and limsup of sets) Let

$$\text{(12)} \qquad \underline{\lim}\, A_n \equiv \bigcup_{n=1}^\infty \bigcap_{k=n}^\infty A_k = \{\omega : \omega \text{ is in all but finitely many } A_n\text{'s}\}$$
$$\equiv \{\omega : \omega \in A_n \text{ a.b.f.}\},$$

where we use *a.b.f.* to abbreviate *in all but finitely many cases.* Let

$$\text{(13)} \qquad \overline{\lim}\, A_n \equiv \bigcap_{n=1}^\infty \bigcup_{k=n}^\infty A_k = \{\omega : \omega \text{ is in an infinite number of } A_n\text{'s}\}$$
$$\equiv \{\omega : \omega \in A_n \text{ i.o.}\},$$

where we use *i.o.* to abbreviate *infinitely often.*

[It is important to learn to read these two mathematical equations in a way that makes it clear that the verbal description is correct.] Note that we always have $\underline{\lim}\, A_n \subset \overline{\lim}\, A_n$. Define

$$\text{(14)} \qquad \lim A_n \equiv \underline{\lim}\, A_n \qquad \text{whenever } \underline{\lim}\, A_n = \overline{\lim}\, A_n.$$

We also let $\liminf A_n \equiv \underline{\lim}\, A_n$ and $\limsup A_n \equiv \overline{\lim}\, A_n$, giving us alternative notations.

Proposition 1.3 Clearly, $\lim A_n$ equals $\cup_1^\infty A_n$ when A_n is an $\nearrow$ sequence, and $\lim A_n$ equals $\cap_1^\infty A_n$ when A_n is a $\searrow$ sequence.

Exercise 1.2 (a) Now $\mu(\liminf A_n) \le \liminf \mu(A_n)$ is always true.
(b) Also, $\limsup \mu(A_n) \le \mu(\limsup A_n)$ holds if $\mu(\Omega) < \infty$. (Why the condition?)

Definition 1.5 (lim inf and lim sup of numbers) Recall that for real number sequences a_n one defines $\underline{\lim}\, a_n \equiv \liminf a_n$ and $\overline{\lim}\, a_n \equiv \limsup a_n$ by

$$(15)\qquad \begin{aligned} \liminf_{n\to\infty} a_n &\equiv \lim_{n\to\infty} (\inf_{k\ge n} a_k) = \sup_{n\ge 1} (\inf_{k\ge n} a_k) \qquad \text{and} \\ \limsup_{n\to\infty} a_n &\equiv \lim_{n\to\infty} \left(\sup_{k\ge n} a_k\right) = \inf_{n\ge 1} \left(\sup_{k\ge n} a_k\right), \end{aligned}$$

and these yield the smallest limit point and the largest limit point, respectively, of the sequence a_n.

Definition 1.6 (Continuity of measures) A set function μ is *continuous from below* (*above*) if $\mu(\lim A_n) = \lim \mu(A_n)$ for all sequences A_n in Ω that are $\nearrow$ (for all sequences A_n in Ω that are $\searrow$, with at least one $\mu(A_n)$ finite). We call μ *continuous* in case it is continuous both from below and from above. If $\lim \mu(A_n) = \mu(A)$ whenever $A_n \nearrow A$, then μ is said to be *continuous from below at* A, etc.

The next result is often used in conjunction with the Carathéodory extension theorem of the next section. View it as a converse to the proposition 1.2.

Proposition 1.4 (Continuity of measures) If a finitely additive measure μ on either a field or σ-field is either continuous from below or has $\mu(\Omega) < \infty$ and is continuous from above at $\emptyset$, then it is a countably additive measure.

Proof. Suppose first that μ is continuous from below. Then

$$\mu\left(\textstyle\sum_1^\infty A_k\right) = \mu\left(\lim \textstyle\sum_1^n A_k\right)$$

$$\text{(a)}\qquad = \lim \mu\left(\textstyle\sum_1^n A_k\right) \qquad \text{by continuity from below}$$

$$\text{(b)}\qquad = \lim \textstyle\sum_1^n \mu(A_k) \qquad \text{by f.a. (where we used only that } \mathcal{A} \text{ is a field)}$$

$$\text{(c)}\qquad = \textstyle\sum_1^\infty \mu(A_k),$$

giving the required countable additivity. Thus μ is a measure.

Suppose next that μ is finite and is also continuous from above at $\emptyset$. Then f.a. (even if $\mathcal{A}$ is only a field) gives

$$\mu\left(\textstyle\sum_1^\infty A_k\right) = \mu\left(\textstyle\sum_1^n A_k\right) + \mu\left(\textstyle\sum_{n+1}^\infty A_k\right) = \textstyle\sum_1^n \mu(A_k) + \mu\left(\textstyle\sum_{n+1}^\infty A_k\right)$$

$$\text{(d)}\qquad \to \textstyle\sum_1^\infty \mu(A_k) + 0,$$

where $\mu(\sum_{n+1}^\infty A_k) \to \mu(\emptyset) = 0$ by continuity from above at $\emptyset$, since $\sum_{n+1}^\infty A_k \searrow \emptyset$ and μ is finite. That is, this f.a. measure is also c.a., and hence it is a measure. $\square$

Definition 1.7 ("Little oh," "big oh," and "at most" $\oplus$) We write:

$$
(16) \qquad \begin{aligned} a_n &\equiv o(r_n) && \text{if} && a_n/r_n \to 0\,, \\ a_n &\equiv O(r_n) && \text{if} && \overline{\lim}\, |a_n/r_n| \le (\text{some } M) < \infty. \end{aligned}
$$

We write

$$
(17) \qquad a_n = b_n \oplus c_n \qquad \text{if} \qquad |a_n - b_n| \le c_n.
$$

This last notation allows us to string inequalities together linearly, instead of having to start a new inequality on a new line. (I use it often.)

Exercise 1.3 (π-systems and λ-systems) Consider a measurable space $(\Omega, \mathcal{A})$. A class $\mathcal{D}$ of subsets is called a λ-*system* if it contains the space Ω and all proper differences ($A \setminus B$, when $B \subset A$ with both $A, B \in \mathcal{D}$) and if it is closed under monotone increasing limits. [Recall that a class is called a π-*system* if it is closed under finite intersections, while $\bar{\pi}$-systems are also required to contain Ω.]
(a) The minimal λ-system generated by a class $\mathcal{C}$ is denoted by $\lambda[\mathcal{C}]$. Show that $\lambda[\mathcal{C}]$ is equal to the intersection of all λ-systems containing $\mathcal{C}$.
(b) A collection $\mathcal{A}$ of subsets of Ω is a σ-field if and only if it is both a π-system and a λ-system.
(c) Let $\mathcal{C}$ be a π-system and let $\mathcal{D}$ be a λ-system. Then $\mathcal{C} \subset \mathcal{D}$ implies that $\sigma[\mathcal{C}] \subset \mathcal{D}$. Note (or, show) that this follows from (19) below.

Proposition 1.5 (Dynkin's π-λ theorem) Let μ and μ' be two finite measures on the measurable space $(\Omega, \mathcal{A})$. Let $\mathcal{C}$ be a $\bar{\pi}$-system, where $\mathcal{C} \subset \mathcal{A}$. Then

$$
(18) \qquad \mu = \mu' \text{ on the } \bar{\pi}\text{-system } \mathcal{C} \qquad \text{implies} \qquad \mu = \mu' \text{ on } \sigma[\mathcal{C}].
$$

Proof. We first show that on *any* measurable space $(\Omega, \mathcal{A})$ we have

$$
(19) \qquad \sigma[\mathcal{C}] = \lambda[\mathcal{C}] \qquad \text{when } \mathcal{C} \text{ is a } \pi\text{-system of subsets of } \mathcal{A}.
$$

Let $\mathcal{D} \equiv \lambda[\mathcal{C}]$. By the easy exercise 1.3(a)(b), it suffices to show that $\mathcal{D}$ is a π-system (that is, that $A, B \in \mathcal{D}$ implies $A \cap B \in \mathcal{D}$). We first go just halfway; let

$$
(a) \qquad \mathcal{E}_C \equiv \{A \in \mathcal{D} : AC \in \mathcal{D}\}, \quad \text{for any fixed } C \in \mathcal{C}.
$$

Then $\mathcal{C} \subset \mathcal{E}_C$, and $\Omega \in \mathcal{E}_C$. Also, for $A, B \in \mathcal{E}_C$ with $B \subset A$ and for $C \in \mathcal{C}$ we have (since both AC and BC are in $\mathcal{D}$) that $(A \setminus B)C = (AC \setminus BC) \in \mathcal{D}$, so that $A \setminus B \in \mathcal{E}_C$. Finally, if A_n is $\nearrow$ in $\mathcal{E}_C$, then $A_n C$ is $\nearrow$ in $\mathcal{D}$; so $A \equiv \lim A_n$ has $AC \in \mathcal{D}$, and $A \in \mathcal{E}_C$. Thus $\mathcal{E}_C$ is a λ-system containing $\mathcal{C}$. Thus $\mathcal{E}_C = \mathcal{D}$, since $\mathcal{D}$ was the smallest such class. We have thus learned of $\mathcal{D}$ that

$$
(b) \qquad AC \in \mathcal{D} \text{ for all } C \in \mathcal{C}, \text{ for each } A \in \mathcal{D}.
$$

To go the rest of the way, we define

(c) $\mathcal{F}_D \equiv \{A \in \mathcal{D} : AD \in \mathcal{D}\}$, for any fixed $D \in \mathcal{D}$.

Then $\mathcal{C} \subset \mathcal{F}_D$, by (b), and $\Omega \in \mathcal{F}_D$. Also, for $A, B \in \mathcal{F}_D$ with $B \subset A$ and for $D \in \mathcal{D}$ we have (since both AD and BD are in $\mathcal{D}$) that $(A \setminus B)D = (AD \setminus BD) \in \mathcal{D}$, so that $A \setminus B \in \mathcal{F}_D$. Finally, if A_n is $\nearrow$ in $\mathcal{F}_D$, then A_nD is $\nearrow$ in $\mathcal{D}$; so $A \equiv \lim A_n$ has $AD \in \mathcal{D}$, and $A \in \mathcal{F}_D$. Thus $\mathcal{F}_D$ is a λ-system containing $\mathcal{C}$. Thus $\mathcal{F}_D = \mathcal{D}$, since $\mathcal{D}$ was the smallest such class. We have thus learned of $\mathcal{D}$ that

(d) $AD \in \mathcal{D}$ for all $A \in \mathcal{D}$, for each $D \in \mathcal{D}$.

That is, $\mathcal{D}$ is closed under intersections; and thus $\mathcal{D}$ is a π-system. Thus (19) holds.

We will now demonstrate that $\mathcal{G} \equiv \{A \in \mathcal{A} : \mu(A) = \mu'(A)\}$ is a λ-system on Ω. First, $\Omega \in \mathcal{G}$, since Ω is in the $\bar{\pi}$-system $\mathcal{C}$. Second, when $A \subset B$ are both in $\mathcal{G}$ we have the equality (since $\mu(A)$ and $\mu'(A)$ are finite)

(e) $\mu(B \setminus A) = \mu(B) - \mu(A) = \mu'(B) - \mu'(A) = \mu'(B \setminus A)$,

giving $B \setminus A \in \mathcal{G}$. Finally, let $A_n \nearrow A$ with all A_n's in $\mathcal{G}$. Then proposition 1.2(i) yields the result

(f) $\mu(A) = \lim \mu(A_n) = \lim \mu'(A_n) = \mu'(A)$,

so that $A \in \mathcal{G}$. Thus $\mathcal{G}$ is a λ-system.

Thus the collection $\mathcal{G}$ on which $\mu = \mu'$ is a λ-system that contains the $\bar{\pi}$-system $\mathcal{C}$. Applying (19) shows that $\sigma[\mathcal{C}] \subset \mathcal{G}$. □

The previous result is very useful in extending the verification of independence from small classes of sets to larger ones. The next proposition is used for both Fubini's theorem and the existence of a regular conditional probability distribution. It could also have been used below to give an alternate proof of uniqueness in the Carathéodory extension theorem.

Proposition 1.6 (Minimal monotone class; Halmos) The minimal monotone class $\mathcal{M} \equiv m[\mathcal{C}]$ containing the field $\mathcal{C}$ and the minimal σ-field $\sigma[\mathcal{C}]$ generated by the same field $\mathcal{C}$ satisfy

(20) $m[\mathcal{C}] = \sigma[\mathcal{C}]$ when $\mathcal{C}$ is a field.

Proof. Since σ-fields are monotone classes, we have that $\sigma[\mathcal{C}] \supset \mathcal{M}$. If we now show that $\mathcal{M}$ is a field, then proposition 1.1(c) will imply that $\sigma[\mathcal{C}] \subset \mathcal{M}$.

To show that $\mathcal{M}$ is a field, it suffices to show that

(a) A, B in $\mathcal{M}$ implies AB, A^cB, AB^c are in $\mathcal{M}$.

Suppose that (a) has been established. We will now show that (a) implies that $\mathcal{M}$ is a field.

Complements: Let $A \in \mathcal{M}$, and note that $\Omega \in \mathcal{M}$, since $\mathcal{C} \subset \mathcal{M}$. Then $A, \Omega \in \mathcal{M}$ implies that $A^c = A^c\Omega \in \mathcal{M}$ by (a).

Unions: Let $A, B \in \mathcal{M}$. Then $A \cup B = (A^c \cap B^c)^c \in \mathcal{M}$.

Thus $\mathcal{M}$ is indeed a field, provided that (a) is true. It thus suffices to prove (a).

For each $A \in \mathcal{M}$, let $\mathcal{M}_A \equiv \{B \in \mathcal{M} : AB, A^cB, AB^c \in \mathcal{M}\}$. Note that it suffices to prove that

(b) $\mathcal{M}_A = \mathcal{M}$ for each fixed $A \in \mathcal{M}$.

We first show that

(c) $\mathcal{M}_A$ is a monotone class.

Let B_n be monotone in $\mathcal{M}_A$, with limit set B. Since B_n is monotone in $\mathcal{M}_A$, it is also monotone in $\mathcal{M}$, and thus $B \equiv \lim_n B_n \in \mathcal{M}$. Since $B_n \in \mathcal{M}_A$, we have $AB_n \in \mathcal{M}$, and since AB_n is monotone in $\mathcal{M}$, we have $AB = \lim_n AB_n \in \mathcal{M}$. In like fashion, A^cB and AB^c are in $\mathcal{M}$. Therefore, $B \in \mathcal{M}_A$, by definition of $\mathcal{M}_A$. That is, (c) holds.

We next show that

(d) $\mathcal{M}_A = \mathcal{M}$ for each fixed $A \in \mathcal{C}$.

Let $A \in \mathcal{C}$ and let $C \in \mathcal{C}$. Then $A \in \mathcal{M}_C$, since $\mathcal{C}$ is a field. But $A \in \mathcal{M}_C$ if and only if $C \in \mathcal{M}_A$, by the symmetry of the definition of $\mathcal{M}_A$. Thus $\mathcal{C} \subset \mathcal{M}_A$. That is, $\mathcal{C} \subset \mathcal{M}_A \subset \mathcal{M}$, and $\mathcal{M}_A$ is a monotone class by (c). But $\mathcal{M}$ is the minimal monotone class containing $\mathcal{C}$, by the definition of $\mathcal{M}$. Thus (d) holds. But in fact, we shall now strengthen (d) to

(e) $\mathcal{M}_B = \mathcal{M}$ for each fixed $B \in \mathcal{M}$.

The conditions for membership in $\mathcal{M}$ imposed on pairs A, B are symmetric. Thus for $A \in \mathcal{C}$, the statement established above in (d) that $B \in \mathcal{M}(= \mathcal{M}_A)$ is true if and only if $A \in \mathcal{M}_B$. Thus $\mathcal{C} \subset \mathcal{M}_B$, where $\mathcal{M}_B$ is a monotone class. Thus $\mathcal{M}_B = \mathcal{M}$, since (as was earlier noted) $\mathcal{M}$ is the smallest such monotone class. Thus (e) (and hence (a)) is established. □

2 Construction and Extension of Measures

Definition 2.1 (Outer extension) Let Ω be arbitrary. Let μ be a measure on a field $\mathcal{C}$ of subsets Ω. For each $A \in 2^{\Omega}$ define

$$(1) \qquad \mu^*(A) \equiv \inf\left\{ \sum_{n=1}^{\infty} \mu(A_n) \; : \; A \subset \bigcup_{n=1}^{\infty} A_n \text{ with all } A_n \in \mathcal{C} \right\}.$$

Now, μ^* is called the *outer extension* of μ. The sequences $A_1, A_2, \ldots$ are called *Carathéodory coverings*. [There is always at least one covering, since $\Omega \in \mathcal{C}$.]

Theorem 2.1 (Carathéodory extension theorem) A measure μ on a field $\mathcal{C}$ can be extended to a measure on the σ-field $\sigma[\mathcal{C}]$ generated by $\mathcal{C}$ by defining

$$(2) \qquad \mu(A) \equiv \mu^*(A) \qquad \text{for each } A \text{ in } \mathcal{A} \equiv \sigma[\mathcal{C}].$$

If μ is σ-finite on $\mathcal{C}$, then the extension is unique on $\mathcal{A} = \sigma[\mathcal{C}]$ and is also σ-finite.

Comment: Let $\mathcal{A}^*$ denote the μ^*-measurable sets, as in (1.1.8). The measure μ on the field $\mathcal{C}$ will, in fact, be extended to $\mathcal{A}^*$. Thus, $\sigma[\mathcal{C}] \subset \mathcal{A}^*$ will also be shown.

Proof. The proof proceeds by a series of claims.

Claim 1: μ^* is an outer measure on $(\Omega, 2^{\Omega})$.
Null: Now, $\mu^*(\emptyset) = 0$, since $\emptyset, \emptyset, \ldots$ is a covering of $\emptyset$.
Monotone: Let $A \subset B$. Then every covering of B is also a covering of A. Thus $\mu^*(A) \le \mu^*(B)$.
Countably subadditive: Let all $A_n \subset \Omega$ be arbitrary. Let $\epsilon > 0$. For each A_n there is a covering $\{A_{nk} : k \ge 1\}$ such that

$$(3) \qquad \textstyle\sum_1^{\infty} \mu(A_{nk}) \le \mu^*(A_n) + \epsilon/2^n, \qquad \text{since } \mu^*(A_n) \text{ is an infimum.}$$

[The choice of a convergent series (like $\epsilon/2^n$) that adds to ϵ is an important technique for the reader to learn.] Now $\cup_n A_n \subset \cup_n(\cup_k A_{nk})$. Thus

$$\begin{aligned}
\mu^*(\textstyle\bigcup_n A_n) &\le \mu^*(\textstyle\bigcup_n \bigcup_k A_{nk}) \qquad \text{since } \mu^* \text{ is monotone} \\
&\le \textstyle\sum_n \sum_k \mu(A_{nk}) \\
&\qquad \text{since the } A_{nk}\text{'s form a covering of the set } \textstyle\bigcup_n \bigcup_k A_{nk} \\
&\le \textstyle\sum_n [\mu^*(A_n) + \epsilon/2^n] \qquad \text{by (3)} \\
&= \textstyle\sum_n \mu^*(A_n) + \epsilon.
\end{aligned}$$

But $\epsilon > 0$ was arbitrary, and thus $\mu^*(\bigcup_n A_n) \le \sum_n \mu^*(A_n)$.

Claim 2: $\mu^*|\mathcal{C} = \mu$ (that is, $\mu^*(C) = \mu(C)$ for all $C \in \mathcal{C}$), and $\mathcal{C} \subset \mathcal{A}^*$.
Let $C \in \mathcal{C}$. Then $\mu^*(C) \le \mu(C)$, since $C, \emptyset, \emptyset, \ldots$ is a covering of C. For the other direction, we let $A_1, A_2, \ldots$ be any covering of C. Since μ is c.a. on $\mathcal{C}$, and since $\cup_1^{\infty}(A_n \cap C) = C \in \mathcal{C}$, we have from proposition 1.1.2(c) that

$$\mu(C) = \mu(\textstyle\bigcup_1^{\infty}(A_n \cap C)) \le \sum_1^{\infty} \mu(A_n \cap C) \le \sum_1^{\infty} \mu(A_n),$$

and thus $\mu(C) \le \mu^*(C)$. Thus $\mu(C) = \mu^*(C)$. We next show that any $C \in \mathcal{C}$ is also in $\mathcal{A}^*$. Let $C \in \mathcal{C}$. Let $\epsilon > 0$, and let a test set T be given. There exists a covering $\{A_n\}_1^{\infty} \subset \mathcal{C}$ of T such that

(a) $\mu^*(T) + \epsilon \geq \sum_1^\infty \mu(A_n)$ since $\mu^*(T)$ is an infimum

$= \sum_1^\infty \mu(CA_n) + \sum_1^\infty \mu(C^c A_n)$

since μ is c.a. on $\mathcal{C}$ with C and A_n in $\mathcal{C}$

(b) $\geq \mu^*(CT) + \mu^*(C^c T)$ since CA_n covers CT and $C^c A_n$ covers $C^c T$.

But $\epsilon > 0$ is arbitrary. Thus $C \in \mathcal{A}^*$. Thus $\mathcal{C} \subset \mathcal{A}^*$.

Claim 3: The class $\mathcal{A}^*$ of μ^*-measurable subsets of Ω is a field that contains $\mathcal{C}$. Now, $A \in \mathcal{A}^*$ implies that $A^c \in \mathcal{A}^*$: The definition of μ^*-measurable is symmetric in A and A^c. And $A, B \in \mathcal{A}^*$ implies that $AB \in \mathcal{A}^*$: For any test set $T \subset \Omega$ we have the required inequality

$\mu^*(T) = \mu^*(TA) + \mu^*(TA^c)$ since $A \in \mathcal{A}^*$

$= \mu^*(TAB) + \mu^*(TAB^c) + \mu^*(TA^cB) + \mu^*(TA^cB^c)$

since $B \in \mathcal{A}^*$ with test set TA and with test set TA^c

$\geq \mu^*(TAB) + \mu^*(TAB^c + TA^cB + TA^cB^c) = \mu^*(TAB) + \mu^*(T(AB)^c)$

since μ^* is countably subadditive. As the reverse inequality is trivial,

(c) $\mu^*(T) = \mu^*(TAB) + \mu^*(T(AB)^c)$, giving $AB \in \mathcal{A}^*$.

Thus $\mathcal{A}^*$ is a field.

Claim 4: μ^* is a f.a. measure on $\mathcal{A}^*$.
Let $A, B \in \mathcal{A}^*$ be disjoint. Finite additivity follows from

$\mu^*(A + B) = \mu^*((A + B)A) + \mu^*((A + B)A^c)$

since $A \in \mathcal{A}^*$ with test set $A + B$

(d) $= \mu^*(A) + \mu^*(B)$.

Trivially, $\mu^*(A) \geq 0$ for all sets A. And $\mu^*(\emptyset) = 0$ was shown in the first claim.

Claim 5: $\mathcal{A}^*$ is a σ-field, and it contains $\sigma[\mathcal{C}]$.
It suffices to show that $A \equiv \sum_1^\infty A_n \in \mathcal{A}^*$ whenever all $A_n \in \mathcal{A}^*$, since $\mathcal{A}^*$ is a field. Now, $B_n \equiv \sum_1^n A_k \in \mathcal{A}^*$, since $\mathcal{A}^*$ is a field. Using $B_n \in \mathcal{A}^*$ for the first step,

$\mu^*(T) = \mu^*(TB_n) + \mu^*(TB_n^c) \geq \mu^*(TB_n) + \mu^*(TA^c)$

since μ^* is monotone and $B_n^c \supset A^c$

$= \mu^*((TB_n)A_1) + \mu^*((TB_n)A_1^c) + \mu^*(TA^c)$ as $A_1 \in \mathcal{A}^*$

$= \mu^*(TA_1) + \mu^*(T \sum_2^n A_k) + \mu^*(TA^c)$

$= \mu^*(TA_1) + \mu^*(T(\sum_2^n A_k)A_2) + \mu^*(T(\sum_2^n A_k)A_2^c) + \mu^*(TA^c)$

$= \mu^*(TA_1) + \mu^*(TA_2) + \mu^*(T \sum_3^n A_k) + \mu^*(TA^c)$

(e) $= \cdots = \sum_1^n \mu^*(TA_k) + \mu^*(TA^c)$.

Letting $n \to \infty$ gives

(f) $\mu^*(T) \geq \sum_1^\infty \mu^*(TA_k) + \mu^*(TA^c)$

(g) $\geq \mu^*(TA) + \mu^*(TA^c)$ since μ^* is countably subadditive.

Thus $A \in \mathcal{A}^*$.

Claim 6: μ^* is c.a. on $\mathcal{A}^*$.
Replace T by A in (f) to get $\mu^*(A) \geq \sum_1^\infty \mu^*(A_n)$, and then countable subadditivity gives the reverse inequality.

Claim 7: When μ is a finite measure, its extension μ^* to $\mathcal{A}^*$ is unique.
Let ν denote any other extension of μ to $\mathcal{A}^*$. Let A in $\mathcal{A}^*$. For any Carathéodory covering $A_1, A_2, \ldots$ of A (with the A_n's in $\mathcal{A}$), countable subadditivity gives

$$\nu(A) \leq \nu(\cup_1^\infty A_n) \leq \textstyle\sum_1^\infty \nu(A_n) = \sum_1^\infty \mu(A_n),$$

since $\mu = \nu$ on $\mathcal{A}$. Thus (recall the definition of μ^* in (1))

(h) $\quad \nu(A) \leq \mu^*(A) \qquad$ for all $A \in \mathcal{A}^*$.

Note that the measures μ^* and ν on $\mathcal{A}^*$ also satisfy

(i) $\quad \nu(A) + \nu(A^c) = \nu(\Omega) = \mu^*(\Omega) = \mu^*(A) + \mu^*(A^c)$

for all A in $\mathcal{A}^*$ (using $\Omega \in \mathcal{A}$ for $\nu(\Omega) = \mu^*(\Omega)$). Since (h) gives both

(j) $\quad \nu(A) \leq \mu^*(A) \qquad$ and $\qquad \nu(A^c) \leq \mu^*(A^c)$

(where all four of these terms are finite), we can infer from (i) that

(k) $\quad \nu(A) = \mu^*(A) \qquad$ for all $A \in \mathcal{A}^*$.

This gives the uniqueness of μ^* on $\mathcal{A}^*$.

Claim 8: Uniqueness of μ^* on $\mathcal{A}^*$ also holds when μ is a σ-finite measure on $\mathcal{A}$.
Label the sets of the measurable partition as D_n, and let $\Omega_n \equiv \sum_1^n D_k$ so that $\Omega_n \nearrow \Omega$. Claim 7 establishes that

(l) $\quad \nu(A\Omega_n) = \mu^*(A\Omega_n) \qquad$ for all $A \in \mathcal{A}^*$.

It follows that

(m) $\quad \nu(A) = \lim_n \nu(A\Omega_n) \qquad$ by proposition 1.1.2

$\qquad = \lim_n \mu^*(A\Omega_n) \qquad$ by (l)

(n) $\quad = \mu^*(A) \qquad$ by proposition 1.1.2,

completing the proof. In fact, the following corollary was established. □

Corollary 1 The μ^*-measurable sets $\mathcal{A}^*$ of (1.1.8) were shown to contain $\sigma[\mathcal{C}]$. Moreover, the measure μ on the field $\mathcal{C}$ was in fact extended to $\mathcal{A}^*$ in the proof above. Also, for μ a σ-finite measure on $\mathcal{C}$, the extension was shown to be unique on $\mathcal{A}^*$ and also to be σ-finite.

Question We extended our measure μ from the field $\mathcal{C}$ to a collection $\mathcal{A}^*$ that is at least as big as the σ-field $\sigma[\mathcal{C}]$. Have we actually gone beyond $\sigma[\mathcal{C}]$? Can we go further? Corollary 2 will show that we can always "complete" such a measure (which may or may not extend it), but definite limitations to extension will be implied by proposition 2.1. The most famous example is the "Lebesgue sets" of proposition 2.3.

Definition 2.2 (Complete measures) Let $(\Omega, \mathcal{A}, \mu)$ denote a measure space. If $\mu(A) = 0$, then A is called a *null set*. We call $(\Omega, \mathcal{A}, \mu)$ *complete* if whenever we have $A \subset$ (some B) $\in \mathcal{A}$ with $\mu(B) = 0$, we necessarily also have $A \in \mathcal{A}$. [That is, all subsets of sets of measure 0 are required to be measurable.]

Exercise 2.1 (Completion) Let $(\Omega, \mathcal{A}, \mu)$ denote a measure space. Show that

$$\hat{\mathcal{A}}_\mu \equiv \{A : A_1 \subset A \subset A_2 \text{ with } A_1, A_2 \in \mathcal{A} \text{ and } \mu(A_2 \setminus A_1) = 0\} \tag{4}$$

$$= \{A \cup N : A \in \mathcal{A}, \text{ and } N \subset (\text{some } B) \in \mathcal{A} \text{ having } \mu(B) = 0\} \tag{5}$$

$$= \{A \triangle N : A \in \mathcal{A}, \text{ and } N \subset (\text{some } B) \in \mathcal{A} \text{ having } \mu(B) = 0\}, \tag{6}$$

and that $\hat{\mathcal{A}}_\mu$ is a σ-field. Define $\hat{\mu}$ on $\hat{\mathcal{A}}_\mu$ by

$$\hat{\mu}(A \cup N) = \mu(A) \tag{7}$$

for all $A \in \mathcal{A}$ and for all $N \subset$ (some B) $\in \mathcal{A}$ having $\mu(B) = 0$. Show that $(\Omega, \hat{\mathcal{A}}_\mu, \hat{\mu})$ is a complete measure space for which $\hat{\mu}|\mathcal{A} = \mu$. [Note: A proof must include a demonstration that definition (7) leads to a well-defined $\hat{\mu}$. That is, whenever $A_1 \cup N_1 = A_2 \cup N_2$ we must have $\mu(A_1) = \mu(A_2)$, so that $\hat{\mu}(A_1 \cup N_1) = \hat{\mu}(A_2 \cup N_2)$.]

Definition 2.3 (Lebesgue sets) The completion of Lebesgue measure on $(R, \mathcal{B}, \lambda)$ is still called Lebesgue measure. The resulting completed σ-field $\hat{\mathcal{B}}_\lambda$ of the Borel sets $\mathcal{B}$ is called the *Lebesgue sets*.

Corollary 2 When we complete a measure μ on a σ-field $\mathcal{A}$, this completed measure $\hat{\mu}$ is the unique extension of μ to $\hat{\mathcal{A}}_\mu$. [It is typical to denote the extension by μ also (rather than $\hat{\mu}$).]

Corollary 3 (Thus when we begin with a σ-finite measure μ on a field $\mathcal{C}$, both the extension to $\mathcal{A} \equiv \sigma[\mathcal{C}]$ and the further extension to $\hat{\mathcal{A}}_\mu \equiv \hat{\sigma}[\mathcal{C}]_\mu$ are unique.) Here, we note that all sets in $\hat{\mathcal{A}}_\mu = \hat{\sigma}[\mathcal{C}]_\mu$ are in the class $\mathcal{A}^*$ of μ^*-measurable sets.

Proof. Consider corollary 2 first. Let ν denote any extension to $\hat{\mathcal{A}}_\mu$. We will demonstrate that

$$\nu(A \cup N) = \mu(A) \qquad \text{for all } A \in \mathcal{A}, \text{ and all null sets } N \tag{a}$$

(that is, $\nu = \hat{\mu}$). Assume not. Then there exist sets $A \in \mathcal{A}$ and $N \subset$ (some B) in $\mathcal{A}$ with $\mu(B) = 0$ such that $\nu(A \cup N) > \mu(A)$ [necessarily, $\nu(A \cup N) \geq \nu(A) = \mu(A)$]. For this A and N we have

$$\mu(A) = \nu(A) < \nu(A \cup N) = \nu(A \cup (A^c N)) \qquad \text{where } A^c N \subset A^c B = (\text{null})$$

$$= \nu(A) + \nu(A^c N) \leq \nu(A) + \nu(B) \tag{b}$$

since ν is a measure on the completion

$$= \mu(A) + \mu(B) \qquad \text{since } \nu \text{ is an extension of } \mu. \tag{c}$$

Hence $\mu(B) > 0$, which is a contradiction. Thus the extension is unique.

We now turn to corollary 3. Only the final claim needs demonstrating. Suppose A is in $\hat{\sigma}[\mathcal{C}]_\mu$. Then $A = A' \cup N$ for some $A' \in \mathcal{A}$ and some N satisfying $N \subset B$ with $\mu(B) = 0$. Since $\mathcal{A}^*$ is a σ-field, it suffices to show that any such N is in $\mathcal{A}^*$. Since μ^* is subadditive and monotone, we have

(d) $$\mu^*(T) \le \mu^*(TN) + \mu^*(TN^c) = \mu^*(TN^c) \le \mu^*(T),$$

because $\mu^*(TN) = 0$ follows from using $B, \emptyset\, \emptyset, \ldots$ to cover TN. Thus equality holds in this last equation, showing that N is μ^*-measurable. □

Exercise 2.2 Let μ and ν be finite measures on $(\Omega, \mathcal{A})$.
(a) Show by example that $\hat{\mathcal{A}}_\mu$ and $\hat{\mathcal{A}}_\nu$ need not be equal.
(b) Prove or disprove each half: $\hat{\mathcal{A}}_\mu = \hat{\mathcal{A}}_\nu$ iff μ and ν have identical null sets.
(c) Give an example of an LS-measure μ on R (see section 1.3) for which $\hat{\mathcal{B}}_\mu = 2^R$.

Exercise 2.3 (Approximation lemma; Halmos) Let the σ-finite measure μ on the field $\mathcal{C}$ be extended to $\mathcal{A} = \sigma[\mathcal{C}]$, and also refer to the extension as μ.
(a) Show that for each $A \in \mathcal{A}$ (or $\hat{\mathcal{A}}_\mu$) having $\mu(A) < \infty$, and for each $\epsilon > 0$,

(8) $$\mu(A \triangle C) < \epsilon \qquad \text{for some set } C \in \mathcal{C}.$$

[Hint. Truncate the sum in (1.2.1) to define C.]
(b) Let μ denote counting measure on the integers. Then $\mathcal{C} \equiv \{C : C \text{ or } C^c \text{ is finite}\}$ is a field. Determine $\sigma[\mathcal{C}]$. Show that (8) fails for the set A of even integers. Show that this $(\Omega, \mathcal{A}, P)$ is σ-finite, and this μ is a Lebesgue–Stieltjes measure (as below).

Definition 2.4 (Regular measures on metric spaces) Let d denote a metric on Ω, let $\mathcal{A}$ denote the Borel sets, and let μ be a measure on $(\Omega, \mathcal{A})$. Suppose that for each set A in $\hat{\mathcal{A}}_\mu$, and for every $\epsilon > 0$, one can find an open set O_ϵ and a closed set C_ϵ for which both $C_\epsilon \subset A \subset O_\epsilon$ and $\mu(O_\epsilon \setminus C_\epsilon) < \epsilon$. Suppose also that if $\mu(A) < \infty$, one then requires that the set C_ϵ be compact. Then μ is called a *regular measure*. [Note exercise 1.3.1 below. Contrast its content with (8).]

Exercise 2.4 (Nonmeasurable sets) Let Ω consist of the sixteen values $1, \ldots, 16$. (Think of them arranged in four rows of four values.) Let

$$C_1 = \{1,2,3,4,5,6,7,8\}, \qquad C_2 = \{9,10,11,12,13,14,15,16\},$$
$$C_3 = \{1,2,5,6,9,10,13,14\}, \quad C_4 = \{3,4,7,8,11,12,15,16\}.$$

Let $\mathcal{C}$ denote the field generated by $\{C_1, C_2, C_3, C_4\}$, and let $\mathcal{A} = \sigma[\mathcal{C}]$.
(a) Show that $\mathcal{A} \equiv \sigma[\mathcal{C}] \ne 2^\Omega$. (Note that 2^Ω contains $2^{16} = 65{,}536$ sets.)
(b) Let $\mu(C_i) = \frac{1}{2}$, $1 \le i \le 4$, with $\mu(C_1C_3) = \frac{1}{4}$. Show $\hat{\mathcal{A}}_\mu = \mathcal{A}$, with $2^4 = 16$ sets.
(c) Let $\mu(C_i) = \frac{1}{2}$, $i = 2,3,4$, with $\mu(C_2C_4) = 0$. Show that $\hat{\mathcal{A}}_\mu$ has $2^{10} = 1024$ sets.
(d) Illustrate proposition 2.1 below in the context of this exercise.

Proposition 2.1 (Not all sets need be measurable) Let μ be a measure on $\mathcal{A} \equiv \sigma[\mathcal{C}]$, with $\mathcal{C}$ a field. If $B \notin \hat{\mathcal{A}}_\mu$, then there are infinitely many measures on $\sigma[\hat{\mathcal{A}}_\mu \cup \{B\}]$ that agree with μ on $\mathcal{C}$. [Thus the σ-field $\hat{\mathcal{A}}_\mu$ is as far as we can go with the unique extension process.] (We merely state this observation for reference, without proof.) [To exhibit a subset of R not in the Borel sets $\mathcal{B}$ requires the axiom of choice.]

Proposition 2.2 (Not all subsets are Lebesgue sets) There is a subset D of R that is not in the Lebesgue sets $\hat{\mathcal{B}}_\lambda$.

Proof. Define the equivalence relation $\sim$ on elements of $[0,1)$ by $x \sim y$ if $x-y$ is a rational number. Use the axiom of choice to specify a set D that contains exactly one element from each equivalence class. Now define $D_z \equiv \{z + x \text{ (modulo 1)} : x \in D\}$ for each rational z in $[0,1)$, so that $[0,1) = \sum_z D_z$ represents $[0,1)$ as a countable union of disjoint sets. Moreover, all D_z must have the same outer measure; call it a. Assume $D = D_0$ is measurable. But then $1 = \lambda([0,1)) = \sum_z \lambda(D_z) = \sum_z a$ gives only $\sum_z a = 0$ (when $a = 0$) and $\sum_z a = \infty$ (when $a > 0$) as possibilities. This is a contradiction. Thus $D \notin \hat{\mathcal{B}}_\lambda$. □

Exercise 2.5 Write out the details that $D = \sum_z D_z$ (with disjoint sets D_z for the rationals z) in the above proof.

Proposition 2.3 (Not all Lebesgue sets are Borel sets) There necessarily exists a set $A \in \hat{\mathcal{B}}_\lambda \setminus \mathcal{B}$ that is a Lebesgue set but not a Borel set.

Proof. This proof follows exercise 6.3.3 below; it requires the axiom of choice. □

Exercise 2.6 Every subset A of Ω having $\mu^*(A) = 0$ is a μ^*-measurable set.

Exercise 2.7* Show that the Carathéodory theorem can fail if μ is not σ-finite.

Coverings

Earlier in this section we encountered Carathéodory coverings.

Exercise 2.8*(Vitali covering) (a) We say that a family $\mathcal{V}$ of intervals I is a *Vitali covering* of a set D if for each $x \in D$ and each $\epsilon > 0$ there exists an interval $I \in \mathcal{V}$ for which $x \in I$ and $\lambda(I) < \epsilon$.
(b) (Vitali covering theorem) Let $D \subset R$ have outer Lebesgue measure $\lambda^*(D) < \infty$. Let $\mathcal{V}$ be a collection of closed intervals that forms a Vitali covering of D. Then there exists a finite number of pairwise disjoint intervals $(I_1, \ldots, I_m)$ in $\mathcal{V}$ whose Lebesgue outer measure λ^* satisfies

(9) $\lambda^*(D \setminus \sum_{j=1}^m I_j) < \epsilon.$

(Compare this "nice approximation" of a set to the nice approximations given in exercise 2.3 and in definition 2.4.) [Lebesgue measure λ will be formally shown to exist in the next section, and λ^* will be discussed more fully.] [Result (9) will be useful in establishing the Lebesgue result that increasing functions on R necessarily have a derivative, except perhaps on a set having Lebesgue measure zero.]

Exercise 2.9*(Heine–Borel) If $\{U_t : t \in T\}$ is an arbitrary collection of open sets that covers a compact subset D of R, then there exists a finite number of them $U_1, \ldots, U_m$ that also covers D.

The familiar Heine–Borel result will be frequently used. It is stated here only to contrast it with the important new ideas of Carathéodory and Vitali coverings.

3 Lebesgue–Stieltjes Measures

At the moment we know only a few measures informally. We now construct the large class of measures that lies at the heart of probability theory.

Definition 3.1 (Lebesgue–Stieltjes measure) A measure μ on the real line R assigning finite values to finite intervals is called a *Lebesgue–Stieltjes measure*. [The measure μ on $(R, 2^R)$ whose value $\mu(A)$ for any set A equals the number of rationals in A is *not* a Lebesgue–Stieltjes measure.]

Definition 3.2 (gdf) A finite $\nearrow$ function F on R that is right-continuous is called a *generalized df* (to be abbreviated *gdf*). Then $F_-(\cdot) \equiv \lim_{y \nearrow \cdot} F(y)$ denotes the left-continuous version of F. The *mass function* of F is defined by

$$\Delta F(\cdot) \equiv F(\cdot) - F_-(\cdot), \qquad \text{while} \qquad F(a,b] \equiv F(b) - F(a) \quad \text{for all } a \le b$$

is called the *increment function* of F. Identify gdfs having the same increment function. Only one member F of each equivalence class so obtained satisfies $F_-(0) = 0$, and this F can (and occasionally will) be used as the *representative member of the class* (also to be called the *representative gdf*).

Example 3.1 We earlier defined three measures on $(R, \mathcal{B})$ informally.
(a) For Lebesgue measure λ, a gdf is the identity function $F(x) = x$.
(b) For counting measure, a gdf is the greatest integer function $F(x) = [x]$.
(c) For unit point mass at x_0, a gdf is $F(x) = 1_{[x_0,\infty)}(x)$. □

Theorem 3.1 (Correspondence theorem; Loève) The relationship

(1) $\qquad \mu((a,b]) \equiv F(a,b] \qquad \text{for all } -\infty \le a \le b \le +\infty$

establishes a 1-to-1 correspondence between the Lebesgue–Stieltjes measures μ on $\mathcal{B}$ and the set of representative members of the equivalence classes of generalized dfs. [Each such μ extends uniquely to $\hat{\mathcal{B}}_\mu$.]

Notation 3.1 We formally establish some notation that will be used throughout. Important classes of sets include:

(2) $\qquad \mathcal{C}_I \equiv \{\text{all intervals } (a,b], (-\infty,b], \text{ or } (a,+\infty) : -\infty < a < b < +\infty\}.$

(3) $\qquad \mathcal{C}_F \equiv \{\text{all finite disjoint unions of intervals in } \mathcal{C}_I\} = \text{ (a field) }.$

(4) $\qquad \mathcal{B} \equiv \sigma[\mathcal{C}_F] \equiv \text{ (the } \sigma\text{-field of Borel sets)}.$

(5) $\qquad \hat{\mathcal{B}}_\mu \equiv \text{ (the } \sigma\text{-field } \mathcal{B} \text{ completed for the measure } \mu).$

(6) $\qquad \bar{\mathcal{B}} \equiv \sigma[\mathcal{B}, \{-\infty\}, \{+\infty\}].$ □

Proof. Given an LS-measure μ, define the increment function $F(a,b]$ via (1). We clearly have $0 \le F(a,b] < \infty$ for all finite a,b, and $F(a,b] \to 0$ as $b \searrow a$, by proposition 1.1.2. Now specify $F_-(0) \equiv 0, F(0) \equiv \mu(\{0\}), F(b) \equiv F(0) + F(0,b]$ for $b > 0$, and $F(a) = F(0) - F(a,0]$ for $a < 0$. This $F(\cdot)$ is the representative gdf.

Given a representative gdf, we define μ on the collection $\mathcal{I}$ of all finite intervals $(a,b]$ via (1). We will now show that μ is a well-defined and c.a. measure on this collection $\mathcal{I}$ of finite intervals.
Nonnegative: $\mu \ge 0$ for any $(a,b]$, since F is $\nearrow$.
Null: $\mu(\emptyset) = 0$, since $\emptyset = (a,a]$ and $F(a,a] = 0$.
Countably additive on $\mathcal{I}$: Let $I \equiv (a,b] = \sum_1^\infty I_n \equiv \sum_1^\infty (a_n, b_n]$. We must show that $\mu(\sum_1^\infty I_n) = \sum_1^\infty \mu(I_n)$.
First, we will show that $\sum_1^\infty \mu(I_k) \le \mu(I)$. Fix n. Then $\sum_1^n I_k \subset I$, so that (relabel if necessary, so that $I_1, \ldots, I_n$ is a left-to-right ordering of these intervals)

(a) $\quad \sum_1^n \mu(I_k) = \sum_1^n F(a_k, b_k] \le F(a,b] = \mu(I).$

Letting $n \to \infty$ in (a) gives the first claim.
Next, we will show that $\mu(I) \le \sum_1^\infty \mu(I_k)$. Suppose $b - a > \epsilon > 0$ (the case $b - a = 0$ is trivial, as $\mu(\emptyset) = 0$). Fix $\theta > 0$. For each $k \ge 1$, use the right continuity of F to choose an $\epsilon_k > 0$ so small that

(b) $\quad F(b_k, b_k + \epsilon_k] < \theta/2^k, \qquad$ and define $J_k \equiv (a_k, c_k) \equiv (a_k, b_k + \epsilon_k)$.

These J_k form an open cover of the compact interval $[a+\epsilon, b]$, so that some finite number of them are known to cover $[a+\epsilon, b]$, by the Heine–Borel theorem. Sorting through these intervals one at a time, choose (a_1, c_1) to contain b, choose (a_2, c_2) to contain a_1, choose (a_3, c_3) to contain $a_2, \ldots$; finally (for some K), choose (a_K, c_K) to contain $a + \epsilon$. Then (relabeling the subscripts, if necessary)

$$F(a+\epsilon, b] \le F(a_K, c_1] \le \sum_1^K F(a_k, c_k] \le \sum_1^K F(a_k, b_k] + \sum_1^K \theta/2^k$$

(c) $$\le \sum_1^\infty \mu(I_k) + \theta.$$

Let $\theta \searrow 0$ and then $\epsilon \searrow 0$ in (c) to obtain the second claim as

(d) $\quad \mu(I) = F(a,b] \le \sum_1^\infty \mu(I_k).$

We will now show that μ is a well-defined c.a. measure on the given field $\mathcal{C}_F$. If $A = \sum_n I_n \in \mathcal{C}_F$ with each I_n of type $(a,b]$, then we define $\mu(A) \equiv \sum_n \mu(I_n)$. If we also have $A = \sum_m I'_m$, then we must show (where the subscripts m and n could take on either a finite or a countably infinite number of values) that

(e) $\quad \sum_m \mu(I'_m) = \sum_n \mu(I_n) = \mu(A).$

Now, $I'_m = A \cap I'_m = \sum_n I_n I'_m$ and $I_n = A I_n = \sum_m I'_m I_n$, so μ is well defined by

(f) $\quad \sum_m \mu(I'_m) = \sum_m \sum_n \mu(I_n I'_m) = \sum_n \sum_m \mu(I_n I'_m) = \sum_n \mu(I_n) = \mu(A).$

The c.a. of μ on $\mathcal{C}_F$ is then trivial; if disjoint $A_n = \sum_m I_{nm}$ for each n, it then follows that $A \equiv \sum_n A_n = \sum_n \sum_m I_{nm}$ with $\mu(A) = \sum_n \sum_m \mu(I_{mn}) = \sum_n \mu(A_n)$.

Finally, a measure μ on $\mathcal{C}_F$ determines a unique measure on $\mathcal{B}$, as is guaranteed by the Carathéodory extension of theorem 1.2.1. □

Exercise 3.1 Show that all Lebesgue–Stieltjes measures on $(R, \mathcal{B})$ are regular measures (recall definition 1.2.4). (Use the open intervals J_n of the previous proof.)

Probability Measures, Probability Spaces, and DFs

Definition 3.3 (Probability distributions $P(\cdot)$ and dfs $F(\cdot)$)
(a) In probability theory we think of Ω as the set of all possible outcomes of some experiment, and we refer to it as the *sample space*. The individual points ω in Ω are referred to as the *elementary outcomes*. The measurable subsets A in the collection $\mathcal{A}$ are referred to as *events*. A measure of interest is now denoted by P; it is called a *probability measure*, and must satisfy $P(\Omega) = 1$. We refer to $P(A)$ as the *probability* of A, for each event A in $\hat{\mathcal{A}}_P$. The triple $(\Omega, \mathcal{A}, P)$ (or $(\Omega, \hat{\mathcal{A}}_P, \hat{P})$, if this is different) is referred to as a *probability space*.
(b) An $\nearrow$ right-continuous function F on R having $F(-\infty) \equiv \lim_{x\to-\infty} F(x) = 0$ and $F(+\infty) \equiv \lim_{x\to+\infty} F(x) = 1$ is called a *distribution function* (which we will abbreviate as *df*). [For probability measures, setting $F(-\infty) = 0$ is used to specify the representative df.]

Corollary 1 (The correspondence theorem for dfs) Defining $P(\cdot)$ on all intervals $(a, b]$ via $P((a, b]\,) \equiv F(b) - F(a)$ for all $-\infty \le a < b \le +\infty$ establishes a 1-to-1 correspondence between all probability distributions $P(\cdot)$ on $(R, \mathcal{B})$ and all dfs $F(\cdot)$ on R.

Exercise 3.2 Prove this simple corollary.

Chapter 2

Measurable Functions and Convergence

1 Mappings and σ-Fields

Notation 1.1 (Inverse images) Suppose X denotes a function mapping some set Ω into the extended real line $\bar{R} \equiv R \cup \{\pm\infty\}$; we denote this by $X : \Omega \to \bar{R}$. Let X^+ and X^- denote the *positive part* and the *negative part* of X, respectively:

$$(1) \qquad X^+(\omega) \equiv \begin{cases} X(\omega) & \text{if } X(\omega) \geq 0, \\ 0 & \text{else,} \end{cases}$$

$$(2) \qquad X^-(\omega) \equiv \begin{cases} -X(\omega) & \text{if } X(\omega) \leq 0, \\ 0 & \text{else.} \end{cases}$$

Note that

$$(3) \qquad X = X^+ - X^- \quad \text{and} \quad |X| = X^+ + X^- = X + 2\,X^- = 2\,X^+ - X\,.$$

We also use the following notation:

$$(4) \qquad [\,X = r\,] \equiv X^{-1}(r) \equiv \{\,\omega : X(\omega) = r\,\} \qquad \text{for all real } r,$$

$$(5) \qquad [\,X \in B\,] \equiv X^{-1}(B) \equiv \{\,\omega : X(\omega) \in B\,\} \qquad \text{for all Borel sets } B,$$

$$(6) \qquad X^{-1}(\mathcal{B}) \equiv \{\,X^{-1}(B) : B \in \mathcal{B}\,\}\,.$$

We call these the *inverse images* of r, B, and $\mathcal{B}$, respectively. We let

$$(7) \qquad \bar{\mathcal{B}} \equiv \sigma[\,\mathcal{B}, \{+\infty\}, \{-\infty\}]\,.$$

Inverse images are also well-defined when $X : \Omega \to \Omega'$ for arbitrary sets Ω and Ω'. □

For $A, B \in \Omega$ we define $A \triangle B \equiv AB^c \cup A^cB$ and $A \setminus B \equiv AB^c$. There is use for the notation

$$(8) \qquad \|X\| \equiv \sup_{\omega \in \Omega} |X(\omega)|\,,$$

and we will also reintroduce this *sup norm* in other contexts below.

Proposition 1.1 Let $X : \Omega \to \Omega'$ and $Y : \Omega' \to \Omega''$. Let T denote an arbitrary index set. Then for all $A, B, A_t \subset \Omega'$ we have

(9) $X^{-1}(B^c) = [\,X^{-1}(B)\,]^c,$ $\qquad X^{-1}(A \setminus B) = X^{-1}(A) \setminus X^{-1}(B),$

(10) $X^{-1}(\bigcup_{t\in T} A_t\,) = \bigcup_{t\in T} X^{-1}(A_t),$ $\qquad X^{-1}(\bigcap_{t\in T} A_t) = \bigcap_{t\in T} X^{-1}(A_t).$

For all sets $A \subset \Omega''$, the composition $Y \circ X$ satisfies

(11) $(Y \circ X)^{-1}(A) = X^{-1}(Y^{-1}(A)) = X^{-1} \circ Y^{-1}(A)\,.$

Proof. Trivial. □

Proposition 1.2 (Preservation of σ-fields) Let $X : \Omega \to \Omega'$. Then:

(12) $\mathcal{A} \equiv X^{-1}(\text{a } \sigma\text{-field } \mathcal{A}' \text{ of subsets of } \Omega'\,) = (\text{a } \sigma\text{-field of subsets of } \Omega\,)\,.$

(13) $X^{-1}(\sigma[\mathcal{C}'\,]\,) = \sigma[X^{-1}(\mathcal{C}'\,)]$ $\qquad$ for any collection $\mathcal{C}'$ of subsets of Ω'.

(14) $\mathcal{A}' \equiv \{A' : X^{-1}(A'\,) \in (\text{a specific } \sigma\text{-field } \mathcal{A} \text{ of subsets of } \Omega)\}$
$\quad = (\text{a } \sigma\text{-field of subsets of } \Omega'\,)\,.$

Proof. Now, (12) is trivial from proposition 1.1. Consider (14). Now:

(a) $A' \in \mathcal{A}'$ implies $X^{-1}(A'\,) \in \mathcal{A}$
implies $X^{-1}(A'^{c}\,) = [X^{-1}(A'\,)]^c \in \mathcal{A}$ $\qquad$ implies $A'^{c} \in \mathcal{A}'\,,$

(b) A'_n's $\in \mathcal{A}'$ implies $X^{-1}(A'_n)$'s $\in \mathcal{A}$
implies $X^{-1}(\bigcup_n A'_n) = \bigcup_n X^{-1}(A'_n) \in \mathcal{A}$ $\qquad$ implies $\bigcup_n A'_n \in \mathcal{A}'$.

This gives (14). Consider (13). Using (12) gives

(c) $X^{-1}(\sigma[\mathcal{C}'\,]\,) = (\text{a } \sigma\text{-field containing } X^{-1}(\mathcal{C}'\,)) \supset \sigma[X^{-1}(\mathcal{C}'\,)]\,.$

Then (14) shows that

(d) $\mathcal{A}' \equiv \{\,A' : X^{-1}(A'\,) \in \sigma[X^{-1}(\mathcal{C}'\,)]\,\} = (\text{a } \sigma\text{-field containing } \mathcal{C}'\,) \supset \sigma[\mathcal{C}'\,]\,,$

so that (using (d) for the second inclusion below)

(e) $X^{-1}(\sigma[\mathcal{C}'\,]\,) \subset X^{-1}(\mathcal{A}'\,) \subset \sigma[X^{-1}(\mathcal{C}'\,)]\,.$

Combining (c) and (e) gives (13). [We apply (13) below to obtain (2.2.6).] □

Roughly, using (12) we will restrict X so that $\mathcal{F}(X) \equiv X^{-1}(\bar{\mathcal{B}}\,) \subset \mathcal{A}$ for our original $(\Omega, \mathcal{A}, P)$, so that we can then "induce" a measure on $(\bar{R}, \bar{\mathcal{B}}\,)$. Or, (14) tells us that the collection $\mathcal{A}'$ is such that we can always induce a measure on $(\Omega', \mathcal{A}'\,)$. We do this in the next section. First, we generalize our definition of Borel sets to n dimensions.

Example 1.1 (Euclidean space) Let

$$R_n \equiv R \times \cdots \times R \equiv \{(r_1, \ldots, r_n) \;:\; \text{each } r_i \text{ is in } R\}\,.$$

Let U_n denote all open subsets of R_n, in the usual Euclidean metric. Then

(15) $\mathcal{B}_n \equiv \sigma[U_n]$ is called the class of *Borel sets* of R_n.

Following the usual notation, $B_1 \times \cdots \times B_n \equiv \{(b_1, \ldots, b_n) \;:\; b_1 \in B_1, \ldots, b_n \in B_n\}$. Now let

(16) $\prod_{i=1}^{n} \mathcal{B} \equiv \mathcal{B} \times \cdots \times \mathcal{B} \equiv \sigma[\{B_1 \times \cdots \times B_n \;:\; \text{all } B_i \text{ are in } \mathcal{B}\}]$.

Now consider

(17) $\sigma[\,\{(-\infty, r_1] \times \cdots \times (-\infty, r_n] \;:\; \text{all } r_i \text{ are in } R\}\,]$.

Note that these three σ-fields are equal. Just observe that each of these three classes generates the generators of the other two classes, and apply exercise 1.1.1. [Surely, we can define a generalization of area λ_2 on $(R_2, \mathcal{B}_2)$ by beginning with $\lambda_2(B_1 \times B_2) = \lambda(B_1) \times \lambda(B_2)$ for all B_1 and B_2 in $\mathcal{B}$, and then extending to all sets in $\mathcal{B}_2$. We will do this in theorem 5.1.1, and we will call it Lebesgue measure on two-dimensional Euclidean space.] □

2 Measurable Functions

We seek a large usable class of functions that is closed under passage to the limit. This is the fundamental property of the class of measurable functions. Propositions 2.2 and 2.3 below will show that the class of measurable functions is also closed under all of the standard mathematical operations. Thus, this class is sufficient for our needs.

Definition 2.1 (Simple functions, etc.) Let the measure space $(\Omega, \mathcal{A}, \mu)$ be given and fixed throughout our discussion. Consider the following classes of functions. The *indicator function* $1_A(\cdot)$ of the set $A \subset \Omega$ is defined by

(1) $$1_A(\omega) \equiv \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{else.} \end{cases}$$

A *simple function* is of the form

(2) $$X(\omega) \equiv \sum_{i=1}^{n} x_i \, 1_{A_i}(\omega) \qquad \text{for } \sum_{1}^{n} A_i = \Omega \ \text{ with all } \ A_i \in \mathcal{A}, \ \text{and} \ \ x_i \in R.$$

An *elementary function* is of the form

(3) $$X(\omega) \equiv \sum_{i=1}^{\infty} x_i \, 1_{A_i}(\omega) \qquad \text{for } \sum_{i=1}^{\infty} A_i = \Omega \ \text{ with all } \ A_i \in \mathcal{A}, \ \text{and} \ \ x_i \in \bar{R}.$$

Definition 2.2 (Measurability) Suppose that $X : \Omega \to \Omega'$, where $(\Omega, \mathcal{A})$ and $(\Omega', \mathcal{A}')$ are both measurable spaces. We then say that X is $\mathcal{A}'$-$\mathcal{A}$-*measurable* if $X^{-1}(\mathcal{A}') \subset \mathcal{A}$. We also denote this by writing either

(4) $$X : (\Omega, \mathcal{A}) \to (\Omega', \mathcal{A}') \qquad \text{or} \qquad X : (\Omega, \mathcal{A}, \mu) \to (\Omega', \mathcal{A}')$$

(or even $X : (\Omega, \mathcal{A}, \mu) \to (\Omega', \mathcal{A}', \mu')$ for the measure μ' "induced" on $(\Omega', \mathcal{A}')$ by the mapping X, as will soon be defined). In the special case $X : (\Omega, \mathcal{A}) \to (\bar{R}, \bar{\mathcal{B}})$, we simply call X *measurable*; and in this special case we let $\mathcal{F}(X) \equiv X^{-1}(\bar{\mathcal{B}})$ denote the *sub σ-field* of $\mathcal{A}$ *generated by* X.

Proposition 2.1 (Measurability criteria) Let $X : \Omega \to \bar{R}$. Suppose $\sigma[\mathcal{C}] = \bar{\mathcal{B}}$. Then measurability can be characterized by either of the following:

(5) X is measurable if and only if $X^{-1}(\mathcal{C}) \subset \mathcal{A}$.

(6) X is measurable if and only if $X^{-1}([-\infty, x]) \in \mathcal{A}$ for all $x \in \bar{R}$.

Note that we could replace $[-\infty, x]$ by any one of $[-\infty, x), [x, +\infty]$, or $(x, +\infty]$.

Proof. Consider (5). Let $X^{-1}(\mathcal{C}) \subset \mathcal{A}$. Then

(a) $X^{-1}(\bar{\mathcal{B}}) = X^{-1}(\sigma[\mathcal{C}]) = \sigma[X^{-1}(\mathcal{C})]$ by proposition 2.1.2

(b) $\subset \mathcal{A}$ since $X^{-1}(\mathcal{C}) \subset \mathcal{A}$, and $\mathcal{A}$ is a σ-field.

The other direction is trivial. Thus (5) holds. To demonstrate (6), we need to show that $\bar{\mathcal{B}}$ satisfies

(c) $\sigma[\,\{[-\infty,\, x] : x \in R\,\}\,] = \bar{\mathcal{B}}.$

Since $\mathcal{B} = \sigma[\mathcal{C}_I]$ for $\mathcal{C}_I$ as in (1.3.2) and since

(d) $(a,b] = [-\infty,b] \cap [-\infty,a]^c, \quad [-\infty,b) = \cup_1^\infty[-\infty, b-1/n],$

(e) $\{-\infty\} = \cap_n[-\infty,-n], \quad \{+\infty\} = \cap_n[-\infty,n]^c,$ etc.,

the equality (c) is obvious. The rest is trivial. □

Proposition 2.2 (Measurability of common functions) Let X, Y, and X_n's be measurable functions. Consider cX with $c > 0$, $-X$, $\inf X_n$, $\sup X_n$, $\liminf X_n$, $\limsup X_n$, $\lim X_n$ if it exists, X^2, $X \pm Y$ if it is well-defined, XY where $0 \cdot \infty \equiv 0$, X/Y if it is well-defined, X^+, X^-, $|X|$, $g(X)$ for continuous g, and the composite function $g(X)$ for all measurable functions g. All of these are measurable functions.

Proposition 2.3 (Measurability via simple functions)

(7) Simple and elementary functions are measurable.

(8) $X : \Omega \to \bar{R}$ is measurable if and only if X is the limit of a sequence of simple functions.

Moreover:

(9) If $X \geq 0$ is measurable, then X is the limit of a sequence of simple functions that are ≥ 0 and $\nearrow$.

[The X_n's and Z_n's that are defined in the proof below are important.]

Proof. The functions in proposition 2.2 are measurable, since:

(a) $[cX < x] = [X < x/c], \qquad [-X < x] = [X > -x].$

(b) $[\inf X_n < x] = \cup[X_n < x], \qquad \sup X_n = -\inf(-X_n).$

(c) $\liminf X_n = \sup_n(\inf_{k\geq n} X_k), \qquad \limsup X_n = -\liminf(-X_n).$

(d) $\lim X_n = \liminf X_n,$ provided that $\lim X_n(\omega)$ exists for all ω.

(e) $[X^2 < x] = [-\sqrt{x} < X < \sqrt{x}\,] = [X < \sqrt{x}\,] \cap [X \leq -\sqrt{x}\,]^c.$

Each of the sets where X or Y equals 0, ∞, or $-\infty$ is measurable; use this below.

(f) $[X > Y] = \cup_r\{X > r > Y : r \text{ is rational}\}$, so $[X > Y]$ is a measurable set.

So, $[X + Y > z] = [X > z - Y] \in \mathcal{A}$ since $z - Y$ is trivially measurable.

(Here $[X = \infty] \cap [Y = -\infty] = \emptyset$ is implied, as $X + Y$ is well defined. Etc., below.)

(g) $X - Y = X + (-Y)$ and $XY = [(X+Y)^2 - (X-Y)^2\,]/4.$

(h) $\quad X/Y = X \times (1/Y),$

since $[1/Y < x] = [Y > 1/x]$ for $x > 0$ in case $Y > 0$, and for general Y one can write $\frac{1}{Y} = \frac{1}{Y}\,1_{[Y>0]} - \frac{1}{-Y}\,1_{[Y<0]}$ with the two indicator functions measurable.

(i) $\quad X^+ = X \vee 0 \quad$ and $\quad X^- = (-X) \vee 0\,.$

For g measurable, $(g \circ X)^{-1}(\bar{\mathcal{B}}) = X^{-1}(g^{-1}(\bar{\mathcal{B}})) \subset X^{-1}(\bar{\mathcal{B}}) \subset \mathcal{A}$. Then continuous g are measurable, since

(j) $\quad g^{-1}(\mathcal{B}) = g^{-1}(\sigma[\text{ open sets }]) = \sigma[g^{-1}(\text{ open sets })] \subset \sigma[\text{ open sets }] \subset \bar{\mathcal{B}}\,,$

and both $g^{-1}(\{+\infty\}) = \emptyset \in \bar{\mathcal{B}}$ and $g^{-1}(\{-\infty\}) = \emptyset \in \bar{\mathcal{B}}$, where we now apply the result for measurable g.

We now prove proposition 2.3. Claim (7) is trivial. Consider (8). Define simple functions X_n by

(10) $$X_n \equiv \sum_{k=1}^{n2^n} \frac{k-1}{2^n} \times \left\{ 1_{[\,\frac{k-1}{2^n} \le X < \frac{k}{2^n}\,]} - 1_{[\,\frac{k-1}{2^n} \le -X < \frac{k}{2^n}\,]} \right\} + n \times \{1_{[\,X \ge n\,]} - 1_{[\,-X \ge n\,]}\}\,.$$

Since $|X_n(\omega) - X(\omega)| \le 2^{-n}$ for $|X(\omega)| < n$, we have

(k) $\quad X_n(\omega) \to X(\omega) \quad$ as $n \to \infty \quad$ for each $\omega \in \Omega$.

Also, the nested subdivisions $k/2^n$ cause X_n to satisfy

(l) $\quad X_n \nearrow \quad$ when $X \ge 0$.

We extend proposition 2.3 slightly by further observing that

(11) $\quad \|X_n - X\| \to 0 \quad$ as $n \to \infty, \qquad$ if X is bounded.

Also, the elementary functions

(12) $$Z_n \equiv \sum_{k=1}^{\infty} \frac{k-1}{2^n} \times \left\{ 1_{[\,\frac{k-1}{2^n} \le X < \frac{k}{2^n}\,]} - 1_{[\,\frac{k-1}{2^n} \le -X < \frac{k}{2^n}\,]} \right\} + \infty \times \{1_{[\,X=\infty\,]} - 1_{[\,X=-\infty\,]}\}$$

are always such that

(13) $\quad \|\,(Z_n - X) \times 1_{[-\infty<X<\infty\,]}\,\| \le 1/2^n \to 0 \quad$ as $n \to \infty$. □

Proposition 2.4 (The discontinuity set is measurable; Billingsley) If (M, d) and (M', d') are metric spaces and $\psi : M \to M'$ is any function (not necessarily a measurable function), then the *discontinuity set* of ψ defined by

(14) $\quad D_\psi \equiv \{\, x \in M : \psi \text{ is not continuous at } x\,\}$

is necessarily in the Borel σ-field $\mathcal{B}_d$ (that is, the σ-field generated by the d-open subsets of M).

Proof. Let

(a) $$A_{\epsilon,\delta} \equiv \{x \in M : \; d(x,y) < \delta, \; d(x,z) < \delta \; \text{ and } \; d'(\psi(y),\psi(z)) \geq \epsilon \; \text{ for distinct } y,z \in M\}.$$

Note that $A_{\epsilon,\delta}$ is an open set, since $\{u \in M : d(x,u) < \delta_0\} \subset A_{\epsilon,\delta}$ will necessarily occur if $\delta_0 \equiv \{\delta - [d(x,y) \vee d(x,z)]\}/2$; that is, the y and z that work for x also work for all u in M that are sufficiently close to x. (Note: The y that worked for x may have been x itself.) Then

(b) $$D_\psi = \bigcup_{i=1}^\infty \bigcap_{j=1}^\infty A_{\epsilon_i,\delta_j} \; \in \; \mathcal{B}_d,$$

where $\epsilon_1, \epsilon_2, \ldots$ and $\delta_1, \delta_2, \ldots$ both denote the positive rationals, since each $A_{\epsilon,\delta}$ is an open set. □

Induced Measures

Example 2.1 (Induced measures) We now turn to the "induced measure" previewed above. Suppose $X : (\Omega, \mathcal{A}, \mu) \to (\Omega', \mathcal{A}')$, so that X is $\mathcal{A}'$-$\mathcal{A}$-measurable. We define $\mu_X \equiv \mu'$ by

(15) $$\mu_X(A') \equiv \mu'(A') \equiv \mu(X^{-1}(A')) \qquad \text{for each } A' \in \mathcal{A}'.$$

Then $\mu_X \equiv \mu'$ is a measure on $(\Omega', \mathcal{A}')$, called the *induced measure*. This is true, since we verify that

(a) $$\mu'(\emptyset) = \mu(X^{-1}(\emptyset)) = \mu(\emptyset) = 0, \qquad \text{and}$$

$$\mu'(\textstyle\sum_1^\infty A'_n) = \mu(X^{-1}(\sum_1^\infty A'_n)) = \mu(\sum_1^\infty X^{-1}(A'_n))$$

(b) $$= \textstyle\sum_1^\infty \mu(X^{-1}(A'_n)) = \sum_1^\infty \mu'(A'_n).$$

Note also that

(c) $$\mu'(\Omega') = \mu(X^{-1}(\Omega')) = \mu(\Omega).$$

Thus if μ is a probability measure, then so is $\mu_X \equiv \mu'$. Note also that we could regard X as an $\mathcal{A}'$-$\mathcal{F}(X)$-measurable transformation from the measure space $(\Omega, \mathcal{F}(X), \mu)$ to $(\Omega', \mathcal{A}', \mu_X)$.

Suppose further that F is a generalized df on the real line R, and that $\mu_F(\cdot)$ is the associated measure on $(R, \mathcal{B})$ satisfying $\mu_F((a,b]) = F(b) - F(a)$ for all a and b (as was guaranteed by the correspondence theorem (theorem 1.3.1)). Thus $(R, \mathcal{B}, \mu_F)$ is a measure space. Define

(16) $$X(\omega) = \omega \qquad \text{for all } \omega \in R.$$

Then X is a measurable transformation from $(R, \mathcal{B}, \mu_F)$ to $(R, \mathcal{B})$ whose induced measure μ_X is equal to μ_F. Thus for any given df F we can always construct a measurable function X whose df is F. □

Exercise 2.1 Suppose $(\Omega, \mathcal{A}) = (R_2, \mathcal{B}_2)$, where $\mathcal{B}_2$ denotes the σ-field generated by all open subsets of the plane. Recall that this σ-field contains all sets $B \times R$ and $R \times B$ for all $B \in \mathcal{B}$; here $B_1 \times B_2 \equiv \{(r_1, r_2) : r_1 \in B_1, r_2 \in B_2\}$. Now define measurable transformations $X_1((r_1, r_2)) = r_1$ and $X_2(r_1, r_2)) = r_2$. Then define $Z_1 \equiv (X_1^2 + X_2^2)^{1/2}$ and $Z_2 \equiv \text{sign}(X_1 - X_2)$, where $\text{sign}(r)$ equals $1, 0, -1$ according as r is $> 0, = 0, < 0$. The exercise is to give geometric descriptions of the σ-fields $\mathcal{F}(Z_1)$, $\mathcal{F}(Z_2)$, and $\mathcal{F}(Z_1, Z_2)$. (Suppose X_1 and X_2 are iid Logistic(0,1).)

Proposition 2.5 (The form of an $\mathcal{F}(X)$-measurable function) Suppose that X is a measurable functions on $(\Omega, \mathcal{A})$ and that Y is $\mathcal{F}(X)$-measurable. Then there must exist a measurable function g on $(\bar{R}, \bar{\mathcal{B}})$ such that $Y = g(X)$.

Proof. [The approach of this proof is to consider indicator functions, simple functions, nonnegative functions, general functions. This approach will be used again and again. Learn it!] Suppose that $Y = 1_D$ for some set $D \in \mathcal{F}(X)$, so that Y is an indicator function that is $\mathcal{F}(X)$-measurable. Then we can rewrite Y as $Y = 1_D = 1_{X^{-1}(B)} = 1_B(X) \equiv g(X)$, for some $B \in \bar{\mathcal{B}}$, where $g(r) \equiv 1_B(r)$. Thus the proposition holds for indicator functions. It holds for simple functions, since when all $B_i \in \bar{\mathcal{B}}$,

$$Y = \textstyle\sum_1^m c_i\, 1_{D_i} = \sum_1^m c_i\, 1_{X^{-1}(B_i)} = \sum_1^m c_i\, 1_{B_i}(X) \equiv g(X).$$

Let $Y \geq 0$ be $\mathcal{F}(X)$-measurable. Then there do exist $\nearrow$simple $\mathcal{F}(X)$-measurable functions Y_n such that $Y \equiv \lim_n Y_n = \lim_n g_n(X)$ for the $\nearrow$simple $\bar{\mathcal{B}}$-measurable functions g_n. Now let $g = \lim g_n$, which is $\bar{\mathcal{B}}$-measurable, and note that $Y = g(X)$. For general $Y = Y^+ - Y^-$, use $g = g^+ - g^-$. $\square$

Exercise 2.2 (Measurability criterion) Let $\mathcal{C}$ denote a $\bar{\pi}$-system of subsets of Ω. Let $\mathcal{V}$ denote a vector space of functions (that is, $X + Y \in \mathcal{V}$ and $\alpha X \in \mathcal{V}$ for all $X, Y \in \mathcal{V}$ and all $\alpha \in R$).
(a) Suppose that:

(17) $\quad 1_C \in \mathcal{V} \quad$ for all $C \in \mathcal{C}$.

(18) $\quad$ If $A_n \nearrow A \quad$ with $1_{A_n} \in \mathcal{V}, \quad$ then $1_A \in \mathcal{V}$.

Show that $1_A \in \mathcal{V}$ for every $A \in \sigma[\mathcal{C}]$.
(b) It then follows trivially that every simple function

(19) $\quad X_n \equiv \sum_1^m \alpha_i\, 1_{A_i}\;$ is in $\mathcal{V}$;

here $m \geq 1$, all $\alpha_i \in R$, and $\sum_1^m A_i = \Omega$ with all $A_i \in \sigma[\mathcal{C}]$.
(c) Now suppose further that $X_n \nearrow X$ for X_n's as in (19) implies that $X \in \mathcal{V}$. Show that $\mathcal{V}$ contains all $\sigma[\mathcal{C}]$-measurable functions.

3 Convergence

Convergence Almost Everywhere

Definition 3.1 ($\to_{a.e.}$) Let $X_1, X_2, \ldots$ denote measurable functions on $(\Omega, \mathcal{A}, \mu)$ to $(\bar{R}, \bar{\mathcal{B}})$. Say that the sequence X_n *converges almost everywhere* to X (denoted by $X_n \to_{a.e.} X$ as $n \to \infty$) if for some $N \in \mathcal{A}$ for which $\mu(N) = 0$ we have $X_n(\omega) \to X(\omega)$ as $n \to \infty$ for all $\omega \notin N$. If for all $\omega \notin N$ the sequence $X_n(\omega)$ is a Cauchy sequence, then we say that the sequence X_n *mutually converges a.e.* and denote this by writing $X_n - X_m \to_{a.e.} 0$ as $m \wedge n \to \infty$.

Exercise 3.1 Let $X_1, X_2, \ldots$ be measurable functions from $(\Omega, \mathcal{A}, \mu)$ to $(\bar{R}, \bar{\mathcal{B}})$.
(a) If $X_n \to_{a.e.} X$, then $X = \tilde{X}$ a.e. for some measurable $\tilde{X}$.
(b) If $X_n \to_{a.e.} X$ and μ is complete, then X itself is measurable.

Proposition 3.1 A sequence of measurable functions X_n that are a.e. finite converges a.e. to a measurable function X that is a.e. finite if and only if these functions X_n converges mutually a.e. [Thus we can redefine such functions on null sets and make them everywhere finite and everywhere convergent.]

Proof. The union of the countable number of null sets on which finiteness or convergence fails is again a null set N. On N^c, the claim is just a property of the real numbers. □

Proposition 3.2 (The convergence and divergence sets are measurable) Consider the finite measurable functions $X, X_1, X_2, \ldots$ (perhaps they have been redefined on null sets to achieve this); thus, they are $\mathcal{B}$-$\mathcal{A}$-measurable. Then the convergence and mutual convergence sets are measurable. In fact, the *convergence set* is given by

$$(1) \qquad [X_n \to X] \equiv \bigcap_{k=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} \left[|X_m - X| < \frac{1}{k} \right] \in \mathcal{A},$$

and the *mutual convergence set* is given by

$$(2) \qquad [X_n - X_m \to 0] \equiv \bigcap_{k=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} \left[|X_m - X_n| < \frac{1}{k} \right] \in \mathcal{A}.$$

Proof. Just read the right-hand side of (1) as, for all $\epsilon \equiv 1/k > 0$ there exists an n such that for all $m \ge n$ we have $|X_m(\omega) - X(\omega)| < 1/k$. (Practice saying this until it makes sense.) □

Taking complements in (1) allows the *divergence set* to be expressed via

$$(3) \qquad [\,X_n \to X]^c = \bigcup_{k=1}^{\infty} \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} \left[|X_m - X| \ge \frac{1}{k} \right] \equiv \bigcup_{k=1}^{\infty} A_k \quad \text{with } A_k \nearrow \text{ in } k,$$

where

$$(4) \qquad A_k = \textstyle\bigcap_{n=1}^{\infty} D_{kn}, \ \text{ and the } \ D_{kn} \equiv \bigcup_{m=n}^{\infty} [\,|X_m - X| \ge 1/k\,] \ \text{ are } \searrow \text{ in } n.$$

Proposition 3.3 Consider finite measurable X_n's and a finite measurable X.
(i) We have

$$X_n \to_{a.e.} \text{(such an } X) \qquad \text{if and only if}$$

$$(5) \qquad \mu\big(\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} [\,|X_m - X_n| \geq \epsilon\,]\,\big) = 0\,, \quad \text{for all } \epsilon > 0.$$

[A *finite* limit X exists if and only if the Cauchy criterion holds; and we want to be able to check for the existence of a finite limit X without knowing its value.]
(ii) **(Most useful criterion for $\to_{a.e.}$)** When $\mu(\Omega) < \infty$, we have

$$X_n \to_{a.e.} \text{(such an } X) \qquad \text{if and only if}$$

$$(6) \qquad \mu\big(\bigcup_{m=n}^{\infty} [\,|X_m - X_n| \geq \epsilon\,]\,\big) \to 0, \quad \text{for all } \epsilon > 0, \qquad \text{if and only if}$$

$$(7) \qquad \mu\big([\,\max_{n \leq m \leq N} |X_m - X_n| \geq \epsilon\,]\,\big) \leq \epsilon \ \text{ for all } N \geq n \geq (\text{some } n_\epsilon), \text{ for all } \epsilon > 0.$$

Proof. This is immediate from (3), (4), and proposition 1.1.2. □

Remark 3.1 (Additional measurability for convergence and divergence) Suppose we still assume that $X_1, X_2, \ldots$ are finite measurable functions. Then the following sets are seen to be measurable:

$$[\omega : X_n(\omega) \to X(\omega) \in \bar{R}\,]^c = [\,\liminf X_n < \limsup X_n]$$

$$(8) \qquad = \bigcup_{\text{rational } r} [\,\liminf X_n < r < \limsup X_n\,] \in \mathcal{A},$$

$$(9) \qquad [\,\limsup X_n = +\infty\,] = \bigcap_{m=1}^{\infty} [\,\limsup X_n > m\,] \in \mathcal{A}.$$

These comments reflect the following fact: If $X_n(\omega)$ does not converge to a finite number, then there are several different possibilities; but these interesting events are all measurable. □

Convergence in Measure

Definition 3.2 ($\to_\mu$) A given sequence of measurable and a.e. finite functions $X_1, X_2, \ldots$ is said to *converge in measure* to the measurable function X taking values in $\bar{R}$ (to be denoted by $X_n \to_\mu X$ as $n \to \infty$) if

$$(10) \qquad \mu([\,|X_n - X| \geq \epsilon]\,) \to 0 \quad \text{as } n \to \infty, \quad \text{for all } \epsilon > 0.$$

[Such convergence implies that X must be finite a.s., as

$$[\,|X| = \infty] \subset \{\bigcup_{k=1}^{\infty} [\,|X_k| = \infty]\} \cup [\,|X_n - X| \geq \epsilon]$$

shows.] We say that these X_n *converge mutually in measure*, which we denote by writing $X_m - X_n \to_\mu 0$ as $m \wedge n \to \infty$, if $\mu([\,|X_m - X_n| \geq \epsilon]\,) \to 0$ as $m \wedge n \to \infty$, for each $\epsilon > 0$.

Proposition 3.4 (a) If $X_n \to_\mu X$ and $X_n \to_\mu \tilde{X}$, then $X = \tilde{X}$ a.e.
(b) On a complete measure space, $X = \tilde{X}$ on N^c, for a null set N.

Proof. For all $\epsilon > 0$

(a) $\mu(\,[\,|X - \tilde{X}| \ge 2\epsilon\,]\,) \le \mu(\,[\,|X_n - X| \ge \epsilon\,]\,) + \mu(\,[\,|X_n - \tilde{X}| \ge \epsilon\,]\,) \to 0,$

giving $\mu(\,[\,|X - \tilde{X}| \ge \epsilon]\,) = 0$ for all $\epsilon > 0$. Thus

(b) $\mu(\,[\,X \ne \tilde{X}\,]\,) = \mu(\,\bigcup_k [\,|X - \tilde{X}| \ge 1/k\,]\,) \le \sum_1^\infty \mu(\,|X - \tilde{X}| \ge 1/k\,) = \sum_1^\infty 0\,,$

as claimed. □

Exercise 3.2 (a) Show that in general $\to_\mu$ does not imply $\to_{a.e.}$.
(b) Give an example with $\mu(\Omega) = \infty$ where $\to_{a.e.}$ does not imply $\to_\mu$.

Theorem 3.1 (Relating $\to_\mu$ to $\to_{a.e.}$) Let X and $X_1, X_2, \ldots$ be measurable and finite a.e. functions. The following are true.

(11) $X_n \to_{a.e.}$ (such an X) if and only if $X_n - X_m \to_{a.e.} 0$.

(12) $X_n \to_\mu$ (such an X) if and only if $X_n - X_m \to_\mu 0$.

(13) Let $\mu(\Omega) < \infty$. Then $X_n \to_{a.e.}$ (such an X) implies $X_n \to_\mu X$.

(14) (Riesz) If $X_n \to_\mu X$, then for some n_k we have $X_{n_k} \to_{a.e.} X$. (See (16)).

(Reducing $\to_\mu$ to $\to_{a.e.}$ by going to subsequences) Suppose $\mu(\Omega) < \infty$. Then

(15) $X_n \to_\mu X$ if and only if
each subsequence n' has a further n'' on which $X_{n''} \to_{a.e.}$ (such an X).

Proof. Now, (11) is proposition 3.1, and (12) is exercise 3.3 below. Result (13) comes from the elementary observation that

(a) $\mu(\,[\,|X_n - X| \ge \epsilon\,]\,) \le \mu(\,\bigcup_{m=n}^\infty [\,|X_m - X| \ge \epsilon\,]\,) \to 0,$ by (6).

To prove (14), choose $n_k \uparrow$ such that

(b) $\mu(A_k) \equiv \mu(\,[\,|X_{n_k} - X| > 1/2^{k+1}\,]\,) < 1/2^{k+1},$

with $\mu(\,[\,|X_n - X| > 1/2^{k+1}\,]\,) < 1/2^{k+1}$ for all $n \ge n_k$. Now let

(c) $B_m \equiv \bigcup_{k=m}^\infty A_k,$ so that $\mu(B_m) \le \sum_{k=m}^\infty 2^{-(k+1)} \le 1/2^m.$

On $B_m^c = \bigcap_m^\infty A_k^c$ we have $|X_{n_k} - X| \le 1/2^{k+1}$ for all $k \ge m$, so that

(d) $|X_{n_k}(\omega) - X(\omega)| \le 1/2^{k+1} \to 0$ as $k \to \infty$, for each $\omega \in B_m^c$,

with $\mu(B_m) \le 1/2^m$. Since convergence occurs on each B_m^c, we have

(e) $X_{n_k}(\omega) \to X(\omega)$ as $k \to \infty$ for each $\omega \in C \equiv \bigcup_{m=1}^\infty B_m^c\,,$

where $B_m = \cup_{k=m}^{\infty} A_k$ is $\searrow$ with $(\cap_{m=1}^{\infty} B_m) \subset$ (every B_m). So

(f) $\quad \mu(C^c) = \mu(\bigcap_{m=1}^{\infty} B_m) \le \limsup \mu(B_m) \le \lim 1/2^m = 0,$

completing the proof of (14). {Recall from (d) that $X_{n_k} \to_{a.e.} X$ *with*

(16) $\quad \mu(\{\, |X_{n_k} - X| \le 1/2^{k+1} \text{ for all } k \ge m\}^c) \le 1/2^m \qquad$ for all $m \ge 1$, and so

(17) $\quad \mu(\{\, |X_{n_k} - X_{n_{k+1}}| \le 1/2^k \text{ for all } k \ge m\}^c) \le 1/2^m \qquad$ for all $m \ge 1$.

For exercise 3.3 below, replace X above by $X_{n_{k+1}}$ in (b), and prove a.s. convergence to some X on this subsequence. Then show that the whole sequence converges in measure to this X. Results (16) and (17) will prove useful.}

Consider the unproven half of (15). Suppose that every n' contains a further n'' as claimed (with a particular X). Assume that $X_n \to_\mu X$ fails. Then for some $\epsilon_o > 0$ and some n'

(g) $\quad \lim_{n'} \mu([\,|X_{n'} - X| > \epsilon_o\,]) = (\text{some } a_o) > 0.$

But we are given that some further subsequence n'' has $X_{n''} \to_{a.e.} X$, and thus $X_{n''} \to_\mu X$ by (13), using $\mu(\Omega) < \infty$. Thus

(h) $\quad \lim_{n''} \mu([\,|X_{n''} - X| > \epsilon_o\,]) = 0;$

but this is a contradiction of (f). □

Exercise 3.3 As in (12), show that $X_n \to_\mu X$ if and only if $X_m - X_n \to_\mu 0$. [Hint. Adapt the proof of (16).]

Exercise 3.4 (a) Suppose that $\mu(\Omega) < \infty$ and g is continuous a.e. μ_X (that is, g is continuous except perhaps on a set of μ_X measure 0). Then $X_n \to_\mu X$ implies that $g(X_n) \to_\mu g(X)$.
(b) Let g be uniformly continuous on the real line. Then $X_n \to_\mu X$ implies that $g(X_n) \to_\mu g(X)$. (Here, $\mu(\Omega) = \infty$ is allowed.)

Exercise 3.5 (a) **(Dini)** If $X_n : \Omega \to R$ are continuous, with Ω compact, and with $X_n(\omega) \searrow X(\omega)$ for each $\omega \in \Omega$, then X_n converges uniformly to X on Ω.
(b) In general, a uniform limit of bounded and continuous functions X_n is also bounded and continuous.

4 Probability, RVs, and Convergence in Law

Definition 4.1 (Random variable and df) (a) A *probability space* $(\Omega, \mathcal{A}, P)$ is just a measure space for which $P(\Omega) = 1$. Now, $X : (\Omega, \mathcal{A}, P) \to (R, \mathcal{B})$ will be called a *random variable* (to be abbreviated *rv*); thus it is a $\mathcal{B}$-$\mathcal{A}$-measurable function. If $X : (\Omega, \mathcal{A}, P) \to (\bar{R}, \bar{\mathcal{B}})$, then we will call X an *extended rv*.
(b) The *distribution function* (to be abbreviated *df*) of a rv is defined by

(1) $\qquad F_X(x) \equiv P(X \le x) \qquad \text{for all } -\infty < x < \infty.$

We recall that $F \equiv F_X$ satisfies

(2) $\qquad F$ is $\nearrow$ and right continuous, with $F(-\infty) = 0$ and $F(+\infty) = 1$.

We let C_F denote the *continuity set* of F that contains all points at which F is continuous. [That $F \nearrow$ is trivial, and the other three properties all follow from the monotone property of measure, since $(\infty, x] = \cap_{n=1}^{\infty}(-\infty, x + a_n]$ for every sequence $a_n \searrow 0$, $\cap_{n=1}^{\infty}(-\infty, -n] = \emptyset$, and $\cup_{n=1}^{\infty}(-\infty, n] = R$.]
(c) If F is $\nearrow$ and right continuous with $F(-\infty) \ge 0$ and $F(+\infty) \le 1$, then F will be called a *sub df*.
(d) The induced measure on $(R, \mathcal{B})$ (or $(\bar{R}, \bar{\mathcal{B}})$) will be denoted by P_X. It satisfies

(3) $\qquad P_X(B) = P(X^{-1}(B)) = P(X \in B) \qquad \text{for all } B \in \mathcal{B}$

(for all $B \in \bar{\mathcal{B}}$ if X is an extended rv). We call this the *induced distribution* of X. We use the notation $X \cong F$ to denote that the induced distribution $P_X(\cdot)$ of the rv X has df F.
(e) We say that rvs X_n (with dfs F_n) *converge in distribution* or *converge in law* to a rv X_0 (with df F_0) if

(4) $\qquad F_n(x) = P(X_n \le x) \to F_0(x) = P(X_0 \le x) \qquad \text{at each } x \in C_{F_0}.$

We abbreviate this by writing either $X_n \to_d X_0$, $F_n \to_d F_0$, or $\mathcal{L}(X_n) \to \mathcal{L}(X_0)$.

Notation 4.1 Suppose now that $\{X_n : n \ge 0\}$ are rvs on $(\Omega, \mathcal{A}, P)$. Then it is customary to write $X_n \to_p X_0$ (in place of $X_n \to_\mu X_0$) and $X_n \to_{a.s.} X_0$ (as well as $X_n \to_{a.e.} X_0$). The "p" is an abbreviation for *in probability*, and the "a.s." is an abbreviation for *almost surely*.
Anticipating the next chapter, we let $Eg(X)$ denote $\int g(X)\,d\mu$, or $\int g(X)\,dP$ when μ is a probability measure P. We say that X_n *converges to* X_0 *in rth mean* if $E|X_n - X_0|^r \to 0$. We denote this by writing $X_n \to_r X_0$ or $X_n \to_{\mathcal{L}_r} X_0$. □

Proposition 4.1 Suppose that the rvs $X \cong F$ and $X_n \cong F_n$ satisfy $X_n \to_p X$. Then $X_n \to_d X$. (Thus, $X_n \to_{a.s.} X$ implies that $X_n \to_d X$.)

Proof. (This result has limited importance. But the technique introduced here is useful; see exercise 4.1 below.) Now,

(a) $\qquad F_n(t) = P(X_n \le t) \le P(X \le t + \epsilon) + P(\,|X_n - X| \ge \epsilon)$

(b) $\qquad\qquad \le F(t + \epsilon) + \epsilon \qquad \text{for all } n \ge \text{some } n_\epsilon.$

Also,

$$F_n(t) = P(X_n \le t) \ge P(X \le t - \epsilon \text{ and } |X_n - X| \le \epsilon) \equiv P(AB)$$
$$\ge P(A) - P(B^c) = F(t-\epsilon) - P(\,|X_n - X| > \epsilon)$$
$$\ge F(t-\epsilon) - \epsilon \qquad \text{for } n \ge (\text{some } n'_\epsilon).$$

Thus for $n \ge (n_\epsilon \vee n'_\epsilon)$ we have

(c) $$F(t-\epsilon) - \epsilon \le \underline{\lim}\, F_n(t) \le \overline{\lim}\, F_n(t) \le F(t+\epsilon) + \epsilon.$$

If t is a continuity point of F, then letting $\epsilon \to 0$ in (c) gives $F_n(t) \to F(t)$. Thus $F_n \to_d F$. □

The following elementary result is extremely useful. Often, one knows that $X_n \to_d X$, but what one is really interested in is a slight variant of X_n, rather than X_n itself. The next result was designed for just such situations.

Definition 4.2 (Type) Two rvs X and Y are of the same *type* if $Y \cong aX + b$.

Theorem 4.1 (Slutsky) Suppose that $X_n \to_d X$, while the rvs $Y_n \to_p a$ and $Z_n \to_p b$ as $n \to \infty$ (here X_n, Y_n, and Z_n are defined on a common probability space, but X need not be). Then

(5) $$U_n \equiv Y_n \times X_n + Z_n \to_d aX + b \qquad \text{as } n \to \infty.$$

Exercise 4.1 Prove Slutsky's theorem. [Hint. Recall the proof of proposition 4.1. Then write $U_n = (Y_n - a)X_n + (Z_n - b) + aX_n + b$ where $Y_n - a \to_p 0$ and $Z_n - b \to_p 0$. Note also that $P(\,|X_n| > (\text{some } M_\epsilon)) < \epsilon$ for all $n \ge (\text{some } n_\epsilon)$.]

Exercise 4.2 Let c be a constant. Show that $X_n \to_d c$ if and only if $X_n \to_p c$.

Remark 4.1 Suppose $X_1, X_2, \ldots$ are independent rvs with a common df F. Then $X_n \to_d X_0$ for any rv X_0 having df F. However, there is no rv X for which X_n converges to X in the sense of $\to_{a.s.}$, $\to_p$, or $\to_r$. (Of course, we are assuming that X is not a *degenerate* rv (that is, that μ_F is not a unit point mass).) □

5 Discussion of Sub σ-Fields

Consider again a sequence of rvs $X_1, X_2, \ldots$ where each quantity X_n is a measurable transformation $X_n : (\Omega, \mathcal{A}, P) \to (R, \mathcal{B}, P_{X_n})$, and where P_{X_n} denotes the induced measure. Each rv X_n is $\mathcal{B}$-$\mathcal{F}(X_n)$-measurable, with $\mathcal{F}(X_n)$ a sub σ-field of $\mathcal{A}$. Even though the intersection of any number of σ-fields is a σ-field, the union of even two σ-fields need not be a σ-field. We thus define the *sub σ-field generated by* $X_1, \ldots, X_n$ as

(1) $\qquad \mathcal{F}(X_1, \ldots, X_n) \equiv \sigma[\bigcup_{k=1}^{n} \mathcal{F}(X_k)\,] = \mathsf{X}^{-1}(\mathcal{B}_n) \qquad$ for $\mathsf{X}_n \equiv (X_1, \ldots, X_n)'$,

where the equality will be shown in the elementary proposition 5.2.1 below.

Note that $\mathcal{F}(X_1, \ldots, X_n) \subset \mathcal{F}(X_1, \ldots, X_n, X_{n+1})$, so that these necessarily form an increasing sequence of σ-fields of $\mathcal{A}$. Also, define

(2) $\qquad \mathcal{F}(X_1, X_2, \ldots) \equiv \sigma[\bigcup_{k=1}^{\infty} \mathcal{F}(X_k)\,]$.

It is natural to say that such $\mathsf{X}_n = (X_1, \ldots, X_n)'$ are adapted to the $\mathcal{F}(X_1, \ldots, X_n)$. In fact, if $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots$ is any sequence of σ-fields for which $\mathcal{F}(X_1, \ldots, X_n) \subset \mathcal{F}_n$ for all n, then we say that the X_n's are *adapted* to the $\mathcal{F}_n$'s.

Think of $\mathcal{F}(X_1, \ldots, X_n)$ as the *amount of information* available at time n from $X_1, \ldots, X_n$; that is, you have available for inspection all of the probabilities

(3) $\qquad P((X_1, \ldots, X_n) \in B_n) = P((X_1, \ldots, X_n)^{-1}(B_n)) = P_{(X_1, \ldots, X_n)}(B_n)$,

for all Borel sets $B_n \in \mathcal{B}_n$. Rephrasing, you have available for inspection all of the probabilities

(4) $\qquad P(A)$, for all $A \in \mathcal{F}(X_1, \ldots, X_n)$.

At stage $n+1$ you have available $P(A)$ for all $A \in \mathcal{F}(X_1, \ldots, X_n, X_{n+1})$; that is, you have more information available. [Think of $\mathcal{F}_n \setminus \mathcal{F}(X_1, \ldots, X_n)$ as the amount of information available to you at time n that goes beyond the information available from $X_1, \ldots, X_n$; perhaps some of it comes from other rvs not yet mentioned, but it is available nonetheless.]

Suppose we are not given rvs, but rather (speaking informally now, based on your general feel for probability) we are given joint dfs $F_n(x_1, \ldots, x_n)$ that we think ought to suffice to construct probability measures on $(R_n, \mathcal{B}_n)$. In (2.2.16) we saw that for $n = 1$ we could just let $(\Omega, \mathcal{A}, \mu) = (R, \mathcal{B}, \mu_F)$ and use $X(\omega) = \omega$ to define a rv that carried the information in the df F. How do we define probability measures P_n on $(R_n, \mathcal{B}_n)$ so that the *coordinate rvs*

(5) $\qquad X_k(\omega_1, \ldots, \omega_n) = \omega_k \qquad$ for all $(\omega_1, \ldots, \omega_n) \in R_n$

satisfy

(6) $\qquad P_n(X_1 \le x_1, \ldots, X_n \le x_n) = F_n(x_1, \ldots, x_n) \qquad$ for all $(x_1, \ldots, x_n) \in R_n$,

and thus carry all the information in F_n? Chapter 5 will deal with this construction. But even now it is clear that for this to be possible, the F_n's will have to satisfy some kind of consistency condition as we go from step n to $n+1$. Moreover, the consistency problem should disappear if the resulting X_n's are "independent."

But we need more. We will let R_∞ denote all infinite sequences $\omega_1, \omega_2, \ldots$ for which each $\omega_i \in R$. Now, the construction of (5) and (6) will determine probabilities on the collection $\mathcal{B}_n \times \prod_{k=n+1}^{\infty} R$ of all subsets of R_∞ of the form

$$(7) \qquad \begin{aligned} &B_n \times \textstyle\prod_{k=n+1}^{\infty} R \\ &\quad \equiv \{(\omega_1, \ldots, \omega_n, \omega_{n+1}, \ldots) : (\omega_1, \ldots, \omega_n) \in B_n, \ \ \omega_k \in R \text{ for } k \geq n+1\}, \end{aligned}$$

with $B_n \in \mathcal{B}_n$. Each of these collections is a σ-field (which within this special probability space can be denoted by $\mathcal{F}(X_1, \ldots, X_n)$) in this overall probability space $(R_\infty, \mathcal{B}_\infty, P_\infty)$, for some appropriate $\mathcal{B}_\infty$. But what is an appropriate σ-field $\mathcal{B}_\infty$ for such a probability measure P_∞? At a minimum, $\mathcal{B}_\infty$ must contain

$$(8) \qquad \sigma[\textstyle\bigcup_{n=1}^{\infty} \{\mathcal{B}_n \times \prod_{k=n+1}^{\infty} R\}] = \sigma[\bigcup_{n=1}^{\infty} \mathcal{F}(X_1, \ldots, X_n)],$$

and indeed, this is what we will use for $\mathcal{B}_\infty$. Of course, we also want to construct the measure P_∞ on $(R_\infty, \mathcal{B}_\infty)$ in such a way that

$$(9) \qquad P_\infty\left(\textstyle\prod_{k=1}^{n} (-\infty, x_k] \times \prod_{k=n+1}^{\infty} R\right) = F_n(x_1, \ldots, x_n) \qquad \text{for all } n \geq 1$$

and for all $x_1, \ldots, x_n$ in R. The details are given in chapter 5.

Until chapter 5 we will assume that we are *given* the rvs $X_1, X_2, \ldots$ on some $(\Omega, \mathcal{A}, P)$, and we will need to deal only with the *known* quantities $\mathcal{F}(X_1, \ldots, X_n)$ and $\mathcal{F}(X_1, X_2, \ldots)$ defined in (1) and (2). This is probability theory: Given $(\Omega, \mathcal{A}, P)$, we study the behavior of rvs $X_1, X_2, \ldots$ that are defined on this space. Now contrast this with statistics: Given a physical situation producing measurements $X_1, X_2, \ldots$, we construct models $\{(R_\infty, \mathcal{B}_\infty, P_\infty^\theta) : \theta \in \Theta\}$ based on various plausible models for $F_n^\theta(x_1, \ldots, x_n)$, $\theta \in \Theta$, and we then use the data $X_1, X_2, \ldots$ and the laws of probability theory to decide which model $\theta_0 \in \Theta$ was most likely to have been correct and what action to take. In particular, the statistician must know that the models to be used are well-defined.

We also need to extend all this to uncountably many rvs $\{X_t : t \in T\}$, for some interval T such as $[a, b]$, or $[a, \infty)$, or $[a, \infty]$, or $(-\infty, \infty)$, ... We say that rvs $X_t : (\Omega, \mathcal{A}, P) \to (R, \mathcal{B})$ for $t \in T$ are *adapted* to an $\nearrow$ sequence of σ-fields $\mathcal{F}_t$ if $\mathcal{F}_s \subset \mathcal{F}_t$ for all $s \leq t$ with both $s, t \in T$ and if each X_t is $\mathcal{F}_t$-measurable. In this situation we typically let $R_T \equiv \prod_{t \in T} R_t$ and then let

$$(10) \qquad \mathcal{F}_t \equiv \mathcal{F}(X_s : s \leq t) \equiv \sigma[\textstyle\bigcup_s X_s^{-1}(\mathcal{B}) : s \leq t \text{ and } s \in T] \qquad \text{for all } t \in T.$$

This is also done in chapter 5 (where more general sets T are, in fact, considered).

The purpose in presenting this section here is to let the reader start now to become familiar and comfortable with these ideas before we meet them again in chapter 5 in a more substantial and rigorous presentation. (The author assigns this as reading at this point and presents only a very limited amount of chapter 5 in his lectures.)

Exercise 5.1 (a) Show that the class $\mathcal{C} \equiv \{X_1^{-1}(B_1) \cap \{X_2^{-1}(B_2) : B_1, B_2 \in \mathcal{B}\}$ is a $\bar{\pi}$-system that generates the σ-field $\mathcal{F}(X_1, X_2)$.
(b) Recall the Dynkin π–λ theorem, and state its implications in this context.
(c) State an extension of this part (a) to $\mathcal{F}(X_1, \ldots, X_n)$ and to $\mathcal{F}(X_1, X_2, \ldots)$.

Chapter 3

Integration

1 The Lebesgue Integral

Let $(\Omega, \mathcal{A}, \mu)$ be a fixed measure space and let $X, Y, X_n, \ldots$ denote measurable functions from $(\Omega, \mathcal{A}, \mu)$ to $(\bar{R}, \bar{\mathcal{B}})$. If $\Omega = \sum_1^n A_i$ where $A_1, \ldots, A_n$ are in $\mathcal{A}$, then $A_1, \ldots, A_n$ is called a *partition* (or *measurable partition*) of Ω.

Definition 1.1 (Lebesgue integral $\int X\,d\mu$ or $\int X$) If $X = \sum_{i=1}^n x_i 1_{A_i} \geq 0$ is a simple function (where all $x_i \geq 0$ and $A_1, \ldots, A_n$ is a partition of Ω), then

$$(1) \qquad \int X\,d\mu \equiv \sum_{i=1}^n x_i\,\mu(A_i).$$

(We must verify that it is well defined.) If $X \geq 0$, then

$$(2) \qquad \int X\,d\mu \equiv \sup\left\{\int Y\,d\mu : 0 \leq Y \leq X \text{ and } Y \text{ is such a simple function}\right\}.$$

[Of course, we must show that the value of $\int X\,d\mu$ in (1) is independent of the representation of X that is specified.] For general measurable X,

$$(3) \qquad \int X\,d\mu \equiv \int X^+\,d\mu - \int X^-\,d\mu,$$

provided that at least one of $\int X^+\,d\mu$ and $\int X^-\,d\mu$ is finite. We let

$$(4) \qquad \begin{aligned} \mathcal{L}_1 &\equiv \mathcal{L}_1(\Omega, \mathcal{A}, \mu) \equiv \{X : \textstyle\int |X|\,d\mu < \infty\}, \\ \mathcal{L}_1^+ &\equiv \mathcal{L}_1^+(\Omega, \mathcal{A}, \mu) \equiv \{X \in \mathcal{L}_1 : X \geq 0\}, \\ \mathcal{L}_r &\equiv \mathcal{L}_r(\Omega, \mathcal{A}, \mu) \equiv \mathcal{L}_r(\mu) \equiv \{X : \textstyle\int |X|^r\,d\mu < \infty\}, \quad \text{for each } r > 0; \end{aligned}$$

in each of these definitions we agree to identify X and X' whenever $X = X'$ a.e. μ. If X (which is not measurable) equals a measurable function Y on a set A having $\mu(A^c) = 0$, then $\int X\,d\mu \equiv \int Y\,d\mu$. [Clearly, $\int X\,d\mu$ is not affected by such Y and A.]

If X is measurable and $\int X\,d\mu$ is finite, then X is called *integrable*. For any $A \in \mathcal{A}$,

(5) $$\int_A X\,d\mu \equiv \int X\,1_A\,d\mu.$$

We also use the notation (especially in proofs, to save space)

(6) $$\int X \equiv \int X\,d\mu \equiv (\text{the } \textit{integral} \text{ of } X) \equiv \mathrm{E}X \equiv (\text{the } \textit{expectation} \text{ of } X)\,.$$

For ordinary Lebesgue measure μ on R, we often write $\int X\,d\mu = \int X(r)\,dr$.

It needs to be demonstrated that the above definition makes sense and that $\int X\,d\mu$ satisfies the following elementary properties.

Proposition 1.1 (Elementary properties of the integral) It holds that definition 1.1 of the integral is unambiguous. Now suppose that the functions X and Y are measurable, that $\int X\,d\mu$ and $\int Y\,d\mu$ are well-defined, and that their sum (the number $\int X\,d\mu + \int Y\,d\mu$) is a well-defined number in $[-\infty, +\infty]$. Then

(7) $$\int (X+Y)\,d\mu = \int X\,d\mu + \int Y\,d\mu \qquad \text{and} \qquad \int cX\,d\mu = c\int X\,d\mu,$$

(8) $$0 \le X \le Y \qquad \text{implies} \qquad 0 \le \int X\,d\mu \le \int Y\,d\mu.$$

Proof. Consider first the case of simple functions.
Claim 1: Defining $\int X\,d\mu = \sum_1^m x_i\,\mu(A_i)$ for simple functions $X = \sum_1^m x_i\,1_{A_i}$ makes $\int X\,d\mu$ well-defined for such simple functions.
Suppose that we also have $X = \sum_1^n z_j\,1_{C_j}$. Then

(a) $$\sum_{i=1}^m x_i \sum_{j=1}^n 1_{A_iC_j} = X = \sum_{j=1}^n z_j \sum_{i=1}^m 1_{A_iC_j}\ ,$$

so that $x_i = z_j$ if $A_iC_j \neq \emptyset$. Thus

$$\sum_{i=1}^m x_i\,\mu(A_i) = \sum_{i=1}^m x_i \sum_{j=1}^n \mu(A_iC_j) = \sum_{i=1}^m \sum_{j=1}^n x_i\,\mu(A_iC_j)$$

$$= \sum_{i=1}^m \sum_{j=1}^n z_j\,\mu(A_iC_j) \qquad \text{since } x_i = z_j \text{ if } A_iC_j \neq \emptyset$$

$$= \sum_{j=1}^n z_j \sum_{i=1}^m \mu(A_iC_j) = \sum_{j=1}^n z_j\,\mu(C_j)\,;$$

and since the two extreme terms that represent the two different definitions of the quantity $\int X\,d\mu$ are equal, we see that $\int X\,d\mu$ is well-defined.
Claim 2: The integral behaves linearly for simple functions.
Suppose $X = \sum_1^m x_i\,1_{A_i}$ and $Y = \sum_1^n y_j\,1_{B_j}$. Then $X+Y = \sum_1^m \sum_1^n (x_i+y_j)\,1_{A_iB_j}$. We thus have

$$\int (X+Y)\,d\mu = \sum_1^m \sum_1^n (x_i+y_j)\,\mu(A_iB_j)$$

$$= \sum_1^m \sum_1^n x_i\,\mu(A_iB_j) + \sum_1^m \sum_1^n y_j\,\mu(A_iB_j)$$

$$= \sum_1^m x_i \sum_1^n \mu(A_i B_j) + \sum_1^n y_j \sum_1^m \mu(A_i B_j)$$

$$= \sum_1^m x_i \, \mu(A_i) + \sum_1^n y_j \, \mu(B_j) = \int X \, d\mu + \int Y \, d\mu,$$

which establishes the additivity.
Claim 3: Even in general, it is trivial that $\int cX = c \int X$ in (7).
Claim 4: Even in general, the monotonicity in (8) is trivial.

The proof for general $X \geq 0$ and $Y \geq 0$ is included in the proof of the monotone convergence theorem (MCT) (that is, the first theorem of the next section). That is, we will prove the MCT using linearity just for simple functions, and then we will use the MCT to obtain the current linearity for any functions $X \geq 0$ and $Y \geq 0$. (The final linearity step is then trivial. Just write $X = X^+ - X^-$ and $Y = Y^+ - Y^-$ and do algebra.) □

Notation 1.1 Let F denote a generalized df and let μ_F denote the associated Lebesgue–Stieltjes measure. Suppose that g is an integrable function on R. We will then freely use the notation

(9) $$\int_R g(x) \, dF(x) \equiv \int_R g(x) \, d\mu_F(x).$$ □

2 Fundamental Properties of Integrals

Theorem 2.1 (MCT, the monotone convergence theorem) Suppose that $X_n \nearrow X$ a.e. for measurable functions $X_n \geq 0$. Then

$$(1) \qquad 0 \leq \int X_n \, d\mu \nearrow \int X \, d\mu \,.$$

Corollary 1 For $X \geq 0$, the simple X_n in (2.2.10) satisfy $\int X_n \, d\mu \nearrow \int X \, d\mu$.

Proof. By redefining on null sets if necessary, we may assume that $X_n \nearrow X$ for all ω. Thus X is measurable, by proposition 2.2.2. Also, $\int X_n$ is $\nearrow$, and so $a \equiv \lim \int X_n$ exists in $[0, \infty]$. Moreover, $X_n \leq X$ implies $\int X_n \leq \int X$; and so we conclude that $a = \lim \int X_n \leq \int X$.

Let $Y \equiv \sum_1^m c_j 1_{D_j}$ be an arbitrary simple function satisfying $0 \leq Y \leq X$. Fix $0 < \theta < 1$. Then note that $A_n \equiv [X_n \geq \theta Y] \nearrow \Omega$ (since $0 \leq \theta Y \leq X$ on $[X = 0]$ and $0 \leq \theta Y < X$ on $[X > 0]$ are both trivial). Claims 3 and 4 of the proposition 3.1.1 proof give

$$\text{(a)} \qquad \theta \int Y \times 1_{A_n} = \int \theta Y \times 1_{A_n} \leq \int X_n \times 1_{A_n} \leq \int X_n \leq a;$$

and letting $n \to \infty$ gives $\theta \int Y \leq a$ (as $Y 1_{A_n} \nearrow Y$ gives $\int Y 1_{A_n} \nearrow \int Y$ since each $c_j \, \mu(D_j A_n) \to c_j \, \mu(D_j)$ by proposition 1.1.2) for each $0 < \theta < 1$, so that $\int Y \leq a$. Since $0 \leq Y \leq X$ is arbitrary, this gives $\int X \leq a = \lim \int X_n$. □

Proof. We now return to the linearity of the integral for general measurable functions $X \geq 0$ and $Y \geq 0$. Let $X_n \nearrow X$ and $Y_n \nearrow Y$ for the measurable simple functions of (2.2.10). Then $X_n + Y_n \nearrow X + Y$. Thus the MCT twice, the linearity of the integral for simple functions, and then the MCT again give the general linearity of the integral via

$$\int X + \int Y = \lim \int X_n + \lim \int Y_n = \lim \left(\int X_n + \int Y_n \right)$$

$$\text{(a)} \qquad = \lim \int (X_n + Y_n) \qquad \text{by simple function linearity}$$

$$\text{(b)} \qquad = \int (X + Y) \qquad \text{by the MCT.}$$

In general, combine the integrals of X^+, X^-, Y^+, and Y^- appropriately. □

Theorem 2.2 (Fatou's lemma) For X_n's measurable,

$$(2) \qquad \int \underline{\lim} \, X_n \, d\mu \leq \underline{\lim} \int X_n \, d\mu, \qquad \text{provided that } X_n \geq 0 \text{ a.e. for all } n.$$

Proof. Redefine on null sets (if necessary) so that all $X_n \geq 0$. Then

$$(3) \qquad Y_n \equiv \inf_{k \geq n} X_k \nearrow \underline{\lim} \, X_n, \qquad \text{or} \qquad \underline{\lim} \, X_n = \lim Y_n \text{ with } Y_n \nearrow,$$

so that

$$\text{(a)} \qquad \int \underline{\lim} \, X_n = \int \lim Y_n = \lim \int Y_n \qquad \text{by the MCT}$$

$$\text{(b)} \qquad = \underline{\lim} \int Y_n \leq \underline{\lim} \int X_n \qquad \text{since } Y_n \leq X_n \,.$$

□

Theorem 2.3 (DCT, the dominated convergence theorem) Suppose now that $|X_n| \le Y$ a.e. for all n, for some *dominating function* $Y \in \mathcal{L}_1$; and suppose either (i) $X_n \to_{a.e.} X$ or (ii) $X_n \to_\mu X$. Then

$$(4) \qquad \int |X_n - X|\, d\mu \to 0 \quad \text{as } n \to \infty \qquad (\text{that is,} \quad X_n \to_{\mathcal{L}_1} X).$$

[Note that $(\sup_{n\ge1} |X_n|\,)$ could make a suitable dominating function.]

Corollary 1 Note that (4) implies

$$(5) \qquad \int X_n\, d\mu \to \int X\, d\mu \qquad\qquad (\text{that is,} \quad \mathrm{E}X_n \to \mathrm{E}X),$$

$$(6) \qquad \sup_{A\in\mathcal{A}} \left| \int_A X_n\, d\mu - \int_A X\, d\mu \right| \to 0.$$

Proof. (i) Suppose that $X_n \to_{a.e.} X$. Then $Z_n \equiv |X_n - X| \to_{a.e.} 0$. (Here, $0 \le Z_n \le 2Y$ a.s., where both of the functions 0 and $2Y$ are in $\mathcal{L}_1$.) Now apply Fatou's lemma to the rvs $2Y - Z_n \ge 0$, and conclude that

$$\text{(a)} \qquad \int (2Y - 0) = \int \underline{\lim}\,(2Y - Z_n) \le \underline{\lim} \int (2Y - Z_n) \qquad \text{by Fatou}$$

$$= \underline{\lim}\,(\textstyle\int 2Y - \int Z_n)$$

$$(7) \qquad = \textstyle\int 2Y - \overline{\lim} \int Z_n\,.$$

Hence, $\limsup \int Z_n \le \int 0 = 0$ (as $\int 2Y$ is finite). Combining the two results gives

$$\text{(b)} \qquad 0 \le \liminf \textstyle\int Z_n \le \limsup \int Z_n \le 0\,;$$

so $\lim \int Z_n = 0$, as claimed.

(ii) Suppose $X_n \to_\mu X$. Let $a \equiv \limsup \int Z_n \ge 0$. Let n' be a subsequence such that $\int Z_{n'} \to a$. But $Z_{n'} \to_\mu 0$, so theorem 2.3.1 gives a further subsequence n'' such that $Z_{n''} \to_{a.e.} 0$, while we still have $\int Z_{n''} \to a$. But $\int Z_{n''} \to 0$ by case (i). Thus $a = 0$. Thus

$$\text{(c)} \qquad 0 \le \limsup \textstyle\int Z_n = a = 0, \quad \text{or} \quad \int Z_n \to 0\,.$$

(iii) Consider the corollary. We have $|\int X_n - \int X| \le \int |X_n - X|$, and thus

$$\text{(d)} \qquad |\textstyle\int_A X_n - \int_A X| \le \int_A |X_n - X| \le \int |X_n - X| \to 0$$

uniformly in all $A \in \mathcal{A}$. □

Theorem 2.4 $\int \sum_1^\infty X_n\, d\mu = \sum_1^\infty \int X_n d\mu$ if $X_n \ge 0$ a.e., for all n.

Proof. Note that $0 \le Z_n \equiv \sum_1^n X_k \nearrow Z \equiv \sum_1^\infty X_k$ a.e., and now apply the MCT to the Z_n's. □

Theorem 2.5 (Absolute continuity of the integral) Fix $X \in \mathcal{L}_1$. Then

$$(8) \qquad \int_A |X|\,d\mu \to 0 \qquad \text{as} \qquad \mu(A) \to 0.$$

That is, $\int_A |X|\,d\mu < \epsilon$, provided only that $\mu(A) <$ (an appropriate δ_ϵ).

Proof. Now, $\int |X|\,1_{[\,|X|\le n]} \nearrow \int |X|$ by the MCT, so we may claim that

(a) $\int |X|\,1_{[\,|X|>n]} \le \epsilon/2$ for $n \ge N \equiv$ (some N_ϵ) .

Thus

(b) $\int_A |X| \le \int_A |X|\,1_{[\,|X|\le N]} + \int |X|\,1_{[\,|X|>N]} \le N \times \mu(A) + \epsilon/2 \le \epsilon,$

provided that $\mu(A) \le \epsilon/(2N)$. □

Exercise 2.1 (Only the zero function) Show that

(9) $X \ge 0$ and $\int X\,d\mu = 0$ implies $\mu(\,[X > 0]\,) = 0$.

Exercise 2.2 (Only the zero function) Show that

$$(10) \qquad \int_A X\,d\mu = \begin{cases} = 0, \\ \ge 0 \end{cases} \quad \text{for all } A \in \mathcal{A} \qquad \text{implies} \qquad X = \begin{cases} = 0 & \text{a.e.,} \\ \ge 0 & \text{a.e.} \end{cases}$$

Exercise 2.3 Consider a measure space $(\Omega, \mathcal{A}, \mu)$. Let $\mu_0 \equiv \mu|\mathcal{A}_0$ for a sub σ-field $\mathcal{A}_0$ of $\mathcal{A}$. Starting with indicator functions, show that $\int X\,d\mu = \int X\,d\mu_0$ for any $\mathcal{A}_0$-measurable function X. Hint: Consider four cases, as in the next proof.

Definition 2.1 (Induced measure) Suppose that $X : (\Omega, \mathcal{A}, \mu) \to (\Omega', \mathcal{A}'\,)$ is a measurable function. Recall from (2.2.15) that

$$(11) \qquad \mu'(A'\,) \equiv \mu_X(A'\,) = \mu(X^{-1}(A'\,)) \qquad \text{for all } A' \in \mathcal{A}',$$

and μ' is a measure on $(\Omega', \mathcal{A}'\,)$, called the *induced measure* of X.

Theorem 2.6 (Theorem of the unconscious statistician) (i) The induced measure $\mu_X(\cdot)$ of the measurable function $X : (\Omega, \mathcal{A}, \mu) \to (\Omega', \mathcal{A}', \mu_X)$ determines the induced measure $\mu_{g(X)}$ for all measurable functions $g : (\Omega', \mathcal{A}') \to (\bar{R}, \bar{\mathcal{B}}\,)$.
(ii) (Change of variable) Then

$$(12) \qquad \int_{X^{-1}(A')} g(X(\omega))\,d\mu(\omega) = \int_{A'} g(x)\,d\mu_X(x) \qquad \text{for all } A' \in \mathcal{A}',$$

in the sense that if either side exists then so does the other and they are equal. So,

$$(13) \qquad \int_{X^{-1}(g^{-1}(B))} g(X(\omega))\,d\mu(\omega) = \int_{g^{-1}(B)} g(x)\,d\mu_X(x) = \int_B y\,d\mu_Y(y) \quad \text{for } B \in \bar{\mathcal{B}}.$$

Proof. (i) Now, $Y \equiv g(X)$ is measurable. By (2.1.11) and (2.2.5) we see that

(a) $\mu_Y(B) = \mu_{g(X)}(B) = \mu(\,[g(X) \in B]\,) = \mu(X^{-1} \circ g^{-1}(B)) = \mu_X(g^{-1}(B))$

is well-defined, since $g^{-1}(B) \in \mathcal{A}'$. Thus the first claim holds.

(ii) We only prove the first equality in (13) when $A' = \Omega'$ and $X^{-1}(\Omega') = \Omega$, since we can replace g by $g \times 1_{A'}$, noting that $1_{A'}(X(\omega)) = 1_{X^{-1}(A')}(\omega)$.
Case 1. $g = 1_{A'}$: Then

(b) $$\int 1_{A'}(X)\,d\mu = \int 1_{X^{-1}(A')}\,d\mu = \mu(X^{-1}(A')) = \mu_X(A') = \int 1_{A'}\,d\mu_X.$$

Case 2. $g = \sum_{i=1}^n c_i\,1_{A'_i}$, where $\sum_{i=1}^n A'_i = \Omega'$ with $A'_i \in \mathcal{A}'$ and all $c_i \geq 0$: Then

$$\int g(X)\,d\mu = \int \sum_{i=1}^n c_i\,1_{A'_i}(X)\,d\mu = \sum_{i=1}^n c_i \int 1_{A'_i}(X)\,d\mu$$

(c) $$= \sum_{i=1}^n c_i \int 1_{A'_i}\,d\mu_X = \int g\,d\mu_X.$$

Case 3. $g \geq 0$: Let $g_n \geq 0$ be simple with $g_n \nearrow g$: Then

$$\int g(X)\,d\mu = \lim \int g_n(X)\,d\mu \qquad \text{by the MCT, since } g_n(X) \nearrow g(X)$$

$$= \lim \int g_n\,d\mu_X \qquad \text{by case 2}$$

(d) $$= \int g\,d\mu_X \qquad \text{by the MCT.}$$

Case 4. g is measurable, and either $\int g(X)^+\,d\mu$ or $\int g(X)^-\,d\mu$ is finite: Using $g = g^+ - g^-$, we note that $g(X)^+ = g^+(X)$ and $g(X)^- = g^-(X)$. Then

$$\int g(X)\,d\mu = \int g(X)^+\,d\mu - \int g(X)^-\,d\mu = \int g^+(X)\,d\mu - \int g^-(X)\,d\mu$$

$$= \int g^+\,d\mu_X - \int g^-\,d\mu_X \qquad \text{by case 3}$$

(e) $$= \int g\,d\mu_X.$$

In the arguments (b), (c), (d), (e) one should start from the end that is assumed to exist, in order to make a logically tight argument. (Note the next exercise.) □

Exercise 2.4 Let $Y \equiv g(X)$ in the context of the theorem 2.6. Verify the second equality in (13).

Exercise 2.5 Let X equal -1,0,1 with probability 1/3 for each possibility. Let $g(x) = x^2$. Then evaluate both sides in (13), and see why such calculations were performed unconsciously for years.

3 Evaluating and Differentiating Integrals

Let $(R, \hat{\mathcal{B}}_\mu, \mu)$ denote a Lebesgue–Stieltjes measure space that has been completed. If g is $\hat{\mathcal{B}}_\mu$-measurable, then $\int g\,d\mu$ is called the *Lebesgue–Stieltjes integral* of g; and if F is the generalized df corresponding to μ, then we also use the notation $\int g\,dF$. Also, $\int_a^b g\,dF \equiv \int_{(a,b]} g\,dF = \int 1_{(a,b]}\,g\,dF$.

Theorem 3.1 (Equality of LS and RS integrals) Let g be continuous on $[a,b]$. Then the Lebesgue–Stieltjes integral $\int_a^b g\,dF$ and the Riemann–Stieltjes integral are equal. [And since the LS-integral and the RS-integral are equal, we can continue to evaluate most LS-integrals using the methods learned in elementary calculus.]

Proof. We first recall the classical setup associated with the definition of the RS-integral. Consider any sequence of partitions $a \equiv x_{n0} < \cdots < x_{nn} \equiv b$ such that the partition $\mathcal{X}_n \equiv \{x_{n0}, x_{n1}, \ldots, x_{nn}\}$ is a *refinement* of $\mathcal{X}_{n-1}$ in the sense that $\mathcal{X}_{n-1} \subset \mathcal{X}_n$. Then if $\text{mesh}_n \equiv \max_{1\le k\le n}(x_{nk} - x_{n,k-1}) \to 0$, and if x_{nk}^*'s are such that $x_{n,k-1} < x_{nk}^* \le x_{nk}$, we have (letting $g_n(a)$ be defined by right continuity)

(a) $$g_n \equiv \sum_{k=1}^n g(x_{nk}^*)\,1_{(x_{n,k-1},x_{nk}]} \to g \qquad \text{uniformly on } [a,b\,],$$

since g is (necessarily) uniformly continuous on $[a,b]$. Thus for all such sequences the LS-integral of section 3.1 satisfies

$$\int_a^b g\,dF \equiv \int_a^b g\,d\mu = \lim \int_a^b g_n\,d\mu \qquad \text{by the DCT, bounded by a constant}$$

$$= \lim \sum_1^n g(x_{nk}^*)\,\mu((x_{n,k-1}, x_{nk}]) = \lim \sum_1^n g(x_{nk}^*)\,F(x_{n,k-1}, x_{nk}]$$

(b) $$= \lim \{\text{a Riemann–Stieltjes sum for the integral of } g\}$$

(c) $$= \{\text{the Riemann–Stieltjes integral of } g\},$$

and this holds for all partitions and x_{nk}^*'s as above, provided only that $\text{mesh}_n \to 0$. Thus the LS-integral $\int_a^b g\,dF$ and the RS-integral are equal for continuous g. □

Exercise 3.1* (RS-integral compared to LS-integral) We state a few additional facts here, just for completeness:

(1) g is RS-integrable with respect to F if and only if g is continuous a.e. $\mu_F(\cdot)$.

(2) If g is RS-integrable with respect to F, then the RS and LS-integrals $\int_a^b g\,dF$ are equal.

Let $D(F)$ and $D(g)$ denote the discontinuity sets of F and g. Then

(3) g is not RS-integrable when $D(F) \cap D(g) \neq \emptyset$.

(Consider $g(\cdot) \equiv 1_{\{0\}}(\cdot)$ and $F \equiv 1_{[0,\infty)}$ regarding (3).)

Exercise 3.2 Suppose that the improper RS-integral of a continuous function g on R, defined by $\mathrm{RS}(\int g\,dF) \equiv \lim_{a\to-\infty,b\to\infty} (\mathrm{RS}\int_a^b g\,dF)$ exists finitely. Then $\lim_{a\to-\infty,b\to\infty} (\mathrm{LS}\int_a^b |g|\,dF)$ need not be finite. Thus the fact that an improper RS-integral exists does not imply that the function is LS-integrable. Construct an example on $[0,\infty)$.

Exercise 3.3 (Differentiation under the integral sign) (a) Suppose that the function $X(t,\cdot)$ is an integrable function on (Ω,μ), for each $t \in [a,b\,]$. Suppose also that for a.e. ω the partial derivative $\frac{\partial}{\partial t}X(t,\omega)$ exists for all t in the nondegenerate interval $[a,b]$ (use one-sided derivatives at the end points), and that n

$$\left|\frac{\partial}{\partial t}X(t,\omega)\right| \le Y(\omega) \qquad \text{for all } t \in [a,b\,], \quad \text{where } Y \in \mathcal{L}_1.$$

Then the derivative and integral may be interchanged, in that

$$(4) \qquad \frac{d}{dt}\int_\Omega X(t,\omega)\,d\mu(\omega) = \int_\Omega \left[\frac{\partial}{\partial t}\,X(t,\omega)\right] d\mu(\omega) \qquad \text{for all } t \in [a,b\,].$$

(b) Fix $t \in (a,b)$. Formulate weaker hypotheses that yield (4) at this fixed t.

4 Inequalities

Convexity We begin by briefly reviewing convexity. A real-valued function f defined on some interval I of real numbers is *convex* if

$$(1) \qquad \begin{array}{l} f(\alpha x + (1-\alpha)y) \le \alpha\, f(x) + (1-\alpha)\, f(y) \\ \text{for all } x, y \text{ in } I \text{ and all } 0 \le \alpha \le 1. \end{array}$$

We will use the following facts. If f is convex on an interval, then f is continuous on the interior I^o of the interval. Also, the left and right derivatives exist and satisfy $D^-(x) \le D^+(x)$ at each point in the interior I^o of the interval. The following is useful. Convexity on the interval I holds if and only if

$$(2) \qquad \begin{array}{l} f((x+y)/2) \le [f(x) + f(y)]/2 \qquad \text{for all } x, y \text{ in } I, \text{ provided that} \\ f \text{ is also assumed to be bounded (or continuous, or measurable) on } I. \end{array}$$

[There exist functions satisfying the inequality in (2) that are not continuous, but they are unbounded in every finite interval. Thus requiring (1) for all $0 \le \alpha \le 1$ is strictly stronger then requiring it to hold only for $\alpha = 1/2$.] We need a simple test for convexity (when f is 'nice'), and so note that f is convex if either

$$(3) \qquad f'(x) \text{ is } \nearrow \text{ for all } x \in I \qquad \text{or} \qquad f''(x) \ge 0 \text{ for all } x \in I\,.$$

We call f *strictly convex* if strict inequality holds in any of the above. If f is convex, then there exists a linear function l such that $f(x) \ge l(x)$ with equality at any prespecified x_0 in the interior I^o of the domain I of f; this function is called the *supporting hyperplane*. (Call f *concave* in $-f$ is convex.) □

Definition 4.1 (Moments) [The following definitions make sense on a general measure space $(\Omega, \mathcal{A}, \mu)$, and are standard notation on a probability space $(\Omega, \mathcal{A}, P)$.] Recall from (3.1.6) that $\mathrm{E}h(X) = \int h(X(\omega))\, d\mu(\omega) = \int h(X)\, d\mu = \int h(X)$. Let

$$(4) \qquad \mu \equiv \mathrm{E}X \equiv \text{ (the } \textit{mean} \text{ of } X), \qquad \text{(Note two different uses of } \mu)$$

$$(5) \qquad \begin{array}{l} \sigma^2 \equiv \mathrm{Var}[X] \equiv \mathrm{E}(X-\mu)^2 = \text{ (the } \textit{variance} \text{ of } X), \\[1ex] \sigma \equiv \mathrm{StDev}[X] \equiv \text{(the } \textit{standard deviation} \text{ of } X)\,. \end{array}$$

We will write $X \cong (\mu, \sigma^2)$ if $\mathrm{E}X = \mu$ and $\mathrm{Var}[X] = \sigma^2 < \infty$. We will then write $X \cong F(\mu, \sigma^2)$ if X also has df $F(\cdot)$. Now let

$$(6) \qquad \mathrm{E}X^k \equiv (k\textit{th moment} \text{ of } X), \qquad \text{for } k \ge 1 \text{ an integer},$$

$$(7) \qquad \mathrm{E}|X|^r \equiv (r\textit{th absolute moment} \text{ of } X), \qquad \text{for } r > 0,$$

$$(8) \qquad \|X\|_r \equiv \{\mathrm{E}|X|^r\}^{1/r} \text{ (or } \mathrm{E}|X|) \equiv (r\textit{th norm} \text{ of } X), \quad \text{for } r \ge 1 \text{ (or, } r < 1),$$

$$(9) \qquad \mu_k \equiv \mathrm{E}(X-\mu)^k \equiv (k\textit{th central moment} \text{ of } X), \qquad \text{for } k \ge 1,$$

$$(10) \qquad \mathrm{Cov}[X, Y] \equiv \mathrm{E}[(X-\mu_X)(Y-\mu_Y)] = \text{ (the } \textit{covariance} \text{ of } X \text{ and } Y).$$

Note that $\mathrm{Cov}[X, X] = \mathrm{Var}[X]$. (Probability theory has $\mu(\Omega) = 1$.)

Throughout this section X and Y will denote measurable functions.

Proposition 4.1 ($\mathcal{L}_s \subset \mathcal{L}_r$) Let $\mu(\Omega) < \infty$. Then $\mathcal{L}_s \subset \mathcal{L}_r$ whenever $0 < r < s$. [So if $\mathrm{E}|X|^s < \infty$, then $\mathrm{E}|X|^r$ and $\mathrm{E}X^k$ are finite for all $0 \le r, k \le s$.]

Proof. Now, $|x|^r \le 1 + |x|^s$; and integrability is equivalent to absolute integrability. Note that $\mu(\Omega) < \infty$ is used to claim $1 \in \mathcal{L}_1$. □

Proposition 4.2 Let $\mu(\Omega) = 1$. Then $\sigma^2 < \infty$ holds if and only if $\mathrm{E}X^2 < \infty$. Moreover, $\sigma^2 = \mathrm{E}X^2 - \mu^2$ when $\mu(\Omega) = 1$.

Proof. Suppose $\mathrm{E}X^2 < \infty$. Then $\mathrm{E}X^2 - \mu^2 = \mathrm{E}(X^2) - \mathrm{E}(2\mu X) + \mathrm{E}(\mu^2) = \mathrm{E}(X-\mu)^2 = \mathrm{Var}[X]$. Note that proposition 4.1 was used for $\mathrm{E}X$. Thus $\mu(\Omega) < \infty$ was used. Suppose that $\sigma^2 < \infty$. Then $\mathrm{E}\{(X-\mu)^2 + 2\mu(X-\mu) + \mu^2\} = \mathrm{E}X^2$. □

Inequality 4.1 (C_r-inequality) $\mathrm{E}|X+Y|^r \le C_r\,\mathrm{E}|X|^r + C_r\,\mathrm{E}|Y|^r$,
where $C_r = 2^{r-1}$ for $r \ge 1$ and $C_r = 1$ for $0 < r \le 1$.

Proof. There are no restrictions on μ. Note that $\mathrm{E}|X+Y|^r \le \mathrm{E}(|X| + |Y|)^r$.
Case 1. $r > 1$: Then $|x|^r$ is convex in x for $x \ge 0$, since its derivative is $\uparrow$. Thus $|(x+y)/2|^r \le [|x|^r + |y|^r]/2$; and now take expectations.

Case 2. $0 < r \le 1$: Now, $|x|^r$ is concave and $\nearrow$ for $x \ge 0$; just examine derivatives. Thus $|x+y|^r - |x|^r \le |0+y|^r - 0^r$ since the increase from x to $x+y$ can not exceed the increase from 0 to y, and now take expectations. □

Inequality 4.2 (Hölder's inequality) $\mathrm{E}|X\,Y| \le \mathrm{E}^{1/r}|X|^r\,\mathrm{E}^{1/s}|Y|^s$ for $r > 1$, and $1/r + 1/s = 1$ (so, $s - 1 = 1/(r-1)$. Alternatively, $\|X\,Y\|_1 \le \|X\|_r \times \|Y\|_s$.
When both expectations $\mathrm{E}|X|^r$ and $\mathrm{E}|Y|^s$ fall in $(0,\infty)$, we have equality if and only if there exists a constant $a > 0$ for which $|Y|^s = a|X|^r$ a.e.; hence, $a = \mathrm{E}|Y|^s/\mathrm{E}|X|^r$.

Proof. The result is trivial if $\mathrm{E}|X|^r = 0$ or ∞. Likewise for $\mathrm{E}|Y|^s$. So suppose that both expectations are in $(0,\infty)$. Since $f(x) = e^x$ is convex by fact (3), it satisfies (1) with $\alpha \equiv 1/r$ and $1-\alpha = 1/s$, $x \equiv r\log|a|$ and $y \equiv s\log|b|$ for some a and b; thus (1) becomes (with equality if and only if $r\log|a| = x = y = s\log|b|$)

(11) $\exp(\frac{1}{r}\,x + \frac{1}{s}\,y) \le \frac{1}{r}\,e^x + \frac{1}{s}\,e^y$, or:

Young's inequality For all a, b we have

(12) $|ab| \le \dfrac{|a|^r}{r} + \dfrac{|b|^s}{s}$ with equality iff $|a|^r = |b|^s$ iff $|b| = |a|^{1/(s-1)} = |a|^{r-1}$.

Now let $a = |X|/\|X\|_r$ and $b = |Y|/\|Y\|_s$, and take expectations. Equality holds if and only if $|Y|/\|Y\|_s = (|X|/\|X\|_r)^{1/(s-1)}$ a.e. (that is, all mass is located at equality in (12)) if and only if

(a) $$\frac{|Y|^s}{\mathrm{E}|Y|^s} = \left(\frac{|X|}{\|X\|_r}\right)^{s/(s-1)} = \frac{|X|^r}{\mathrm{E}|X|^r} \quad \text{a.e.}$$ □

Exercise 4.1 (Convexity inequality) Show that

(13) $u^\alpha v^{1-\alpha} \le \alpha\, u + (1-\alpha)\, v$ for all $0 \le \alpha \le 1$ and all $u, v \ge 0$.

Use this to reprove Hölder's inequality.

Inequality 4.3 (Cauchy–Schwarz) $\{\mathrm{E}(XY)\}^2 \le (\,\mathrm{E}|XY|\,)^2 \le \mathrm{E}X^2\, \mathrm{E}Y^2$.
If both $\mathrm{E}X^2$ and $\mathrm{E}Y^2$ take values in $(0,\infty)$, then equality holds throughout both of the inequalities if and only if either $Y = aX$ a.e. or $Y = -aX$ a.e., for some $a > 0$; in fact, $a^2 = \mathrm{E}Y^2/\mathrm{E}X^2$. (Only $Y^2 = cX^2$ a.e. for some $c > 0$ is required for equality in the rightmost inequality above.)

Example 4.1 (Correlation inequality) For rvs X and Y (on a probability space) having positive and finite variances, it holds that

(14) $-1 \le \rho \le 1$,

for the *correlation* ρ of X and Y defined by

(15) $$\rho \equiv \rho_{X,Y} \equiv \mathrm{Corr}\,[X,Y] \equiv \frac{\mathrm{Cov}[X,Y]}{\sqrt{\mathrm{Var}[X]\,\mathrm{Var}[Y]}}\,.$$ □

Exercise 4.2 Consider rvs X and Y having $\mathrm{E}X^2$ and $\mathrm{E}Y^2$ in $(0,\infty)$. Show that

$\rho = +1$ if and only if $X - \mu_X = a\,(Y - \mu_Y)$ a.e. for some $a > 0$,

$\rho = -1$ if and only if $X - \mu_X = a\,(Y - \mu_Y)$ a.e. for some $a < 0$.

Thus ρ measures linear dependence, not general dependence.

Inequality 4.4 (Liapunov's inequality) (a) It holds that

(16) $\mathrm{E}^{1/r}\,|X|^r$ is $\nearrow$ in r for all $r \ge 0$, provided $\mu(\Omega) = 1$.

(b) Let $\mu(\Omega)$ be finite. Then $\|X\|_r \le \|X\|_s \times \{\mu(\Omega)\}^{\frac{1}{r}-\frac{1}{s}}$ for all $0 < r < s$.

(c) $h(r) \equiv \log \mathrm{E}|X|^r$ is convex on $[a,b]$ if $X \in \mathcal{L}_a \cap \mathcal{L}_b$ ($0 < a$ and any $\mu(\Omega)$ value).

Proof. (c) Apply Hölder to $|X|^{\alpha a}$ and $|X|^{(1-\alpha)b}$ with $r = 1/\alpha$ and $s = 1/(1-\alpha)$ and obtain the inequality

(p) $\mathrm{E}|X|^{\alpha a + (1-\alpha)b} \le (\mathrm{E}|X|^{\alpha a \cdot \frac{1}{\alpha}})^\alpha (\mathrm{E}|X|^{(1-\alpha)b\cdot\frac{1}{1-\alpha}})^{1-\alpha}$

(q) $= (\mathrm{E}|X|^a)^\alpha (\mathrm{E}|X|^b)^{1-\alpha}$.

(All expectations are finite if $X \in \mathcal{L}_a \cap \mathcal{L}_b$; since $a \le r \le b$ and $c > 0$ implies that $c^r \le c^b$ or $c^r \le c^a$ as $c \ge 1$ or $c \le 1$.) Taking logarithms gives the convexity

(r) $h(\alpha a + (1-\alpha)b) \le \alpha h(a) + (1-\alpha)h(b)$.

(b) Finiteness of $\mathrm{E}|X|^b$ gives the finiteness of $\mathrm{E}|X|^r$ on $(0,b]$ via proposition 4.1. Then apply Hölder to $|X|^r \cdot 1$ with $a = \frac{s}{r}$ and $\frac{1}{a} + \frac{1}{b} = 1$ (and $\mathrm{E}1 = \mu(\Omega)$). □

Exercise 4.3 (Littlewood's inequality) Define $m_r \equiv \mathrm{E}|X|^r$. Show that for $0 \le r \le s \le t$ we have (write $m_s = \mathrm{E}(\,|X|^{\lambda s} \cdot |X|^{(1-\lambda)s})$, and apply Hölder)

(17) $m_s^{t-r} \le m_r^{t-s}\; m_t^{s-r}$ (thus, $m_2^3 \le m_1^2\, m_4$).

Inequality 4.5 (Minkowski's inequality) $\mathrm{E}^{1/r}|X+Y|^r \le \mathrm{E}^{1/r}|X|^r + \mathrm{E}^{1/r}|Y|^r$ for all $r \ge 1$. That is, $\|X+Y\|_r \le \|X\|_r + \|Y\|_r$ for $r \ge 1$.
(Recall that $\|X+Y\|_r \le \|X\|_r + \|Y\|_r$ for $0 < r \le 1$, by the C_r-inequality and (8).)

Proof. This is trivial for $r = 1$. Suppose $r > 1$, and note that $s = r/(r-1)$. Then for any measure μ we have

(a) $\mathrm{E}\{|X+Y|^r\} \le \mathrm{E}\{|X|\,|X+Y|^{r-1}\} + \mathrm{E}\{|Y|\,|X+Y|^{r-1}\}$

$\le (\,\|X\|_r + \|Y\|_r\,)\ \|\,|X+Y|^{r-1}\,\|_s$ by Hölder's inequality twice

$= (\,\|X\|_r + \|Y\|_r)\ \mathrm{E}^{1/s}\,|X+Y|^{(r-1)s} = (\,\|X\|_r + \|Y\|_r)\ \mathrm{E}^{1/s}\,|X+Y|^r.$

If $\mathrm{E}|X+Y|^r = 0$, the result is trivial. If not, we divide to get the result. □

Inequality 4.6 (Basic inequality) Let $g \ge 0$ be $\nearrow$ on $[0,\infty)$ and even. Then for all measurable X we have

(18) $\mu(\,|X| \ge \lambda\,) \le \mathrm{E}\,g(X)/g(\lambda)$ for all $\lambda > 0$.

Proof. Now,

(a) $\mathrm{E}\,g(X) = \int_{[\,|X|\ge\lambda]} g(X)\,d\mu + \int_{[\,|X|<\lambda]} g(X)\,d\mu \ge \int_{[\,|X|\ge\lambda]} g(X)\,d\mu$

(b) $\ge g(\lambda)\int_{[\,|X|\ge\lambda]} 1\,d\mu = g(\lambda)\ \mu(\,|X| \ge \lambda\,).$ □

The next two inequalities are immediate corollaries.

Inequality 4.7 (Markov's inequality) $\mu(|X| \ge \lambda) \le \mathrm{E}|X|^r/\lambda^r$ for all $\lambda > 0$.

Inequality 4.8 (Chebyshev's inequality) If $\mathrm{E}|X| < \infty$, then

(19) $\mu(\,|X-\mu| \ge \lambda) \le \mathrm{Var}[X]/\lambda^2$ for all $\lambda > 0$.

Inequality 4.9 (Paley–Zygmund) If $\mathrm{E}|X| < \infty$ for a rv $X \ge 0$, then

(20) $P(X > \lambda) \ge [(\mathrm{E}X - \lambda)^+]^2/\mathrm{E}X^2$ for each $\lambda > 0$.

Proof. Now,

$$\mathrm{E}X = \mathrm{E}(X\,1_{[X\le\lambda]}) + \mathrm{E}(X\,1_{[X>\lambda]}) \le \lambda + \sqrt{\mathrm{E}(X^2)\,P(X>\lambda)}$$

by Cauchy–Schwarz. Rearranging gives the inequality. □

Inequality 4.10 (Jensen's inequality) Let $X : (\Omega, \mathcal{A}, P) \to (I, \mathcal{B}, P_X)$, where $\mathrm{E}|X| < \infty$ and $\mathrm{E}X$ is in the interior I^o of the interval I. (Any $-\infty \le a < b \le \infty$ is permissible for $I \subset R$.) Let g be convex on this interval I. Then the rv X satisfies

(21) $g(\,\mathrm{E}X\,) \le \mathrm{E}\,g(X)$ (here, $P(\Omega) = 1$ is required).

For strictly convex g, equality holds if and only if $X = \mathrm{E}X$ a.e.
[Comment. Useful g include t^r on $[0,\infty)$ for any $r \ge 1$, $|t|$, and $-\log t$ on $(0,\infty)$.]

Proof. Let $l(\cdot)$ be a supporting hyperplane to $g(\cdot)$ at $\mathrm{E}X$. Then

(a) $\mathrm{E}\, g(X) \geq \mathrm{E}\, l(X)$

(b) $= l(\mathrm{E}X)$ since $l(\cdot)$ is linear and $\mu(\Omega) = 1$

(c) $= g(\mathrm{E}X)$ since $g(\cdot) = l(\cdot)$ at $\mathrm{E}X$.

Now $g(X) - l(X) \geq 0$. Thus $\mathrm{E}\, g(X) = \mathrm{E}\, l(X)$ if and only if $g(X) = l(X)$ a.e. μ if and only if $X = \mathrm{E}X$ a.e. μ . □

Inequality 4.11 (Bonferroni) For any events A_k on a probability space $(\Omega, \mathcal{A}, P)$,

$$\sum_{i=1}^n P(A_i) \geq P(\cup_{i=1}^n A_i) \geq \sum_{i=1}^n P(A_i) - \sum\sum_{i \neq j} P(A_i\, A_j).$$

Exercise 4.4 ($\tilde{\text{W}}$insorized variance) (a) Let the rv X have finite mean μ. Fix c, d with $c \leq \mu \leq d$. Let $\tilde{X}$ equal c, X, d according as $[X \leq c]$, $[c < X \leq d]$, $[d < X]$, and set $\tilde{\mu} \equiv \mathrm{E}\tilde{X}$. Show that $\mathrm{E}|\tilde{X} - \tilde{\mu}|^2 \leq \mathrm{E}|\tilde{X} - \mu|^2 \leq \mathrm{E}|X - \mu|^2$.
(b)* (Chow and Teicher) Given both a rv X with finite mean μ and a number $r \geq 1$, show how to choose c, d so that $\mathrm{E}|\tilde{X} - \tilde{\mu}|^r \leq \mathrm{E}|X - \mu|^r$.

Exercise 4.5*(Hardy) Let $h \in \mathcal{L}_2(R^+, \mathcal{B}, \lambda)$ and define $\bar{H}(u) = u^{-1} \int_0^u h(s)\, ds$ for $u > 0$. Let $r > 1$. Use the Hölder inequality to show that

(22) $\int_0^\infty \bar{H}^r(u)\, du \leq (\frac{r}{r-1})^r \int_0^\infty h^r(u)\, du$ with equality if and only if $h = 0$ a.e.

Let $r > 1$. Then $\sum_1^\infty (\frac{1}{n} \sum_1^n x_k)^r \leq (\frac{r}{r-1})^r \sum_1^\infty x_n^r$ when all $x_n \geq 0$.
[Hint. Write $\bar{H}(u) = u^{-1} \int_0^u h(s) s^\alpha s^{-\alpha}\, ds$ for some α. Also, first consider $x_n \searrow$.]

Exercise 4.6 (Wellner) Let $T \cong \text{Binomial}(n, p)$, so $P(T = k) = \binom{n}{k} p^k (1-p)^{n-k}$ for $0 \leq k \leq n$. The measure associated with T has mean np and variance $np(1-p)$. Then use inequality 4.6 with $g(x) = \exp(rx)$ and $r > 0$, to show that

(23) $P(T/n \geq p\epsilon) \leq \exp(-np\, h(\epsilon))$, where $h(\epsilon) \equiv \epsilon(\log(\epsilon) - 1) + 1$.

Exercise 4.7 (Geometric mean) Show that $(x_1 \times \cdots \times x_n)^{1/n} \leq (x_1 + \cdots + x_n)/n$ whenever all $x_k \geq 0$.

Exercise 4.8* Let $X, Y \geq 0$ with $XY \geq 1$ and $P(\Omega) = 1$. Show that

(24) $\mu_X \times \mu_Y \geq 1$ and $\{1 + \mu_X^2\}^{1/2} \leq \mathrm{E}\{(1 + X^2)^{1/2}\} \leq (1 + \mu_X)$.

Exercise 4.9*(Clarkson's inequality) Let X, Y in $\mathcal{L}_r(\Omega, \mathcal{A}, \mu)$. Show that

(25) $\mathrm{E}|X + Y|^r + \mathrm{E}|X - Y|^r \leq 2^{r-1} \{\mathrm{E}|X|^r + \mathrm{E}|Y|^r\}$ provided $r \geq 2$.

Exercise 4.10 Let $\mathcal{L}\,\mathrm{Log}\,\mathcal{L}$ denote all measurable X having $\mathrm{E}\{|X| \times \mathrm{Log}(|X|)\}$ finite, with $\mathrm{Log}(x) \equiv (1 \vee \log x)$. Show that $\mathcal{L}_r \subset \mathcal{L}\,\mathrm{Log}\,\mathcal{L} \subset \mathcal{L}_1$, for all $r > 1$.

Exercise 4.11 Show that for all a, b we have

(26) $$\big||a| - |b|\big|^r \le \big||a|^r - |b|^r\big| \qquad \text{for } r \ge 1,$$

with the reverse inequality for $0 < r < 1$.

Exercise 4.12 Let $(\Omega, \mathcal{A}, \mu)$ have $\mu(\Omega) < \infty$. Then $P(A) \equiv \mu(A)/\mu(\Omega)$, for all $A \in \mathcal{A}$, is a probability measure P.
(a) Restate Jensen's inequality (21) in terms of μ.
(b) Restate Liapunov's inequality inequality 4.4(b) in terms of P.

5 Modes of Convergence

Definition 5.1 (Modes of convergence) Let X and X_n's be measurable and a.e. finite from the measure space $(\Omega, \mathcal{A}, \mu)$ to $(\bar{R}, \bar{B})$.

(a) Recall that X_n *converges a.e.* to X (denoted by $X_n \to_{a.e.} X$) if

(1) $$X_n(\omega) \to X(\omega) \quad \text{for all } \omega \in A, \qquad \text{where } \mu(A^c) = 0.$$

(b) Also, recall that X_n *converges in measure* to X (denoted by $X_n \to_\mu X$) if

(2) $$\mu(\,[\,\omega : |X_n(\omega) - X(\omega)\,| \ge \epsilon\,]\,) \to 0 \qquad \text{for each } \epsilon > 0.$$

(c) Now (rigorously for the first time), X_n *converges in rth mean* to X (denoted by $X_n \to_r X$ or $X_n \to_{\mathcal{L}_r} X$) if

(3) $$\mathrm{E}|X_n - X|^r \to 0 \qquad \text{for } X_n\text{'s and } X \text{ in } \mathcal{L}_r;$$

here, $r > 0$ is fixed. [Note from the C_r-inequality that if $X_n - X$ and one of X or X_n is in $\mathcal{L}_r$, then the other of X_n or X is also in $\mathcal{L}_r$.]

Recall from chapter 2 that $X_n \to_{a.e.}$ (some a.e. finite X) holds if and only if $X_n - X_m \to_{a.e.} 0$ as $m \wedge n \to \infty$. Likewise, in chapter 2 we had $X_n \to_\mu$ (some X) if and only if $X_n - X_m \to_\mu 0$ as $m \wedge n \to \infty$. Next, we will consider $X_n \to_r X$. (First, note that $X_n \to_r X$ trivially implies $X_n \to_\mu X$, using the Markov inequality.)

Exercise 5.1 (Completeness of $\mathcal{L}_r$) (I) Let X_n's be in any $\mathcal{L}_r$, for $r > 0$.

(a) (Riesz–Fischer) $X_n \to_r$ (some $X \in \mathcal{L}_r$) if and only if $X_n - X_m \to_r 0$.

That is, $\mathcal{L}_r$ is *complete* with respect to $\to_r$. Prove (a), using (2.3.14). [Show that $(\mathcal{L}_r, \|\cdot\|_r)$ is a complete metric space (when $r > 0$), provided that we identify X and X' whenever $X = X'$ a.e.] (Note theorem 5.8 below regarding separability.)

(II) Let $\mu(\Omega) < \infty$. Then:

(b) If $X_n \to_s X$, then $X_n \to_r X$ for all $0 < r \le s$.

(c) Show by example that $X_n \to_{\mathcal{L}_r} X$ does not imply that $X_n \to_{a.e.} X$.

(d) Show by example that $X_n \to_{a.e.} X$ does not imply that $X_n \to_{\mathcal{L}_1} X$.

[Hint: Use Fatou's lemma in (a) and Hölder's inequality in (b).]

Summary Let X and X_n's be as in definition 5.1. Then

(4)
X_n converges a.e., in measure, or in $\mathcal{L}_r$ to such an X
if and only if
X_n is Cauchy a.e., Cauchy in measure, or Cauchy in $\mathcal{L}_r$.

Consequences of Convergence in Distribution on $(\Omega, \mathcal{A}, P)$

Notation 5.1 Suppose now that μ really denotes a probability measure, and so we will label it P. Recall that X_n *converges in distribution* to X (denoted by $X_n \to_d X$, $F_n \to_d F$ or $\mathcal{L}(X_n) \to \mathcal{L}(X)$ with $\mathcal{L}(\cdot)$ referring to "law") when the dfs F and F_n of X and X_n satisfy (recall (2.4.4))

$$(5) \qquad F_n(x) \to F(x) \quad \text{as } n \to \infty \quad \text{for each continuity point } x \in C_F \text{ of } F(\cdot)\,.$$

[Note that $F_n \equiv 1_{[1/n,\infty)} \to_d F \equiv 1_{[0,\infty)}$, even though $F_n(0) = 0 \not\to 1 = F(0)$.] The statement $\to_d$ will carry with it the implication that F corresponds to a probability measure P, which can be viewed as the $P_X = \mu_X$ of an appropriate rv X. □

Theorem 5.1 (Helly–Bray) Consider the rvs X and X_n on some $(\Omega, \mathcal{A}, P)$. Suppose $F_n \to_d F$, and suppose that g is bounded and is continuous a.s. F. Then

$$(6) \qquad \int g\,dF_n = \mathrm{E}\,g(X_n) = \int g(X_n)\,dP \to \int g(X)\,dP = \mathrm{E}\,g(X) = \int g\,dF.$$

Conversely, $\mathrm{E}g(X_n) \to \mathrm{E}g(X)$ for all bounded, continuous g implies that $F_n \to_d F$.

Theorem 5.2 (Mann–Wald) Consider the rvs X and X_n on some $(\Omega, \mathcal{A}, P)$. Suppose $X_n \to_d X$, and let g be continuous a.s. F. Then $g(X_n) \to_d g(X)$.

Proof. We ask for a proof for continuous g in the next exercise, but we give a "look-ahead" proof now. (See theorem 6.3.2 below for the Skorokhod proof.)

Skorokhod theorem If $X_n \to_d X$, there are Y and Y_n on some $(\Omega, \mathcal{A}, P)$ having

$$(7) \qquad Y_n \cong X_n \text{ and } Y \cong X \text{ and } \textit{especially} \quad Y_n \to Y \text{ a.s. } P_Y(\cdot)\,.$$

Note that $A_1 \equiv \{\omega : Y_n(\omega) \to Y(\omega)\}$ has $P(A_1) = 1$. Also,

$$\text{(a)} \qquad P(A_2) \equiv P(\{\omega : g \text{ is continuous at } Y(\omega)\})$$

$$\text{(b)} \qquad\quad = P_Y(\{y : g \text{ is continuous at } y\}\,) = P_Y(C_g) = 1.$$

Thus $A \equiv A_1 \cap A_2 = A_1 \cap Y^{-1}(C_g)$ has $P(A) = 1$. Especially,

$$\text{(c)} \qquad g(Y_n(\omega)) \to g(Y(\omega)) \qquad \text{for all } \omega \in A, \quad \text{with } P(A) = 1\,.$$

Since g is bounded, applying the DCT to (7) gives the Helly–Bray claim that

$$\text{(d)} \qquad \textstyle\int g\,dF_n \equiv \int g\,d\mu_{F_n} = \int g(Y_n)\,dP \to \int g(Y)\,dP = \int g\,d\mu_F \equiv \int g\,dF.$$

We note additionally that since (7) implies $g(Y_n) \to_{a.s.} g(Y)$, it also implies $g(Y_n) \to_d g(Y)$. Since $g(X_n) \cong g(Y_n)$ and $g(X) \cong g(Y)$, we can conclude that $g(X_n) \to_d g(X)$. This argument did not use the boundedness of g, and so proves the Mann–Wald theorem. [The Helly–Bray theorem will be used later in this section (in proving Vitali's theorem). Proving it as indicated in the next exercise would have been possible now, but the proof based on Skorokhod's theorem is more in keeping with the spirit of this book.] (Theorem 3.2.6 was used twice in (d).)

Consider the converse. Let $g_\epsilon(\cdot)$ equal 1, be linear, equal 0 on $(-\infty, x-\epsilon]$, on $[x-\epsilon, x]$, on $[x,\infty)$; and let $h_\epsilon(\cdot)$ equal 1, be linear, equal 0 on $(-\infty, x]$, on $[x, x+\epsilon]$, on $[x+\epsilon, \infty)$, with g_ϵ and h_ϵ both continuous. Then

$$F(x-\epsilon) \le \mathrm{E}\, g_\epsilon(X) = \lim \mathrm{E}\, g_\epsilon(X_n) \le \underline{\lim}\, F_n(x)$$

(e) $$\le \overline{\lim}\, F_n(x) \le \lim \mathrm{E}\, h_\epsilon(X_n) = \lim \mathrm{E}\, h_\epsilon(X) \le F(x+\epsilon),$$

so that $F_n(x) \to F(x)$ at all continuity points of F. □

Definition 5.2 (Determining class) Let $\mathcal{G}$ denote a collection of bounded and continuous functions g on the real line R. If for any rvs X and Y the condition

$$\mathrm{E}\, g(X) = \mathrm{E}\, g(Y) \text{ for all } g \in \mathcal{G} \qquad \text{implies} \qquad X \cong Y\,,$$

then call $\mathcal{G}$ a *determining class*. [The proof of the converse half of Helly–Bray exhibited one such class of particularly simple functions. (See also section 9.2 for further examples which will prove particularly useful.)]

Exercise 5.2 (a) Prove the Helly–Bray result $\int g\, dF_n \to \int g\, dF$ for all bounded and continuous g, without appeal to theorem 6.3.2 of Skorokhod. [Truncate the real line at large continuity points $\pm M$ of F, and then use the uniform continuity of g on the interval $[-M, M]$ to obtain a simple proof in this case. (Note exercise 9.1.1.)
(b) Alter your proof to be valid when g is bounded and merely continuous a.s. μ_F.

General Moment Convergence on $(\Omega, \mathcal{A}, \mu)$

Theorem 5.3 (Moment convergence under $\to_r$) Let $X_n \to_r X$, $r > 0$. Then

(8) $$\mathrm{E}|X_n|^r \to \mathrm{E}|X|^r.$$

Moreover, $X_n^+ \to_r X^+$, $X_n^- \to_r X^-$, and $|X_n| \to_r |X|$. (See also exercise 9.1.1.)

Proof. Let $(\Omega, \mathcal{A}, \mu)$ be arbitrary and $0 < r \le 1$. The C_r-inequality gives

(a) $$\mathrm{E}|X_n|^r \le \mathrm{E}|X_n - X|^r + \mathrm{E}|X|^r \quad \text{and} \quad \mathrm{E}|X|^r \le \mathrm{E}|X - X_n|^r + \mathrm{E}|X_n|^r,$$

so that

(9) $$|\,\mathrm{E}|X_n|^r - \mathrm{E}|X|^r\,| \le \mathrm{E}|X_n - X|^r \to 0 \qquad \text{when } 0 < r \le 1.$$

Suppose $r \ge 1$. Then using Minkowski's inequality twice (as in (a)) gives

(10) $$\left|\mathrm{E}^{1/r}|X_n|^r - \mathrm{E}^{1/r}|X|^r\right| \le \mathrm{E}^{1/r}|X_n - X|^r \to 0, \qquad \text{when } r \ge 1.$$

Combining (9) and (10) shows that $\mathrm{E}|X_n|^r \to \mathrm{E}|X|^r$. (Recall exercise 5.1(b).)

Now, $|X_n^+ - X^+|$ equals $|X_n - X|, |X_n - 0|, |0 - X|, |0-0|$ just as $[X_n \ge 0, X \ge 0]$ $[X_n \ge 0, X < 0], [X_n < 0, X \ge 0], [X_n < 0, X < 0]$. Thus

(11) $$|X_n^+ - X^+| \le |X_n - X|, \qquad \text{and} \qquad |X_n^- - X^-| \le |X_n - X|$$

also. Hence $X_n^+ \to_r X^+$, so that $\mathrm{E}(X_n^+)^r \to \mathrm{E}(X^+)^r$. Likewise for X_n^-. Cross-product terms are 0, since $X^+(\omega)X^-(\omega) = 0$; so if $r = k$ is integral, then

(b) $$\mathrm{E}X_n^k = \mathrm{E}(X_n^+)^k + (-1)^k \mathrm{E}(X_n^-)^k \to \mathrm{E}(X^+)^k + (-1)^k \mathrm{E}(X^-)^k = \mathrm{E}(X^k)\,.$$ □

Uniform Integrability and Vitali's Theorem

Definition 5.3 (Uniformly integrable) A collection of measurable X_t's is called *integrable* if $\sup_t \mathrm{E}|X_t| < \infty$. Further, a collection of rvs $\{X_t : t \in T\}$ is said to be *uniformly integrable* (which is abbreviated *u.i.*) if

(12) $$\sup_{t\in T} \mathrm{E}\{\,|X_t| \times 1_{[\,|X_t|\ge\lambda\,]}\} \to 0 \qquad \text{as } \lambda \to \infty\,.$$

(We will see below that u.i. implies 'integrable' when $\mu(\Omega) < \infty$.)

Remark 5.1 (Dominated X_t's are u.i.) Suppose these $|X_t| \le Y$ a.s. for some $Y \in \mathcal{L}_1$. Then these X_t's are integrable, in that $\sup_t \mathrm{E}|X_t| \le \mathrm{E}Y < \infty$. But, more is true. For some null sets N_t, we have $[|X_t| \ge \lambda] \subset [|Y| \ge \lambda] \cup N_t$. It follows that $\mu(|X_t| \ge \lambda) \le \mu(|Y| \ge \lambda) \to 0$ uniformly in t as $\lambda \to \infty$ (use Markov's inequality). Then for each fixed t,

$$\int_{[\,|X_t|\ge\lambda\,]} |X_t|\,d\mu \le \int_{[\,|Y|\ge\lambda\,]} Y\,d\mu \to 0 \qquad \text{uniformly in } t \text{ as } \lambda \to \infty$$

by the absolute continuity of the integral of Y in theorem 3.2.5. Thus:

(13) If $|X_t| \le Y$ for some $Y \in \mathcal{L}_1$, then these X_t's are uniformly integrable.

The functions $X_n(t) \equiv \frac{1}{n}\,1_{[-n^2,n^2]}(t)$ on $(R, \mathcal{B}, \lambda)$ are u.i. but not integrable. □

Exercise 5.3 (a) Now $\mathrm{E}Y = \int_0^\infty P(Y \ge y)\,dy = \int_0^\infty [1 - F(y)]\,dy$ for any rv $Y \ge 0$ with df F (as will follow from Fubini's theorem below). Sketch a proof.
(b) In fact, this formula can also be established rigorously now. Begin with simple functions Y and sum by parts. Then apply the MCT for the general result.
(c) Use the result of (a) to show that for $Y \ge 0$ and $\lambda \ge 0$ we have

$$\int_{[Y\ge\lambda]} Y\,dP = \lambda\,P(Y \ge \lambda) + \int_\lambda^\infty P(Y \ge y)\,dy. \qquad \text{(Draw pictures.)}$$

(d) Suppose there is a $Y \in \mathcal{L}_1$ such that $P(\,|X_n| \ge y) \le P(\,Y \ge y)$ for all $y > 0$ and all $n \ge 1$. Then use (b) to show that $\{X_n : n \ge 1\}$ is uniformly integrable.

Exercise 5.4 (Uniform integrability criterion) If $\sup_t \mathrm{E}|X_t|^r \le M < \infty$ for some $r > 1$, then the X_t's are uniformly integrable. (Compare this to theorem 5.6 of de la Vallée Poussin below, by letting $G(x) = x^r$.)

Theorem 5.4 (Uniform absolute continuity of integrals) Let $\mu(\Omega) < \infty$. A family of measurable X_t's is uniformly integrable if and only if both

(14) $$\sup_t \mathrm{E}|X_t| \le (\text{some } M) < \infty \qquad \text{(the collection is integrable)} \quad \text{and}$$

(15) $$\mu(A) < \delta_\epsilon \quad \text{implies} \quad \sup_t \int_A |X_t|\,d\mu < \epsilon \qquad \text{(uniform absolute continuity).}$$

Proof. Suppose these conditions hold. Then Markov's inequality and (14) give

(a) $$\mu(\,|X_t| \ge \lambda) \le \mathrm{E}|X_t|\,/\lambda \le \sup_t \mathrm{E}|X_t|\,/\lambda < \delta_\epsilon \qquad \text{uniformly in } t$$

for λ large enough. Then (15) applied to the sets $[\,|X_t| \ge \lambda]$ yields (12). (Note that $\mu(\Omega) < \infty$ was not used.)

Suppose the u.i. condition (12) holds. If $\mu(A) < \delta$, then

(b) $\int_A |X_t|\, d\mu = \int_A |X_t| \times 1_{[\,|X_t|<\lambda\,]}\, d\mu + \int_A |X_t| \times 1_{[\,|X_t|\geq\lambda\,]}\, d\mu$

(c) $\leq \lambda \times \mu(A) + \int |X_t|\, 1_{[\,|X_t|\geq\lambda\,]}\, d\mu \leq \epsilon/2 + \epsilon/2 = \epsilon$ using (12)

for a sufficiently large fixed λ and for $\delta \leq \epsilon/(2\lambda)$. (We have not yet used $\mu(\Omega) < \infty$.) Moreover, for λ large enough, (12) again gives

(d) $\mathrm{E}|X_t| \leq \lambda\, \mu(\Omega) + \int_{[\,|X_t|\geq\lambda\,]} |X_t|\, d\mu \leq \lambda\, \mu(\Omega) + 1$ for all t;

thus the collection is integrable. Thus (15) holds. □

Theorem 5.5 (Vitali) (i) Let $\mu(\Omega) < \infty$ and $r > 0$. Suppose that $X_n \to_\mu X$. The following are equivalent:

(16) $\{|X_n|^r : n \geq 1\}$ are uniformly integrable.

(17) $X_n \to_r X$.

(18) $\mathrm{E}|X_n|^r \to \mathrm{E}|X|^r < \infty$.

(19) $\overline{\lim}_n\, \mathrm{E}|X_n|^r \leq \mathrm{E}|X|^r < \infty$.

(20) The uniform absolute continuity of (15) holds for the $|X_n|^r$, and $X \in \mathcal{L}_r$.

(ii) Let $\mu(\Omega) = \infty$ and $r \geq 1$. Suppose $X_n \to_{\mu \text{ or } a.e.} X$ with all X_n in $\mathcal{L}_r$.
(a) Then (17), (18), and (19) are equivalent—and they imply (16).
(b) Suppose (15) holds and that for each $\epsilon > 0$ there exists a set A_ϵ for which both $\mu(A_\epsilon) < \infty$ and $\sup_n \int_{A_\epsilon^c} |X_n|^r\, d\mu \leq \epsilon$. This holds if and only if (17)–(19) hold.

Corollary 1 ($\mathcal{L}_r$-convergence) Let $\mu(\Omega) < \infty$. Let $r > 0$. Let all $X_n \in \mathcal{L}_r$. Then $X_n \to_r X$ (or, $\mathrm{E}|X_n - X|^r \to 0$) if and only if both $X_n \to_\mu X$ and one (hence both) of the two families of functions $\{|X_n|^r : n \geq 1\}$ or $\{|X_n - X|^r : n \geq 1\}$ is u.i.

Remark 5.2 (Vitali's theorem) Let $X_n \to_\mu X$ throughout, with $\mu(\Omega)$ arbitrary. Fatou and $X_{n''} \to_{a.e.} X$ on some further subsequence n'' of any n' *always* yield

(21) $\mathrm{E}|X|^r = \mathrm{E}\lim |X_{n''}|^r \leq \underline{\lim}\, \mathrm{E}|X_{n''}|^r \leq \overline{\lim}\, \mathrm{E}|X_{n''}|^r \leq \overline{\lim}\, \mathrm{E}|X_n|^r$.

So Vitali (ii)(a) yields $\mathrm{E}|X_n|^r \to \mathrm{E}|X|^r < \infty$, if $r \geq 1$. Vitali thus gives (for $r \geq 1$)

(22) $\mathrm{E}|X_n - X|^r \to 0$ if and only if $\mathrm{E}|X_n|^r \to \mathrm{E}|X|^r$ (any $\mu(\Omega)$ value)
if and only if (in case $\mu(\Omega) < \infty$) the rvs $\{|X_n|^r : n \geq 1\}$ are u.i. □

Exercise 5.5 Consider Vitali's theorem. In the proof that follows we will show that (17) ⇒ (18) ⇒ (19) ⇒ (16) for $r > 0$ and any $\mu(\Omega)$ value.
(p) Prove Vitali's (ii)(a) that (19) implies (17) when $r \geq 1$. (See exercise 5.10.)
(q) Prove the "true" Vitali theorem in (ii)(b). (Find a hint in exercise 5.10.)
(r) Give an example to demonstrate the implication that just (16) can hold in (ii).
(s) Note that $t \in [0, \infty)$ may replace $n \in \{1, 2, \ldots\}$ in all of Vitali's theorem.
(t) Let $r \geq 1$. Let $X_n \to_\mu X$, where $|X_n|^r \leq Y_n$ with $Y_n \to_\mu Y$ and $\mathrm{E}|Y_n| \to \mathrm{E}|Y|$ provides a bound. Show that $X_n \to_r X$.

Proof. Suppose (16); show (17). Now, $X_{n'} \to_{a.s.} X$ for some subsequence by theorem 2.3.1. Thus $\mathrm{E}|X|^r = \mathrm{E}(\underline{\lim}\,|X_{n'}|^r) \le \underline{\lim}\,\mathrm{E}|X_{n'}|^r \le M < \infty$ using Fatou and (14). Thus $X \in \mathcal{L}_r$. The C_r-inequality gives $|X_n - X|^r \le C_r\{|X_n|^r + |X|^r\}$. The $|X_n - X|^r$ are easily shown u.i. as in the theorem 5.4 proof. So for large n,

(j) $\quad \mathrm{E}|X_n - X|^r = \mathrm{E}\{|X_n - X|^r \times 1_{[\,|X_n - X|>\epsilon\,]}\} + \mathrm{E}\{|X_n - X|^r \times 1_{[\,|X_n - X|\le\epsilon\,]}\}$

(k) $\quad \le \epsilon + \epsilon^r \times \mu(\Omega)$ (only here (and for $\mathrm{E}|X|^r$ finite) was $\mu(\Omega) < \infty$ needed);

the ϵ in (k) is from (15), since $\mu(|X_n - X| \ge \epsilon) \to 0$ by hypothesis. Thus (17) holds.

Now (16) implies (20) by theorem 5.4, with $X \in \mathcal{L}_r$ by Fatou (as in the previous paragraph) and using $\mu(\Omega) < \infty$. We will not use $\mu(\Omega) < \infty$ again. Next, (20) implies (16) since $\mu(|X_n| \ge \lambda) \le \mu(|X_n - X| \ge \lambda/2) \cup \mu(|X| \ge \lambda/2) < \epsilon + \epsilon$, by first specifying λ large (as $X \in \mathcal{L}_r$) and then n large (as $X_n \to_\mu X$).

Now, (17) implies (18) by theorem 5.3. Also (18) trivially implies (19).

Suppose (19) holds. Define f_λ to be a continuous function on $[0, \infty)$ that equals $|x|^r$, 0, or is linear, according as $|x|^r \le \lambda$, $|x|^r \ge \lambda + 1$, or $\lambda \le |x|^r \le \lambda + 1$. Then (graphing $f_\lambda(x)$ and x^r on $[0, \lambda + 1]$), we have $Y_n \equiv f_\lambda(X_n) \to_\mu Y \equiv f_\lambda(X)$ by the uniform continuity of each f_λ. (See exercise 2.3.4(b).) Let n' denote any subsequence of n, and let n'' denote a further subsequence on which $X_{n''} \to_{a.e.} X$ (see (2.3.14) in theorem 2.3.1 of Riesz). On the subsequence n'' we then have

$$\overline{\lim}_{n''} \int_{[\,|X_{n''}|^r > \lambda+1]} |X_{n''}|^r \, d\mu$$

$$= \overline{\lim}_{n''} \{\textstyle\int |X_{n''}|^r \, d\mu - \int_{[\,|X_{n''}|^r \le \lambda+1]} |X_{n''}|^r \, d\mu\}$$

(l) $\quad \le \mathrm{E}|X|^r - \underline{\lim}_{n''} \int_{[\,|X_{n''}|^r \le \lambda+1]} |X_{n''}|^r \, d\mu \qquad$ by (19)

(m) $\quad \le \mathrm{E}|X|^r - \underline{\lim}_{n''}\, \mathrm{E}\, f_\lambda(X_{n''}) \le \mathrm{E}|X|^r - \mathrm{E}\, f_\lambda(X) \qquad$ by Fatou

(n) $\quad \le \int_{[\,|X|^r \ge \lambda\,]} |X|^r \, d\mu \to 0 \qquad$ as $\lambda \to \infty$, since $X \in \mathcal{L}_r$.

Thus, $\overline{\lim}_n \int_{[|X_n|^r > \lambda+1]} |X_n|^r \, d\mu \le \int_{[|X|^r \ge \lambda]} |X|^r \, d\mu \to 0$, which implies (16). □

Theorem 5.6 (de la Vallée Poussin) Let $\mu(\Omega) < \infty$. A family of $\mathcal{L}_1$-integrable functions X_t is uniformly integrable if and only if there exists a convex function G on $[0, \infty)$ for which $G(0) = 0$, $G(x)/x \to \infty$ as $x \to \infty$ and

(23) $\quad \sup_t \mathrm{E}\, G(|X_t|) < \infty$.

Proof. For λ so large that $G(x)/x \ge c$ for all $x \ge \lambda$ we have

(a) $\quad \int_{[\,|X_t| \ge \lambda\,]} |X_t| \, d\mu \le \frac{1}{c} \int_{[\,|X_t| \ge \lambda\,]} G(|X_t|)\, d\mu \le \frac{1}{c} \sup_t \mathrm{E}G(|X_t|\,) < \epsilon$

for c sufficiently large. Thus (23) implies $\{X_t : t \in T\}$ is uniformly integrable.

Now we show that $\{X_t : t \in T\}$ u.i. implies (23) for some G. We define $G(x) = \int_0^x g(y)\, dy$ where (with a sequence $b_n \nearrow$ having $b_0 = 0$, to be specified below) $g(x) \equiv b_n$ for all $n \le x < n + 1$, $n \ge 0$. Define $a_n(t) \equiv \mu(\,|X_t| \ge n)$. Note,

$$\mathrm{E}\, G(\,|X_t|\,) \le b_1\, \mu(1 \le |X_t| < 2) + (b_1 + b_2)\, \mu(2 \le |X_t| < 3) + \cdots$$

(b) $\quad = \sum_{n=1}^\infty b_n\, a_n(t)$.

Note that $G(n+2) \geq [n/2]\, g([n/2]) = [n/2]\, b_{[n/2]}$, so that $G(x)/x \to \infty$ as $x \to \infty$. It thus suffices to choose $b_n \nearrow \infty$ such that $\sup_t \sum_1^\infty b_n\, a_n(t) < \infty$. By the definition of uniform integrability, we can choose integers $c_n \uparrow \infty$ such that

(c) $\quad \sup_t \int_{[\,|X_t| \geq c_n]} |X_t|\, d\mu \leq 1/2^n$.

Thus for all t we have

$$1/2^n \geq \int_{[\,|X_t|\geq c_n]} |X_t|\, d\mu \geq \sum_{i=c_n}^\infty i\ \mu(i \leq |X_t| < i+1)$$

$$= \sum_{i=c_n}^\infty \sum_{j=1}^i \mu(i \leq |X_t| < i+1\,)$$

$$\geq \sum_{j=c_n}^\infty \sum_{i=j}^\infty \mu(i \leq |X_t| < i+1) = \sum_{j=c_n}^\infty \mu(\,|X_t| \geq j\,)$$

(d) $$= \sum_{j=c_n}^\infty a_j(t).$$

Thus, interchanging the order of summation,

(e) $\quad 1 = \sum_{n=1}^\infty 2^{-n} \geq \sup_t \sum_{n=1}^\infty \sum_{j=c_n}^\infty a_j(t) = \sup_t \sum_{j=1}^\infty b_j\, a_j(t)$

for $b_j \equiv$ (the number of integers n such that c_n is $\leq j$).

While u.i. yields a convex G in (23), u.i. follows from (23) without convexity. □

Exercise 5.6 Consider only the definition of u.i. Do not appeal to Vitali.
(a) Let $\xi \cong \text{Uniform}(0,1)$, and let $X_n \equiv (n/\log n)1_{[0,1/n]}(\xi)$ for $n \geq 3$. Show that these X_n are uniformly integrable and $\int X_n\, dP \to 0$, even though these rvs are not dominated by any fixed integrable rv Y.
(b) Let $Y_n \equiv n1_{[0,1/n]}(\xi) - n1_{[1/n,2/n]}(\xi)$. Show that these Y_n are not uniformly integrable, but that $\int Y_n\, dP \to 0$. However, $\int |Y_n|\, dP \not\to 0$.

Summary of Modes of Convergence Results

Theorem 5.7 (Convergence implications) Let X and X_n's be measurable and a.e. finite. (Note figure 5.1.)

(i) If $X_n \to_{a.e.} X$ and $\mu(\Omega) < \infty$, then $X_n \to_\mu X$.

(ii) If $X_n \to_\mu X$, then $X_{n'} \to_{a.e.} X$ on some subsequence n'.

(iii) If $X_n \to_r X$, then $X_n \to_\mu X$ and $\{|X_n|^r : n \geq 1\}$ are uniformly integrable.

(iv) Let $r \geq 1$. If $X_n \to_{\mu \text{ or } a.e.} X$ and $\overline{\lim}\ \mathrm{E}|X_n|^r \leq \mathrm{E}|X|^r < \infty$, then $X_n \to_r X$.

Let $\mu(\Omega) < \infty$. If $X_n \to_\mu X$ and $\{|X_n|^r : n \geq 1\}$ are u.i., then $X_n \to_r X$.

(v) If $X_n \to_r X$ and $\mu(\Omega) < \infty$, then $X_n \to_{r'} X$ for all $0 < r' < r$.

(vi) If $X_n \to_p X$, then $X_n \to_d X$.

(vii) Let $\mu(\Omega) < \infty$. Then $X_n \to_\mu X$ if and only if every subsequence $\{n'\}$ contains a further subsequence $\{n''\}$ for which $X_{n''} \to_{a.e.} X$.

(viii) If $X_n \to_d X$, then $Y_n \to_{a.e.} Y$ for Skorokhod rvs with $Y_n \cong X_n$ and $Y \cong X$.

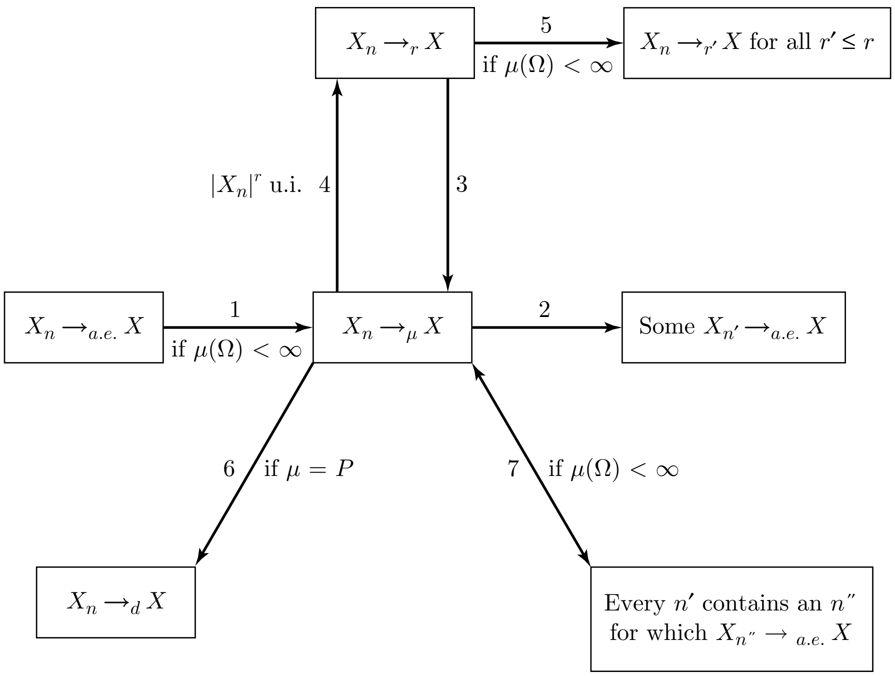


Figure 5.1 Convergence implications.

Proof. See theorem 2.3.1 for (i) and (ii). Markov's inequality gives (iii) via

$$\mu(\,|X_n - X| \geq \epsilon) \leq \mathrm{E}|X_n - X|^r/\epsilon^r \to 0.$$

Vitali's theorem includes both halves of (iv). Hölder's inequality gives (v) via

(a) $$\mathrm{E}|X_n - X|^{r'} \leq \{\mathrm{E}|X_n - X|^{r'(r/r')}\}^{r'/r}\,(\mathrm{E}1)^{1-r'/r}, \quad (\text{with } \mathrm{E}1 = \mu(\Omega));$$

note also exercise 5.1(b) and the proof of inequality 3.4.4(b). Proposition 2.4.1 gives (vi). Theorem 2.3.1 then gives (vii). The Skorokhod construction (to appear more formally as theorem 6.3.2 below) was stated above in (7); (7) gives (viii). □

Approximation of Functions in $\mathcal{L}_r$ by Continuous Functions

Let C_c denote the class of continuous functions on R that vanish outside a compact set, and then let $C_c^{(\infty)}$ denote the subclass that has an infinite number of continuous derivatives. Let $\mathcal{S}_c$ denote the class of all *step functions* on R, where such a step function is of the form $\sum_1^m y_j 1_{I_j}$ for disjoint finite intervals I_j. Further, let F denote a generalized df, and let $\mu(\cdot) \equiv \mu_F(\cdot)$ denote the associated Lebesgue–Stieltjes measure. Let X denote a measurable function on $(\Omega, \mathcal{A}, \mu) = (R, \mathcal{B}, \mu_F)$.

Theorem 5.8 (The continuous functions are dense in $\mathcal{L}_r(R, \mathcal{B}, \mu_F)$, $r \geq 1$) Suppose throughout that $X \in \mathcal{L}_r$, for some fixed $1 \leq r < \infty$.
(a) (Continuous functions) Then for each $\epsilon > 0$ there is a bounded and continuous function Y_ϵ in C_c for which $\int |X - Y_\epsilon|^r \, d\mu_F < \epsilon$. Thus the class C_c is ϵ-dense within the class $\mathcal{L}_r$ under the $\|\cdot\|_r$–norm.
(b) We may insist that $Y_\epsilon \in C_c^{(\infty)}$. (The Y_ϵ of exercise 5.17 has $\sup |Y_\epsilon| \leq \sup |X|$.)
(c) (Step functions) Such a close approximation may also be found within the step functions $\mathcal{S}_c$, making them ϵ-dense also.
(d) All this extends to rvs on $(R_n, \mathcal{B}_n)$ (or on locally compact Hausdorff spaces).
(e) All these spaces $\mathcal{L}_r$ are separable, provided μ is σ-finite and $\mathcal{A}$ is *countably generated* (that is, $\mathcal{A} = \sigma[\mathcal{C}]$ with $\mathcal{C}$ a countable collection of sets).

Proof. Let $r = 1$. Consider only X^+. Approximate it by a simple function $X_\epsilon = \sum_1^\kappa x_i 1_{A_i}$ of (2.2.10) so closely that $\int |X^+ - X_\epsilon| \, d\mu_F < \epsilon/3$. (We can require that all $A_i \subset$ (some $[-M_\epsilon, M_\epsilon]$) for which $\int_{[|X^+| \geq M_\epsilon]} X^+ \, d\mu_F < \epsilon/3$, and that each $x_i > 0$.) Now, the Halmos approximation lemma of exercise 1.2.3 guarantees sets $B_1, \ldots, B_n$ made up of a finite disjoint union of intervals of the form $(a, b]$ (with a and b finite continuity points of F, as in (b) of the proof of theorem 1.3.1) for which

(p) $\quad \mu_F(A_i \triangle B_i) < \epsilon/(3\kappa |x_i|)$, and so $X'_\epsilon \equiv \sum_1^\kappa x_i 1_{B_i}$ satisfies
$\quad \int |X_\epsilon - X'_\epsilon| \, d\mu_F < \epsilon/3$. (note that these B_i need not be disjoint).

This X'_ϵ is the step function called for in part (c). Rewrite this $X'_\epsilon = \sum_1^m y_j 1_{C_j}$ with disjoint $C_j = (a_j, b_j]$. Now approximate 1_{C_j} by the continuous function W_j that equals 0, is linear, equals 1 according as $x \in [a_j, b_j]^c$, as $x \in [a_j, a_j + \delta] \cup [b_j - \delta, b_j]$, as $x \in [a_j + \delta, b_j - \delta]$. (We require that δ be specified so small that the combined μ_F measure of all $2m$ sets of the type $x \in [a_j, a_j + \delta]$ and $[b_j - \delta, b_j]$ is at most $\theta \equiv \epsilon/(6 \sum_1^m y_j)$. Then let $Y_\epsilon \equiv \sum_1^m c_j W_j$, which has $\int |X'_\epsilon - Y_\epsilon| \, d\mu_F < \epsilon/3$. Thus $\int |X - Y_\epsilon| \, d\mu_F < \epsilon$, as called for in part (a). For (b), the function $\psi(x/\delta)$ [where

$$\psi(x) \equiv \frac{\int_x^1 \exp(-1/((s(1-s))) \, ds}{\int_0^1 \exp(-1/((s(1-s))) \, ds} \qquad \text{for } 0 \leq x \leq 1, \tag{24}$$

with $\psi(x)$ equal to 1 or 0 according as $x \leq 0$ or $x \geq 1$ is easily seen to have an infinite number of continuous derivatives on R (with all said derivatives equal to 0 when x equals 0 or 1)]. Use $\psi(-x/\delta)$ on $[a_j, a_j + \delta]$ and $\psi(x/\delta)$ on $[b_j - \delta, b_j]$ to connect values 0 to 1, instead of linear connections. The result is a function in $C_c^{(\infty)}$.

For $r > 1$, write $X = X^+ - X^-$ and use the C_r-inequality and $|a-b|^r \le |a^r - b^r|$ for all $a, b \ge 0$. For example, make $\mathrm{E}|X^+ - Y_\epsilon^+|^r \le \mathrm{E}|(X^+)^r - (Y_\epsilon^+)^r| < \epsilon$ by the case $r = 1$. (Exercise 5.18 asks for a proof of (e).) □

Miscellaneous Results

Exercise 5.7 (Scheffé's theorem) Let $f_o, f_1, f_2, \ldots$ be ≥ 0 on $(\Omega, \mathcal{A}, \mu)$. Prove the following without resort to Vitali. Then prove them via Vitali.
(a) Suppose $\int_\Omega f_n \, d\mu = 1$ for all $n \ge 0$, and $f_n \to_{a.e.} f_o$ with respect to μ, then

$$(25) \qquad \sup_{A \in \mathcal{A}} \left| \int_A f_n \, d\mu - \int_A f_o \, d\mu \right| \le \int_\Omega |f_n - f_o| \, d\mu \to 0 \quad \text{as } n \to \infty.$$

(Think of this as the uniform convergence of measures with densities f_n.)
[Hint. Integrate $(f_o - f_n)^+$ and $(f_o - f_n)^-$ separately. Note that $(f_o - f_n)^+ \le f_o$.]
(b) Show that $\overline{\lim} \int_\Omega f_n \, d\mu \le \int_\Omega f_o \, d\mu < \infty$ and $f_n \to_{\mu \text{ or } a.e.} f_o$ is sufficient for (25).

Exercise 5.8*($\to_{a.u.}$, Egorov) (i) We define $X_n \to_{a.u.} X$ (which is used as an abbreviation for *almost uniform convergence*) to mean that for all $\epsilon > 0$ there exists an A_ϵ with $\mu(A_\epsilon) < \epsilon$ such that $X_n \to_{\text{uniformly}} X$ on A_ϵ^c. Recall (2.3.6) to show

(26) **(Egorov)** If $\mu(\Omega) < \infty$ and $X_n \to_{a.e.}$ (some X), then $X_n \to_{a.u.} X$.

If $|X_n| \le Y$ a.e. for all n where $Y \in \mathcal{L}_r$ with $r > 1$, then $\mu(\Omega) < \infty$ is not needed.
(ii) (a) If $X_n \to_{a.u.} X$, then both $X_n \to_{a.e.} X$ and $X_n \to_\mu X$.
(b) If $X_n \to_\mu X$, then $X_{n'} \to_{a.u.} X$ on some subsequence n'.

Exercise 5.9 (a) (Approximating integrable functions) Suppose $\mu(\Omega) \in [0, \infty]$ and $\int_\Omega |X| \, d\mu < \infty$. Fix $\epsilon > 0$. Show the existence of a set A_ϵ with $\mu(A_\epsilon) < \infty$ for which both $|X| \le$ (some M_ϵ) on A_ϵ and $\int_{A_\epsilon^c} |X| \, d\mu < \epsilon$.
(b) Let $\mu(\Omega) < \infty$. Let X be measurable and finite a.e. For any $\epsilon > 0$, specify a finite number M_ϵ and a set A_ϵ having $\mu(A_\epsilon^c) < \epsilon$ and $|X| \le M_\epsilon$ on A_ϵ.

Exercise 5.10 Verify Vitali's theorem 5.5(ii)(a) when $\mu(\Omega) = \infty$ is allowed. [Hint. Apply exercise 5.9, Scheffé's theorem, absolute continuity of the integral, Egorov's theorem, and exercise 3.4.12.]

Exercise 5.11 (ℓ_r-Spaces) Let Ω be an arbitrary set and consider the class of all subsets A. Let $\mu(A)$ denote the cardinality of A when this is finite, and let it equal ∞ otherwise. This is *counting measure* on Ω. Let $0 < r < \infty$. Let $\ell_r(\Omega)$ denote all functions $X : \Omega \to R$ for which $\sum_{\omega \in \Omega} |X(\omega)|^r < \infty$. Then

$$(27) \qquad \|X\|_r = \{\textstyle\sum_{\omega \in \Omega} |X(\omega)|^r\}^{1/r}, \qquad \text{for all } r \ge 1,$$

defines a norm on $\ell_r(\Omega)$ (see (3.4.8) for $0 < r < 1$). This is just a special case of an $\mathcal{L}_r$-space, and it is important enough to deserve its specialized notation. Show that

$$(28) \qquad \ell_r \subset \ell_s \qquad \text{for all } 0 < r < s < \infty.$$

This set inclusion is proper if Ω has infinitely many points.

The exercises below presented for "flavor" or as tools, rather than to be worked.

Exercise 5.12*(Weak $\mathcal{L}_r$-convergence; and in $\mathcal{L}_\infty$) Let $X_n, X \in \mathcal{L}_r$, with $r \geq 1$. Let $1/r + 1/s = 1$ define s for $r > 1$. Let $s = \infty$ when $r = 1$, and $\mathcal{L}_\infty$ denotes all bounded $\mathcal{A}$-measurable Y on Ω, and let $\|X\|_\infty \equiv \inf\{c : \mu(\{\omega : |X(\omega)| > c\}) = 0\}$ denote the *essential supremum* of such functions X. Let $s = 1$ when $r = \infty$. (The following results can be compared with Vitali's theorem.)

(A) (a) Fix $1 \leq r < \infty$. Let $X_n \to_r X$ on $\mathcal{L}(\Omega, \mathcal{A}, \mu)$. Show that X_n *converges weakly in* $\mathcal{L}_r$ (denoted by $X_n \to_{w-\mathcal{L}_r} X$) in that

(29) $\quad \int_\Omega X_n\, Y\, d\mu \to \int_\Omega X\, Y\, d\mu \qquad$ for all $Y \in \mathcal{L}_s$.

(b) (Radon-Reisz) Conversely, suppose that $X_n \to_{w-\mathcal{L}_r} X$ and additionally that the moments satisfy $\mathrm{E}|X_n|^r \to \mathrm{E}|X|^r$, where $1 < r < \infty$. Show that $X_n \to_{\mathcal{L}_r} X$.

(B) (c) Let $X_n \to_{\mu \text{ or } a.e.} X$. Let $1 < r < \infty$ and $\sup_n \mathrm{E}|X_n|^r < \infty$. Show (29). (Recall Scheffé's theorem regarding $r = 1$.)

(C) (d) (Lehmann) Fix M. Let $\mathcal{F}_M \equiv \{X : \|X\|_\infty \leq M < \infty\}$. Let $X, X_1, X_2, \ldots$ denote specific functions in $\mathcal{F}_M$. Then (29) holds for all $Y \in \mathcal{L}_1$ if and only if (29) holds for all Y in the subclass $\{1_A : \mu(A) < \infty\}$. (Note also exercise 5.21 below.)

Exercise 5.13* (a) $\mathcal{L}_\infty(\Omega, \sigma[\{\text{all open sets}\}], \mu)$ is a complete metric space under the essential sup norm $\|\cdot\|_\infty$ whenever Ω is a locally compact Hausdorff space.
(b) The set $\mathcal{S}_c$ of simple functions that vanish off of compact sets is dense in this complete metric space $(\mathcal{L}_\infty, \|\cdot\|_\infty)$. (Recall theorem 5.8.)
(c) No family of continuous functions is dense in $(\mathcal{L}_\infty([0,1], \mathcal{B}, \lambda), \|\cdot\|_\infty)$, and the space $\mathcal{L}_\infty$ is not separable under the norm $\|\cdot\|_\infty$.

Exercise 5.14*(Loève) Suppose $X_1, X_2, \ldots$ are integrable on $(\Omega, \mathcal{A}, \mu)$. Define $\phi_n(A) \equiv \int_A X_n\, d\mu$ for all $A \in \mathcal{A}$, and suppose $\phi_n(A)$ converges to a finite number for all $A \in \mathcal{A}$. Define $|\phi|_n(A) \equiv \int_A |X_n|\, d\mu$. Then $\sup_n |\phi|_n(\Omega) < \infty$. Moreover,

(30) $\quad \sup_n |\phi|_n(A) \to 0 \qquad$ as either $\mu(A) \to 0$ or $A \searrow 0$.

Finally, there exists an integrable function X (that is unique a.e. μ) for which $\phi_n(A) \to \phi(A)$ for all $A \in \mathcal{A}$, with $\phi(A) \equiv \int_A X\, d\mu$. (Relate this to the absolute continuity of measures introduced in the next chapter.)

Exercise 5.15*(Lusin) Suppose $X_n \to_{a.e.} X$, with μ being σ-finite on $(\Omega, \mathcal{A})$. Determine a measurable decomposition $\Omega = A_0 + A_1 + A_2 + \cdots$ for which $\mu(A_0) = 0$ and $X_n \to X$ uniformly on each of $A_1, A_2, \ldots$.

Exercise 5.16*(Lusin) Let X be an $(R, \mathcal{B})$ measurable function on R.
(a) Let $\epsilon > 0$. Show that there exists a continuous function Y_ϵ on R and a closed set D_ϵ such that $\lambda(D_\epsilon^c) < \epsilon$ and $X = Y_\epsilon$ on D_ϵ.
(b) Show that a function $X : R \to R$ is $\mathcal{B}$-measurable if and only if there exists a sequence of continuous function $Y_n : R \to R$ for which $Y_n \to_{a.e.} X$.
[Hint. (a) Begin with simple functions like those in (2.2.10). Consider each $[n, n+1]$ separately. Apply Egorov's theorem.]

Exercise 5.17*(Lusin) Let X be measurable on $(\Omega, \mathcal{A}, \mu)$, where Ω is a locally compact Hausdorff space [every point has a neighborhood whose closure is compact, such as the real line R with the usual Euclidean metric] and $\mathcal{A} = \sigma[\{\text{open sets}\}]$. Suppose $X(\omega) = 0$ for all $\omega \in A^c$, where $\mu(A) < \infty$. Let $\epsilon > 0$. Then there exists Y_ϵ, where $Y_\epsilon(\omega) = 0$ for all $\omega \in B^c$, with the set B compact, and where Y_ϵ is continuous, $\sup |Y_\epsilon| \le \sup |X|$, and $\mu(\{\omega : X(\omega) \neq Y_\epsilon(\omega)\}) < \epsilon$. [That is, a measurable function is "almost equal" to a continuous function.] (Note exercise **??.??** below.)

Exercise 5.18* Prove the separability of $\mathcal{L}_r$ in theorem 5.8(e).

Exercise 5.19*(Halmos) Let $(\Omega, \mathcal{A}, \mu)$ be σ-finite. For each A_1, A_2 in $\mathcal{A}$ define $\rho(A_1, A_2) = \mu(A_1 \triangle A_2)$. We agree to identify all members of the equivalence classes of subsets $A \equiv \{A' : \rho(A, A') = 0\}$.) Let $\mathcal{A}_0$ denote the collection of all of these equivalence classes A that satisfy $\mu(A) < \infty$.
(a) Show that $(\mathcal{A}_0, \rho)$ is a metric space.
(b) Show that the metric space $(\mathcal{A}_0, \rho)$ is separable whenever $\mathcal{A} = \sigma[\mathcal{C}]$ for some countable collection $\mathcal{C}$ (that is, whenever $\mathcal{A}$ is countably generated).

Definition 5.4 (Dominated families of measures) Suppose that $\mathcal{M}$ is a family of measures μ on some $(\Omega, \mathcal{A})$ having $\mu \ll \mu_0$ for some σ-finite measure $(\Omega, \mathcal{A}, \mu_0)$. Denote this by $\mathcal{M} \ll \mu_0$, and say that $\mathcal{M}$ is *dominated* by μ_0. Show that there exists a probability distribution P_0 on $(\Omega, \mathcal{A})$ for which $\mu \ll P_0$ for all $\mu \in \mathcal{M}$; that is, for which $\mathcal{M} \ll P_0$. (Note definition **??.??**.)

Exercise 5.20*(Berger) Let $\mathcal{P}$ denote a collection of probability measures P on the measurable space $(\Omega, \mathcal{A})$. Suppose $\mathcal{A} = \sigma[\mathcal{C}]$ for some countable collection $\mathcal{C}$; that is, $\mathcal{A}$ is countably generated. Let d_{TV} denote the total variation metric on $\mathcal{P}$; see exercise 4.2.10 below. Show that

(31) $\mathcal{P}$ is dominated if and only if $(\mathcal{P}, d_{TV})$ is a separable metric space.

(For example, let $\mathcal{P}$ denote all Poisson(λ) distributions on $0, 1, 2, \ldots$ having $\lambda > 0$. The countable collection of distributions with λ is rational is dense in $(\mathcal{P}, d_{TV})$.) [Hint. Use the previous exercise.]

Exercise 5.21*(Lehmann) Suppose that $(\Omega, \mathcal{A}, \mu)$ is σ-finite and $\mathcal{A}$ is countably generated. Let Φ denote the set of all $\mathcal{A}$-measurable ϕ for which $0 \le \phi(\omega) \le 1$ for all $\omega \in \Omega$. Consider an arbitrary sequence $\phi_n \in \Phi$. Show that a subsequence n' and a function $\phi \in \Phi$ must exist for which

(32) $\int_\Omega \phi_{n'} \, f \, d\mu \to \int_\Omega \phi \, f \, d\mu$ for all $f \in \mathcal{L}_1(\Omega, \mathcal{A}, \mu)$

(that is, $\phi_{n'} \to_{w-\mathcal{L}_1} \phi$, or $\phi_{n'}$ converges weakly in $\mathcal{L}_1$ to ϕ in the sense of (29)). [Hint. By exercise 5.12(d) it suffices to verify (32) for all $f = 1_A$ with $\mu(A) < \infty$.]

Exercise 5.22 (An added touch) Let $a_n \ge 0$ satisfy $\sum_1^\infty a_n < \infty$. Show that there necessarily exists a sequence $c_n \uparrow \infty$ for which $\sum_1^\infty c_n a_n < \infty$.

Chapter 4

Derivatives via Signed Measures

0 Introduction

In a typical calculus class the derivative $F'(x)$ of a function F at x is defined as the limit of the difference quotients $[F(x+h) - F(x)]/h$ as $h \to 0$. One of the major theorems encountered is then the Fundamental Theorem of Calculus that expresses F as the integral of its derivative (with this result formulated on some interval $[a, b]$ with respect to ordinary Lebesgue measure $d\lambda = dx$). We can thus write $F(x) - F(a) = \int_a^x F'(y)\, dy$ under appropriate hypothesis on F. In the context of an elementary probability class we let $f \equiv F'$ and rewrite the fundamental result as $P(\,[a, x]\,) = \int_{[a,x]} f(y)\, dy = \int_a^x f(y)\, dy$ for all $a \le x \le b$, or even as

(1) $\qquad P(A) = \int_A f(y)\, dy \qquad$ for all events A of the form $[a, x]$.

Let us now turn this order around and begin by defining one function ϕ as the "indefinite integral" of another function X, and do it on an arbitrary measure space $(\Omega, \mathcal{A}, \mu)$. Thus for a fixed $X \in \mathcal{L}_1(\Omega, \mathcal{A}, \mu)$, define

(2) $\qquad \phi(A) = \int_A X\, d\mu \qquad$ for all $A \in \mathcal{A}$.

As example 4.1.1 will show, if $X \ge 0$ then this ϕ is a measure on $(\Omega, \mathcal{A})$; in general $\phi(A) \equiv \int_A X\, d\mu = \int_A X^+\, d\mu - \int_A X^-\, d\mu$ is the difference of two measures, and is thus called a "signed measure." As (2) suggests, we can think of X as a derivative of the signed measure ϕ with respect to the measure μ. This is the so called "Radon-Nikodym derivative." In this context it is possible to formulate important general questions that have clean conclusions via straight forward and/or clever proofs. This is done in section 4.1 and section 4.2, and this gives us most of what we need as we go forward. But before going on, in section 4.3 and section 4.4 we relate this new approach back to the more pedantic and more limited standard approach represented by (1). Of course, the f in (1) must equal the Radon-Nikodym derivative (viewed in the new context); but much is gained by this new perspective.

1 Decomposition of Signed Measures

Definition 1.1 (Signed measure) A *signed measure* on a σ-field (or a field) $\mathcal{A}$ is a set function $\phi : \mathcal{A} \to (-\infty, +\infty]$ for which $\phi(\emptyset) = 0$ and $\phi(\sum A_n) = \sum \phi(A_n)$ for all countable disjoint sequences of A_n's in $\mathcal{A}$ (requiring $\sum A_n$ in $\mathcal{A}$ in the case of a field). When additivity is required only on finite unions, then ϕ is called a *finitely additive* (f.a.) *signed measure*. (If $\phi \geq 0$, then ϕ is a measure or f.a. measure.) If $|\phi(\Omega)| < \infty$, then ϕ is called *finite*. If $\Omega = \sum_1^\infty \Omega_n$ with all components $\Omega_n \in \mathcal{A}$ and all values $|\phi(\Omega_n)| < \infty$, then ϕ is called σ-*finite*.

Proposition 1.1 (Elementary properties) (i) If $\phi(A)$ is finite and $B \subset A$, then $\phi(B)$ is finite. Thus $\phi(\Omega)$ finite is equivalent to $\phi(A)$ being finite for all $A \in \mathcal{A}$.
(ii) If $|\phi(\sum_1^\infty A_n)| < \infty$, then $\sum_1^\infty |\phi(A_n)| < \infty$ (so, it is absolutely convergent).

Proof. (i) Now,

(a) $\quad$ (a finite number) $= \phi(A) = \phi(B) + \phi(A \setminus B)$

implies that $\phi(B)$ and $\phi(A \setminus B)$ are both finite numbers.
(ii) Let A_n^+ equal A_n or $\emptyset$ as $\phi(A_n)$ is ≥ 0 or < 0. And let A_n^- equal A_n or $\emptyset$ as $\phi(A_n)$ is ≤ 0 or > 0. Then $\sum \phi(A_n^+) = \phi(\sum A_n^+) < \infty$ by (i), since $\sum A_n^+ \subset \sum A_n$. Likewise, $\sum \phi(A_n^-) = \phi(\sum A_n^-)$. Now, convergent series of numbers in $[0, \infty)$ may be rearranged at will. Thus $\sum |\phi(A_n)| = \sum \phi(A_n^+) - \sum \phi(A_n^-)$ is finite. $\square$

Example 1.1 (The prototypical example) Let X be measurable. Then

(1) $\quad \phi(A) \equiv \int_A X \, d\mu \quad$ is a signed measure if $X^- \in \mathcal{L}_1$.

Note that ϕ is finite if $X \in \mathcal{L}_1$. Also, ϕ is σ-finite if X is a.e. finite and μ is σ-finite.

Proof. Now, $\phi(\emptyset) = \int_\emptyset X \, d\mu = \int X \cdot 1_\emptyset \, d\mu = \int 0 \, d\mu = 0$. Also,

(a) $\quad \phi(A) = \int_A X^+ \, d\mu - \int_A X^- \, d\mu \geq -\int_A X^- \, d\mu > -\infty$

for all $A \in \mathcal{A}$. Finally,

$$\phi(\textstyle\sum_1^\infty A_n) = \int_{\Sigma A_n} X = \int_{\Sigma A_n} X^+ - \int_{\Sigma A_n} X^- \quad \text{with the } X^-\text{-term finite}$$

$$= \textstyle\sum \int_{A_n} X^+ - \sum \int_{A_n} X^- \qquad \text{by the MCT, twice}$$

(b) $$= \textstyle\sum (\int_{A_n} X^+ - \int_{A_n} X^-) = \sum \int_{A_n} X = \sum_1^\infty \phi(A_n).$$

Thus ϕ is a signed measure.

Note that $|\phi(A)| = |\int_A X| \leq \int_A |X| \leq \int |X| < \infty$ for all A, if $X \in \mathcal{L}_1$.

Let $\Omega \equiv \sum_n \Omega_n$ be a measurable decomposition for the σ-finite μ. Then the sets $\Omega_{nm} \equiv \Omega_n \cap [m \leq X < m+1]$ and $\Omega_{n,\pm\infty} \equiv \Omega_n \cap [X = \pm\infty]$, for $n \geq 1$ and for all integers m, is a decomposition showing ϕ to be σ-finite. $\square$

Definition 1.2 (Continuous signed measure) A signed measure ϕ is *continuous from below* (*above*) if $\phi(\lim A_n) = \lim \phi(A_n)$ for all $A_n \nearrow$ (for all $A_n \searrow$, with at least one $\phi(A_n)$ finite). We call ϕ *continuous* in case it is continuous both from below and from above.

Proposition 1.2 (Continuity of signed measures)
A signed measure on either a field or a σ-field is countably additive and continuous. Conversely, if a finitely additive signed measure on either a field or σ-field is either continuous from below or is finite and continuous from above at $\emptyset$, then it is a countably additive signed measure.

Proof. This result has nearly the same proof as does the corresponding result for measures; see proposition 1.1.4. □

Exercise 1.1 (a) Actually write out all details of the proof of proposition 1.2.
(b) If ϕ and ψ are signed measures, then so is $\phi + \psi$.

Theorem 1.1 (Jordan–Hahn decomposition) Let ϕ be a signed measure on the measurable space $(\Omega, \mathcal{A})$, having events $\mathcal{A}$. Then Ω can be decomposed into events as $\Omega = \Omega^+ + \Omega^-$, where

(2) Ω^+ is a *positive set* for ϕ, in that $\phi(A) \geq 0$ for all events $A \subset \Omega^+$,

(3) Ω^- is a *negative set* for ϕ, in that $\phi(A) \leq 0$ for all events $A \subset \Omega^-$.

Moreover, we obtain measures on the measurable space $(\Omega, \mathcal{A})$ via the definitions

$$\phi^+(A) \equiv \phi(A \cap \Omega^+) \quad \text{and} \quad \phi^-(A) \equiv -\phi(A \cap \Omega^-) \qquad (\text{with} \quad \phi = \phi^+ - \phi^-), \tag{4}$$

with ϕ^+ a measure and ϕ^- a finite measure on $(\Omega, \mathcal{A})$. Of course, $\phi^+(\Omega^-) = 0$ and $\phi^-(\Omega^+) = 0$. We will call ϕ^+, ϕ^-, and $|\phi|(\cdot) \equiv \phi^+ + \phi^-$ the *positive part*, the *negative part*, and the *total variation measure* associated with ϕ; thus

$$|\phi|(\cdot) \equiv \phi^+(\cdot) + \phi^-(\cdot) \qquad \text{is the total variation measure on } (\Omega, \mathcal{A}), \tag{5}$$

Moreover, the following relationships hold:

$$\begin{aligned} \phi^+(A) &= \sup\{\phi(B) : B \subset A, B \in \mathcal{A}\}\,, \\ \phi^-(A) &= -\inf\{\phi(B) : B \subset A, B \in \mathcal{A}\}. \end{aligned} \tag{6}$$

Exercise 1.2 Identify ϕ^+, ϕ^-, $|\phi|$, and $|\phi|(\Omega)$ in the context of the prototypical situation of example 1.1. Be sure to specify Ω^+ and Ω^-.

Proof. Let us note first that (6) follows from the previous parts of the theorem. If $B \subset A$ then $\phi(B) = \phi(B\Omega^+) + \phi(B\Omega^-) \leq \phi(B\Omega^+) = \phi^+(B) \leq \phi^+(A)$, while equality is actually achieved for the particular subset $A\Omega^+$. Thus, (6) holds.

Consider claims (2) and (3). Let B denote some set having $\phi(B) < 0$. [That $\phi(B) > -\infty$ is crucial; this proof will not work on the positive side.] (If no such set exists, let $\Omega^+ \equiv \Omega$, giving $|\phi| = \phi^+ = \phi$ and $\phi^- \equiv 0$.) We now show that

(a) B contains a negative set C.

If B is a negative set, use it for C. If not, then we will keep removing sets A_k with $\phi(A_k) > 0$ from B until only a negative set C is left. We will remove disjoint sets A_k with $\phi(A_k) \geq 1$ as many times as we can, then sets with $\phi(A_k) \geq \frac{1}{2}$ as many times as we can, ... To this end, let

(b)
$$\begin{aligned} &n_1 \equiv \min\{i : \phi(A_1) \geq 1/i \text{ for some } A_1 \subset B, \text{ with } A_1 \in \mathcal{A}\}, \\ &\cdots \\ &n_k \equiv \min\{i : \phi(A_k) \geq 1/i \text{ for some } A_k \subset B \setminus \textstyle\sum_{j=1}^{k-1} A_j, \text{ with } A_k \in \mathcal{A}\} \\ &\cdots . \end{aligned}$$

(If $(n_1, n_2, \ldots) = (1, 1, 2, 2, 2, 3, \ldots)$, then $\phi(A_i) \geq 1$ for $1 \leq i \leq 2$, $\frac{1}{2} \leq \phi(A_i) < 1$ for $3 \leq i \leq 5$, ...) Let $C \equiv B \setminus \sum_k A_k$, where the union is infinite (unless the process of choosing n_k's terminates) and where only finitely many sets A_k exist for each $1/i$ [else proposition 1.1(i) would be violated]. The c.a. of ϕ then gives

(c) $0 > \phi(B) = \phi(C) + \sum_k \phi(A_k) \geq \phi(C) > -\infty.$

Moreover, C is a negative set, since no subset can have measure exceeding $1/i$ for any i. Now we know that we have at least one negative set. So we let

(d) $d \equiv \inf\{\phi(C) : C \text{ is a negative set }\} < 0$, and define $\Omega^- \equiv \cup_k C_k$,

where C_k denotes a sequence of negative sets for which $\phi(C_k) \searrow d$. Now, Ω^- is also a negative set (else one of the C_k's would not be), and thus $\phi(\Omega^-) \geq d$, because it must exceed the infimum of such values. But $\phi(\Omega^-) \leq d$ also holds, since $\phi(\Omega^-) = \phi(C_k) + \phi(\Omega^- \setminus C_k) \leq \phi(C_k)$ for all k gives $\phi(\Omega^-) \leq d$. Thus $\phi(\Omega^-) = d$; so, d must be finite. Then Ω^+ is a positive set, since if $\phi(A) < 0$ for some $A \subset \Omega^+$, then the set $\Omega^- \cup A$ would have $\phi(A \cup \Omega^-) < d$ (which is a contradiction). □

Exercise 1.3 The set Ω^+ is essentially unique, in that if Ω_1^+ and Ω_2^+ both satisfy the theorem, then $|\phi|(\Omega_1^+ \triangle \Omega_2^+) = 0$.

Lebesgue Decomposition

Definition 1.3 (Absolute continuity of measures) Let μ and ϕ denote a measure and a signed measure on a σ-field $\mathcal{A}$. Call ϕ *absolutely continuous with respect to* μ, denoted by $\phi \ll \mu$, if $\phi(A) = 0$ for each $A \in \mathcal{A}$ having $\mu(A) = 0$. We say ϕ is *singular with respect to* μ, denoted by $\phi \perp \mu$, if there exists a set $N \in \mathcal{A}$ for which $\mu(N) = 0$ while $|\phi|(N^c) = 0$.

Exercise 1.4 Let μ be a measure and let ϕ be signed measures on $(\Omega, \mathcal{A})$. Show that the following are equivalent: (a) $\phi \ll \mu$. (b) $\phi^+ \ll \mu$ and $\phi^- \ll \mu$. (c) $|\phi| \ll \mu$.

Theorem 1.2 (Lebesgue decomposition) Let μ denote any σ-finite measure on the measurable space $(\Omega, \mathcal{A})$. Let ϕ be any other σ-finite signed measure on this space $(\Omega, \mathcal{A})$. Then there exists a unique decomposition of ϕ with respect μ as

(7) $\qquad \phi = \phi_{ac} + \phi_s \qquad \text{where} \quad \phi_{ac} \ll \mu \quad \text{and} \quad \phi_s \perp \mu,$

with ϕ_{ac} and ϕ_s being σ-finite signed measures. Moreover,

(8) $\qquad \phi_{ac}(A) = \int_A Z_0 \, d\mu \qquad \text{for all } A \in \mathcal{A}$

for some finite $\mathcal{A}$-measurable function Z_0, which is unique a.e. μ.

Proof. By σ-finiteness and the Jordan–Hahn decomposition, we need only give the proof if μ and ϕ are finite measures; just separately consider $\phi^+_{\Omega_n}$ and $\phi^-_{\Omega_n}$ $(n = 1, 2, \ldots)$ for a joint σ-finite decomposition $\Omega = \sum_1^\infty \Omega_n$ of μ and $|\phi|$. (To give the details would be pedantic.) We now establish the existence of the decomposition in the reduced problem when ϕ and μ are finite measures. Let

(a) $\qquad \mathcal{Z} \equiv \{Z : Z \geq 0, Z \in \mathcal{L}_1 \text{ and } \int_A Z \, d\mu \leq \phi(A) \text{ for all } A \in \mathcal{A}\}.$

Now, $\mathcal{Z} \neq \emptyset$, since $Z \equiv 0$ is in $\mathcal{Z}$.

Case 1. $\phi \ll \mu$: The first step is to observe that

(b) $\qquad Z_1, Z_2 \in \mathcal{Z} \text{ implies } Z_1 \vee Z_2 \in \mathcal{Z}.$

With $A_1 \equiv \{\omega \in A : Z_1(\omega) > Z_2(\omega)\}$ and $A_2 \equiv AA_1^c$, we have

(c) $\qquad \int_A (Z_1 \vee Z_2) \, d\mu = \int_{A_1} Z_1 \, d\mu + \int_{A_2} Z_2 \, d\mu \leq \phi(A_1) + \phi(A_2) = \phi(A).$

Thus (b) holds. Now choose a sequence $Z_n \in \mathcal{Z}$ such that

(d) $\qquad \int_\Omega Z_n \, d\mu \to c \equiv \sup_{Z \in \mathcal{Z}} \int_\Omega Z \, d\mu \leq \sup_{Z \in \mathcal{Z}} \phi(\Omega) \leq \phi(\Omega) < \infty.$

We may replace Z_n by $\tilde{Z}_n \equiv Z_1 \vee \cdots \vee Z_n$ in (d). These $\tilde{Z}_n$ in (d) are an $\nearrow$ sequence of functions. Then let $Z_0 \equiv \lim \tilde{Z}_n$. The MCT then gives (for any $A \in \mathcal{A}$)

(e) $\qquad \int_A Z_0 \, d\mu = \lim \int_A \tilde{Z}_n \, d\mu \leq \lim \phi(A) \leq \phi(A), \quad \text{so that} \quad Z_0 \in \mathcal{Z}, \qquad \text{and}$

(f) $\qquad \int_\Omega Z_0 \, d\mu = \lim \int_\Omega \tilde{Z}_n \, d\mu = c < \infty, \qquad \text{showing that} \quad Z_0 \geq 0 \text{ is a.e. finite.}$

(Redefine Z_0 on a null set so that it is always finite.)

We now define

(g) $\qquad \phi_{ac}(A) \equiv \int_A Z_0 \, d\mu \quad \text{and} \quad \phi_s(A) \equiv \phi(A) - \phi_{ac}(A) \qquad \text{for all } A \in \mathcal{A}.$

Then ϕ_{ac} is a finite measure, which can be seen by applying example 1.1 with c finite; and $\phi_{ac} \ll \mu$. Moreover,

(h) $\qquad \phi_s \equiv \phi - \phi_{ac} \geq 0$

(since $Z_0 \in \mathcal{Z}$), so that ϕ_s is a finite measure by exercise 1.1. If $\phi_s(\Omega) = 0$, then $\phi = \phi_{ac}$ and we are done, with $\phi_s \equiv 0$. [In the next paragraph we verify that $\phi_s \equiv 0$ always holds in Case 1; that is, we will verify that $\phi_s(\Omega) = 0$.]

Assume $\phi_s(\Omega) > 0$. Then (since $\mu(\Omega)$ is finite) there is some $\theta > 0$ for which

(i) $\qquad \phi_s(\Omega) > \theta\, \mu(\Omega)$.

Let Ω^+ and Ω^- denote the Jordan–Hahn decomposition for $\phi^* \equiv \phi_s - \theta\, \mu$. Then

(j) $\qquad \mu(\Omega^+) > 0 \qquad$ must follow (while (i) is being assumed).

[Assume (j) is not true, so that $\mu(\Omega^+) = 0$. This implies $\phi_{ac}(\Omega^+) = \int_{\Omega^+} Z_o\, d\mu = 0$. It further implies that $\phi_s(\Omega^+) = 0$ (since $\phi_s = \phi - \phi_{ac} \ll \mu$, as $\phi \ll \mu$ is assumed for Case 1 and as $\phi_{ac} \ll \mu$ is obvious from example 4.1.1). But $\phi_s(\Omega^+) = 0$ contradicts (i) by implying that

$$\phi_s(\Omega) - \theta\, \mu(\Omega) \equiv \phi^*(\Omega) = -\phi^*(\Omega^-) + \phi^*(\Omega^+)$$

(k) $\qquad = -\phi^*(\Omega^-) + [\phi_s(\Omega^+) - \theta\mu(\Omega^+)] = -\phi^*(\Omega^-) \le 0.$

Thus (j) must also hold, under the assumption made above that inequality (i) holds.] Now, $\phi_s(A\Omega^+) \ge \theta\, \mu(A\Omega^+)$ (by the definition of Ω^+ below (i)). Thus (as $\phi_s \ge 0$ by (h) gives the inequality $\phi_s(A\Omega^-) \ge 0$),

$$\phi(A) = \phi_{ac}(A) + \phi_s(A) = \int_A Z_0\, d\mu + \phi_s(A\Omega^+) + \phi_s(A\Omega^-)$$

$$\ge \int_A Z_0\, d\mu + \phi_s(A\Omega^+)$$

$$\ge \int_A Z_0\, d\mu + \theta\, \mu(A\Omega^+) \qquad \text{as } \Omega^+ \text{ is a positive set for } \phi^* \equiv \phi_s - \theta\, \mu$$

(l) $\qquad = \int_A (Z_0 + \theta\, 1_{\Omega^+})\, d\mu \qquad$ for all $A \in \mathcal{A}$.

This implies both $Z_\theta \equiv Z_0 + \theta\, 1_{\Omega^+} \in \mathcal{Z}$ and $\int_\Omega Z_\theta\, d\mu = c + \theta\, \mu(\Omega^+) > c$. But this is a contradiction. Thus $\phi_s(\Omega) = 0$. Thus ϕ equals ϕ_{ac} and satisfies (8), and the theorem holds in Case 1. The a.s. μ uniqueness of Z_0 follows from exercise 3.2.2. (This also establishes the Radon–Nikodym theorem below.)

Case 2. General ϕ: Let $\nu \equiv \phi + \mu$, and note that both $\phi \ll \nu$ and $\mu \ll \nu$. Then by Case 1 we can infer that

(m) $\qquad \phi(A) = \int_A X\, d\nu \quad$ and $\quad \mu(A) = \int_A Y\, d\nu \qquad$ for all $A \in \mathcal{A}$

for finite ν-integrable functions $X \ge 0$ and $Y \ge 0$ that are unique a.e. ν. Let $D \equiv \{\omega : Y(\omega) = 0\}$, and then $D^c = \{\omega : Y(\omega) > 0\}$. Define

(n) $\qquad \phi_s(A) \equiv \phi(AD) \qquad$ and $\qquad \phi_{ac}(A) = \phi(AD^c)$.

Now $\mu(D) = \int_D Y\, d\nu = \int_D 0\, d\nu = 0$, and (n) gives $\phi_s(D^c) = \phi(D^cD) = \phi(\emptyset) = 0$; thus $\phi_s \perp \mu$. Is $\phi_{ac} << \mu$? Let $\mu(A) = 0$. Then $\mu(AD^c) = 0$. Then by (m), $0 = \mu(AD^c) = \int_{AD^c} Y\, d\nu$; and thus $Y = 0$ a.e. ν in AD^c, by exercise 3.2.1. But $Y > 0$ on AD^c, and so $\nu(AD^c) = 0$. Then (n) and (m) give $\phi_{ac}(A) = \phi(AD^c) = \int_{AD^c} X\, d\nu = 0$, since $\nu(AD^c) = 0$. So, $\mu(A) = 0$ implies $\phi_{ac}(A) = 0$. Thus $\phi_{ac} \ll \mu$.

Consider the uniqueness of the decomposition. If $\phi = \phi_{ac} + \phi_s = \bar{\phi}_{ac} + \bar{\phi}_s$, then $\psi \equiv \phi_{ac} - \bar{\phi}_{ac} = \bar{\phi}_s - \phi_s$ satisfies both $\psi \perp \mu$ and $\psi \ll \mu$. Thus $\psi \equiv 0$. $\square$

Exercise 1.5 Verify the following elementary facts for signed measures ϕ_1, ϕ_2, ϕ and a measure μ on some measurable space $(\Omega, \mathcal{A})$.
(a) If $\phi_1 \ll \mu$ and $\phi_2 \ll \mu$, then $\phi_1 + \phi_2 \ll \mu$
(b) If $\phi_1 \perp \mu$ and $\phi_2 \perp \mu$, then $\phi_1 + \phi_2 \perp \mu$
(c) If $\phi \ll \mu$ and $\phi \perp \mu$, then $\phi \equiv 0$. (This was used in the previous proof.)

2 The Radon–Nikodym Theorem

Recall that the absolute continuity $\phi \ll \mu$ means that

(1) $\qquad \phi(A) = 0 \qquad$ whenever $\qquad \mu(A) = 0$ with $A \in \mathcal{A}$.

Theorem 2.1 (Radon–Nikodym) Suppose both the signed measure ϕ and the measure μ are σ-finite on a measurable space $(\Omega, \mathcal{A})$. Then $\phi \ll \mu$ if and only if there exists uniquely a.e. μ a finite-valued $\mathcal{A}$-measurable function Z_0 on Ω for which

(2) $$\phi(A) = \int_A Z_0 \, d\mu \qquad \text{for all } A \in \mathcal{A}.$$

Moreover, ϕ is finite if and only if Z_0 is integrable.

The function Z_0 of (2) is often denoted by $\frac{d\phi}{d\mu}$, so that we also have the following very suggestive notation:

$$\phi(A) = \int_A \frac{d\phi}{d\mu} \, d\mu \qquad \text{for all } A \in \mathcal{A}.$$

We call Z_0 the *Radon–Nikodym derivative* (or the *density*) of ϕ with respect to μ.

Proof. The Lebesgue decomposition theorem shows that such a Z_0 necessarily exists. The sufficiency is just the trivial example 4.1.1. The "moreover" part is also a trivial result. □

Theorem 2.2 (Change of variable theorem) Let $\mu \ll \nu$ where μ and ν are σ-finite measures on $(\Omega, \mathcal{A})$. If $\int X \, d\mu$ has a well-defined value in $[-\infty, \infty]$, then

(3) $$\int_A X \, d\mu = \int_A X \left[\frac{d\mu}{d\nu}\right] d\nu \qquad \text{for all } A \in \mathcal{A},$$

One useful special case results from

(4) $$\int_a^b f \, dG = \int_a^b fg \, dH, \qquad \text{with } G \equiv \int_a^{\cdot} g \, dH \text{ for a generalized df } H,$$

where $g \geq 0$ on $(a, b]$ is measurable, and where we agree that $\int_a^b \equiv \int_{(a,b]}$.

Proof. Case 1. $X = 1_B$, for $B \in \mathcal{A}$: Then the Radon–Nikodym theorem gives

(a) $$\int_A 1_B \, d\mu = \mu(AB) = \int_{AB} \frac{d\mu}{d\nu} \, d\nu = \int_A 1_B \frac{d\mu}{d\nu} \, d\nu.$$

Case 2. $X = \sum_1^n c_i 1_{B_i}$, for a partition B_i: Case 1 and linearity of the integral give

(b) $$\int_A X \, d\mu = \sum_{i=1}^n c_i \int_A 1_{B_i} \, d\mu = \sum_1^n c_i \int_A 1_{B_i} \frac{d\mu}{d\nu} \, d\nu = \int_A X \frac{d\mu}{d\nu} \, d\nu.$$

Case 3. $X \geq 0$: Let $X_n \geq 0$ be simple functions that $\nearrow$ to X. Then the MCT twice gives

(c) $$\int_A X \, d\mu = \lim \int_A X_n \, d\mu = \lim \int_A X_n \frac{d\mu}{d\nu} \, d\nu = \int_A X \frac{d\mu}{d\nu} \, d\nu.$$

Case 4. X measurable and at least one of X^+, X^- in $\mathcal{L}_1$: Then

$$\int_A X\,d\mu = \int_A X^+\,d\mu - \int_A X^-\,d\mu$$

$$\text{(d)} \qquad = \int_A X^+ \frac{d\mu}{d\nu}\,d\nu - \int_A X^- \frac{d\mu}{d\nu}\,d\nu = \int_A X\,\frac{d\mu}{d\nu}\,d\nu\,,$$

so long as one of $\int_A X^+\,d\mu$ and $\int_A X^-\,d\mu$ is finite. □

Exercise 2.1 (Derivative of a sum, and a chain rule) Let μ and ν be σ-finite measures on $(\Omega, \mathcal{A})$. Let ϕ and ψ be σ-finite signed measures on $(\Omega, \mathcal{A})$. Then

$$\text{(5)} \qquad \frac{d(\phi+\psi)}{d\mu} = \frac{d\phi}{d\mu} + \frac{d\psi}{d\mu} \ \text{ a.e. } \mu \qquad \text{if } \phi \ll \mu \text{ and } \psi \ll \mu\,,$$

$$\text{(6)} \qquad \frac{d\phi}{d\nu} = \frac{d\phi}{d\mu}\cdot\frac{d\mu}{d\nu} \ \text{ a.e. } \nu \qquad \text{if } \phi \ll \mu \text{ and } \mu \ll \nu.$$

Show that $\frac{d\mu}{d\nu} = 1/\frac{d\nu}{d\mu}$ holds a.e. μ and a.e. ν if $\mu \ll \nu$ and $\nu \ll \mu$.

Note that theorem 3.2.6 (of the unconscious statistician) is another change of variable theorem. That is, if $X : (\Omega, \mathcal{A}) \to (\bar{\Omega}, \bar{\mathcal{A}})$ and $g : (\bar{\Omega}, \bar{\mathcal{A}}) \to (\bar{R}, \bar{\mathcal{B}})$, then

$$\text{(7)} \qquad \int_{(g\circ X)^{-1}(B)} g(X)\,d\mu = \int_{g^{-1}(B)} g\,d\mu_X = \int_B y\,d\mu_{g(X)}(y) \qquad \text{for all } B \in \mathcal{B},$$

when one of the these integrals is well-defined. (See also exercise 6.3.3 below.)

Exercise 2.2 Let P_{μ,σ^2} denote the $N(\mu, \sigma^2)$ distribution. Let P have the density $f \equiv dP/d\lambda$ with respect to Lebesgue measure λ for which $f > 0$.
(a) Show that $\lambda \ll P$ with density $1/f$.
(b) Show that $P_{\mu,1} \ll P_{0,1}$ and compute $dP_{\mu,1}/dP_{0,1}$.
(c) Show that $P_{0,\sigma^2} \ll P_{0,1}$ and compute $dP_{0,\sigma^2}/dP_{0,1}$.
(d) Compute $dP/dP_{0,1}$ and $dP_{0,1}/dP$ when P denotes the Cauchy distribution.

Exercise 2.3 Flip a coin. If heads results, let X be a Uniform$(0,1)$ outcome; but if tails results, let X be a Poisson(λ) outcome. The resulting distribution on R is labeled ϕ.
(a) Let μ denote Lebesgue measure on R. Find the Lebesgue decomposition of ϕ with respect to this μ; that is, write $\phi = \phi_{ac} + \phi_s$.
(b) Let μ be counting measure on $\{0, 1, 2, \ldots\}$. Find the Lebesgue decomposition of ϕ with respect to this μ.
[If need be, see the definitions of various distributions in chapter 9.]

Exercise 2.4 Let μ be a σ-finite measure on $(\Omega, \mathcal{A})$. Define $\phi(A) \equiv \int_A X\,d\mu$ for all $A \in \mathcal{A}$ for some μ-integrable function X. Show that

$$|\phi|(A) = \int_A |X|\,d\mu \qquad \text{for all } A \in \mathcal{A}.$$

Exercise 2.5 (Alternative definition of absolute continuity) Let ϕ be finite and let μ be σ-finite, for measures on $(\Omega, \mathcal{A})$. Then $\phi \ll \mu$ if and only if for every $\epsilon > 0$ there exists $\delta_\epsilon > 0$ such that $\mu(A) < \delta_\epsilon$ implies $\phi(A) < \epsilon$. (Show that if μ is finite and ϕ is not, then the claim could fail.)

Exercise 2.6 (Domination) If $\mu_1, \mu_2, \ldots$ are finite measures on some $(\Omega, \mathcal{A})$, then there exists a finite measure μ on $(\Omega, \mathcal{A})$ such that $\mu_k \ll \mu$ for each $k \geq 1$.

Exercise 2.7 (Halmos) Suppose $\mu_1, \mu_2, \ldots$ and $\nu_1, \nu_2, \ldots$ are finite measures on some $(\Omega, \mathcal{A})$ for which $\mu_k \ll \nu_k$ for each $k \geq 1$. Suppose also that

$$\mu(A) = \lim_{n\to\infty} \textstyle\sum_{k=1}^n \mu_k(A) \qquad \text{and} \qquad \nu(A) = \lim_{n\to\infty} \textstyle\sum_{k=1}^n \nu_k(A)$$

for all $A \in \mathcal{A}$. Show that the following hold a.e. ν:

$$(8) \qquad \begin{aligned} &\{d \textstyle\sum_{k=1}^n \nu_k\}/\{d\nu\} \nearrow 1 \quad \text{and} \quad \{d \textstyle\sum_{k=1}^n \mu_k\}/\{d\nu\} \nearrow \{d\mu/d\nu\}, \\ &\{d \textstyle\sum_{k=1}^n \mu_k\}/\{d \textstyle\sum_{k=1}^n \nu_k\} \to \{d\mu/d\nu\}. \end{aligned}$$

These can be thought of as theorems about Radon–Nikodym derivatives, about absolute continuity of measures, or about change of variables.

Exercise 2.8 Let $\mathcal{A}$ denote the collection of all subsets A of an uncountable set Ω for which either A or A^c is countable. Let $\mu(A)$ denote the cardinality of A. Define $\phi(A)$ to equal 0 or ∞ according as A is countable or uncountable. Show that $\phi \ll \mu$. Then show that the Radon–Nikodym theorem fails.

Exercise 2.9 For a σ-finite measure μ and a finite measure ν on $(\Omega, \mathcal{A})$, let

$$\phi(A) \equiv \mu(A) - \nu(A) \qquad \text{for all } A \in \mathcal{A}.$$

(a) Show that ϕ is a signed measure. (b) Show that

$$\phi(A) = \textstyle\int_A (f - g)\, d(\mu + \nu),$$

for measurable functions f and g with $g \in \mathcal{L}_1^+(\mu + \nu)$. (Note example 4.1.1.)
(c) Determine ϕ^+, ϕ^-, and $|\phi|$; and determine $|\phi|(\Omega)$ in case μ is also a finite measure.

Exercise 2.10 (Total variation distance between probability measures) Define P and Q to be probability measures on $(\Omega, \mathcal{A})$.
(a) Show that the *total variation distance* $d_{TV}(P, Q)$ between P and Q satisfies

$$(9) \qquad d_{TV}(P, Q) \equiv \{\sup_{A\in\mathcal{A}} |P(A) - Q(A)|\} = \tfrac{1}{2} \textstyle\int |\, p - q\,|\, d\mu$$

for any σ-finite measure μ dominating both P and Q (that is, $P \ll \mu$ and $Q \ll \mu$).
(b) Use part (a) to show that $d_{TV}(P, Q) = |P - Q|\,(\Omega)/2$.
(c) Note specifically that the choice of dominating measure μ does not affect the value of $d_{TV}(P, Q)$. (Note section 14.2 below.)

Exercise 2.11 (Hellinger distance between probability measures) Let P and Q denote probability measures on $(\Omega, \mathcal{A})$. Define the *Hellinger distance* $H(P,Q)$ by

(10) $$H^2(P,Q) \equiv \frac{1}{2}\int[\sqrt{p}-\sqrt{q}\,]^2\,d\mu$$

for any measure μ dominating both P and Q. Show that the choice of dominating measure μ does not affect the value of $H(P,Q)$. (Note section 14.2 below.)

Exercise 2.12 Let ϕ be a σ-finite signed measure. Define

$$\int X\,d\phi = \int X\,d\phi^+ - \int X\,d\phi^-$$

when this is finite. Show that $|\int X\,d\phi| \le \int |X|\,d|\phi|$.

Exercise 2.13 Let $(\Omega, \mathcal{A})$ be a measurable space, and let $\mathcal{M}$ denote the collection of all finite signed measures μ on $(\Omega, \mathcal{A})$. Let $\|\mu\| \equiv |\mu|(\Omega)$. Thus $\|\mu_1 - \mu_2\| = |\mu_1 - \mu_2|(\Omega)$. Show that $(\mathcal{M}, \|\cdot\|)$ is a complete metric space.

3 Lebesgue's Theorem

Theorem 3.1 (Lebesgue) Suppose F is an $\nearrow$ function on $[a, b]$. Then F has an integrable derivative F' that exists and is finite a.e. λ on $[a, b]$.

Proof. Consider the *Dini derivates*

$$D^+F(x) \equiv \limsup_{h\to 0+} [F(x+h) - F(x)]/h\,,$$
$$D^-F(x) \equiv \limsup_{h\to 0+} [F(x) - F(x-h)]/h\,,$$
$$D_+F(x) \equiv \liminf_{h\to 0+} [F(x+h) - F(x)]/h\,,$$
$$D_-F(x) \equiv \liminf_{h\to 0+} [F(x) - F(x-h)]/h\,.$$

Trivially, $D^+F(x) \ge D_+F(x)$ and $D^-F(x) \ge D_-F(x)$. All four derivates having the same finite value is (of course) the definition of F being *differentiable* at x, with the common value of the derivates being called the derivative of F at x and being denoted by $F'(x)$. Let

$$A \equiv \{x : D^+F(x) > D_-F(x)\}$$

(a) $$\equiv \bigcup_{r,s} A_{rs} \equiv \bigcup_{r,s}\{x : D^+F(x) > s > r > D_-F(x)\},$$

where the union is over all rational r and s. To show that $\lambda(A) = 0$, it suffices to show that all A_{rs} have outer Lebesgue measure zero, in that $\lambda^*(A_{rs}) = 0$. To this end, let U be an open set for which $A_{rs} \subset U$ with $\lambda(U) < \lambda^*(A_{rs}) + \epsilon$. For each $x \in A_{rs}$ we can specify infinitely many and arbitrarily small h for which $[x-h, x] \subset U$ and $[F(x) - F(x-h)]/h < r$. This collection of closed intervals covers A_{rs} in the sense of Vitali (see exercise 1.2.8). Thus some finite disjoint collection of them has interiors $I_1 \equiv (x_1 - h_1, x_1), \ldots, I_m \equiv (x_m - h_m, x_m)$ for which $B_{rs} \equiv A_{rs} \cap (\sum_{i=1}^m I_i)$ has $\lambda^*(B_{rs}) > \lambda^*(A_{rs}) - \epsilon$. Then

(b) $$\sum_{i=1}^m [F(x_i) - F(x_i - h_i)] < r\sum_{i=1}^m h_i \le r\,\lambda(U) < r\,[\lambda^*(A_{rs}) + \epsilon\,]\,.$$

For each $y \in B_{rs}$ we can specify infinitely many and arbitrarily small h for which $[y, y+h] \subset$ (some I_i) and $[F(y+h) - F(y)]/h > s$. This collection covers B_{rs} in the sense of Vitali. Thus some finite disjoint collection of them has interiors $J_1 \equiv (y_1, y_1 + h_1), \ldots, J_n \equiv (y_n, y_n + h_n)$ for which $C_{rs} \equiv B_{rs} \cap (\sum_{j=1}^n J_j)$ has $\lambda^*(C_{rs}) > \lambda^*(B_{rs}) - \epsilon$. Then

(c) $$\sum_{j=1}^n [F(y_j + h_j) - F(y_j)] > s\sum_{j=1}^n h_j \ge s\,[\lambda^*(B_{rs}) - \epsilon\,] > s\,[\lambda^*(A_{rs}) - 2\epsilon\,].$$

Moreover, since the disjoint union of the J_j's is a subset of the disjoint union of the I_i's, results (b) and (c) yield

$$r\,[\lambda^*(A_{rs}) + \epsilon\,] > \sum_{i=1}^m [F(x_i) - F(x_i - h_i)]$$

(d) $$\ge \sum_{j=1}^n [F(y_j + h_j) - F(y_j)] > s\,[\lambda^*(A_{rs}) - 2\epsilon\,]$$

for every $\epsilon > 0$. That is, $r\,\lambda^*(A_{rs}) \ge s\,\lambda^*(A_{rs})$. But $r < s$. Thus $\lambda^*(A_{rs}) = 0$ for all rational r and s. Thus $\lambda^*(A) = 0$. Analogously,, $\lambda(\{x : D^-F(x) > D_+F(x)\}) = 0$. So $D^-F(x) \ge D_-F(x) \ge D^+F(x) \ge D_+F(x) \ge D^-F(x)$ a.e. λ. Thus F' exists a.e. λ.

Now, the measurable function difference quotients

(e) $\quad D_nF(x) \equiv n[F((x+1/n)\wedge b) - F(x)]$

on $[a,b]$ converge a.e. λ to $F'(x)$ on $[a,b]$, so that $F'(x)$ is measurable. Applying Fatou's lemma to the D_nF (which are ≥ 0, since F is $\nearrow$) gives

(f) $\quad \int_a^b F'(x)\,dx = \int_a^b [\,\underline{\lim}\, D_nF(x)]\,dx \leq \underline{\lim} \int_a^b D_nF(x)\,dx$

$\quad\quad = \underline{\lim} \int_a^b n[F((x+1/n)\wedge b) - F(x)]\,dx$

(g) $\quad\quad = \underline{\lim}\, [\, \int_b^{b+1/n} n\,F(b)\,dx \; - \int_a^{a+1/n} n\,F(x)\,dx\,] \leq \underline{\lim}\, \{F(b) - F(1+1/n)\}$

(h) $\quad\quad \leq F(b) - F(a),$ $\quad$ using the monotonicity of F.

Thus F' is integrable, and hence F' is also finite a.e. λ. (We now summarize this result as a corollary, as situations with strict inequality are very revealing.) □

Corollary 1 Suppose F is an $\nearrow$ function on $[a,b]$. Then F' exists a.e. λ and

(1) $\quad \int_a^b F'(x)\,dx \leq F(b) - F(a).$

So, F is differentiable a.e. λ, and its derivative F' is finite a.e. λ and satisfies (1).

The Lebesgue singular df in example 6.1.1 below will show that equality need not hold in (1); this continuous df is constant valued on a collection of disjoint intervals of total length 1. An example in Hewitt and Stromberg (1965, p. 278) shows that $F'(x) = 0$ is possible for *all* x, even with a $\uparrow F$.

Theorem 3.2 (Term-by-term differentiation of series) Let g_k be $\nearrow$ on $[a,b]$ for each $k \geq 1$, and suppose that $S_n(x) \equiv \sum_{k=1}^n g_k(x)$ converges at $x = a$ and $x = b$. Then $S_n(x) \to S(x)$ for all x in $[a,b]$, for some finite-valued measurable function $S(x)$. Mainly, $S'(\cdot)$ exists a.s. λ and is given by

(2) $$S'(x) = \sum_{k=1}^{\infty} g_k'(x).$$

Corollary 1 If the power series $S(x) \equiv \sum_{n=1}^{\infty} a_n(x-a)^n$ converges absolutely for $x = a + R$, then for all $|x - a| < R$ we may differentiate $S(x)$ term by term. Moreover, this is true for any number of derivatives of S.

Proof. Note that $S_n(a)$ is a convergent sum. Now write

$$S_n(x) = S_n(a) + [S_n(x) - S_n(a)] = S_n(a) + \textstyle\sum_{k=1}^n [g_k(x) - g_k(a)]\,.$$

Since $\nearrow$ sequences bounded above converge, the convergence at $x = a$ and $x = b$ gives convergence at all x in the interval. We may assume all $g_k \geq 0$ on $[a,b]$ with $g_k(a) = 0$; else replace g_k by $g_k(\cdot) - g_k(a)$. Since S and all S_n are $\nearrow$, the derivatives S' and all S_n' exist a.e. λ by theorem 3.1 (of Lebesgue). Now,

(a) $\quad S_n'(x) \leq S_{n+1}'(x) \leq S'(x) \quad$ a.e. $\lambda\,;$

both essentially follow from

$$\text{(b)}\qquad \frac{S(x+h)-S(x)}{h} = \frac{S_n(x+h)-S_n(x)}{h} + \sum_{n+1}^{\infty} \frac{g_k(x+h)-g_k(x)}{h}$$

$$\geq \frac{S_n(x+h)-S_n(x)}{h}\,.$$

From (a) we see (without having made use of $g_k(a)=0$) that

(c) $S_n'(\cdot)$ converges a.e. λ with $\lim S_n' \leq S'$ a.e. λ.

Because $S_n' \nearrow$, it suffices to show that $S_{n_i}' \to_{a.e.} S'$ for some subsequence n_i. Since $S_n(b) \nearrow S(b)$, we may specify n_i so large that $0 \leq S(b) - S_{n_i}(b) < 2^{-i}$, and then

(d) $0 \leq S(x) - S_{n_i}(x) = \sum_{n_i+1}^{\infty} g_k(x) \leq \sum_{n_i+1}^{\infty} g_k(b) = S(b) - S_{n_i}(b) < 2^{-i}$,

for all $x \in [a,b]$. Thus

(e) $0 \leq \sum_{i=1}^{\infty} [S(x) - S_{n_i}(x)] \leq \sum_{i=1}^{\infty} 2^{-i} = 1$ for all $x \in [a,b]$,

where the series in (e) has summands

(f) $h_i(x) \equiv S(x) - S_{n_i}(x)$ that are $\nearrow$ in x.

Thus conclusion (c) also applies to these h_i's (not just the g_k's), and we thus conclude from (c) that the series

(g) $T_n' \equiv \sum_{i=1}^{n} h_i'$ converges a.e. λ.

But a series of real numbers can converge only if its nth term goes to 0; that is,

(h) $S'(x) - S_{n_i}'(x) = h_i'(x) \to 0$ a.e. λ.

As noted above, this suffices for the theorem. $\square$

Exercise 3.1 Prove the corollary.

Example 3.1 (Taylor's expansion) Suppose $g(\cdot)$ is defined in a neighborhood of a. Let x^* denote a generic point between x and a. Let

(3) $P_1(x) \equiv g(a) + g'(a)(x-a)\,,$

(4) $P_2(x) \equiv P_1(x) + g''(a)(x-a)^2/2!\,,$

(5) $P_3(x) \equiv P_2(x) + g'''(a)(x-a)^3/3!\,, \ldots\,,$

(6) $R_1(x) \equiv [g(x) - g(a)]/(x-a)$ or $g'(a)$, as $x \neq a$ or $x = a$,

(7) $R_2(x) \equiv 2!\,[g(x) - P_1(x)]/(x-a)^2$ or $g''(a)$, as $x \neq a$ or $x = a$,

(8) $R_3(x) \equiv 3!\,[g(x) - P_2(x)]/(x-a)^3$ or $g'''(a)$, as $x \neq a$ or $x = a$.

Then l'Hospital's rule gives (provided $g'(a)$, $g''(a)$, $g'''(a)$, ... exist, respectively)

$$\lim_{x\to a} R_1(x) = g'(a) = R_1(a), \tag{9}$$

$$\lim_{x\to a} R_2(x) = \lim_{x\to a} \frac{g'(x) - P_1'(x)}{x-a} = \lim_{x\to a} \frac{g'(x) - g'(a)}{x-a} = g''(a) = R_2(a), \tag{10}$$

$$\lim_{x\to a} R_3(x) = \lim_{x\to a} \frac{2!\,[g'(x) - P_2'(x)]}{(x-a)^2} = \lim_{x\to a} \frac{g''(x) - P_2''(x)}{x-a}$$

$$= \lim_{x\to a} \frac{g''(x) - g''(a)}{x-a} = g'''(a). \tag{11}$$

Thus we find it useful to use the representations (with $g^{(k)}(a)$ abbreviating that $g^{(k)}(\cdot)$ exists at a, and with $g^{(k)}(\cdot)$ abbreviating that $g^{(k)}(x)$ exists for all x in a neighborhood of a)

$$g(x) = \begin{cases} P_1(x) + [R_1(x) - g'(a)](x-a) & \text{if } g'(a), \\ P_1(x) + [g'(x^*) - g'(a)](x-a) & \text{if } g'(\cdot)\,, \end{cases} \tag{12}$$

$$g(x) = \begin{cases} P_2(x) + [R_2(x) - g''(a)](x-a)/2! & \text{if } g''(a), \\ P_2(x) + [g''(x^*) - g''(a)](x-a)^2/2! & \\ \quad = P_1(x) + g''(x^*)(x-a)^2/2! & \text{if } g''(\cdot)\,, \end{cases} \tag{13}$$

$$g(x) = \begin{cases} P_3(x) + [R_2(x) - g''(a)](x-a)^3/3! & \text{if } g'''(a), \\ P_3(x) + [g'''(x^*) - g'''(a)](x-a)^3/3! & \\ \quad = P_2(x) + g'''(x^*)(x-a)^3/3! & \text{if } g'''(\cdot)\,. \end{cases} \tag{14}$$

□

Exercise 3.2 (a) Show that if $g''(x)$ exists, then

$$g''(x) = \lim_{h\to 0} \tfrac{1}{h^2}\left\{g(x+h) - 2g(x) + g(x-h)\right\}. \tag{15}$$

(b) An analogous result holds for any $g^{(2k)}(x)$.

Exercise 3.3 Prove the Vitali covering theorem. (See exercise 1.2.6.)

Exercise 3.4 Let $f(x) = \sum_0^\infty a_k\, x^k / \sum_1^\infty b_k\, x^k$ in some interval. Suppose that all $a_k, b_k > 0$ and $a_k/b_k \uparrow$. Then $f'(x) > 0$ for all x in that interval. (This result is useful in conjunction with the monotone likelihood ratio principle.)

4 The Fundamental Theorem of Calculus

Definition 4.1 (Bounded variation) Let F denote a real-valued function on $[a, b]$. The *total variation* of F over $[a, b]$ is defined by

$$V_a^b F \equiv V_{[a,b]} F$$

$$\equiv \sup\Big\{ \sum_{k=1}^{n} |F(x_k) - F(x_{k-1})| \,:\, a \equiv x_0 < x_1 < \cdots < x_n \equiv b,\ n \geq 1 \Big\}. \tag{1}$$

We say that F is of *bounded variation* (BV) on $[a, b]$ if $V_a^b F < \infty$.

It is clear that

$$V_a^b F = V_a^c F + V_c^b F \qquad \text{for } a \leq c \leq b \text{ and } F \text{ of BV}. \tag{2}$$

Definition 4.2 (Absolutely continuous functions) A real-valued function F on any subinterval I of the line R is said to be *absolutely continuous* if for all $\epsilon > 0$ there exists a $\delta_\epsilon > 0$ such that

$$\sum_{k=1}^{n} |F(d_k) - F(c_k)| < \epsilon \qquad \text{whenever} \qquad \sum_{k=1}^{n} (d_k - c_k) < \delta_\epsilon \tag{3}$$

with $n \geq 1$ and with disjoint subintervals (c_k, d_k) contained in I.

Definition 4.3 (Lipschitz condition) A real-valued function F on any subinterval I of R is said to be *Lipschitz* if for some finite constant M we have

$$|F(y) - F(x)| \leq M\, |y - x| \qquad \text{for all } x \text{ and } y \text{ in } I. \tag{4}$$

We first establish some elementary relationships among the Lipschitz condition, absolute continuity, bounded variation, and the familiar property of being $\nearrow$. These concepts have proven to be important in the study of differentiation. We will soon proceed further in this direction, and we will also consider the relationship between ordinary derivatives and Radon–Nikodym derivatives. We first recall from theorem 1.3.1 (the correspondence theorem) that every generalized df F can be associated with a Lebesgue–Stieltjes measure μ_F via $\mu_F((a, b]) \equiv F(b) - F(a)$.

Proposition 4.1 (The basics) Let λ denote Lebesgue measure.
(i) If F is of BV on $[a, b]$, then

$$F(x) = F_1(x) - F_2(x) \qquad \text{with} \quad F_1(x) \equiv V_a^x F \quad \text{and} \quad F_2(x) \equiv V_a^x F - F(x) \tag{5}$$

both being $\nearrow$ on $[a, b]$. Also, $F' = F_1' - F_2'$ a.e. λ, with F_1' and F_2' both integrable.
(ii) If F is absolutely continuous, then it is of BV. The F_1 and F_2 in (i) are both absolutely continuous and $\nearrow$.
(iii) Lipschitz functions are absolutely continuous.

Proof. Consider (i). Now, $F_1(x) = V_a^x F$ is obviously $\nearrow$; use (2). Then F_2 is also $\nearrow$, since for $x \le y$ we have

$$F_2(y) - F_2(x) = [V_a^y F - F(y)] - [V_a^x F - F(x)]$$

(a) $$= V_x^y F - [F(y) - F(x)] \ge 0.$$

Since F_1 and F_2 are $\nearrow$, their derivatives F_1' and F_2' exist a.e. λ and are integrable by theorem 4.3.1 (Lebesgue's theorem).

Consider (ii). Suppose that $F(\cdot)$ is absolutely continuous. Letting $\epsilon = 1$ and choosing δ_1 so small that the equally spaced values $a \equiv x_0 < x_1 < \cdots < x_n \equiv b$ have mesh $\equiv (b-a)/n < \delta_1$, we have from (2) that

(b) $$V_a^b F = \sum_{k=1}^n V_{[x_{k-1},x_k]} F \le \sum_{k=1}^n 1 = n;$$

and thus F is of BV. But we must still show that F_1 is absolutely continuous if F is. So we suppose that F is absolutely continuous, and specify that $\sum_1^n (d_k - c_k) < \delta_{\epsilon/2}$ for some choice of n, c_k's, and d_k's. We now show that these same n, c_k, d_k 'work' for F_1. Well, for each fixed k with $1 \le k \le n$ and the tiny number $\epsilon/(2n)$, the definition of the BV of F gives

(c) $$F_1(d_k) - F_1(c_k) = V_{[c_k,d_k]} F < \sum_{j=1}^{m_k} |F(a_{n,k,j}) - F(a_{n,k,j-1})| + (\epsilon/2n)$$

for some choice of $c_k \equiv a_{n,k,0} < \cdots < a_{n,k,m_k} \equiv d_k$. These add to give

$$\sum_{k=1}^n |F_1(d_k) - F_1(c_k)| = \sum_{k=1}^n |V_{[a,d_k]} - V_{[a,c_k]}| = \sum_{k=1}^n V_{[c_k,d_k]} F$$

(d) $$\le \sum_{k=1}^n \left(\sum_{j=1}^{m_k} |F(a_{n,k,j}) - F(a_{n,k,j-1})| + (\epsilon/2n)\right)$$

(e) $$\le (\epsilon/2) + (\epsilon/2) = \epsilon$$

by absolute continuity of F, since it follows from above that

(f) $$\sum_{k=1}^n \sum_{j=1}^{m_k} (a_{n,k,j} - a_{n,k,j-1}) = \sum_{k=1}^n (d_k - c_k) < \delta_{\epsilon/2}.$$

Consider (iii). Being Lipschitz implies absolute continuity with $\delta_\epsilon = \epsilon/M$. □

Exercise 4.1 For F of BV on $[a,b]$, let $F_1(x) \equiv V_a^x F^+$ and $F_2(x) \equiv V_a^x F^-$, where

$$V_a^x F^\pm \equiv \sup\{\sum_{k=1}^n [F(x_k) - F(x_{k-1})]^\pm : a \equiv x_0 < \cdots < x_n \equiv x, n \ge 1\}.$$

Verify that $F - F(a) = F_1 - F_2$ with F_1 and F_2 both $\nearrow$ (an alternative to (5)). Example: Let $F(x)$ equal x, $2-x$, $x-4$ on $[0,1]$, $[1,3]$, $[3,4]$. Determine the decomposition of (5) for this F, as well as the decomposition of this exercise.

Exercise 4.2 Let f be continuous on $[a,b]$, and define $F(x) = \int_a^x f(y)\,dy$ for each $a \le x \le b$. Then F is differentiable at each $x \in (a,b)$ and $F' = f$ on (a,b). [Since f is continuous, we need only the Riemann integral. Can we extend this to the Lebesgue integral? Can we reverse the order, and first differentiate and then integrate? The next theorem answers these questions.]

Theorem 4.1 (Fundamental theorem of calculus) (i) Let F be absolutely continuous on $[a, b]$, and let λ denote Lebesgue measure. Then F' exists a.e. λ and

$$(6)\qquad F(x) - F(a) = \int_a^x F'\,d\lambda \qquad \text{for all } x \in [a, b]; \quad \text{also, } F' = \frac{d\mu_F}{d\lambda} \text{ a.e. } \lambda.$$

(ii) If $F(x) - F(a) = \int_a^x f\,d\lambda$ for some f that is integrable with respect to λ on $[a, b]$, then F is absolutely continuous on $[a, b]$. Moreover, $f = F' = \frac{d\mu_F}{d\lambda}$ a.e. λ.

Remark 4.1 (a) The fundamental theorem of calculus can be summarized by saying that F is absolutely continuous if and only if it is the integral of its derivative. The ordinary derivative F' is, in fact, also a Radon–Nikodym derivative of the signed measure μ_F naturally associated with F; proposition 4.2 below makes this clear.
(b) If F is of BV on $[a, b]$, then the derivative F' exists a.e. λ on $[a, b]$ and is integrable with respect to Lebesgue measure λ and $\int_a^b F'(x)\,d\lambda(x) \le F(b) - F(a)$; see (4.3.1). The Lebesgue singular df F of (6.1.9) below yields a strict inequality.]
(c) The Lipschitz condition represents "niceness with a vengeance," as it guarantees that all difference quotients are uniformly bounded. □

Proof. Consider the converse. If $F(x) \equiv F(a) + \int_a^x f(y)dy$ for $a \le x \le b$, then F is absolutely continuous by the absolute continuity of the integral theorem. Then F is of bounded variation on $[a, b]$ and F' exists a.e. λ in $[a, b]$, by proposition 4.1(ii). Moreover, F' is integrable, using (4.3.1). But does $F' = f$ a.e. λ?
Case 1: Suppose $|f|$ is bounded by some finite M on $[a, b]$. We could consider f^+ and f^- separately, but we will simply assume without loss of generality that $f \ge 0$. Then the difference quotient $D_nF(x) \equiv n\int_x^{x+1/n} f(y)dy$ of F also satisfies $|D_nF| \le M$ on $[a, b]$, and $D_nF(x) \to F'(x)$ a.e. Applying the DCT (with dominating function identically equal to M) once for each fixed $x \in (a, b)$ gives

$$\int_a^x F'(y)\,dy = \int_a^x \lim D_nF(y)\,dy = \lim \int_a^x n\,[F(y + 1/n) - F(y)]\,dy$$

$$= \lim\,[n\int_x^{x+1/n} F(y)\,dy - n\int_a^{a+1/n} F(y)\,dy\,]$$

$$= F(x) - F(a) \qquad \text{by continuity of } F$$

$$(a)\qquad = \int_a^x f(y)dy.$$

Thus $F'(y) = f(y)$ a.e. on $[a, b]$ by exercise 3.2.2 (refer also to the prototypical example 4.1.1).
Case 2: Suppose f is integrable. Again, $f \ge 0$ may be assumed. Let $f_n(\cdot) \equiv n \wedge f(\cdot)$, with $f - f_n \ge 0$. Now, $\int_a^x f_n$ has derivative f_n a.e. on $[a, b]$, by case 1. Thus

$$(b)\qquad F'(x) = \tfrac{d}{dx}\int_a^x f(y)dy = \tfrac{d}{dx}\int_a^x f_n(y)dy + \tfrac{d}{dx}\int_a^x [f(y) - f_n(y)]dy \ge f_n(x) + 0$$

for all n, and hence $F'(x) \ge f(x)$ a.e. on $[a, b]$. Thus

$$(c)\qquad \int_a^b F'(x)dx \ge \int_a^b f(x)dx = F(b) - F(a), \quad \text{which is } \ \ge \int_a^b F'(x)dx$$

by (4.3.1). The two inequalities in (c) combine to give

(d) $\int_a^b [F'(x) - f(x)]\, dx = 0$ with $F'(x) - f(x) \ge 0$ a.e.,

so that $F' = f$ a.e. on $[a, b]$ by exercise 3.2.1.

Consider the direct half when F is absolutely continuous on $[a, b]$. Without loss, suppose that F is $\nearrow$ (by proposition 4.1(ii)), so that F' exists a.e. on $[a, b]$ (see theorem 4.3.1) and that F' is integrable (see (4.3.1)). Use

(e) $\mu_F((a, x]) \equiv F(x) - F(a)$ for all $x \in [a, b]$

and the correspondence theorem to associate a Lebesgue–Stieltjes measure μ_F with F (which is a generalized df). We will show that $\mu_F << \lambda$ in proposition 4.2 below. Then the Radon–Nikodym will give

(f) $F(x) - F(a) = \int_a^x f\, d\lambda$ for all $x \in [a, b]$, with $f \equiv d\mu_F/d\lambda$. □

Now apply the converse half of the fundamental theorem of calculus to conclude that $F' = f \equiv d\mu_F/d\lambda$ a.e. on $[a, b]$. □

Proposition 4.2 (i) Let F be $\nearrow$ and absolutely continuous on a subinterval $[a, b]$ of the line R. Then the Lebesgue–Stieltjes measure μ_F (as in (e) above) satisfies $\mu_F \ll \lambda$ (name its Radon–Nikodym derivative $d\mu_F/d\lambda$), and

(7) $F(x) - F(a) = \int_a^x f d\lambda$ for all $x \in [a, b]$, with $f \equiv d\mu_F/d\lambda =_{a.e.} F'$.

(ii) Let F be absolutely continuous on R, and fix a. Then (7) holds for all x in R.

Proof. Let $\mu \equiv \mu_F$ and fix the finite interval $[a, b]$. Given $\epsilon > 0$, let $\delta_\epsilon > 0$ be as in the definition (3) of absolute continuity. Let $A \in \mathcal{B}$ be a subset of $[a, b]$ having $\lambda(A) < \delta_\epsilon/2$. Recalling our definition (1.2.1) of λ via Carathéodory coverings, we can claim that Lebesgue measure satisfies

(a) $\lambda(A) = \inf\{\sum_{n=1}^\infty \lambda(A_n) : A \subset \bigcup_{n=1}^\infty A_n$ for A_n's in the field $\mathcal{B}_F\}$.

Thus (for some fixed choice of sets in the field $\mathcal{B}_F$), we can write

(b) $A \subset \bigcup_1^\infty (c_n, d_n]$, where $\sum_1^\infty (d_n - c_n) < \lambda(A) + \delta_\epsilon/2 < \delta_\epsilon$

(recall that each A_n in the field $\mathcal{B}_F$ is a finite disjoint union of intervals $(c, d]$ with c and d finite, with finiteness following from $A \subset [a, b]$). Thus

(c) $\mu_F(A) \le \mu_F(\bigcup_1^\infty (c_n, d_n]) \le \sum_1^\infty \mu_F((c_n, d_n])$

$= \sum_1^\infty [F(d_n) - F(c_n)] = \lim_m \sum_1^m [F(d_n) - F(c_n)]$

(d) $\le \lim_m \epsilon = \epsilon$ (since F is absolutely continuous)

with $\sum_1^m (d_n - c_n) < \delta_\epsilon$. Thus $\mu_F(A) < \epsilon$ whenever $\lambda(A) < \delta_\epsilon$, so that $\mu_F(A) = 0$ whenever $\lambda(A) = 0$. Now apply the Radon–Nikodym theorem to obtain f as in (7). (Now, let F be $\nearrow$ and absolutely continuous on R. Just apply (7) with $A = [k, k+1]$ for every $-\infty < k < \infty$. Adding up on k gives $\mu_F(A) = 0$ whenever $\lambda(A) = 0$, for any $A \in \mathcal{B}$. Thus $\mu_F \ll \lambda$.) □

Exercise 4.3 (Absolutely continuous dfs) Let F be $\nearrow$, right continuous and bounded on R, with $F(-\infty) = 0$. Define μ_F via $\mu_F((a, b]\,) = F(b) - F(a)$ for all $a < b$. Show that $\mu_F \ll \lambda$ if and only if F is an absolutely continuous function on R.

Exercise 4.4 (a) Show that the composition $g(h)$ of two absolutely continuous functions is absolutely continuous when h is monotone.
(b) Show that $g(h)$ need not be absolutely continuous without restrictions on h.
(c) Define a continuous function on $[0, 1]$ that is not absolutely continuous.
(d) The functions $g + h$ and $g \cdot h$ are absolutely continuous when both f and g are.

Exercise 4.5* Suppose that $h : [a, b] \to (0, \infty)$ is absolutely continuous on $[a, b]$. Show that $\log h$ is also absolutely continuous on $[a, b]$.

Example 4.1 (Change of variable; densities of transformed rvs) Let X be a rv on $(\Omega, \mathcal{A}, P)$ with df $F_X \ll \lambda \equiv$ (Lebesgue measure) and density f_X. Let

(8) $\quad Y \equiv g(X) \qquad$ where g^{-1} is $\uparrow$ and absolutely continuous.

Then

$$F_Y(y) \equiv P(Y \le y) = P(g(X) \le y) = P(X \le g^{-1}(y)) = F_X(g^{-1}(y))\,,$$

where the composition $F_Y = F_X(g^{-1})$ of these absolutely continuous functions is absolutely continuous. So the fundamental theorem of calculus tells us that F_Y is the integral of its derivative. We can then compute this derivative from the ordinary chain rule. Thus

$$F_Y(b) - F_Y(a) = \int_a^b F_Y'(r)\, d\lambda(r) = \int_a^b [F_X'(g^{-1}(r))\, \tfrac{d}{dr}\, g^{-1}(r)]\, d\lambda(r)$$

for all $a \le b$. Thus $F_Y \ll \lambda$ with density

(9) $\quad f_Y(y) = f_X(g^{-1}(y))\, \frac{d}{dy}\, g^{-1}(y)$

on the real line. Call $(d/dy)\, g^{-1}(y)$ the *Jacobian* of the transformation. □

Exercise 4.6 (Specific step functions that are dense in $\mathcal{L}_2$) Let $h \in \mathcal{L}_2([0, 1], \mathcal{B}, \lambda)$. Consider the following two approximations to $h(\cdot)$. Let

$$\bar{h}_m(t) \equiv m\int_{(i-1)/m}^{i/m} h(s)\, ds \qquad \text{and} \qquad \check{h}_m(t) \equiv h(i/(m + 1))$$

for $(i - 1)/m < t \le i/m$ and $m \ge 1$. Show that:

(10) $\quad \bar{h}_m \to h \quad$ a.s. and $\mathcal{L}_2$.

(11) $\quad \check{h}_m \to h \quad$ a.s. and $\mathcal{L}_2 \qquad$ provided that h is $\nearrow$.

[Hint. Show that $0 \le \int_0^1 (\bar{h}_m - h)^2\, dt = \int_0^1 (h^2 - \bar{h}_m^2)\, dt$, and then

$$\bar{h}_m(t) = m(i/m - t)\,\{\int_t^{i/m} h\, ds/(i/m - t)\}$$
$$\qquad + m(t - (i - 1)/m)\,\{\int_{(i-1)/m}^t h\, ds/(t - (i - 1)/m)\} \to h(t) \quad \text{a.s.}$$

Alternatively, use the fact that the continuous functions are dense in $\mathcal{L}_2$.]

Exercise 4.7* (Another characterization of absolute continuity)
(a) F is Lipschitz on $[a, b]$ iff F is differentiable a.e. λ on $[a, b]$ with F' bounded.
(b) Absolutely continuous functions on R map $\mathcal{B}$ into $\mathcal{B}$ and null sets into null sets.
(c) A continuous function of BV is absolutely continuous iff it maps $\mathcal{B}$ into $\mathcal{B}$.

Chapter 5

Measures and Processes on Products

1 Finite-Dimensional Product Measures

Definition 1.1 (Product spaces) Suppose $(\Omega, \mathcal{A})$ and $(\Omega', \mathcal{A}')$ are measurable spaces. Define

$$(1) \qquad \mathcal{A} \times \mathcal{A}' \equiv \sigma[\mathcal{F}] \ \text{ where } \ \mathcal{F} \equiv \Big\{ \sum_{i=1}^{m} (A_i \times A_i') \,:\, m \geq 1, A_i \in \mathcal{A} \text{ and } A_i' \in \mathcal{A}' \Big\},$$

$$(2) \qquad \mathcal{F}_0 \equiv \{A \times A' \,:\, A \in \mathcal{A} \text{ and } A' \in \mathcal{A}'\}.$$

Here $A \times A' \equiv \{(\omega, \omega') : \omega \in A, \omega' \in A'\}$ is called a *measurable rectangle*. The σ-field $\mathcal{A} \times \mathcal{A}' \equiv \sigma[\mathcal{F}]$ is called the *product σ-field*. $(\Omega \times \Omega', \mathcal{A} \times \mathcal{A}')$ is called the *product measurable space*. The sets $A \times \Omega'$ and $\Omega \times A'$ are called *cylinder sets*.

Proposition 1.1 $\mathcal{F}$ is a field. (See figure 5.1, and write the displayed set as a disjoint union of sets in $\mathcal{F}_0$.)

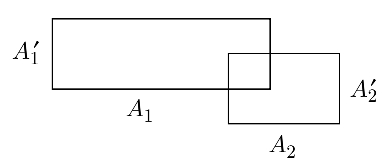


Figure 1.1 The field $\mathcal{F}$.

Theorem 1.1 (Existence of the product measure) Let $(\Omega, \mathcal{A}, \mu)$ and $(\Omega', \mathcal{A}', \nu)$ be σ-finite measure spaces. Define ϕ on the field $\mathcal{F}$ via

$$(3) \qquad \phi\Big(\sum_{i=1}^{m} (A_i \times A_i') \Big) = \sum_{i=1}^{m} \mu(A_i) \times \nu(A_i').$$

Then ϕ is a well-defined and σ-finite measure on the field $\mathcal{F}$. Moreover, ϕ extends uniquely to a σ-finite measure, called the *product measure* and also denoted by ϕ, on $(\Omega \times \Omega', \mathcal{A} \times \mathcal{A}')$. Even when completed, this measure is still unique and is still referred to as the product measure ϕ.

Proof. (See the following exercise; it mimicks the proof of the correspondence theorem. Here, $\mathcal{F}_0$ and $\mathcal{F}$ play the roles of all finite intervals $\mathcal{I}$ and the field $\mathcal{C}_F$. Although the proof asked for in exercise 1.1 below is "obvious," it still requires some tedious detail.) We will give a better proof herein very soon. □

Exercise 1.1 Verify that ϕ is well-defined on $\mathcal{F}_0$, and that ϕ is countably additive on $\mathcal{F}_0$. Then verify that ϕ is well-defined on $\mathcal{F}$, and that ϕ is countably additive on $\mathcal{F}$. Thus ϕ is a σ-finite measure on $\mathcal{F}$, so that the conclusion of theorem 1.1 follows from the Carathéodory extension of theorem 1.2.1 and its corollary.

Exercise 1.2* Use induction to show that theorem 1.1 extends to n-fold products.

Example 1.1 (Lebesgue measure in n dimensions, etc.) (a) We define

$$(R_n, \mathcal{B}_n) = \prod_{i=1}^{n}(R, \mathcal{B}) \qquad \text{and} \qquad (\bar{R}_n, \bar{\mathcal{B}}_n) \equiv \prod_{i=1}^{n}(\bar{R}, \bar{\mathcal{B}})$$

to be the n-fold products of the real line R with the Borel sets $\mathcal{B}$ and of the extended real line $\bar{R}$ with the σ-field $\bar{\mathcal{B}} \equiv \sigma[\mathcal{B}, \{+\infty\}, \{-\infty\}]$, respectively. Recall from example 2.1.1 that $\mathcal{B}_n = \sigma[\mathcal{U}_n]$, where $\mathcal{U}_n$ denotes all open subsets of R_n. We will refer to both $\mathcal{B}_n$ and $\bar{\mathcal{B}}_n$ as the *Borel sets*.
(b) Let λ denote Lebesgue measure on $(R, \mathcal{B})$, as usual. We extend λ to $(\bar{R}, \bar{\mathcal{B}})$ by the convention that $\lambda(\{+\infty\}) = 0$ and $\lambda(\{-\infty\}) = 0$. Then

$$(4) \qquad (R_n, \mathcal{B}_n, \lambda_n) \equiv \prod_{i=1}^{n}(R, \mathcal{B}, \lambda) \qquad \text{and} \qquad (\bar{R}_n, \bar{\mathcal{B}}_n, \lambda_n) \equiv \prod_{i=1}^{n}(\bar{R}, \bar{\mathcal{B}}, \lambda)$$

provides us with a definition of *n-dimensional Lebesgue measure* λ_n as the natural generalization of the concept of volume. It is clear that

$$(5) \qquad (R_m \times R_n, \mathcal{B}_m \times \mathcal{B}_n, \lambda_m \times \lambda_n) = (R_{m+n}, \mathcal{B}_{m+n}, \lambda_{m+n}),$$

and that this holds on the extended Euclidean spaces as well. (It is usual not to add the $\hat{}$ symbol in dealing with the completions of these particular measures.)
(c) Now, λ is just a particular Lebesgue–Stieltjes measure on $(R, \mathcal{B})$. Any Lebesgue–Stieltjes measure μ_F on $(R, \mathcal{B})$ or $(\bar{R}, \bar{\mathcal{B}})$ yields an obvious n-fold product on either $(R_n, \mathcal{B}_n)$ or $(\bar{R}_n, \bar{\mathcal{B}}_n)$, which could appropriately be denoted by $\mu_F \times \cdots \times \mu_F$. Further, we will let $\mathcal{F}_n$ denote the field consisting of all finite disjoint unions of sets of the form $I_1 \times \cdots \times I_n$ where each I_k is of the form $(a, b]$,$(-\infty, b]$ or $(a, +\infty)$ when considering $(R_n, \mathcal{B}_n)$ (or of the form $(a, b]$, $[-\infty, b]$, or $(a, +\infty]$ when considering $(\bar{R}_n, \bar{\mathcal{B}}_n)$). (That is, in the case of $(R_n, \mathcal{B}_n)$ there is the alternative field $\mathcal{F}_n$ that also generates the σ-field $\mathcal{B}_n$; and this $\mathcal{F}_n$ is made up of simpler sets than is the field $\mathcal{B} \times \cdots \times \mathcal{B}$ used in definition 1.1.)
(d) The Halmos approximation lemma now shows that if $(\mu_F \times \cdots \times \mu_F)(A) < \infty$ and if $\epsilon > 0$ is given, then $(\mu_F \times \cdots \times \mu_F)(A \triangle C_\epsilon) < \epsilon$ for some C_ϵ in (the simpler field) $\mathcal{F}_n$. That is, the simpler field gives us a nicer conclusion in this example, because its sets C are simpler. (Or, use A_ϵ in the field $\mathcal{F}$ of (1) in place of C_ϵ.) □

Exercise 1.3 Any Lebesgue–Stieltjes measure $\mu_{F_1} \cdot\cdot \times \mu_{F_n}$ on $(R_n, \mathcal{B}_n)$, in example 1.1 is a regular measure. Show this for $n = 2$ (appeal to theorem 1.3.1).

Definition 1.2 (Sections) (a) Let X denote a function on $\Omega \times \Omega'$. For each ω in Ω, the function $X_\omega(\cdot)$ on Ω' defined by $X_\omega(\omega') \equiv X(\omega, \omega')$ for each ω' in Ω' is called an ω-*section* of $X(\cdot, \cdot)$. An ω'-*section* $X_{\omega'}(\cdot)$ of $X(\cdot, \cdot)$ is defined analogously. (b) Let C be a subset of $\Omega \times \Omega'$. For each ω in Ω, the set $C_\omega = \{\omega' : (\omega, \omega') \text{ is in } C\}$ is called the ω-*section* of C. An ω'-*section* of C is defined analogously.

Theorem 1.2 (Product measure) Let $(\Omega, \mathcal{A}, \mu)$ and $(\Omega', \mathcal{A}', \nu)$ denote finite measure spaces. Let $C \in \mathcal{A} \times \mathcal{A}'$. Then:

(6) $\quad$ Every $C_{\omega'} \in \mathcal{A}$ and every $C_\omega \in \mathcal{A}'$ $\qquad$ whenever $C \in \mathcal{A} \times \mathcal{A}'$,

(7) $$\phi(C) \equiv \int_{\Omega'} \mu(C_{\omega'})\, d\nu(\omega') = \int_\Omega \nu(C_\omega)\, d\mu(\omega) \qquad \text{for every } C \in \mathcal{A} \times \mathcal{A}',$$

and this ϕ is exactly the product measure $\phi = \mu \times \nu$ of theorem 1.1 .

Proof. We first show (6). This result is trivial for any C in $\mathcal{F}_0$, or any C in $\mathcal{F}$. Now let $\mathcal{S}$ denote the class of all sets C in $\mathcal{A} \times \mathcal{A}'$ for which (6) is true. Then $\mathcal{S}$ is trivially seen to be a σ-field, using

(a) $\quad (\cup_n C_n)_{\omega'} = \cup_n C_{n,\omega'}$ and $(C^c)_{\omega'} = (C_{\omega'})^c$.

But since $\mathcal{F} \subset \mathcal{S}$, we have that $\mathcal{A} \times \mathcal{A}' = \sigma[\mathcal{F}]$ equals $\mathcal{S}$.

Consider (7). Note that if the sets C_n converge monotonically to some set C, then 1_{C_n} converges monotonically to 1_C and

(b) $\quad$ every section of 1_{C_n} converges monotonically to the corresponding section of 1_C.

Let $\mathcal{M}$ denote the collection of all sets C in $\mathcal{A} \times \mathcal{A}'$ for which (7) holds. Clearly, $\mathcal{M}$ contains $\mathcal{F}_0$ and $\mathcal{F}$. We now use (b) to show that $\mathcal{M}$ is a monotone class; it will then follow by proposition 1.1.6 that $\mathcal{M} = \sigma[\mathcal{F}] = \mathcal{A} \times \mathcal{A}'$. Let C_n denote a sequence of sets in the class $\mathcal{M}$ that converge monotonically (we will consider only the $\nearrow$ case, since we only need to take complements in the $\searrow$ case), and we give the name C to the limiting set. Since $1_{C_n} \nearrow 1_C$, the function 1_C is $(\mathcal{A} \times \mathcal{A}')$-measurable, and thus every section of 1_C is measurable by (6). Now, for fixed ω' the number $h(\omega') \equiv \mu(C_{\omega'}) = \int_\Omega 1_{C_{\omega'}}(\omega) d\mu(\omega)$ is (by the MCT and (b)) the $\nearrow$ limit of the sequence of numbers $h_n(\omega') \equiv \mu(C_{n,\omega'}) = \int_\Omega 1_{C_{n,\omega'}}(\omega) d\mu(\omega)$, for each ω' in Ω'. Thus the function h on Ω' is the limit of the functions h_n on Ω'; and since C_n is in $\mathcal{M}$, the functions h_n are $\mathcal{A}'$-measurable by (7); thus h is $\mathcal{A}'$-measurable by proposition 2.2.2. Moreover, the finite $\nearrow$ numbers $\phi(C_n)$ are bounded above by $\mu(\Omega)\, \nu(\Omega')$, and thus converge to some number; call it $\phi(C)$. That is,

(c) $\quad \phi(C) \equiv \lim_n \phi(C_n) = \lim_n \int_{\Omega'} \{\int_\Omega 1_{C_{n,\omega'}}(\omega) d\mu(\omega)\}\, d\nu(\omega')$ for $C_n \in \mathcal{M}$

(d) $\qquad = \lim_n \int_{\Omega'} h_n d\nu = \int_{\Omega'} \{\lim_n h_n\}\, d\nu = \int_{\Omega'} h d\nu \quad$ by the MCT and $h_n \nearrow h$

$\qquad = \int_{\Omega'} \{\int_\Omega 1_{C_{\omega'}}(\omega) d\mu(\omega)\}\, d\nu(\omega') \qquad$ by the definition of h

(e) $\qquad = \int_{\Omega'} \mu(C_{\omega'})\, d\nu(\omega')\,.$

(Since $\phi(C)$ is finite, we see that h is ν-integrable. Thus $h(\omega')$ is finite for a.e. $[\nu]$ ω'.) The argument for each fixed ω is symmetric, and it gives the second equality in (7). Thus C is in $\mathcal{M}$, making $\mathcal{M}$ the monotone class $\mathcal{A} \times \mathcal{A}'$; and (b) holds. [Thus the result (7) holds for the set function ϕ. But is ϕ a measure?]

In this paragraph we will show that the product measure ϕ of theorem 1.1 exists, and is defined by (e). To this end, let $D_1, D_2, \ldots$ be pairwise disjoint sets in $\mathcal{A} \times \mathcal{A}'$, and let $C_n \equiv \sum_1^n D_k \nearrow C \equiv \sum_1^\infty D_k$. Then linearity of both single integrals shows (in the second equality) that

(f) $\qquad \sum_1^\infty \phi(D_k) = \lim_n \sum_1^n \phi(D_k) = \lim_n \phi(\sum_1^n D_k) = \lim_n \phi(C_n)$

(g) $\qquad = \phi(C) = \phi(\sum_1^\infty D_k), \qquad$ by (c) through (e)

so that ϕ is c.a., and a measure on $\mathcal{A} \times \mathcal{A}'$. We have just verified that the product measure of (3) exists on $\mathcal{A} \times \mathcal{A}'$, and is given by (7). That is, we have just proven theorem 1.1 and given the representation (7) for $\phi(C)$. Note that the product measure ϕ also satisfies $\phi(C) = \int_{\Omega\times\Omega'} 1_C(\omega, \omega')\, d\phi(\omega, \omega')$. □

Exercise 1.4 Give the details to verify that $\sum_1^n \phi(D_k) = \phi(\sum_1^n D_k)$ in line (f) of the proof above of the product measure theorem.

Theorem 1.3 (Fubini) Let $(\Omega, \mathcal{A}, \mu)$ and $(\Omega', \mathcal{A}', \nu)$ be σ-finite measure spaces. Let $\phi = \mu \times \nu$ on $(\Omega \times \Omega', \mathcal{A} \times \mathcal{A}')$. Suppose that $X(\omega, \omega')$ is ϕ-integrable (i.e., $X^{-1}(\bar{\mathcal{B}}\,) \subset \mathcal{A} \times \mathcal{A}'$ and $\int_{\Omega\times\Omega'} X\, d\phi$ is finite). Then:

(8) All ω'-sections $X_{\omega'}(\cdot)$ of X are $\mathcal{A}$-measurable functions on Ω.

(9) For a.e. $[\nu]$ fixed ω', the function $X_{\omega'}(\cdot) = X(\cdot, \omega')$ is μ-integrable.

(10) The function $h(\omega') \equiv \displaystyle\int_\Omega X_{\omega'}(\omega)\, d\mu(\omega)$ is a ν-integrable function of ω'.

(11) $$\int_{\Omega\times\Omega'} X(\omega, \omega') d\phi(\omega, \omega') = \int_{\Omega'} [\int_\Omega X(\omega, \omega') d\mu(\omega)] d\nu(\omega') = \int_{\Omega'} h(\omega') d\nu(\omega').$$

[Setting X equal to 1_C in (11) for $C \in \mathcal{A} \times \mathcal{A}'$ shows how the value $\phi(C)$ of the product measure ϕ at C was defined as an iterated integral; recall (7).]

Corollary 1 (Tonelli) Let X be $\mathcal{A} \times \mathcal{A}'$-measurable and suppose either

(12) $$\int \Big[\int |X|\, d\mu \Big]\, d\nu < \infty \qquad \text{or} \qquad \int \Big[\int |X|\, d\nu \Big]\, d\mu < \infty \qquad \text{or} \qquad X \geq 0.$$

Then the claims of Fubini's theorem are true, including

(13) $$\int X\, d\phi = \int \Big[\int X\, d\mu \Big]\, d\nu = \int \Big[\int X\, d\nu \Big]\, d\mu.$$

Corollary 2 ($\mu \times \nu$ null sets) A set C in $\mathcal{A} \times \mathcal{A}'$ is $(\mu \times \nu)$-null if and only if almost every ω-section of C is a ν-null set. That is, for $C \in \mathcal{A} \times \mathcal{A}'$ we have

$$(14) \qquad \mu \times \nu(C) = 0 \quad \text{if and only if} \quad \nu(C_\omega) = 0 \quad \text{for a.e. } [\mu]\ \omega \text{ in } \Omega.$$

Proof. By using the σ-finiteness of the two measures to decompose both Ω and Ω', we may assume in this proof that both μ and ν are finite measures. We begin by discussing only measurability questions.

We will first show that

(a) all ω'-sections of an $(\mathcal{A} \times \mathcal{A}')$-measurable function X are A-measurable.

The previous theorem shows that

(b) all ω'-sections $X_{\omega'}$ of X are $\mathcal{A}$-measurable whenever $X = 1_C$ for some $C \in \mathcal{A} \times \mathcal{A}'$.

Now let X denote any $(\mathcal{A} \times \mathcal{A}')$-measurable function. Then for any B in $\bar{\mathcal{B}}$,

$$(c) \qquad X_{\omega'}^{-1}(B) = \{\omega : X(\omega, \omega') \in B\} = \{\omega : (\omega, \omega') \in X^{-1}(B)\}$$

is the ω'-section of the indicator function of the set $C = X^{-1}(B)$; so (b) shows that any arbitrary ω'-section of this X is $\mathcal{A}$-measurable, and so establishes (a) and (8).

We now turn to all the other claims of the Fubini and Tonelli theorems. By theorem 1.2 they hold for all $(\mathcal{A} \times \mathcal{A}')$-measurable indicator functions. Linearity of the various integrals shows that the theorems also hold for all simple functions. Applying the MCT to the various integrals shows that the theorems also hold for all $(\mathcal{A} \times \mathcal{A}')$-measurable $X \geq 0$. Then linearity of the integral shows that the theorems also hold for all X for whichever of the three integrals exists finitely (the double integral or either iterated integral).

Corollary 2 follows immediately by applying (13) and exercise 3.2.2 (only the zero function) to the integral of the function 1_C. □

Corollary 3 All this extends naturally to n dimensions.

Exercise 1.5 (Fubini's (11) can fail if X is not ϕ-integrable) Let $\Omega = (0, 1)$ and $\Omega' = (1, \infty)$, both equipped with the Borel sets and Lebesgue measure.
(i) Let $f(x, y) = e^{-xy} - 2e^{-2xy}$ for all $x \in \Omega = (0, 1)$ and $y \in \Omega' = (1, \infty)$. Show that:

$$(a) \qquad \int_0^1 [\int_1^\infty f(x, y)\, dy]\, dx = \int_0^1 \tfrac{1}{x} [e^{-x} - e^{-2x}]\, dx \quad \text{is } > 0.$$

$$(b) \qquad \int_1^\infty [\int_0^1 f(x, y)\, dx]\, dy = \int_1^\infty \tfrac{1}{y} [e^{-2y} - e^{-y}]\, dy \quad \text{is } < 0.$$

(ii) Why does Fubini's theorem fail here? (Solve $f(x, y) = 0$, and use this to divide the domain of f. Integrate over each of these two regions separately.)
(iii) Construct another example of this type.

2 Random Vectors on $(\Omega, \mathcal{A}, P)$

We will now treat measurable functions from a probability space $(\Omega, \mathcal{A}, P)$ to a Euclidean space $(R_n, \mathcal{B}_n)$, with $n \geq 1$. Let $\mathsf{x} \equiv (x_1, \ldots, x_n)'$ denote a generic vector in the Euclidean space R_n.

Definition 2.1 (Random vectors) Suppose $\mathsf{X} \equiv (X_1, \ldots, X_n)'$ is such that $\mathsf{X} : \Omega \to R_n$ is $\mathcal{B}_n$-$\mathcal{A}$-measurable. Then X is called a *random vector* (which is also abbreviated *rv*). Define the *joint distribution function* (or just *df*) of X by

$$F(\mathsf{x}) \equiv F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = P(\, \bigcap_{i=1}^n [X_i \leq x_i]\,)\,.$$

Write $\mathsf{x} \leq \mathsf{y}$ to denote that $x_i \leq y_i$ for all $1 \leq i \leq n$; and now define the basic rectangles $(\mathsf{x}, \mathsf{y}\,] \equiv \times_{i=1}^n (x_i, y_i]$ whenever $\mathsf{x} \leq \mathsf{y}$. Let

(1) $$F(\mathsf{x}, \mathsf{y}\,] \equiv P(\, \bigcap_{i=1}^n [x_i < X_i \leq y_i]\,) \qquad \text{for all } \mathsf{x} \leq \mathsf{y}.$$

Proposition 2.1 (Measurability) Now, $\mathsf{X} \equiv (X_1, \ldots, X_n)' : \Omega \to R_n$ or $\bar{R}_n$ is such that

$$\mathsf{X} \text{ is } \begin{cases} \mathcal{B}_n\text{-}\mathcal{A}\text{-measurable} \\ \bar{\mathcal{B}}_n\text{-}\mathcal{A}\text{-measurable} \end{cases} \quad \text{if and only if} \quad \text{each } X_i \text{ is } \begin{cases} \mathcal{B}\text{-}\mathcal{A}\text{-measurable} \\ \bar{\mathcal{B}}\text{-}\mathcal{A}\text{-measurable.} \end{cases}$$

Thus, a random vector is measurable if and only if each coordinate rv is measurable.

Proof. We give the details for finite-valued functions. $(\Rightarrow)$ Now,

$$[X_i \leq x_i] = X_i^{-1}((-\infty, x_i]\,)$$
$$= \mathsf{X}^{-1}(R \times \cdots \times R \times (-\infty, x_i] \times R \times \cdots \times R) \in \mathcal{A}.$$

$(\Leftarrow)$ Also, $[\mathsf{X} \leq \mathsf{x}] = \cap_{i=1}^n [X_i \leq x_i] \in \mathcal{A}$, since each X_i is measurable, where

$$\sigma[\mathcal{H}_n] \equiv \sigma[\,\{\text{all } (-\infty, x_1] \times \cdots \times (-\infty, x_n]\}\,] = \mathcal{B}_n\,.$$

Moreover, using (2.1.12) for the final equality,

$$\mathcal{F}(\mathsf{X}_n) \equiv \mathsf{X}_n^{-1}(\mathcal{B}_n) = \mathsf{X}_n^{-1}(\sigma[\mathcal{H}_n]) = \sigma[\mathsf{X}_n^{-1}(\mathcal{H}_n)] \subset \mathcal{A},$$

with the set inclusion shown in the first line. That is, $\mathsf{X}^{-1}(\mathcal{B}_n) \subset \mathcal{A}$. □

Exercise 2.1 (Joint df) A joint df F is $\nearrow$ and right continuous and satisfies

(2) $$\text{All } F(x_1, \ldots, x_{i-1}, -\infty, x_{i+1}, \ldots, x_n) = 0 \qquad \text{and} \qquad F(\infty, \ldots, \infty) = 1,$$

$$F(x_1, \ldots, x_{i-1}, +\infty, x_{i+1}, \ldots, x_n)$$

(3) $$= F_{X_1,\ldots,X_{i-1},X_{i+1},\ldots,X_n}(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$$

for all $j = 1, \ldots, n$ and $x_1, \ldots, x_n$.

Exercise 2.2 Suppose $F : R_n \to R$ is $\nearrow$ and right continuous and satisfies (2) and (3). Then there exists a unique probability measure $P \equiv P_F$ on $\mathcal{B}_n$ that satisfies

(4) $\quad P((\mathsf{x}, \mathsf{y}]) = F(\mathsf{x}, \mathsf{y}] \qquad$ for all $\mathsf{x} \le \mathsf{y}$.

This is a generalization of the correspondence theorem to $n > 1$. Now note that the identity function $\mathsf{X}(\omega) \equiv \omega$, for each $\omega \in R_n$, is a random vector on $(R_n, \mathcal{B}_n)$ that has as its joint df the function F above. Thus, given any joint df F, there is a random vector X having F as its joint df. This is in the spirit of example 2.2.1.

Definition 2.2 (Joint density of rvs) Let $\mathsf{X} \equiv (X_1, \ldots, X_n)'$ denote a rv. Define $P_n(B) \equiv P(\mathsf{X} \in B)$ for all $B \in \mathcal{B}_n$, so that P_n defines the induced distribution of X on $(R_n, \mathcal{B}_n)$. Let λ_n denote Lebesgue measure on $(R_n, \mathcal{B}_n)$. If $P_n \ll \lambda_n$, then a finite-valued Radon-Nikodym derivative $f_n \equiv dP_n/d\lambda_n$ exists (and is unique a.e. λ_n) for which

(5) $\quad P(\mathsf{X} \in B) = \int_B \cdots \int f_n(x_1, \ldots, x_n)\, dx_1 \times \cdots \times dx_n \qquad$ for all $B \in \mathcal{B}_n$.

When this is true, $f_n(\cdots)$ is called the *joint density* (or, the *density*) of the rv X. (For one-dimensional rvs, we often denote the distribution, df, and density of X by $P_X(\cdot)$, $F_X(\cdot)$, and $f_X(\cdot)$. For two-dimensional rvs $(X, Y)'$, we often use $P_{X,Y}(\cdot)$, $F_{X,Y}(\cdot, \cdot)$, and $f_{X,Y}(\cdot, \cdot)$.)

Exercise 2.3 (Marginal densities) Suppose that $(X_1, \ldots, X_n)'$ has the induced distribution P_n, and $P_n \ll \lambda_n$ with joint density f_n (as in the previous definition). Let $1 \le i_1 < \cdots < i_m \le n$, with $m \le n$, and let $1 \le j_1 < \cdots < j_{n-m} \le n$ denote the complementary indices. Show that the induced distribution P_m of $(X_{i_1}, \ldots, X_{i_m})'$ satisfies $P_m \ll \lambda_m$, and that its joint density is given by

(6) $\quad f_m(x_{i_1}, \ldots, x_{i_m}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_n(x_1, \ldots, x_n)\, dx_{j_1} \times \cdots \times\, dx_{j_{n-m}}$

on R_m. We also call f_m the *marginal density* of $(X_{i_1}, \ldots, X_{i_m})'$.

Exercise 2.4 (Hoeffding–Fréchet bounds)
(i) Show that $H_a(u, v) \equiv u \wedge v$ and $H_b(u, v) \equiv (u + v - 1)^+$ are dfs on $[0, 1]^2$ with Uniform$(0, 1)$ marginal dfs. Determine the sets on which the densities of H_a and H_b are positive. Draw suitable pictures of both of the joint dfs H_a and H_b. Make qualitative comments about the correlation of these two distributions.
(ii) Suppose that $F(x, y)$ is a df on R_2 with marginal dfs $G(x) \equiv F(x, \infty)$ and $H(y) \equiv F(\infty, y)$, respectively. Show that

$$H_b(G(x), H(y)) \le F(x, y) \le H_a(G(x), H(y))$$

for all $(x, y) \in R_2$.

3 Countably Infinite Product Probability Spaces

We now begin to carry out the program discussed in section 2.5. That is, we will extend the notion of rvs and product probability measures to a countably infinite number of dimensions.

Notation 3.1 (R_∞ and $\mathcal{B}_\infty$) Let

(1) $R_\infty \equiv \prod_{n=1}^{\infty} R \equiv \{(x_1, x_2, \ldots) : x_n \in R \text{ for all } n \geq 1\}$.

Let I denote an interval of the type $(c, d\,]$, $(-\infty, d]$, $(c, +\infty)$, or $(-\infty, \infty)$. An *n-dimensional rectangle* will mean any set of the form $I_1 \times \cdots \times I_n \times R \times R \times \cdots$, where each interval I_i is of the type above. A *finite-dimensional rectangle* is an n-dimensional rectangle, for some $n \geq 1$. A *cylinder set* is defined as a set of the form $B_n \times R \times R \times \cdots$ with B_n in $\mathcal{B}_n$ for some $n \geq 1$. Thus:

(2) $\mathcal{C}_I \equiv \{\text{all finite-dimensional rectangles}\}$

$$= \{I_1 \times \cdots \times I_n \times R \times R \times \cdots \; : \; n \geq 1, \text{ all } I_i \text{ as above}\},$$

(3) $\mathcal{C}_F \equiv \{\text{all finite disjoint unions of finite-dimensional rectangles}\}$,

(4) $\mathcal{C}_\infty \equiv \{\text{all cylinder sets}\} \equiv \{B_n \times R \times R \times \cdots \; : \; n \geq 1, B_n \in \mathcal{B}_n\}$.

Both $\mathcal{C}_F$ and $\mathcal{C}_\infty$ are fields, and a trivial application of exercise 1.1.1 shows that

(5) $\mathcal{B}_\infty \equiv \sigma[\mathcal{C}_I] = \sigma[\mathcal{C}_F] = \sigma[\mathcal{C}_\infty]$.

Thus, extending a measure from $\mathcal{C}_I$ to $\mathcal{B}_\infty$ will be of prime interest to us. We first extend the criterion for measurability from n dimensions to a countably infinite number of dimensions. □

Proposition 3.1 (Measurability on $\mathcal{B}_\infty$) (a) Now, $\mathsf{X} \equiv (X_1, X_2, \ldots)' : \Omega \to R_\infty$ is $\mathcal{B}_\infty$-$\mathcal{A}$-measurable if and only if each X_n is $\mathcal{B}$-$\mathcal{A}$-measurable.
(b) If X is $\mathcal{B}_\infty$-$\mathcal{A}$-measurable and if $(i_1, i_2, \ldots)$ is an arbitrary sequence of integers, then $\mathsf{Y} \equiv (X_{i_1}, X_{i_2}, \ldots)'$ is $\mathcal{B}_\infty$-$\mathcal{A}$-measurable.

Exercise 3.1 Prove proposition 3.1.

Notation 3.2 We will use the notation

(6) $\mathcal{F}(X_i) \equiv X_i^{-1}(\mathcal{B})$ and $\mathcal{F}(X_{i_1}, X_{i_2}, \ldots) \equiv \mathsf{Y}^{-1}(\mathcal{B}_\infty) = \sigma[\bigcup_{n=1}^{\infty} X_{i_n}^{-1}(\mathcal{B})]$

to denote the minimal sub σ-fields of $\mathcal{A}$ relative to which the quantities X_i and $\mathsf{Y} \equiv (X_{i_1}, X_{i_2}, \ldots)$ are measurable. □

Now suppose that P_n is a probability measure on $(R_n, \mathcal{B}_n)$, for each $n \geq 1$. The question is: When can we extend the collection $\{P_n : n \geq 1\}$ to a measure on $(R_\infty, \mathcal{B}_\infty)$? Reasoning backwards to see what conditions the family of finite-dimensional distributions should satisfy leads to the following definition.

Definition 3.1 (Consistency) Finite-dimensional distributions $\{(R_n, \mathcal{B}_n, P_n)\}_{n=1}^{\infty}$ are *consistent* if for every $n \geq 1$, every $B_1, \ldots, B_n \in \mathcal{B}$, and every $1 \leq i \leq n$,

$$P_{n-1}((X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n) \in B_1 \times \cdots \times B_{i-1} \times B_{i+1} \cdots \times B_n)$$

$$= P_n((X_1, \ldots, X_{i-1}, X_i, X_{i+1}, \ldots, X_n) \in B_1 \times \cdots \times B_{i-1} \times R \times B_{i+1} \times \cdots \times B_n). \tag{7}$$

Theorem 3.1 (Kolmogorov's extension theorem) An extension of any consistent family of probability measures $\{(R_n, \mathcal{B}_n, P_n)\}_{n=1}^{\infty}$ to a probability $P(\cdot)$ on $(R_\infty, \mathcal{B}_\infty)$ necessarily exists, and it is unique.

We will first summarize the main part of this proof as a separately stated result that seems of interest in its own right.

Theorem 3.2 ($(R_\infty, \mathcal{B}_\infty)$ extension theorem; Breiman) Let P on $\mathcal{C}_I$ satisfy:
(a) $P \geq 0$ and $P(R_\infty) = 1$.
(b) If $D = \sum_{j=1}^m D_j$ for n-fold rectangles D and D_j, then $P(D) = \sum_1^m P(D_j)$.
(c) If D denotes any fixed n-dimensional rectangle, then there exists a sequence of compact n-dimensional rectangles D_j for which $D_j \nearrow D$ and $P(D_j) \nearrow P(D)$.
[That is, P is well-defined and additive on n-dimensional rectangles and satisfies something like continuity from below.]
Then there exists a unique extension of P to $\mathcal{B}_\infty$.

Proof. (Recall the continuity result of proposition 1.1.3.) Now,

$$\mathcal{C}_F \equiv \{\text{all finite disjoint unions of finite-dimensional rectangles}\}$$

$$= \{\text{a field generating } \mathcal{B}_\infty\}. \tag{p}$$

For $A = \sum_1^m D_j \in \mathcal{C}_F$, define $P(A) \equiv \sum_{j=1}^m P(D_j)$.

First, we will show that P is well-defined on $\mathcal{C}_F$. Let $A = \sum_1^m D_j = \sum_1^{m'} D'_k$. Now, $D'_k = D'_k A = \sum_1^m D'_k D_j$ and $D_j = D_j A = \sum_1^{m'} D_j D'_k$. Thus

$$P(A) = \sum_1^m P(D_j) = \sum_1^m P(\sum_1^{m'} D_j D'_k) = \sum_1^m \sum_1^{m'} P(D_j D'_k) \tag{q}$$

$$= \sum_1^{m'} \sum_1^m P(D'_k \cap D_j) = \sum_1^{m'} P(\sum_1^m D'_k \cap D_j) = \sum_1^{m'} P(D'_k) = P(A). \tag{r}$$

Next, we will show that P is f.a. on $\mathcal{C}_F$. So we let $A_1, \ldots, A_m \in \mathcal{C}_F$ be such that $A \equiv \sum_1^m A_i \in \mathcal{C}_F$ also. Then, writing $A_i = \sum_1^{m_i} D_{ij}$ with $D_{i1}, \ldots, D_{i,m_i}$ disjoint,

$$P(A) = P(\sum_1^m A_i) = P(\sum_1^m \sum_1^{m_i} D_{ij}) = \sum_1^m \sum_1^{m_i} P(D_{ij}) = \sum_1^m P(A_i), \tag{s}$$

(using condition (b) in each of the last two equalities), since P is well-defined.

We will now show that P is continuous from above at $\emptyset$. Let A_n's in $\mathcal{C}_F$ be such that $A_n \searrow \emptyset$. We must show that $P(A_n) \searrow 0$. Assume not. Then $P(A_n) \searrow \epsilon > 0$; and by going to subsequences, we may assume that $A_n = A_n^* \times \prod_{n+1}^{\infty} R$, where each A_n^* is a finite union of disjoint rectangles (repeat some members of the sequence if necessary in order to have $A_n^* \subset R_n$). By condition (c), choose $B_n^* \subset A_n^*$ such that B_n^* is a finite union of compact disjoint rectangles in R_n with

(t) $\quad P(A_n \setminus B_n) < \epsilon/2^{n+1}, \quad$ where $\quad B_n \equiv B_n^* \times \prod_{n+1}^{\infty} R.$

Let $C_n = \bigcap_1^n B_k \equiv C_n^* \times \prod_{n+1}^{\infty} R$, with C_n^* compact in R_n (the B_n's need not be $\searrow$, but the C_n's are). Then we observe that $C_n \searrow \emptyset$, since $C_n \subset B_n \subset A_n$ with $A_n \searrow \emptyset$; but we also have $P(C_n) \geq \epsilon/2$, since

(u) $\quad P(A_n \setminus C_n) \leq \sum_{k=1}^n P(A_n \setminus B_k) \leq \sum_{k=1}^n P(A_k \setminus B_k) \leq \sum_1^n \epsilon/2^{k+1} \leq \epsilon/2.$

But $C_n \searrow$ with $P(C_n) \geq \epsilon/2$ for all n is not compatible with the conclusion that $C_n \searrow \emptyset$: Let $\mathsf{x}^{(1)} \in C_1, \ldots, \mathsf{x}^{(n)} \in C_n, \ldots$, where $\mathsf{x}^{(n)} \equiv (x_1^{(n)}, x_2^{(n)}, \ldots)$. Choose an initial subsequence N_1 such that $x_1^{(N_1)} \to$ (some x_1) $\in C_1^*$; then choose a further subsequence N_2 such that $(x_1^{(N_2)}, x_2^{(N_2)}) \to$ (some (x_1, x_2)) $\in C_2^*; \ldots$. Along the diagonal subsequence, say N, we have $x_j^{(N)} \to x_j$, for all j. Now, $\mathsf{x} = (x_1, x_2, \ldots) \in C_n$ for all n. Hence $C_n \not\searrow \emptyset$. But this is a contradiction, and thus allows us to claim that $P(A_n) \searrow 0$ for any A_n's in $\mathcal{C}_F$ that satisfy $A_n \searrow \emptyset$.

Now apply the continuity of measures in proposition 1.1.3, and then apply the Carathéodory extension of theorem 1.2.1 to complete the proof. □

Proof. We now turn to the Kolmogorov extension theorem. The P defined by

(v) $\quad P(B_1 \times \cdots \times B_n \times \cdots) \equiv P_n(B_1 \times \cdots \times B_n) = P_{m+n}(B_1 \times \cdots \times B_n \times R \times \cdots \times R)$

is a well-defined f.a. probability on $\mathcal{C}_I = \{$all finite-dimensional rectangles$\}$; this follows from the consistency condition (7). Thus (a) and (b) of theorem 3.2 hold.

We will now verify (c). Fix n. Let D_n be an arbitrary but fixed n-dimensional rectangle. It is clearly possible to specify compact n-dimensional rectangles D_{nj} for which $D_{nj} \nearrow D_n$ as $j \to \infty$. Write $D_j = D_{nj} \times \prod_{n+1}^{\infty} R$ and $D = D_n \times \prod_{n+1}^{\infty} R$, so that $D_j \nearrow D$. Thus, by the continuity of signed measures in proposition 1.1.3,

(w) $\quad P(D_j) = P_n(D_{nj}) \nearrow P_n(D_n) = P(D),$

since P_n is a measure on $(R_n, \mathcal{B}_n)$. Thus (c) holds. The conclusion follows from theorem 3.2. □

Example 3.1 (Coordinate rvs) Once consistent probability measures $P_n(\cdot)$ on $(R_n, \mathcal{B}_n)$ have been extended to a probability measure $P(\cdot)$ on $(R_\infty, \mathcal{B}_\infty)$, it is appropriate then to define $X_n(x_1, x_2, \ldots) = x_n$, for each $n \geq 1$. These are rvs on the probability space $(\Omega, \mathcal{B}, P) \equiv (R_\infty, \mathcal{B}_\infty, P)$. Moreover,

$$P((X_1, \ldots, X_n) \in B_n) = P((X_1, \ldots, X_n)^{-1}(B_n)) = P(\mathsf{X}^{-1}(B_n \times \textstyle\prod_{n+1}^{\infty} R))$$

(8) $$= P(B_n \times \textstyle\prod_{n+1}^{\infty} R) = P_n(B_n)$$

for all $B_n \in \mathcal{B}_n$. We thus have a realization of $\mathsf{X} \equiv (X_1, X_2, \ldots) : \Omega \to R_\infty$ that is $\mathcal{B}_\infty$-$\mathcal{A}$-measurable, and each $(X_1, \ldots, X_n)$ induces the distribution P_n on $(R_n, \mathcal{B}_n)$. This is the natural generalization of example 2.2.1 and the comment below exercise 5.2.2. □

Theorem 3.3 (The finite dimensional dfs define probability theory) Let $\mathsf{X} = (X_1, X_2, \ldots)'$ denote any random element on $(R_\infty, \mathcal{B}_\infty)$. Then P_{X} can be determined solely by examination of the finite-dimensional distributions of X. Also, whether or not there exists a finite rv X such that X_n converges to X in the sense of $\to_{a.s.}$, $\to_p$, $\to_r$, or $\to_d$ can be similarly determined.

Proof. Let $\mathcal{C}$ denote the $\bar{\pi}$-system consisting of R_∞ and of all sets of the form $\prod_1^n(-\infty, x_i] \times \prod_{n+1}^\infty R$, for some $n \geq 1$ and all $x_i \in R$. The finite-dimensional distributions (even the finite-dimensional dfs) determine P_∞ on $\mathcal{C}$, and hence on $\mathcal{B}_\infty = \sigma[\mathcal{C}]$ (appeal to Dynkin's π-λ theorem of proposition 1.1.5). To emphasize the fact further, we now consider each convergence mode separately.
$\to_d$: Obvious.
$\to_r$: $\mathrm{E}|X_n - X|^r \to 0$ if and only if $\mathrm{E}|X_n - X_m|^r < \epsilon$ for all $n, m \geq$ some N_ϵ.
$\to_p$: $X_n \to_p X$ if and only if $P(|X_n - X_m| > \epsilon) < \epsilon$ for all $n, m \geq$ some N_ϵ.
$\to_{a.s.}$: $X_n \to_{a.s.} X$ if and only if

$$\begin{aligned} 1 &= P(\cup_{n=1}^\infty \cap_{m=n}^\infty [|X_m - X_n| \leq \epsilon]) \qquad \text{for all } \epsilon > 0 \\ &= \lim_n \lim_N P(\cap_{m=n}^N [|X_m - X_n| \leq \epsilon]) = \lim_n \lim_N \{\text{a function of } F_{X_n,\ldots,X_N}\}. \end{aligned}$$

The proof is complete □

Example 3.2 (Equivalent experiments) Perhaps I roll an ordinary die n times with the appearance of an even number called "success." Perhaps I draw a card at random n times, each time from a freshly shuffled deck of standard playing cards, with "red" called "success." Perhaps I flip a fair coin n times with "heads" called "success." Note that $(X_1, \ldots, X_n)$ has the same distribution in all three cases. Thus, if I report only the data from one of these experiments, you can not hope to determine which of the three experiments was actually performed. These are called *equivalent experiments.* □

4 Random Elements and Processes on $(\Omega, \mathcal{A}, P)$

Definition 4.1 (Projections and finite-dimensional subsets) Let M_T denote a collection of functions that associate with each t of some set T a real number denoted by either x_t or $x(t)$. [T is usually a Euclidean set such as $[0,1]$, R, or $[0,1] \times R$. The collection $\mathcal{M}_T$ is often a collection of "nice" functions, such as the continuous functions on T.] For each integer k and all $(t_1, \ldots, t_k)$ in T we let $\pi_{t_1,\ldots,t_k}$ denote the *projection mapping* of M_T into k-dimensional space R_k defined by

$$(1) \qquad \pi_{t_1,\ldots,t_k}(x) \equiv (x(t_1), \ldots, x(t_k)).$$

Then for any B in the set of all k-dimensional Borel subsets $\mathcal{B}_k$ of R_k, the set $\pi^{-1}_{t_1,\ldots,t_k}(B)$ is called a *finite-dimensional subset* of M_T.

Exercise 4.1 Show that the collection $\mathcal{M}^0_T$ of all finite-dimensional subsets of M_T is necessarily a field. (This is true no matter what collection M_T is used.)

Definition 4.2 (Measurable function spaces, finite-dimensional distributions, random elements, and normal processes) We let $\mathcal{M}_T$ denote the σ-field generated by the field $\mathcal{M}^0_T$. We call $\mathcal{M}^0_T$ and $\mathcal{M}_T$ the *finite-dimensional field* and the *finite-dimensional σ-field*, respectively. Call the measurable space $(M_T, \mathcal{M}_T)$ a *measurable function space* over T.

Given any probability space $(\Omega, \mathcal{A}, P)$ and any measurable space $(\Omega^*, \mathcal{A}^*)$, an $\mathcal{A}^*$-$\mathcal{A}$-measurable mapping $X : \Omega \to \Omega^*$ will be called a *random element.* We denote this by $X : (\Omega, \mathcal{A}) \to (\Omega^*, \mathcal{A}^*)$ or by $X : (\Omega, \mathcal{A}, P) \to (\Omega^*, \mathcal{A}^*)$, or even by $X : (\Omega, \mathcal{A}, P) \to (\Omega^*, \mathcal{A}^*, P^*)$, where P^* denotes the induced probability on the image space.

A random element $X : (\Omega, \mathcal{A}, P) \to (M_T, \mathcal{M}_T, P^*)$ in which the image space is a measurable function space will be called a *process.* The *finite-dimensional distributions* of a process are the distributions induced on $(R_k, \mathcal{B}_k)$ by the projection mappings $\pi_{t_1,\ldots,t_k} : (M_T, \mathcal{M}_T, P^*) \to (R_k, \mathcal{B}_k)$. If all of the finite-dimensional distributions of a process X are multivariate normal (see section A.3 below), then we call X a *normal process.*

Definition 4.3 (Realizations and versions) If two random elements X and Y (possibly from different probability spaces to different measurable function spaces) have identical induced finite-dimensional distributions, then we refer to X and Y as different *realizations* of the same random element and we call them *equivalent* random elements. We denote this by agreeing that

$$X \cong Y \qquad \text{means that } X \text{ and } Y \text{ are equivalent random elements.}$$

[We will see in chapter 12 that a process called Brownian motion can be realized on both the $(R_{[0,1]}, \mathcal{B}_{[0,1]})$ of (3) and $(C, \mathcal{C})$, where $C \equiv C_{[0,1]}$ denotes the space of all continuous functions on $[0,1]$ and $\mathcal{C} \equiv \mathcal{C}_{[0,1]}$ denotes its finite-dimensional σ-field.]

If X and Y are defined on the same probability space and $P(X_t = Y_t) = 1$ for all $t \in T$, then X and Y are called *versions* of each other. (In chapter 12 we will see versions X and Y of Brownian motion where $X : (\Omega, \mathcal{A}, P) \to (R_{[0,1]}, \mathcal{B}_{[0,1]})$ and $Y : (\Omega, \mathcal{A}, P) \to (C_{[0,1]}, \mathcal{C}_{[0,1]})$. Of course, this X and Y are also different realizations of Brownian motion.)

Definition 4.4 (Finite-dimensional convergence, $\to_{fd}$) Suppose $X, X_1, X_2, \ldots$ denote processes with image space $(M_T, \mathcal{M}_T)$. If the convergence in distribution

(2) $\pi_{t_1,\ldots,t_k}(X_n) = (X_n(t_1), \ldots, X_n(t_k)) \to_d (X(t_1), \ldots, X(t_k)) = \pi_{t_1,\ldots,t_k}(X)$

holds for all $k \geq 1$ and all $t_1, \ldots, t_k$ in T, then we write $X_n \to_{fd} X$ as $n \to \infty$, and we say that the *finite-dimensional distributions* of X_n *converge* to those of X.

The General Stochastic Process

Notation 4.1 $(R_T, \mathcal{B}_T)$ We now adopt the convention that

(3) $(R_T, \mathcal{B}_T)$ denotes the measurable function space with $R_T \equiv \prod_{t \in T} R_t$,

where each R_t is a copy of the real line. Thus R_T consists of all possible real-valued functions on T, and $\mathcal{B}_T$ is the smallest σ-field with respect to which all π_t are measurable. We call a process $X : (\Omega, \mathcal{A}, P) \to (R_T, \mathcal{B}_T)$ a *general stochastic process.* We note that a general stochastic process is also a process. But we do not yet know what $\mathcal{B}_T$ looks like.

A set $B_T \in \mathcal{B}_T$ is said to have *countable base* $t_1, t_2, \ldots$ if

(4) $B_T = \pi^{-1}_{t_1,t_2,\ldots}(B_\infty)$ for some $B_\infty \in \mathcal{B}_\infty$;

here $\mathcal{B}_\infty$ is the countably infinite-dimensional σ-field of section 5.3. Let $\mathcal{B}_\mathcal{C}$ denote the *class of countable base sets* defined by

(5) $\mathcal{B}_\mathcal{C} \equiv \{B_T \in \mathcal{B}_T : B_T \text{ has a countable base}\}$.

[Recall $\mathcal{F}(X_1, X_2, \ldots)$ and $\mathcal{F}(X_s : s \leq t)$ measurability from section 2.5.] □

Proposition 4.1 (Measurability in $(R_T, \mathcal{B}_T)$) Now, $\mathcal{B}_\mathcal{C}$ is a σ-field. In fact, $\mathcal{B}_\mathcal{C}$ is the smallest σ-field relative to which all π_t are measurable; that is,

(6) $\mathcal{B}_T = \mathcal{B}_\mathcal{C}$.

Also (generalizing proposition 5.2.1),

(7) X is $\mathcal{B}_T$-$\mathcal{A}$-measurable if and only if X_t is $\mathcal{B}$-$\mathcal{A}$-measurable for each $t \in T$.

Proof. Clearly, $\mathcal{B}_T$ is the smallest σ-field containing $\mathcal{B}_\mathcal{C}$; so (6) will follow from showing that $\mathcal{B}_\mathcal{C}$ is a σ-field. Now, $\mathcal{C}$ is closed under complements, since $\pi^{-1}_{t_1,t_2,\ldots}(B_\infty)^c = \pi^{-1}_{t_1,t_2,\ldots}(B^c_\infty)$. Suppose that $B_1, B_2, \ldots$ in $\mathcal{B}_\mathcal{C}$ have countable bases $T_1, T_2, \ldots$, and let $T_0 = \cup_{m=1}^\infty T_m$. Then using the countable set of distinct coordinates in T_0, reexpress each B_m as $B_m = \pi^{-1}_{T_0}(B^\infty_m)$ for some $B^\infty_m \in \mathcal{B}_\infty$. Then $\cup_{m=1}^\infty B_m = \pi^{-1}_{T_0}(\cup_{m=1}^\infty B^\infty_m)$ is in $\mathcal{B}_\mathcal{C}$. Thus $\mathcal{B}_\mathcal{C}$ is closed under countable unions. Thus $\mathcal{B}_\mathcal{C}$ is a σ-field.

Now to establish (7): Suppose X is $\mathcal{B}_T$-$\mathcal{A}$-measurable. Then

(a) $X_t^{-1}(B) = X^{-1}(\pi_t^{-1}(B)) \in \mathcal{A}$ for $B \in \mathcal{B}$,

so that each X_t is $\mathcal{B}$-$\mathcal{A}$-measurable. Suppose that each X_t is $\mathcal{B}$-$\mathcal{A}$-measurable. Then exercise 5.3.1 shows that $(X_{t_1}, X_{t_2}, \ldots)$ is $\mathcal{B}_\infty$-$\mathcal{A}$-measurable for all sequences $t_1, t_2, \ldots$ of elements of T. That is, $X^{-1}(\mathcal{B}_\mathcal{C}) \subset \mathcal{A}$. Since $\mathcal{B}_T = \mathcal{B}_\mathcal{C}$, we thus have $X^{-1}(\mathcal{B}_T) \subset \mathcal{A}$, and hence X is $\mathcal{B}_T$-$\mathcal{A}$-measurable. □

Remark 4.1 (Consistency of induced distributions in $(R_T, \mathcal{B}_T)$) Any general stochastic process $X : (\Omega, \mathcal{A}, P) \to (R_T, \mathcal{B}_T)$ has a *family of induced distributions*

$$(8) \qquad P^*_{t_1,\ldots,t_k}(B_k) = P(X^{-1} \circ \pi^{-1}_{t_1,\ldots,t_k}(B_k)) \qquad \text{for all } B_k \in \mathcal{B}_k$$

for all $k \geq 1$ and all $t_1, \ldots, t_k \in T$. These distributions are necessarily *consistent* in the sense that

$$(9) \qquad P^*_{t_1,\ldots,t_k}(B_1 \times \cdots \times B_{i-1} \times R \times B_{i+1} \times \cdots \times B_k)$$
$$= P^*_{t_1,\ldots,t_{i-1},t_{i+1},\ldots,t_k}(B_1 \times \cdots \times B_{i-1} \times B_{i+1} \times \cdots \times B_k)$$

for all $k \geq 1$, all $B_1, \ldots, B_k \in \mathcal{B}$, all $1 \leq i \leq k$, and all $t_1, \ldots, t_k \in T$. [The next result gives a converse. It is our fundamental result on the existence of stochastic processes with specified distributions.] □

Theorem 4.1 (Kolmogorov's consistency theorem) Given a consistent set of distributions as in (9), there exists a distribution P on $(R_T, \mathcal{B}_T)$ such that the identity map $X(\omega) = \omega$, for all $\omega \in R_T$, is a general stochastic process $X : (R_T, \mathcal{B}_T, P) \to (R_T, \mathcal{B}_T)$ whose family of induced distributions is the $P^*_{t_1,\ldots,t_k}$ of (9).

Exercise 4.2 Prove theorem 4.1. [Define $P^*(B) = P(\pi^{-1}_{T_i}(B))$ for $B \in \mathcal{B}_{\mathcal{C}}$ and each countable subset T_i of T. Use notational ideas from the proof of proposition 4.1 to show easily that $P^*(\cdot)$ is well-defined and countably additive.]

Example 4.1 (Comment on $(R_{[0,1]}, \mathcal{B}_{[0,1]})$) The typical function x in R_T has no smoothness properties. Let $T = [0,1]$ and let $\mathcal{C}$ denote the subset of $R_{[0,1]}$ that consists of all functions that are continuous on $[0,1]$. We now show that

$$(10) \qquad \mathcal{C} \notin \mathcal{B}_{[0,1]} \,.$$

Let $(\Omega, \mathcal{A}, P)$ denote Lebesgue measure on the Borel subsets of $[0,1]$. Let $\xi(\omega) = \omega$. Now let $X : (\Omega, \mathcal{A}, P) \to (R_{[0,1]}, \mathcal{B}_{[0,1]})$ via $X_t(\omega) = 0$ for all $\omega \in \Omega$ and for all $t \in T$. Let $Y : (\Omega, \mathcal{A}, P) \to (R_{[0,1]}, \mathcal{B}_{[0,1]})$ via $Y_t(\omega) = 1_{\{t\}}(\xi(\omega))$. Now, all finite-dimensional distributions of X and Y are identical. Note, however, that $[\omega : X(\omega) \in \mathcal{C}] = \Omega$, while $[\omega : Y(\omega) \in \mathcal{C}] = \emptyset$. Thus $\mathcal{C}$ cannot be in $\mathcal{B}_{[0,1]}$. □

Smoother Realizations of General Stochastic Processes

Suppose now that X is a process of the type $X : (\Omega, \mathcal{A}, P) \to (R_T, \mathcal{B}_T, P^*)$. As the previous example shows, X is not the unique process from $(\Omega, \mathcal{A}, P)$ that induces the distribution P^* on $(R_T, \mathcal{B}_T)$. We now let M_T denote a proper subset of R_T and agree that $\mathcal{M}_T$ denotes the σ-field generated by the finite-dimensional subsets of M_T. Suppose now that $X(\omega) \in M_T$ for all $\omega \in \Omega$. Can X be viewed as a process $X : (\Omega, \mathcal{A}, P) \to (M_T, \mathcal{M}_T, \tilde{P})$ such that $(M_T, \mathcal{M}_T, \tilde{P})$ has the same finite-dimensional distributions as does $(R_T, \mathcal{B}_T, P^*)$? We now show that the answer is necessarily yes. Interesting cases arise when the functions of the M_T above have smoothness properties such as continuity. The next result is very important and useful.

Theorem 4.2 (Smoother realizations of processes) Consider an arbitrary measurable mapping $X : (\Omega, \mathcal{A}, P) \to (R_T, \mathcal{B}_T, P^*)$.
(i) Let $M_T \subset R_T$. Then we can view X as a process $X : (\Omega, \mathcal{A}) \to (M_T, \mathcal{M}_T)$ if and only if every sample path $X_\cdot(\omega) = X(\cdot, \omega)$ is in M_T and every $X_t(\cdot) \equiv X(t, \cdot)$ is a random variable.
(ii) Let $X(\Omega) \subset M_T \subset R_T$. Then $X : (\Omega, \mathcal{A}, P) \to (M_T, \mathcal{M}_T, \tilde{P})$, where the finite-dimensional distributions of $(M_T, \mathcal{M}_T, \tilde{P})$ are the same as those of $(R_T, \mathcal{B}_T, P^*)$.
(iii) Comment: All this is true even when M_T is not in the class $\mathcal{B}_T$.

Proof. (i) ($\Leftarrow$) Note first that $M_T \cap \mathcal{B}_T = \mathcal{M}_T$ (recall definition 4.2). Moreover, when $X(\Omega) \subset M_T$, it necessarily follows that

$$X^{-1}(\mathcal{M}_T) = X^{-1}(M_T \cap \mathcal{B}_T) = X^{-1}(M_T) \cap X^{-1}(\mathcal{B}_T) = \Omega \cap X^{-1}(\mathcal{B}_T)$$

$$\text{(a)} \qquad = X^{-1}(\mathcal{B}_T).$$

Since each X_t is a rv, we have $X^{-1}(\mathcal{B}_T) \subset \mathcal{A}$ by (7). Thus $X^{-1}(\mathcal{M}_T) \subset \mathcal{A}$, and we see that X is indeed an $\mathcal{M}_T$–$\mathcal{A}$–measurable mapping from Ω to M_T. Note further that the natural pairs of generator sets ($\pi^{-1}_{t_1,\ldots,t_k}((-\infty, r_1] \times \cdots \times (-\infty, r_k])$ in $\mathcal{B}_T$, or $\pi^{-1}_{t_1,\ldots,t_k}((-\infty, r_1] \times \cdots \times (-\infty, r_k]) \cap M_T$ in $\mathcal{M}_T$) have the same inverse images under X; thus the finite dimensional distributions induced from $(\Omega, \mathcal{A})$ to $(R_T, \mathcal{B}_T)$ and to $(M_T, \mathcal{M}_T)$ are identical.
($\Rightarrow$) Clearly, $X : \Omega \to M_T$ implies $X : \Omega \to R_T$. Also, for any $t \in T$ and any $B \in \mathcal{B}$,

$$X^{-1}(\pi_t^{-1}(B)) = X^{-1}(M_T \cap \pi_t^{-1}(B)) \qquad \text{since } X : \Omega \to M_T$$

$$\in X^{-1}(\mathcal{M}_T) \qquad \text{since } \mathcal{M}_T = M_T \cap \mathcal{B}_T$$

$$\text{(b)} \qquad \in \mathcal{A} \qquad \text{since } X \text{ is } \mathcal{M}_T\text{–}\mathcal{A}\text{–measurable.}$$

Thus each X_t is a rv, and so X is $\mathcal{B}_T$–$\mathcal{A}$–measurable by (7).
(ii) This is now clear, and it summarizes the most useful part of this theorem. □

Exercise 4.3 Let $\mathcal{M}$ denote any non-void class of subsets of Ω, and let M denote any non-void subset. Show that

$$\text{(11)} \qquad \sigma[\mathcal{M}] \cap M = \sigma[\mathcal{M} \cap M].$$

Remark 4.2 It is interesting to consider the case where M_T is a countable or finite set. The resulting $(M_T, \mathcal{M}_T, \tilde{P})$ is the natural probability space. □

Chapter 6

Distribution and Quantile Functions

1 Character of Distribution Functions

Let $X : (\Omega, \mathcal{A}, P) \to (R, \mathcal{B}, P_X)$ be a rv with *distribution function (df)* F_X, where

(1) $\qquad F_X(x) \equiv P(X \le x) = P_X((-\infty, x]) \qquad$ for $-\infty < x < \infty$.

Then $F \equiv F_X$ was seen earlier to satisfy

(2) $\qquad F$ is $\nearrow$ and right continuous, with $F(-\infty) = 0$ and $F(+\infty) = 1$.

Because of the proposition below, any function F satisfying (2) will be called a df. [If F is $\nearrow$, right continuous, $0 \le F(-\infty)$, and $F(+\infty) \le 1$, we earlier agreed to call F a *sub-df.* As usual, $F(a, b] \equiv F(b) - F(a)$ denotes the increments of F, and $\Delta F(x) \equiv F(x) - F_-(x) = F(x) - F(x-)$ is the mass of F at x.]

(a) Call F *discrete* if F is of the form $F(\cdot) = \sum_j b_j\, 1_{[a_j, \infty)}(\cdot)$ with $\sum_j b_j = 1$, where the a_j form a non-void finite or countable set. Such measures μ_F have Radon–Nikodym derivative $\sum_j b_j\, 1_{\{a_j\}}(\cdot)$with respect to counting measure on the a_j's.

(b) A df F is called *absolutely continuous* if $F(\cdot) = \int_{\infty}^{\cdot} f(y)\, d\lambda(y)$ for some $f \ge 0$ that integrates to 1 over R. The corresponding measure has Radon-Nikodym derivative f with respect to Lebesgue measure λ; this f is also called a *probability density.* Moreover, F is an absolutely continuous function and the ordinary derivative F' of the df F exists a.e. λ and satisfies $F' = f$ as λ.

(c) A df F is called *singular* if $\mu_F(N^c) = 0$ for a λ-null set N.

Proposition 1.1 (There exists an X with df F) If F satisfies (2), then there exists a probability space $(\Omega, \mathcal{A}, P)$ and a rv $X : (\Omega, \mathcal{A}, P) \to (R, \mathcal{B})$ for which the df of X is F. We write $X \cong F$.

Proof. Example 2.2.1 shows that $X(r) = r$ on $(R, \mathcal{B}, \mu_F)$ is one example. $\square$

Theorem 1.1 (Decomposition of a df) Any df F can be decomposed as

(3) $$F = F_d + F_c = F_d + F_s + F_{ac} = (F_d + F_s) + F_{ac}\,,$$

where F_d, F_c, F_s, and F_{ac} are the unique sub-dfs of the following types (unique among those sub-dfs equal to 0 at $-\infty$):

(4) F_d is a step function of the form $\sum_j b_j 1_{[a_j,\infty)}$ (with all $b_j > 0$).

(5) F_c is continuous.

(6) F_s and $F_s + F_d$ are both singular with respect to Lebesgue measure λ.

(7) $F_{ac}(\cdot) = \int_{-\infty}^{\cdot} f_{ac}(y)\, d\lambda(y)$
for some $f_{ac} \ge 0$ that is finite, measurable, and unique a.e. λ,
and this $F_{ac}(\cdot)$ is absolutely continuous on the whole real line.

Proof. Let $\{a_j\}$ denote the set of all discontinuities of F, which can only be jumps; and let $b_j \equiv F(a_j) - F_-(a_j)$. There can be only a countable number of jumps, since the number of jumps of size exceeding size $1/n$ is certainly bounded by n. Now define $F_d \equiv \sum_j b_j 1_{[a_j,\infty)}$, which is obviously $\nearrow$ and right continuous, since $F_d(x,y] \le F(x,y] \searrow 0$ as $y \searrow x$ (the inequality holds, since the sum of jump sizes over every finite number of jumps between a and b is clearly bounded by $F(x,y]$, and then just pass to the limit). Define $F_c = F - F_d$. Now, F_c is $\nearrow$, since for $x \le y$ we have $F_c(x,y] = F(x,y] - F_d(x,y] \ge 0$. Now, F_c is the difference of right-continuous functions, and hence is right continuous; it is left continuous, since for $x \nearrow y$ we have

(a) $$F_c(x,y] = F(x,y] - \textstyle\sum_{x<a_j\le y} b_j = F_-(y) - F(x) - \sum_{x<a_j<y} b_j \to 0 - 0 = 0.$$

We turn to the uniqueness of F_d. Assume that $F_c + F_d = F = G_c + G_d$ for some other $G_d \equiv \sum_j \bar{b}_j 1_{[\bar{a}_j,\infty)}$ with distinct $\bar{a}_j$'s and $\sum_j \bar{b}_j \le 1$. Then $F_d - G_d = G_c - F_c$ is continuous. If $G_d \ne F_d$, then either some jump point or some jump size disagrees. No matter which disagrees, at some a we must have

(b) $$\Delta F_d(a) - \Delta G_d(a) \ne 0,$$

contradicting the continuity of $G_c - F_c = F_d - G_d$. Thus $G_d = F_d$, and hence $F_c = G_c$. This completes the first decomposition.

We now turn to the further decomposition of F_c. Associate a measure μ_c with F_c via $\mu_c((-\infty,x]) = F_c(x)$. Then the Lebesgue decomposition theorem shows that $\mu_c = \mu_s + \mu_{ac}$, where $\mu_s(B) = 0$ and $\mu_{ac}(B^c) = 0$ for some $B \in \mathcal{B}$; we say that μ_s and μ_{ac} are orthogonal. Moreover, this same Lebesgue theorem implies the claimed uniqueness and shows that f_{ac} exists with the uniqueness claimed. Now, $F_{ac}(x) \equiv \mu_{ac}((-\infty,x]) = \int_{-\infty}^{x} f_{ac}(y)\,dy$ is continuous by $F_{ac}(x,y] \le \mu_{ac}(x,y] \to 0$ as $y \to x$ or as $x \to y$. Thus $F_s \equiv F_c - F_{ac}$ is continuous, and $F_s(x) = \mu_s((-\infty,x])$. In fact, F_{ac} is absolutely continuous on R by the absolute continuity of the integral. (Now $F_c = F_s + F_{ac}$ decomposes F_c with respect to λ, while $F = (F_d + F_s) + F_{ac}$ decomposes F with respect to λ.) $\square$

Example 1.1 (Lebesgue singular df) Define the *Cantor set* C by

(8) $\quad C \equiv \{x \in [0,1] : x = \sum_{n=1}^{\infty} 2a_n/3^n, \text{ with all } a_n \text{ equal to 0 or 1}\}.$

[Thus the Cantor set is obtained by removing from $[0,1]$ the open interval $(\frac{1}{3}, \frac{2}{3})$ at stage one, then the open intervals $(\frac{1}{9}, \frac{2}{9})$ and $(\frac{7}{9}, \frac{8}{9})$ at stage two, ...] Finally, we define F on C by

(9) $\quad F(\sum_{n=1}^{\infty} 2a_n/3^n) = \sum_{n=1}^{\infty} a_n/2^n.$

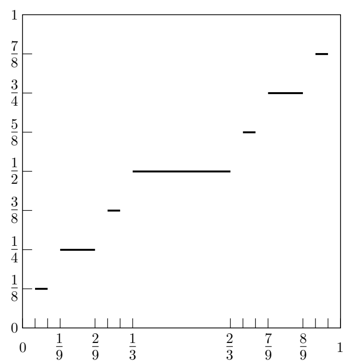


Figure 1.1 Lebesgue singular function.

Now note that $\{F(x) : x \in C\} = [0,1]$, since the right-hand side of (9) represents all of $[0,1]$ via dyadic expansion. We now define F "linearly" on C^c (the first three "components" are shown in figure 1.1 above). Since the resulting F is $\nearrow$ and achieves every value in $[0,1]$, it must be that F is continuous. Now, F assigns no mass to the "flat spots" whose lengths sums to 1 since $\frac{1}{3} + \frac{2}{9} + \frac{4}{27} + \cdots = \frac{1/3}{1-2/3} = 1$. Thus F is singular with respect to Lebesgue measure λ, using $\lambda(C^c) = 1$ and $\mu_F(C^c) = 0$. Call this F the *Lebesgue singular df*. [The theorem in the next section shows that removing the flat spots does, even for a general df F, leave only the essentials.] We have, in fact, shown that

(10) $\quad F : C \to [0,1]$ is 1-1, is $\uparrow$, and is continuous; so $F^{-1} : [0,1] \to C$ is 1-1.□

Exercise 1.1 Let $X \cong N(0,1)$ (as in (A.1.22) below), and let $Y \equiv 2X$.
(a) Is the df $F(\cdot,\cdot)$ of (X,Y) continuous?
(b) Does the measure μ_F on R_2 have a density with respect to two-dimensional Lebesgue measure? [Hint. Appeal to corollary 2 to Fubini's theorem.]

Exercise 1.2 Show that the Cantor set C is *perfect* (thus, each $x \in \mathcal{C}$ is an accumulation point of $\mathcal{C}$) and *totally disconnected* (between any $c_1 < c_2$ in $\mathcal{C}$ there is an interval that lies entirely in $\mathcal{C}^c$). (Note that the cardinality of $\mathcal{C}$ equals that of $[0,1]$.) At which points is $1_C(\cdot)$ continuous?

Definition 1.1 Two rvs X and Y are said to be of the same *type* if $Y \cong aX + b$ for some $a > 0$. Their dfs are also said to be of the same type.

2 Properties of Distribution Functions

Definition 2.1 The *support* of a given df $F \equiv F_X$ is defined to be the minimal closed set C having $P(X \in C) = 1$. A point x is a *point of increase* of F if every open interval U containing x has $P(X \in U) > 0$. A *realizable t-quantile of F*, for $0 < t < 1$, is any value z for which $F(z) = t$. (Such a z need not exist.) Define U_t to be the maximal open interval of x's for which $F(x) = t$ (this flat spot will be an interval as F is $\nearrow$).

Theorem 2.1 (Jumps and flat spots) Let C denote the support of F. Then:
(a) $C \equiv (\bigcup_{0 \le t \le 1} U_t)^c$ is a closed set having $P(C) = 1$.
(b) C is equal to the set of all points of increase.
(c) C is the support of F.
(d) F has at most a countable number of discontinuities, and these discontinuities are all discontinuities of the jump type.
(e) F has an at most countable number of flat spots (the nonvoid U_t's). These are exactly those t's that have more than one realizable t-quantile.
[We will denote jump points and jump sizes of F by a_i's and b_i's. The t values and the $\lambda(U_t)$ values of the multiply realizable t-quantiles will be seen in the proof of proposition 6.3.1 below to correspond to the jump points c_j and the jump values d_j of the function $K(\cdot) \equiv F^{-1}(\cdot)$, and there are at most countably many of them.]

Proof. (a) For each t there is a maximal open interval U_t (possibly void) on which F equals t, and it is bounded for each $0 < t < 1$. Now, $P(X \in U_t) = 0$ using proposition 1.1.2. Note that $C \equiv (\cup_t U_t)^c$ is closed (since the union of an arbitrary collection of open sets is open). Hence $C^c = \cup_{0 \le t \le 1} U_t = \cup(a_n, b_n)$, where $(a_1, b_1), \ldots$ are (at most countably many) disjoint open intervals, and all those with $0 < t < 1$ must be finite. Now, by proposition 1.1.2, for the finite intervals we have $P(X \in (a_n, b_n)) = \lim_{\epsilon \to 0} P(X \in [a_n + \epsilon, b_n - \epsilon]) = \lim_{\epsilon \to 0} 0 = 0$, where $P(X \in [a_n + \epsilon, b_n - \epsilon]) = 0$ holds since this finite closed interval must have a finite subcover by U_t sets. If $(a_n, b_n) = (-\infty, b_n)$, then $P(X \in (-\infty, b_n)) = 0$, since $P(X \in [-1/\epsilon, b_n - \epsilon]) = 0$ as before. An analogous argument works if $(a_n, b_n) = (a_n, \infty)$. Thus $P(X \in C^c) = 0$ and $P(X \in C) = 1$. Note that the U_t's are just the (a_n, b_n)'s in disguise; each $U_t \subset$ some (a_n, b_n), and hence $U_t =$ that (a_n, b_n). Thus U_t is nonvoid for at most countably many t's.

(b) Let $x \in C$. We will now show that it is a point of increase. Let U denote a neighborhood of x, and let $t \equiv F(x)$. Assume $P(U) = 0$. Then $x \in U \subset U_t \subset C^c$, which is a contradiction of $x \in C$. Thus all points $x \in C$ are points of increase. Now suppose conversely that x is a point of increase. Assume $x \notin C$. Then $x \in$ some (a_n, b_n) having $P(X \in (a_n, b_n)) = 0$, which is a contradiction. Thus $x \in C$. Thus the closed set C is exactly the set of points of increase.

(c) Assume that C is not the minimal closed set having probability 1. Then $P(\tilde{C}) = 1$ for some closed $\tilde{C} \subsetneq C$. Let $x \in C \setminus \tilde{C}$ and let $t = F(x)$. Since $\tilde{C}^c$ is open, there is an open interval V_x with $x \in V_x \subset \tilde{C}^c$ and $P(X \in V_x) = 0$. Thus $x \in V_x \subset (\text{ some } U_t) \subset C^c$. So $x \notin C$, which is a contradiction. Thus C is minimal.

So, (d) and (e) follow. See also the summary following proposition 6.3.1. □

3 The Quantile Transformation

Definition 3.1 (Quantile function) For any df $F(\cdot)$ we define the *quantile function (qf)* (which is the inverse of the df) by

(1) $\quad K(t) \equiv F^{-1}(t) \equiv \inf\{x : F(x) \geq t\} \qquad$ for $0 < t < 1$.

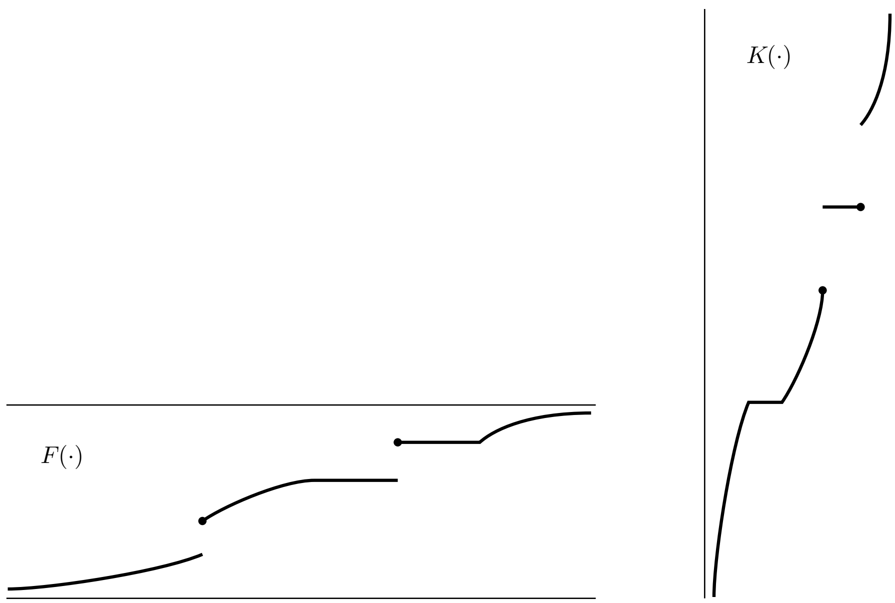


Figure 3.1 The df $F(\cdot)$ and the qf $K(\cdot) = F^{-1}(\cdot)$.

Theorem 3.1 (The inverse transformation) Let

(2) $\quad X \equiv K(\xi) \equiv F^{-1}(\xi), \qquad$ where $\xi \cong$ Uniform$(0,1)$.

The following are all true.

(3) $\quad [X \leq x] = [\xi \leq F(x)] \qquad$ for every real x.

(4) $\quad 1_{[X \leq \cdot]} = 1_{[\xi \leq F(\cdot)]} \qquad$ on R, for every ω.

(5) $\quad X \equiv K(\xi) \equiv F^{-1}(\xi) \qquad$ has df F .

(6) $\quad 1_{[X < \cdot]} = 1_{[\xi < F_-(\cdot)]} \qquad$ on R, for a.e. ω;

failure occurs if and only if $\xi(\omega)$ equals the height of a flat spot of F.

Proof. Fix an arbitrary x. Now, $\xi \le F(x)$ implies $X = F^{-1}(\xi) \le x$ by (1). If $X = F^{-1}(\xi) \le x$, then $F(x+\epsilon) \ge \xi$ for all $\epsilon > 0$; so right continuity of F implies $F(x) \ge \xi$. Thus (3) holds; (4) and (5) are then immediate.

If $\xi(\omega) = t$ where t is not in the range of F, then (6) holds. If $\xi(\omega) = t$ where $F(x) = t$ for exactly one x, then (6) holds. If $\xi(\omega) = t$ where $F(x) = t$ for at least two distinct x's, then (6) fails; theorem 6.2.1 shows that this can happen for at most a countable number of t's. (Or: Graph a df F that exhibits the three types of points t, and the rest is trivial with respect to (6), since the value of F at any other point is immaterial. Specifically, (6) holds for ω unless F has a flat spot at height $t \equiv \xi(\omega)$. Note figure 3.1.) □

Definition 3.2 (Convergence in quantile) Let K_n denote the qf associated with df F_n, for each $n \ge 0$. We write $K_n \to_d K_0$ to mean that $K_n(t) \to K_0(t)$ at each continuity point t of K_0 in $(0,1)$. We then say that K_n *converges in quantile* to K_0, or K_n *converges in distribution* to K.

Proposition 3.1 (Convergence in distribution equals convergence in quantile)

(7) $\quad F_n \to_d F \quad$ if and only if $\quad K_n \to_d K.$

Proof. Suppose $F_n \to_d F$. Let $t \in (0,1)$ be such that there is at most one value x having $F(x) = t$ (that is, this is not a multiply realizable t-quantile). Let $z \equiv F^{-1}(t)$.

First: We have $F(x) < t$ for $x < z$, where x will always denote a continuity point of F. Thus $F_n(x) < t$ for $n \ge$ (some N_x). Thus $F_n^{-1}(t) \ge x$ for $n \ge N_x$. Thus $\liminf F_n^{-1}(t) \ge x$, when $x < z$ is a continuity point of F. Thus $\liminf F_n^{-1}(t) \ge z$, since there are continuity points x that $\nearrow z$. Second: We also have $F(x) > t$ for $z < x$, with x a continuity point. Thus $F_n(x) > t$, and hence $F_n^{-1}(t) \le x$ for $n \ge$ (some N_x). Thus $\limsup F_n^{-1}(t) \le x$. Thus $\limsup F_n^{-1}(t) \le z$, since there are continuity points x that $\searrow z$. Thus $K_n(t) = F_n^{-1}(t) \to z = K(t)$. The proof of the converse is virtually identical. □

Exercise 3.0 Give the proof of the converse for the previous proposition.

Summary $F_n^{-1}(t) \to F^{-1}(t)$ for all but at most a countably infinite number of t's (namely, for all but those t's that have multiply realizable t-quantiles; these correspond to the heights of flat spots of F, and these flat spot heights t are exactly the discontinuity points of K).

Exercise 3.1 (Left continuity of K) Show that $K(t) = F^{-1}(t)$ is left continuous on $(0,1)$. [Note that K is discontinuous at $t \in (0,1)$ if and only if the corresponding U_t is nonvoid (see theorem 6.2.1). Likewise, the jump points c_j and the jump sizes d_j of $K(\cdot)$ are equal to the t values and the $\lambda(U_t)$ values of the multiply realizable t-quantiles.] [We earlier agreed to use a_i and b_i for the jump points and jump sizes of the associated df F.]

Exercise 3.2 (Properties of dfs) (i) For any df F we have

$$F \circ F^{-1}(t) \geq t \qquad \text{for all } 0 \leq t \leq 1,$$

and equality fails if and only if $t \in (0,1)$ is not in the range of F on $[-\infty, \infty]$.
(ii) (The probability integral transformation) If X has a continuous df F, then $F(X) \cong \text{Uniform}(0,1)$. In fact, for any df F,

$$P(F(X) \leq t) \leq t \qquad \text{for all } 0 \leq t \leq 1\,,$$

with equality failing if and only if t is not in the closure of the range of F.
(iii) For any df F we have

$$F^{-1} \circ F(x) \leq x \qquad \text{for all } -\infty < x < \infty\,,$$

and equality fails if and only if $F(y) = F(x)$ for some $y < x$. Thus

$$P(F^{-1} \circ F(X) \neq X) = 0 \qquad \text{whenever } X \cong F.$$

(iv) If F is a continuous df and $F(X) \cong \text{Uniform}(0,1)$, then $X \cong F$.
(v) Graph $F \circ F^{-1}$ and $F^{-1} \circ F$ for the df F in figure 8.4.1.

Proposition 3.2 (The randomized probability integral transformation) Let X denote an arbitrary rv. Let F denote its df, and let (a_j, b_j)'s denote an enumeration of whatever pairs (jump point, jump size) the df F possesses. Let $\eta_1, \eta_2, \ldots$ denote iid Uniform$(0,1)$ rvs (that are also independent of X). Then both

$$\text{(8)} \qquad \dot{\xi} \equiv F(X) - \textstyle\sum_j b_j \eta_j \, 1_{[X=a_j]} \cong \text{Uniform}(0,1) \qquad \text{and}$$

$$\text{(9)} \qquad X = F^{-1}(\dot{\xi}) = K(\dot{\xi}).$$

[We have reproduced the original X from a Uniform$(0,1)$ rv that was defined using both X and some independent extraneous variation. Note figure 3.1.]

Proof. We have merely smoothed out the mass b_j that $F(X)$ placed at $F(a_j)$ by subtracting the random fractional amount $\eta_j b_j$ of the mass b_j. □

Exercise 3.3 (Change of variable) Suppose $Y \cong G$ and $X = H^{-1}(Y) \cong F$, where H is $\nearrow$ and right continuous on the real line with left-continuous inverse H^{-1}.
(a) Then set g, X, μ, μ_X, A' in the theorem of the unconscious statistician equal to $g, H^{-1}, G, F, (-\infty, x]$ to conclude that

$$\text{(10)} \qquad \int_{(-\infty, H(x)]} g(H^{-1})\, dG = \int_{(-\infty, x]} g\, dF,$$

since $(H^{-1})^{-1}((-\infty, x]) = \{t : H^{-1}(t) \leq x\} = \{r : r \leq H(x)\}$ as in (3). Check it.
(b) Let $G = I$, $H = F$, and $Y = \xi \cong \text{Uniform}(0,1)$ above, for any df F. Let g denote any measurable function. Then (via part (a), or via (2) and (3))

$$\text{(11)} \qquad \int_0^{F(x)} g(F^{-1}(t))\, dt = \int_{(-\infty, x]} g\, dF, \quad \text{and} \quad \int_{F_-(x)}^1 g(F^{-1}(t))\, dt = \int_{[x, \infty)} g\, dF.$$

(c) Exercise 3.5.3 established $\mathrm{E}|X| = \int_0^\infty P(|X| > y)\, dy$ (note (6.4.11) below) and

$$\text{(12)} \qquad \int_{[x, \infty)} y\, dF_{|X|}(y) = xP(|X| \geq x) + \int_x^\infty P(|X| \geq y)\, dy\,; \qquad \text{so}$$

$$\text{(13)} \qquad \int_{[0, x]} y\, dF_{|X|}(y) = \int_0^x P(|X| > y)\, dy - xP(|X| > x)\,.$$

Interpret these formulas in terms of various shaded areas shown in figure 8.4.1. Also, relate these areas to (11) with $g(y) = y$ and $X \geq 0$.

Exercise 3.4 Let h be measurable on $[0,1]$. Then

$$(14)\qquad \int_{(-\infty,x]} h(F_-)\,dF \le \int_0^{F(x)} h(t)\,dt \le \int_{(-\infty,x]} h(F)\,dF \qquad \text{if } h \nearrow .$$

Reverse the inequalities if $h \searrow$.

Proof. We now prove proposition 1.2.3. Let D be a subset of $[0,1]$ that is not Lebesgue measurable; its existence is guaranteed by proposition 1.2.2. Let $B \equiv F^{-1}(D)$ for the Lebesgue singular df F. Then (6.1.10) shows that B is a subset of the Cantor set C. Since $\lambda(C) = 0$ and $B \subset C$, then B is a Lebesgue set with $\lambda(B) = 0$; that is, $B \in \hat{\mathcal{B}}_\lambda$. We now assume that B is Borel set (and seek a contradiction). Now F^{-1} is measurable by (6.1.10), and so $(F^{-1})^{-1}(B) \in \mathcal{B}$. But

$$\text{(a)}\qquad (F^{-1})^{-1}(B) = \{r : F^{-1}(r) \in B\} = \{r : F^{-1}(r) \in F^{-1}(D)\} = D \notin \mathcal{B},$$

since F^{-1} is one-to-one on $[0,1]$. This is a contradiction. Thus $B \in \hat{\mathcal{B}}_\lambda \setminus \mathcal{B}$. □

The Elementary Skorokhod Construction Theorem

Let $X_0, X_1, X_2, \ldots$ be iid F. Then $X_n \to_d X_0$, but the X_n do not converge to X_0 in the sense of $\to_{a.s.}$, $\to_p$, or $\to_r$. However, when general $X_n \to_d X_0$, it is possible to replace the X_n's by rvs Y_n having the same (marginal) dfs, for which the stronger result $Y_n \to_{a.s.} Y_0$ holds.

Theorem 3.2 (Skorokhod) Suppose that $X_n \to_d X_0$. Define $\xi(\omega) = \omega$ for each $\omega \in [0,1]$ so that $\xi \cong \text{Uniform}(0,1)$ on $(\Omega, \mathcal{A}, P) \equiv ([0,1], \mathcal{B} \cap [0,1], \lambda)$, for Lebesgue measure λ. Let F_n denote the df of X_n, and define $Y_n \equiv F_n^{-1}(\xi)$ for all $n \ge 0$. Let D_{K_0} denote the at most countable discontinuity set of K_0. Then both

$$(15)\qquad \begin{aligned} &Y_n \equiv K_n(\xi) \equiv F_n^{-1}(\xi) \cong X_n \cong F_n \quad \text{for all } n \ge 0 \qquad \text{and} \\ &Y_n(\omega) \to Y_0(\omega) \quad \text{for all } \omega \notin D_{K_0}. \end{aligned}$$

Proof. This follows trivially from proposition 3.1. □

Exercise 3.5 (Wasserstein distance) For $k = 1$ or 2, define

$$\mathcal{F}_k \equiv \{F : F \text{ is a df, and } \textstyle\int |x|^k\,dF(x) < \infty\}, \qquad \text{and}$$

$$d_k(F_1, F_2) \equiv \{\textstyle\int_0^1 |F_1^{-1}(t) - F_2^{-1}(t)|^k\,dt\}^{1/k} \qquad \text{for all } F_1, F_2 \in \mathcal{F}_k .$$

(a) Show that all such $(\mathcal{F}_k, d_k)$ spaces are complete metric spaces, and that

$$(16)\qquad \begin{aligned} &d_k(F_n, F_0) \to 0 \qquad (\text{with all } \{F_n\}_0^\infty \in \mathcal{F}_k) \qquad \text{if and only if} \\ &F_n \to_d F_0 \quad \text{and} \quad \textstyle\int |x|^k\,dF_n(x) \to \int |x|^k\,dF_0(x) . \end{aligned}$$

(The rvs $Y_n \equiv F_n^{-1}(\xi)$ of (15) satisfy $Y_n \to_{\mathcal{L}_2} Y_0$ if $\sigma_n^2 \to \sigma_0^2$.)
(b) Apply these conclusions to the empirical df and qf.

4 Integration by Parts Applied to Moments

Integration by Fubini's theorem or "integration by parts" formulas are useful in many contexts. Here we record a few of the most useful ones.

Integration by Parts

Proposition 4.1 (Integration by parts formulas) Suppose that both the left-continuous function U and the right-continuous function V are monotone functions. Then for any $a \le b$ we have both

$$(1) \qquad U_+(b)V(b) - U(a)V_-(a) = \int_{[a,b]} U\, dV + \int_{[a,b]} V\, dU \qquad \text{and}$$

$$(2) \qquad U(b)V(b) - U(a)V(a) = \int_{(a,b]} U\, dV + \int_{[a,b)} V\, dU,$$

where $U_+(x) \equiv \lim_{y \searrow x} U(y)$ and $V_-(x) \equiv \lim_{y \nearrow x} V(y)$. [Symbolically, written as $d(UV) = U_- dV + V_+ dU$, it implies also that $\int h\, d(UV) = \int h\, [U_-\, dV + V_+\, dU]$ for any measurable $h \ge 0$.]

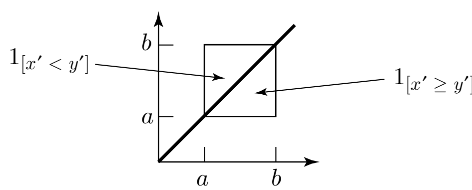


Figure 4.1 Integration by parts.

Proof. We can apply Fubini's theorem at steps (a) and (b) to obtain

$$[U_+(b) - U(a)]\,[V(b) - V_-(a)] = \int_{[a,b]} \{\int_{[a,b]} dU\}\, dV$$

$$(a) \qquad = \int_{[a,b]} \int_{[a,b]} [1_{[x'<y']}(x,y) + 1_{[x' \ge y']}(x,y)]\, dU(x)\, dV(y)$$

$$(b) \qquad = \int_{[a,b]} [U(y) - U(a)]\, dV(y) + \int_{[a,b]} [V(x) - V_-(a)]\, dU(x)$$

$$= \int_{[a,b]} U dV - U(a)\,[V(b) - V_-(a)] + \int_{[a,b]} V dU - V_-(a)\,[U_+(b) - U(a)].$$

Algebra now gives (1). Add $U(a)\,[V(b) - V_-(a)] + V_-(a)\,[U_+(b) - U(a)]$ to each side of (1) to obtain (2). □

Exercise 4.1 How should the left side of (1) be altered if we replace $[a,b]$ in both places on the right side of (1) by (a,b), or by $(a,b]$, or by $[a,b)$? [Just plug in a_+ or a_- as well as b_+ or b_- on both sides of the equation $d(UV) = U_- dV + V_+ dU$ so as to include or exclude that endpoint; this will give the proper formulation.]

Useful Formulas for Means, Variances, and Covariances

If $\xi \cong$ Uniform(0, 1) and F is an arbitrary df, then the rv $X \equiv F^{-1}(\xi)$ has df F. **Thinking of X as $F^{-1}(\xi)$ presents alternative ways to approach problems**. *We will do this often*! Note that this $X = F^{-1}(\xi)$ satisfies both

$$(3) \qquad X = \int_{(0,1)} F^{-1}(t)\, d1_{[\xi \le t]} \qquad \text{and} \qquad X = \int_{(-\infty,\infty)} x\, d1_{[X \le x]}\,,$$

where $1_{[\xi \le t]}$ is a random df that puts mass 1 at the point $\xi(\omega)$ and $1_{[X \le x]}$ is a random df that puts mass 1 at the point x. If X has a finite mean μ, then (depending on which representation of X we use)

$$(4) \qquad \mu = \int_{(0,1)} F^{-1}(t)\, dt \qquad \text{and} \qquad \mu = \int_{(-\infty,\infty)} x\, dF(x).$$

Moreover, when μ is finite we can combine the two previous formulas to write

$$(5) \qquad X - \mu = \int_{(0,1)} F^{-1}(t)\, d(1_{[\xi \le t]} - t) = -\int_{(0,1)} (1_{[\xi \le t]} - t)\, dF^{-1}(t)$$

or

$$(6) \qquad X - \mu = \int_{(-\infty,\infty)} x\, d(1_{[X \le x]} - F(x)) = -\int_{(-\infty,\infty)} (1_{[X \le x]} - F(x))\, dx.$$

The first formula in each of (5) and (6) is trivial; the second follows from integration by parts. For example, (5) is justified by $|t\, F^{-1}(t)| \le |\int_0^t F^{-1}(s)\, ds| \to 0$ as $t \to 0$ when $\mathrm{E}|X| = \int_0^1 |F^{-1}(t)|\, dt < \infty$, and the analogous result $(1-t)F^{-1}(t) \to 0$ as $t \to 1$. For (6), we note that $x[1 - F(x)] \le \int_{(x,\infty)} y\, dF(y) \to 0$ as $x \to \infty$ if $\mathrm{E}|X| < \infty$. So when $\mathrm{E}|X| < \infty$, Fubini's theorem seems to give (see exercise 4.2 for the rigorous proof of (8))

$$\text{(a)} \qquad \mathrm{Var}[X] = \mathrm{E}\left\{\int_{(0,1)} (1_{[\xi \le s]} - s)\, dF^{-1}(s) \int_{(0,1)} (1_{[\xi \le t]} - t)\, dF^{-1}(t)\right\}$$

$$\text{(b)} \qquad = \int_{(0,1)} \int_{(0,1)} \mathrm{E}\{(1_{[\xi \le s]} - s)(1_{[\xi \le t]} - t)\}\, dF^{-1}(s)\, dF^{-1}(t)$$

$$(7) \qquad = \int_{(0,1)} \int_{(0,1)} [s \wedge t - st]\, dF^{-1}(s)\, dF^{-1}(t) \qquad (\text{when } \mathrm{E}|X| < \infty)$$

$$(8) \qquad = \int_{(0,1)} \int_{(0,1)} [s \wedge t - st]\, dF^{-1}(s)\, dF^{-1}(t) \qquad (\text{even if } \mathrm{E}|X| = \infty)$$

via (5), and the parallel formula

$$(9) \qquad \mathrm{Var}[X] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [F(x \wedge y) - F(x)F(y)]\, dx\, dy$$

via (6). Of course, we already know, when $\mathrm{E}(X^2) < \infty$, that

$$(10) \qquad \mathrm{Var}[X] = \int_0^1 [F^{-1}(t) - \mu]^2\, dt \quad \text{and} \quad \mathrm{Var}[X] = \int_{-\infty}^{\infty} (x - \mu)^2\, dF(x).$$

Proposition 4.2 (Other formulas for means, variances, and covariances)
(i) If $X \geq 0$ has df F, then

$$(11)\qquad \int_0^\infty P(X > x)\,dx = \mathrm{E}X = \int_0^\infty (1 - F(x))\,dx \quad\text{and}\quad \mathrm{E}X = \int_0^1 F^{-1}(t)\,dt.$$

(ii) If $\mathrm{E}|X| < \infty$, then

$$(12)\qquad \mathrm{E}(X) = -\int_{-\infty}^0 F(x)\,dx + \int_0^\infty (1 - F(x))\,dx = \int_0^1 F^{-1}(t)\,dt.$$

(iii) Let $r > 0$. If $X \geq 0$, then

$$(13)\qquad \int_0^\infty P(X^r > x)\,dx = \mathrm{E}(X^r) = \int_0^\infty r\,x^{r-1}(1 - F(x))\,dx = \int_0^1 [F^{-1}(t)]^r\,dt$$

In fact, one of the two integrals is finite if and only if the other is finite.
(iv) Let (X, Y) have joint df F with marginal dfs F_X and F_Y. Let G and H be $\nearrow$ and left continuous. Then

$$(14)\qquad \mathrm{Cov}[G(X), H(Y)] = \int_{-\infty}^\infty \int_{-\infty}^\infty [F(x, y) - F_X(x)F_Y(y)]\,dG(x)\,dH(y)$$

whenever this covariance is finite. Note the special case $G = H = I$ for $\mathrm{Cov}[X, Y]$.
Hint. Without loss, $G_-(0) = G_+(0) = H_-(0) = H_+(0)$. Make use of the fact that $G(x) = \int_{[0,\infty)} 1_{[0,x)}(s)\,dG_-(s)$ in the first quadrant, etc.
(v) Let K be $\nearrow$ and left continuous and $\xi \cong \mathrm{Uniform}(0, 1)$ (perhaps $K = h(F^{-1})$ for an $\nearrow$ left-continuous function h, and for $X \equiv F^{-1}(\xi)$ for a df F). When finite,

$$(15)\qquad \mathrm{Var}[K(\xi)] = \int_0^1 \int_0^1 [s \wedge t - st]\,dK(s)\,dK(t) \qquad \text{and}$$

$$(16)\qquad \mathrm{Var}[K(\xi)] = \int_{-\infty}^\infty \int_{-\infty}^\infty [F(x \wedge y) - F(x)F(y)]\,dh(x)\,dh(y) = \mathrm{Var}[h(X)]$$

follow from (8) and (14).
(vi) If $X \geq 0$ is integer-valued, then

$$(17)\qquad \mathrm{E}X = \textstyle\sum_{k=1}^\infty P(X \geq k) \qquad\text{and}\qquad \mathrm{E}X^2 = \sum_{k=1}^\infty (2k - 1)\,P(X \geq k).$$

Exercise 4.2 ($\tilde{\text{W}}$insorized X) Let $\tilde{X}_{a,a'} \equiv \tilde{K}_{a,a'}(\xi)$, where $\xi \cong$ Uniform(0,1). Here, $\tilde{K}_{a,a'}$ equals $K_+(a)$, $K(t)$, $K(1 - a')$ according as $0 < t \leq a$, $a < t < 1 - a'$, $1 - a' \leq t < 1$. We say that X has been *$\tilde{W}$insorized outside* $(a, 1 - a')$.
(a) Use the Fubini/Tonelli combination (as above) to check that

$$\begin{aligned}\mathrm{E}(\tilde{K}^2_{a,a'}(\xi)) - (\mathrm{E}(\tilde{K}_{a,a'}(\xi))^2 &= \mathrm{Var}[\tilde{K}_{a,a'}(\xi)] \\ &= \textstyle\int_0^1 \int_0^1 1_{(a,1-a')}(s)\,1_{(a,1-a')}(t)\,(s \wedge t - st)\,dK(s)\,dK(t);\end{aligned}$$

essentially, obtain (8) for $\tilde{X}_{a,a'}$. Then let $(a \vee a') \to 0$, and apply the MCT, to obtain (7) for general X. (Use (6.6.2) to see that (8) holds even if $\mathrm{E}|X| = \infty$.)
(b) Establish (9) using similar methods.

Exercise 4.3 (a) Prove formulas (11)–(13). [Hint. Use integration by parts.]
(b) Prove the formula (14).

Exercise 4.4 Prove the formulas in (17).

Exercise 4.5 Give an extension of (13) to arbitrary rvs.

Exercise 4.6 (a) Use Fubini and use integration by parts to show twice that for arbitrary F and for every $x \geq 0$ we have

(18) $$\int_{[0,x]} y^2\, dF(y) = 2\int_0^x t\, P(X > t)\, dt - x^2 P(X > x).$$

(b) Verify (6.3.12) and (6.3.13) once again, with the current methods.

Exercise 4.7 (Integration by parts formulas) We showed in proposition 4.1 earlier that $d(UV) = U_- dV + V_+ dU$ (with left continuous U and right continuous V).
(i) Now show (noting that $dU_- = dU_+$) that

(19) $$dU^2 = d(U_- U_+) = U_- dU + U_+ dU = (2U + \Delta U) dU \qquad \text{for } \Delta U \equiv U - U_-.$$

(ii) Apply proposition 4.1 to $1 = U \cdot (1/U)$ to obtain

(20) $$d(1/U) = -\{1/(U_+ U_-)\}\, dU = -\{1/(U(U + \Delta U))\}\, dU.$$

(iii) Show by induction that for $k = 1, 2, \ldots$ we have

(21) $$dU^k = \left(\sum_{i=0}^{k-1} U_+^i U_-^{k-i-1}\right) dU.$$

Exercise 4.8 Show that for an arbitrary df F we have

(22) $$d(F/(1-F)) = \{1/((1-F)(1-F_-))\}\, dF.$$

Exercise 4.9 For any df F we have

$$\int [F(x+\theta) - F(x)]\, dx = \theta \qquad \text{for each } \theta \geq 0.$$

Exercise 4.10 (Stein) Suppose $X \cong (0, \sigma^2)$ with df F. Then $g(x) \equiv \int_x^\infty y\, dF(y)/\sigma^2$ is a density. (And $g(x) = -\int_{-\infty}^x y\, dF(y)/\sigma^2$ is also true.)

Exercise 4.11 (a) Show that $\int_0^\infty \{P(|X| > x)\}^{1/2}\, dx < \infty$ implies $\mathrm{E}X^2 < \infty$.
(b) Show that $\{\int_0^\infty \{P(|X| > x)\}^{1/2}\, dx \leq \frac{r}{r-2}\, \|X\|_r$ for any $r > 2$, so that the integral on the left is finite whenever $X \in \mathcal{L}_r$ for any $r > 2$.
Hint. Verify (a) when X is bounded, via (13) and Markov. Then apply the MCT. Consider (b). Bound $\int_0^\infty = \int_0^c + \int_c^\infty \leq c + \int_c^\infty$ via Markov, and then choose "c" to minimize the bound.

5 Important Statistical Quantities

Notation 5.1 ($\check{\text{T}}$rimming, $\tilde{\text{W}}$insorizing, and $\check{\text{T}}$runcating, and dom(a, a')) Let dom(a, a') denote $[0, 1-a')$ if $X \geq 0$, or $(a, 1]$ if $X \leq 0$, or $(a, 1-a')$ otherwise. Let $\tilde{K}_{a,a'}(\cdot)$ denote $K(\cdot)$ *Winsorized outside* the domain of Winsorization dom(a, a'). Thus when X takes both positive and negative values and we suppose that "a" and "a'" are specified so small that $K_+(a) < 0 < K(1-a')$, it follows that

(1) $$\tilde{K}_{a,a'}(t) \text{ equals } K_+(a),\ K(t),\ K(1-a') \quad \text{as } 0 < t \leq a,\ a < t < 1-a',\ 1-a' \leq t < 1$$

(while $a \equiv 0$ and $\tilde{K}(a) \equiv K(0)$ if $X \geq 0$, etc). Let ξ denote a Uniform$(0,1)$ rv. Let

(2) $$\tilde{\mu}(a, a') \equiv \tilde{\mu}_K(a, a') \equiv \mathrm{E}\tilde{K}_{a,a'}(\xi) \equiv \int_0^1 \tilde{K}_{a,a'}(t)\,dt,$$

which is the (a, a')-*Winsorized mean* of the rv $K(\xi)$, and let

$$\tilde{\sigma}^2(a, a') \equiv \tilde{\sigma}_K^2(a, a') \equiv \mathrm{Var}[\tilde{K}_{a,a'}(\xi)] = \int_0^1 \tilde{K}_{a,a'}^2(t)\,dt - \tilde{\mu}(a, a')^2$$

(3) $$= \int_0^1\int_0^1 [s \wedge t - st]\,d\tilde{K}_{a,a'}(s)\,d\tilde{K}_{a,a'}(t)$$

denote the (a, a')-*Winsorized variance* (recall (6.4.8)). For general X, let

(4) $$\tilde{\mu}(a) \equiv \tilde{\mu}(a, a)\,, \qquad \tilde{\sigma}^2(a) \equiv \tilde{\sigma}^2(a, a)\,, \qquad \text{and} \qquad \tilde{K}_a(\cdot) \equiv \tilde{K}_{a,a}(\cdot)\,;$$

but $\tilde{\mu}(a) \equiv \tilde{\mu}_{0,a}$ if $X \geq 0$, etc.

We now let $0 \leq k_n < n - k_n' \leq n$ denote integers, and then let

$$a_n \equiv k_n/n \ \text{ and } \ a_n' \equiv k_n'/n\,, \qquad \text{so that } 0 \leq a_n < 1 - a_n' \leq 1\,.$$

Let $\tilde{K}_n(\cdot)$ denote $K(\cdot)$ *Winsorized outside* dom(a_n, a_n'). Let

(5) $$\check{\mu}_n \equiv \check{\mu}_K(a_n, a_n') \equiv \int_{a_n}^{1-a_n'} K(t)\,dt, \quad \check{\mu}_n \equiv \check{\mu}_K(a_n, a_n') \equiv \check{\mu}_n/(1 - a_n - a_n'),$$
$$\tilde{\mu}_n \equiv \tilde{\mu}_K(a_n, a_n') \equiv \mu_{\tilde{K}_n} \equiv \mathrm{E}\tilde{K}_n(\xi) \equiv \int_0^1 \tilde{K}_n(t)\,dt,$$

so that $\check{\mu}_n$ is the (a_n, a_n')-*trimmed mean*, $\tilde{\mu}_n$ is the (a_n, a_n')-*Winsorized mean*, and $\check{\mu}_n$ is herein called the (a_n, a_n')-*truncated mean* of the rv $K(\xi)$. Then let

(6) $$\tilde{\sigma}_n^2 \equiv \tilde{\sigma}_K^2(a_n, a_n') \equiv \sigma_{\tilde{K}_n}^2 \equiv \mathrm{Var}[\tilde{K}_n(\xi)] = \int_0^1\int_0^1 [s \wedge t - st]\,d\tilde{K}_n(s)\,d\tilde{K}_n(t)$$

denote the (a_n, a_n')-*Winsorized variance*. When they are finite, the *mean* μ and *variance* σ^2 satisfy

$$\mu \equiv \mu_K = \int_0^1 K(t) = \int x\,dF(x) = \mathrm{E}(X) = \mathrm{E}K(\xi),$$
$$\sigma^2 \equiv \sigma_K^2 = \int_0^1\int_0^1 [s \wedge t - st]\,dK(s)\,dK(t) = \mathrm{E}X^2 - \mu^2 = \mathrm{E}K^2(\xi) - \mu^2.$$

Let $a_. \equiv \inf\{t : K(t) \geq 0\}$, and let $a_\circ \equiv a_. \wedge (1 - a_.)$. (But $a_. \equiv 0$ if $X \geq 0$, and $a_. \equiv 1$ if $X \leq 0$.) Now, $(K - \tilde{\mu}_n)^+$ and $(K - \tilde{\mu}_n)^-$ denote the positive and negative parts of $K - \tilde{\mu}_n$, and let

(7) $$\bar{K}_n \equiv [K - \tilde{\mu}_n] \quad \text{and} \quad \bar{K}_n^2 \equiv -[(K - \tilde{\mu}_n)^-]^2 + [(K - \tilde{\mu}_n)^+]^2 \quad \text{on } (0,1).$$

In this context, we may wish to assume that both

(8) $\qquad (k_n \wedge k'_n) \to \infty \qquad \text{and} \qquad (a_n \vee a'_n) \to 0;$

and perhaps we will also assume

$$a'_n/a_n \to 1 \quad \text{and/or} \quad (k_n - k'_n)/(k_n \wedge k'_n) \to 0\,.$$

We will refer to k_n, k'_n as the *trimming/Winsorizing numbers* and a_n, a'_n as the *trimming/Winsorizing fractions*. Describe the case of (8) as *slowly growing to* ∞.

Now suppose that $X_{n1}, \ldots, X_{nn}$ is an iid sample with df F and qf K. Let $X_{n:1} \le \cdots \le X_{n:n}$ denote the *order statistics* (that is, they are the ordered values of $X_{n1}, \ldots, X_{nn}$). Let $\mathbb{K}_n(\cdot)$ on $[0,1]$ denote the *empirical qf* that equals $X_{n:i}$ on $((i-1)/n, i/n]$, for $1 \le i \le n$, and that is right continuous at zero. Now let

(9) $$\bar{X}_n \equiv \frac{1}{n}\sum_{k=1}^{n} X_{nk} = \mu_{\mathbb{K}_n}(0,0) \qquad \text{and} \qquad S_n^2 \equiv \frac{1}{n}\sum_{k=1}^{n}(X_{nk} - \bar{X}_n)^2 = \sigma^2_{\mathbb{K}_n}(0,0)$$

denote the *sample mean* and the "*sample variance*." We also let

(10) $$\check{X}_n \equiv \frac{1}{n}\sum_{i=k_n+1}^{n-k'_n} X_{n:i} = \check{\mu}_{\mathbb{K}_n}(a_n, a'_n)\,, \qquad \check{X}_n \equiv \frac{1}{n - k_n - k'_n}\sum_{i=k_n+1}^{n-k'_n} X_{n:i}\,,$$

(11) $$\tilde{X}_n \equiv \frac{1}{n}\Big[k_n X_{n:k_n+1} + \sum_{i=k_n+1}^{n-k'_n} X_{n:i} + k'_n X_{n-k'_n}\Big] = \tilde{\mu}_{\mathbb{K}_n}(a_n, a'_n)$$

denote the *sample* (a_n, a'_n)*-truncated mean*, the *sample* (a_n, a'_n)*-trimmed mean*, and the *sample* (a_n, a'_n)*-Winsorized mean*. Let $\tilde{X}_{n:1}, \ldots, \tilde{X}_{n:n}$ denote the (a_n, a'_n)-*Winsorized order statistics*, whose empirical qf is $\tilde{\mathbb{K}}_n$. Now note that

(12) $$\tilde{X}_n = \frac{1}{n}\sum_{i=1}^{n} \tilde{X}_{n:i} = \mu_{\tilde{\mathbb{K}}_n}; \qquad \text{let} \qquad \tilde{S}_n^2 \equiv \frac{1}{n}\sum_{i=1}^{n} (\tilde{X}_{n:i} - \tilde{X}_n)^2 = \sigma^2_{\tilde{\mathbb{K}}_n}$$

denote the *sample* (a_n, a'_n)*-Winsorized variance*. Let

$$\check{\sigma}_n^2 \equiv \tilde{\sigma}_n^2/(1 - a_n - a'_n)^2 \qquad \text{and} \qquad \check{S}_n^2 \equiv \tilde{S}_n^2/(1 - a_n - a'_n)^2.$$

Of course, $\bar{X}_n$, S_n, $\check{X}_n$, $\tilde{S}_n$ estimate μ, σ, $\check{\mu}_n$, $\tilde{\sigma}_n$. We also define the standardized estimators

(13) $$Z_n \equiv \frac{\sqrt{n}\,(\bar{X}_n - \mu)}{\sigma} \qquad \text{and} \qquad \check{Z}_n \equiv \frac{\sqrt{n}\,(\check{X}_n - \check{\mu}_n)}{\check{\sigma}_n} = \frac{\sqrt{n}\,(\check{X}_n - \check{\mu}_n)}{\tilde{\sigma}_n}\,,$$

and the Studentized estimators

(14) $$T_n \equiv \frac{\sqrt{n}\,(\bar{X}_n - \mu)}{S_n} \qquad \text{and} \qquad \check{T}_n \equiv \frac{\sqrt{n}\,(\check{X}_n - \check{\mu}_n)}{\check{S}_n} = \frac{\sqrt{n}\,(\check{X}_n - \check{\mu}_n)}{\tilde{S}_n}\,.$$

[The first formula for $\check{T}_n$ is for statistical application, while the second formula is for probabilistic theory.] We will very often assume that these independent rvs

$X_{n1}, \ldots, X_{nn}$ having df F and qf K are defined in terms of independent Uniform$(0,1)$ rvs $\xi_{n1}, \ldots, \xi_{nn}$ via (see above (6.4.3))

$$(15) \qquad X_{nk} \equiv K(\xi_{nk}) \qquad \text{for } 1 \le k \le n.$$

(If we start with iid K rvs $X_1, \ldots, X_n$ and then define Uniform(0,1) rvs $\dot{\xi}_1, \ldots \dot{\xi}_n$ via (6.3.8), then the $X_k \equiv F^{-1}(\dot{\xi}_k)$ (as in (6.3.9)) are just the original X_k's.) Thus the device of (15) is broadly useful.

We define the *sample median* $\ddot{X}_n$ to equal $X_{n:(n+1)/2}$ or $(X_{n:n/2} + X_{n:n/2+1})/2$, according as n is odd or even.

In the previous context, let $\xi_{n:1} < \cdots < \xi_{n:n}$ denote the order statistics of iid Uniform$(0,1)$ rvs. Let $\mathsf{R}_n \equiv (R_{n1}, \ldots, R_{nn})'$ denote the *ranks* of these $\xi_{n1}, \ldots, \xi_{nn}$, and let $\mathsf{D}_n \equiv (D_{n1}, \ldots, D_{nn})'$ denote their *antiranks*. Thus the rank vector R_n is a random permutation of the vector $(1, 2, \ldots, n)'$, while D_n is the inverse permutation; and these satisfy

$$(16) \qquad \xi_{nD_{nk}} = \xi_{n:k} \qquad \text{and} \qquad \xi_{nk} = \xi_{n:R_{nk}}\,.$$

We will learn later that

$$(17) \qquad (\xi_{n:1}, \ldots, \xi_{n:n}) \text{ and } (R_{n1}, \ldots, R_{nn}) \text{ are independent random vectors.}$$

Such notation is sometimes used throughout the remainder of this book. □

The Empirical DF

Notation 5.2 (Empirical dfs and processes) Let $X_1, X_2, \ldots$ be iid with df F and qf K. The *empirical df* $\mathbb{F}_n$ of $(X_1, \ldots, X_n)$ is defined by

$$(18) \qquad \mathbb{F}_n(x) \equiv \frac{1}{n}\sum_{k=1}^{n} 1_{(-\infty,x]}(X_k) = \frac{1}{n}\sum_{k=1}^{n} 1_{[X_k \le x]} \qquad \text{for } -\infty < x < \infty.$$

This is a step function on the real line R that starts at height 0 and jumps by height $1/n$ each time the argument reaches another observation as it moves from left to right along the line. We can think of $\mathbb{F}_n$ as an estimate of F. The important study of the *empirical process*

$$(19) \qquad \mathbb{E}_n(x) \equiv \sqrt{n}\,[\mathbb{F}_n(x) - F(x)] \qquad \text{for } x \in R$$

will allow us to determine how this estimator $\mathbb{F}_n$ of F performs.

We also let $\xi_1, \xi_2, \ldots$ be iid Uniform$(0,1)$, with true df the identity function I on $[0,1]$ and with *uniform empirical df*

$$(20) \qquad \mathbb{G}_n(t) \equiv \frac{1}{n}\sum_{k=1}^{n} 1_{[0,t]}(\xi_k) = \frac{1}{n}\sum_{k=1}^{n} 1_{[\xi_k \le t]} \qquad \text{for } 0 \le t \le 1.$$

The corresponding *uniform empirical process* is given by

$$(21) \qquad \mathbb{U}_n(t) \equiv \sqrt{n}\,[\mathbb{G}_n(t) - t] \qquad \text{for } t \in [0,1].$$

If we now define an iid F sequence $X_1, X_2, \ldots$ via $X_k \equiv F^{-1}(\xi_k) = K(\xi_k)$, then the empirical df and empirical process of these $(X_1, \ldots, X_n)$ satisfy

$$(22)\qquad \begin{array}{l} (\mathbb{F}_n - F) = [\mathbb{G}_n(F) - I(F)] \quad \text{on } R \qquad \text{and} \qquad \mathbb{E}_n = \mathbb{U}_n(F) \quad \text{on } R, \\ \text{valid for every } \omega, \end{array}$$

as follows by (6.3.3). (If we use the $\dot{\xi}_k$'s of (6.3.8), then the $\mathbb{F}_n$ on the left in (22) is everywhere equal to the $\mathbb{F}_n$ of the original X_k's.) Thus our study of properties of $\mathbb{E}_n$ can proceed via a study of the simpler $\mathbb{U}_n$, which is then evaluated at a deterministic F. (Recall also in this regard theorem 5.3.3 about probability being determined by the finite dimensional distributions.) □

6 Infinite Variance

Whenever the variance is infinite, the Winsorized variance $\tilde{\sigma}^2$ of the previous section completely dominates the square $\tilde{\mu}^2$ of the Winsorized mean. Let $\tilde{K}_{a,a'}$ denote K Winsorized outside $(a, 1-a')$.

Theorem 6.1 (Gnedenko–Kolmogorov) Every nondegenerate qf K satisfies

$$\text{(1)} \qquad \limsup_{a\vee a'\to 0} \{\textstyle\int_a^{1-a'} |K(t)|\,dt\}^2 / \int_a^{1-a'} K^2(t)\,dt = 0 \qquad \text{whenever} \quad \mathrm{E}K^2(\xi) = \infty,$$

$$\text{(2)} \qquad \mathrm{Var}[\tilde{K}_{a,a'}(\xi)]/\mathrm{E}\tilde{K}^2_{a,a'}(\xi) \to 1 \quad \text{as } (a\vee a') \to 0 \qquad \text{whenever} \quad \mathrm{E}K^2(\xi) = \infty\,.$$

Proof. Let $h > 0$ be continuous, symmetric about $t = 1/2$, $\uparrow$ to ∞ on $[1/2, 1)$, and suppose it satisfies $C_h \equiv \int_0^1 h^2(t)\,dt < \infty$. Let $b \equiv 1 - a'$. Then Cauchy–Schwarz provides the bound

$$\text{(a)} \qquad \{\textstyle\int_a^b |K(t)|\,dt\}^2 = \{\int_a^b h(t)|K(t)/h(t)|\,dt\}^2 \le \int_a^b h^2(t)\,dt \int_a^b [K^2(t)/h^2(t)]\,dt.$$

Fix $c \equiv c_\epsilon$ so close to zero that $C_h/h^2(c) < \epsilon$. Fix c_ϵ, and let $a \vee a' \to 0$. Then

$$\{\textstyle\int_a^b |K(t)|\,dt\}^2 / \int_a^b K^2(t)\,dt \le C_h \int_a^b [K^2(t)/h^2(t)]\,dt / \int_a^b K^2(t)\,dt$$

$$\text{(b)} \qquad \le C_h \left\{ \frac{\int_c^{1-c} K^2(t)\,dt}{h^2(1/2)} + \frac{(\int_a^c + \int_{1-c}^b)\,K^2(t)\,dt}{h^2(c)} \right\} \Big/ \textstyle\int_a^b K^2(t)\,dt$$

$$\text{(c)} \qquad \le C_h \{\textstyle\int_c^{1-c} K^2(t)\,dt / h^2(1/2)\} / \int_a^b K^2(t)\,dt + \epsilon$$

$$\text{(d)} \qquad < 2\epsilon \qquad \text{for } a \text{ and } b \text{ near enough to 0 and 1, since } \mathrm{E}K^2(\xi) = \infty.$$

Then (2) follows from $[aK_+(a) + a'K(1-a')]^2 / [aK_+^2(a) + a'K^2(1-a')] \to 0$. $\square$

Exercise 6.1 (Comparing contributions to the variance) Let $K(\cdot)$ be arbitrary. Establish the following elementary properties of qfs (of non-trivial rvs):

$$\text{(a)} \qquad \limsup_{a\vee a'\to 0} [aK_+^2(a) + a'K^2(1-a')]/\tilde{\sigma}^2 \quad \begin{cases} = 0 & \text{if } \mathrm{E}K^2(\xi) < \infty, \\ \le 1 & \text{if } \mathrm{E}K^2(\xi) = \infty. \end{cases}$$

$$\text{(b)} \qquad \limsup_{a\vee a'\to 0} \textstyle\int_a^{1-a'} K^2(t)\,dt/\tilde{\sigma}^2 \quad \begin{cases} < \infty & \text{always}, \\ \le 1 & \text{if } \mathrm{E}K^2(\xi) = \infty. \end{cases}$$

$$\text{(c)} \qquad \limsup_{a\vee a'\to 0} \{a|K_+(a)| + a'|K(1-a')|\}/\tilde{\sigma} \quad \begin{cases} = 0 & \text{always}, \\ = 0 & \text{if } \mathrm{E}K^2(\xi) = \infty. \end{cases}$$

$$\text{(d)} \qquad \limsup_{a\vee a'\to 0} \textstyle\int_a^{1-a'} |K(t)|\,dt/\tilde{\sigma} \quad \begin{cases} < \infty & \text{always}, \\ = 0 & \text{if } \mathrm{E}K^2(\xi) = \infty. \end{cases}$$

$$\text{(e)} \qquad \limsup_{a\vee a'\to 0} \mathrm{E}|\tilde{K}_{a,a'}(\xi)|/\tilde{\sigma} \quad \begin{cases} < \infty & \text{always}, \\ = 0 & \text{if } \mathrm{E}K^2(\xi) = \infty. \end{cases}$$

Exercise 6.2 Let $0 < r < s$. Show that for every nondegenerate qf K we have

$$(3)\qquad \limsup_{a\vee a'\to 0} \{\int_a^{1-a'} |K(t)|^r \, dt\}^{s/r} \Big/ \int_a^{1-a'} |K(t)|^s \, dt = 0 \qquad \text{if } \mathrm{E}|K(\xi)|^s = \infty.$$

Exercise 6.3 (An added touch) Let $K \geq 0$ be $\searrow$ on $(0,1)$ with $\int_0^1 K(t)\, dt < \infty$. Then there exists a positive $\searrow$ function $h(\cdot)$ on $(0,1)$ with $h(t) \to \infty$ as $t \to 0$ for which $\int_0^1 h(t)K(t)\, dt < \infty$. (Note exercise 3.5.22.)

Proposition 6.1 Let X have sth absolute moment $\mathrm{E}|X|^s = \infty$, and let $0 < r < s$. Let $\overline{|X|_n^s} \equiv \frac{1}{n}\sum_1^n |X_k|^s$ for iid rvs $X_1, X_2, \ldots$ distributed as X, etc. Then

$$(4)\qquad [\, \overline{|X|_n^r}\,]^{s/r} / \, \overline{|X|_n^s} \to_{a.s.} 0 \qquad \text{(using remark 6.1).}$$

Remark 6.1 It is useful to give this proof now even though it will not be until the SLLN (theorem 8.4.2) that we prove $\mathrm{E}\,|X|^s < \infty$ yields $\overline{|X|_n^s} \to_{a.s.} \mathrm{E}\,|X|^s$. Likewise, $\mathrm{E}|X|^s = \infty$ yields $\overline{|X|_n^s} \to_{a.s.} \infty$. These will be used in the proofs below.

Proof. We follow the proof of Gnedenko–Kolmogorov's theorem 6.6.1, as we give the proof of (4) for $s = 2$ and $r = 1$. Let $\mathbb{K}_n$ denote the empirical qf. Let h be positive, continuous, symmetric about $t = \frac{1}{2}$, $\uparrow$ to ∞ on $[\frac{1}{2}, 1)$ and suppose it satisfies $C_h \equiv \int_0^1 h^2(t)\, dt < \infty$. Define $a = 1 - b = 1/(2n)$. Then Cauchy–Schwarz gives the bound

$$(\int_a^b |\mathbb{K}_n(t)|\, dt)^2 = (\int_a^b h(t)|\mathbb{K}_n(t)/h(t)|\, dt)^2 \leq \int_a^b h^2(t)\, dt \int_a^b [\mathbb{K}_n^2(t)/h^2(t)]\, dt$$

$$(a)\qquad = C_h \int_a^b [\mathbb{K}_n^2(t)/h^2(t)]\, dt.$$

Let $c \equiv c_\epsilon$ be fixed so close to zero that $C_h/h^2(c) < \epsilon/8$. Then

$$\{\tfrac{1}{n}\textstyle\sum_1^n |X_k|\,\}^2 / \overline{X_n^2} \leq 4\{\int_a^b |\mathbb{K}_n(t)|\, dt\}^2 / \overline{X_n^2}$$

(the "4" comes from the definition of a and b,

which gives only half of the two end intervals)

$$\leq 4\, C_h \int_a^b [\mathbb{K}_n^2(t)/h^2(t)]\, dt / \overline{X_n^2} \qquad \text{by (a)}$$

$$(b)\qquad \leq 4\, C_h \left\{ \frac{\int_c^{1-c} \mathbb{K}_n^2(t)\, dt}{h^2(1/2)} + \frac{(\int_a^c + \int_{1-c}^b)\, \mathbb{K}_n^2(t)\, dt}{h^2(c)} \right\} / \, \overline{X_n^2}$$

$$(c)\qquad \leq \frac{4\, C_h}{h^2(1/2)} \{\int_c^{1-c} \mathbb{K}_n^2(t)\, dt / \overline{X_n^2}\} + \frac{4\, C_h}{h^2(c)} \cdot \{\overline{X_n^2} / \, \overline{X_n^2}\}$$

$$(d)\qquad < \tfrac{\epsilon}{2} + \tfrac{\epsilon}{2} \cdot 1 = \epsilon \qquad \text{for all } n \text{ exceeding some } n_\epsilon,$$

using remark 6.1 for $\overline{X_n^2} \to_{a.s.} \infty$ in the final step. (Thus the numerator of the leading term in (c) converges a.s. to a finite number, while the denominator has an a.s. limit exceeding $2/\epsilon$ times the numerator.) □

Exercise 6.4 (a) Prove proposition 6.1 for general $0 < r < s$.
(b) Prove that (4) holds in the sense of $\to_p 0$ if all $X_{n1}, \ldots, X_{nn}$ are iid as X.

Proposition 6.2 (Equivalent versions of negligibility)
For any vector $\mathsf{X} \equiv (X_1, \ldots, X_n)$, let

(5) $\quad \mathcal{D}_n^2 \equiv [\max_{1\le k\le n} \ \frac{1}{n} |X_k - \bar{X}_n|^2] \,/\, S_n^2 \qquad$ where $\ S_n^2 \equiv \overline{X_n^2} - (\bar{X}_n)^2.$

Let $X_1, X_2, \ldots$ be iid as X, and set $\mathsf{X}_n = (X_1, \ldots, X_n)'$. Then

(6) $\quad \mathcal{D}_n^2 \to_{a.s.} 0 \qquad$ if and only if $\qquad [\max_{1\le k\le n} \ \frac{1}{n} X_k^2] \,/\, \overline{X_n^2} \to_{a.s.} 0\,.$

Let $X_{n1}, \ldots, X_{nn}$ be iid as X, for each $n \ge 1$. Set $\mathsf{X} = (X_{n1}, \ldots, X_{nn})'$. Then

(7) $\quad \mathcal{D}_n^2 \to_p 0 \qquad$ if and only if $\qquad [\max_{1\le k\le n} \ \frac{1}{n} X_{nk}^2] \,/\, \overline{X_n^2} \to_p 0\,.$

Proof. Consider (6). Note that

$$\text{(a)} \qquad \mathcal{D}_n^2 = \left\{ \max \left| \frac{X_k}{\sqrt{n\overline{X_n^2}}} - \frac{\bar{X}_n}{\sqrt{n\overline{X_n^2}}} \right| \right\}^2 /\{1 - (\bar{X}_n)^2 / \,\overline{X_n^2}\,\}.$$

Since $0 \le (\bar{X}_n)^2 / \,\overline{X_n^2} \le 1$ always holds by the Liapunov inequality, we have

$$\text{(b)} \qquad |\bar{X}_n| \,/ \sqrt{n\,\overline{X_n^2}} \le 1/\sqrt{n} \to_{a.s.} 0 \qquad \text{always holds for } \textit{all} \text{ rvs.}$$

Thus the second term in the numerator always goes to zero for all rvs (independent, or not). Now consult remark 6.1 for the following two claims. The denominator of (a) converges a.s. to $1 - 0 = 1$ if $\mathrm{E}X^2 = \infty$ (by (4)), while the denominator of (a) converges a.s. to $(1 - \mathrm{E}^2 X / \,\mathrm{E}X^2) < 1$ if $\mathrm{E}X^2 < \infty$. Thus $\mathcal{D}_n^2(\omega) \to 0$ for a.e. fixed $(X_1(\omega), X_2(\omega), \ldots)$ if and only if the lead term in the numerator of (a) goes to zero for a.e. fixed ω; that is, if and only if $[\max |X_k|\,]/\{n\overline{X_n^2}\}^{1/2} \to 0$ for a.e. fixed ω. This gives (6). Then (7) follows by going to subsequences, using exercise 6.4. □

Proposition 6.3 Let $X \ge 0$ with $\mathrm{E}X^s = \infty$ have df F and qf K. The rth *partial absolute moment* $M_r(\cdot)$ is defined on $[0, 1]$ by

(8) $\quad M_r(t) \equiv \int_0^{1-t} K^r(u)\, du = \int_t^1 m_r(u)\, du\,, \qquad$ where $\qquad m_r(t) \equiv [K(1-t)]^r.$

Then, for $0 < r < s$,

(9) $\quad t\, m_s(t)/M_s(t) \to 0 \quad$ implies $\quad t\, m_r(t)/M_r(t) \to 0.$

(The result (C.2.20) below, especially (and all the rest of sections C.2-C.3,are very much in the spirit of (9) and (2).)

Proof. Raising to a power increases a maximum more than an average, and so

$$\text{(10)} \qquad \frac{t\, m_r(t)}{M_r(t)} = \frac{t\, K^r(1-t)}{\int_{(t,1]} K^r(1-u)\, du} \le \frac{t}{1-t} \ \frac{K^s(1-t)}{\frac{1}{1-t} \int_{(t,1]} K^s(1-u)\, du} = \frac{t\, m_s(t)}{M_s(t)}$$

for all t so close to 1 that $K(1-t) > 1$. This establishes (9). □

The CLT and Slowly Varying Functions

Let $X_{n1}, \ldots, X_{nn}$ be iid with df F and qf $K \equiv F^{-1}$. The central limit theorem (CLT) states that $\bar{Z}_n \equiv \sqrt{n}\,[\bar{X}_n - \mu]/\sigma \to_d N(0,1)$ when σ^2 is finite. In this case

(11) $\quad t\,[K_+^2(ct) \vee K^2(1-ct)]\,/\,\sigma^2 \to 0 \qquad$ as $t \to 0, \qquad$ for each fixed $c > 0$;

just extend the calculation (6.4.6) to the second moment case.

What if σ^2 is infinite? Let ξ denote a Uniform$(0,1)$ rv. Let $X \equiv K(\xi)$, so that X has df F. Let $\tilde{X}(a) \equiv \tilde{K}_{a,a}(\xi)$, with mean $\tilde{\mu}(a) \equiv \int_0^1 \tilde{K}_{a,a}(t)\,dt$ and variance

(12) $\quad \tilde{\sigma}^2(a) = \int_0^1 \int_0^1 [r \wedge s - rs]\, d\tilde{K}_{a,a}(r)\, d\tilde{K}_{a,a}(s),$

which increases to σ^2 as $a \searrow 0$, whether σ^2 is finite or infinite. This expression *makes no reference* to any mean (such as μ or $\tilde{\mu}(a)$). This is nice! It turns out that

(13) $\quad \bar{Z}_n \equiv \sqrt{n}\,[\tilde{X}_n - \tilde{\mu}(1/n)]\,/\,\tilde{\sigma}(1/n) \to_d N(0,1)$

if and only if (exercise 10.1.8 will provide good motivation for this condition)

(14) $\quad t\,[K_+^2(ct) \vee K^2(1-ct)]\,/\,\tilde{\sigma}^2(t) \to 0 \qquad$ as $t \to 0, \qquad$ for each fixed $c > 0$.

This (14) holds if and only if $\tilde{\sigma}^2(t)$ does not grow too fast; that is, if and only if

(15) $\quad \tilde{\sigma}^2(ct)/\tilde{\sigma}^2(t) \to 1 \qquad$ as $t \to 0, \qquad$ for each fixed $c > 0$.

It is appropriate to examine such *slow variation* of $\tilde{\sigma}^2(t)$ in the infinite-variance case. We shall do so very carefully in appendix C, in both the df domain and the quantile domain. (This appendix C could have appeared as $*$-starred sections at the end of chapter 8 on the weak law of large numbers (WLLN).) Often this slow variation question is made rather difficult, by examining this problem *in the context* of the CLT and by treating it *in the context* of the general theory of slowly varying functions. But the equivalence of all the conditions in the appendix will follow from a careful treatment of the easier WLLN and treating slowly varying functions in an elementary fashion from simple pictures and a dash of Cauchy–Schwarz. The connection to the CLT will not be made until the $*$-starred sections 10.5–10.5. When all is said and done, we will have established the following theorem.

Theorem 6.2 (A studentized CLT) Let $X_{n1}, \ldots, X_{nn}$'s be row independent, all with nondegenerate df F. Conditions (14) and (15) are each equivalent to any of:

(16) $\quad \sqrt{n}\,[\bar{X}_n - \tilde{\mu}(1/n)]/\tilde{\sigma}(1/n) \to_d N(0,1)\,.$

(17) $\quad S_n^2/\tilde{\sigma}^2(1/n) \to_p 1\,.$

(18) $\quad \mathcal{D}_n^2 \equiv [\max_{1\le k\le n} \frac{1}{n}\,(X_{nk} - \bar{X}_n)^2\,]\,/\,S_n^2 \to_p 0 \qquad$ (OR, $\ [\max_{1\le k\le n} \frac{1}{n}\,X_{nk}^2]\,/\,\overline{X_n^2} \to_p 0.$)

(19) $\quad U(x) \equiv \int_{[\,y^2 \le x]} y^2\, dF(y) \qquad$ satisfies $\qquad U(cx)/U(x) \to 1,$ for each $c > 1$.

(20) $\quad R(x) \equiv x^2\, P(\,|X| > x)/\int_{[\,y^2\le x]} y^2\, dF(y) \to 0 \qquad$ as $x \to \infty$.

(To apply this theorem, just verify (19) or (20) and then claim that (16), (17), and (18) hold. These trivially imply that $\sqrt{n}\,[\bar{X}_n - \tilde{\mu}(1/n)]/S_n \to_d N(0,1)$, which leads to an asymptotically valid confidence interval for $\tilde{\mu}(1/n)$.)

Chapter 7

Independence and Conditional Distributions

1 Independence

The idea of independence of events A and B is that the occurrence or nonoccurrence of A has absolutely nothing to do with the occurrence or nonoccurrence of B. It is customary to say that A and B are *independent events* if

(1) $\qquad P(AB) = P(A)P(B).$

Classes $\mathcal{C}$ and $\mathcal{D}$ of events are called *independent classes* if (1) holds for all $A \in \mathcal{C}$ and all $B \in \mathcal{D}$. We need to define the independence of more complicated objects.

Definition 1.1 (Independence) Consider a fixed probability space $(\Omega, \mathcal{A}, P)$.
(a) Consider various sub σ-fields of $\mathcal{A}$. Call such σ-fields $\mathcal{A}_1, \ldots, \mathcal{A}_n$ *independent σ-fields* if they satisfy

(2) $\qquad P(A_1 \cap \cdots \cap A_n) = \prod_1^n P(A_i) \qquad$ whenever $A_i \in \mathcal{A}_i$ for $1 \le i \le n$.

The σ-fields $\mathcal{A}_1, \mathcal{A}_2, \ldots$ are called *independent σ-fields* if $\mathcal{A}_1, \ldots, \mathcal{A}_n$ are independent for each $n \ge 2$. (Use this definition for arbitrary classes $\mathcal{A}_1, \ldots, \mathcal{A}_n$, too.)
(b) Rvs $X_1, \ldots, X_n$ are called *independent rvs* if the σ-fields $\mathcal{F}(X_i) \equiv X_i^{-1}(\mathcal{B})$ (for $1 \le i \le n$) are independent. Rvs $X_1, X_2, \ldots$ are called *independent rvs* if all $X_1, \ldots, X_n$ are independent.
(c) Events $A_1, \ldots, A_n$ are *independent events* if $\sigma[A_1], \ldots, \sigma[A_n]$ are independent σ-fields; here note that

(3) $\qquad \sigma[A_i] = \{\phi, A_i, A_i^c, \Omega\}.$

The next exercise is helpful because it will relate the rather formidable definition of independent events in (3) back to the simple definition (1).

Exercise 1.1 (a) Show that $P(AB) = P(A)P(B)$ if and only if $\{\emptyset, A, A^c, \Omega\}$ and $\{\emptyset, B, B^c, \Omega\}$ are independent σ-fields. [Thus we maintain the familiar (1).]
(b) Show that $A_1, \ldots, A_n$ are independent if and only if

$$P(A_{i_1} \cdots A_{i_k}) = \textstyle\prod_{j=1}^{k} P(A_{i_j}) \quad \text{whenever } 1 \le i_1 < \cdots < i_k \le n \text{ with } 1 \le k \le n. \tag{4}$$

Remark 1.1 When discussing a pair of possibly independent events, one should draw the Venn diagram as a square representing Ω divided into half vertically (with respect to A, A^c) and into half horizontally (with respect to B, B^c) creating four cells (rather than as the familiar two-circle picture). Also, if one writes on the table the probability of each of the four combinations AB, AB^c, A^cB, A^cB^c, one has the contingency table superimposed on the picture. (See figure 1.1.) [This extends to two partitions $(A_1, \ldots, A_m)$ and $(B_1, \ldots, B_n)$, but not to three events.] □

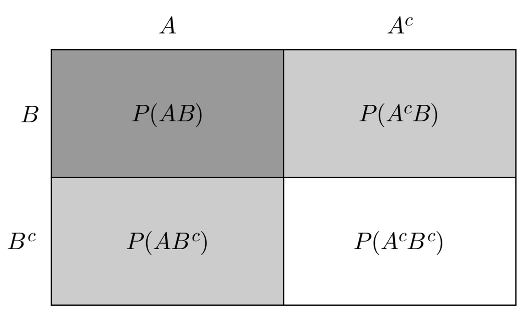


Figure 1.1 The 2 × 2 table.

Theorem 1.1 (Expectation of products) Suppose X and Y are independent rvs for which $g(X)$ and $h(Y)$ are integrable. Then $g(X)\,h(Y)$ is integrable, and

$$\mathrm{E}[g(X)\,h(Y)] = \mathrm{E}g(X)\,\mathrm{E}h(Y)\,. \tag{5}$$

Proof. The assertion is obvious for $g = 1_A$ and $h = 1_B$. Now hold $g = 1_A$ fixed, and proceed through simple and nonnegative h. Then with h held fixed, proceed through simple, nonnegative, and integrable g. Then extend to integrable h. □

Proposition 1.1 (Extending independence on π-systems)
(a) Suppose the π-system $\mathcal{C}$ and a class $\mathcal{D}$ are independent. Then

$$\sigma[\mathcal{C}] \quad \text{and} \quad \mathcal{D} \quad \text{are independent.}$$

(b) Suppose the π-systems $\mathcal{C}$ and $\mathcal{D}$ are independent. Then

$$\sigma[\mathcal{C}] \quad \text{and} \quad \sigma[\mathcal{D}] \quad \text{are independent } \sigma\text{-fields.}$$

(c) If $\mathcal{C}_1, \ldots, \mathcal{C}_n$ are independent π-systems (see (2)), then

$$\sigma[\mathcal{C}_1], \ldots, \sigma[\mathcal{C}_n] \qquad \text{are independent } \sigma\text{-fields.}$$

Proof. (a) Fix $D \in \mathcal{D}$, and define

(i) $\quad \mathcal{C}_D \equiv \{A \in \sigma[\mathcal{C}] : P(AD) = P(A)P(D)\}.$

We now demonstrate that $\mathcal{C}_D$ is a λ-system (that trivially contains $\mathcal{C}$). Trivially, $\Omega \in \mathcal{C}_D$. If $A, B \in \mathcal{C}_D$ with $A \subset B$, then

$$P((A \setminus B)D) = P(AD \setminus BD) = P(AD) - P(BD)$$

(j) $\quad = P(A)P(D) - P(B)P(D) = P(A \setminus B)P(D);$

and this implies that $A \setminus B \in \mathcal{C}_D$. If $A_n \nearrow A$ with all $A_n \in \mathcal{C}_D$, then

(k) $\quad P(AD) = P(\lim A_n D) = \lim P(A_n D) = \lim P(A_n)P(D) = P(A)P(D);$

and this implies that $A \in \mathcal{C}_D$. Thus $\mathcal{C}_D$ is a λ-system, and it trivially contains the π-system $\mathcal{C}$. Thus $\mathcal{C}_D \supset \lambda[\mathcal{C}] = \sigma[\mathcal{C}]$, using (1.1.19) for the equality. Finally, this is true for every $D \in \mathcal{D}$.

Just apply part (a) to the π-system $\mathcal{D}$ and the arbitrary class $\sigma[\mathcal{C}]$ to obtain (b). The now minor (c) is left to exercise 1.3. □

Theorem 1.2 Let $X_1, X_2, \ldots$ be independent rvs on $(\Omega, \mathcal{A}, P)$. Let $\mathsf{i} \equiv (i_1, i_2, \ldots)$ and $\mathsf{j} \equiv (j_1, j_2, \ldots)$ be disjoint sets of integers. (a) Then

(6) $\quad \mathcal{F}(X_{i_1}, X_{i_2}, \ldots) \quad \text{and} \quad \mathcal{F}(X_{j_1}, X_{j_2}, \ldots) \quad$ are independent σ-fields.

(b) This extends immediately to countably many disjoint sets of integers.

Corollary 1 (Preservation of independence)
Any rvs $h_1(X_{i_1}, X_{i_2}, \ldots), h_2(X_{j_1}, X_{j_2}, \ldots), \ldots$ (that are based on disjoint sets of the underlying independent rvs X_k) are themselves independent rvs, for any choice of the $\mathcal{B}$-$\mathcal{B}_\infty$-measurable functions $h_1, h_2, \ldots$.

Proof. Let $\mathcal{C}$ denote all sets of the form $C \equiv [X_{i_1} \in B_1, \ldots, X_{i_m} \in B_m]$, for some $m \geq 1$ and for $B_1, \ldots, B_m$ in $\mathcal{B}$. Let $\mathcal{D}$ denote all sets of the form $D \equiv [X_{j_1} \in B_1', \ldots, X_{j_n} \in B_n']$ for some $n \geq 1$ and sets $B_1', \ldots, B_n'$ in $\mathcal{B}$. Both $\mathcal{C}$ and $\mathcal{D}$ are $\bar{\pi}$-systems, while $\sigma[\mathcal{C}] = \mathcal{F}(X_{i_1}, X_{i_2}, \ldots)$ and $\sigma[\mathcal{D}] = \mathcal{F}(X_{j_1}, X_{j_2}, \ldots)$. In fact, $\mathcal{F}(X_{i_1}, X_{i_2}, \ldots) = \mathsf{X}_{\mathsf{i}}^{-1}(\mathcal{B}_\infty) = \mathsf{X}_{\mathsf{i}}^{-1}(\sigma[\mathcal{C}_\infty]) = \sigma[\mathsf{X}_{\mathsf{i}}^{-1}(\mathcal{C}_\infty)] = \sigma[\mathcal{C}]$. Thus

(p) $\quad P(C \cap D) = P(\{\cap_{k=1}^{m}[X_{i_k} \in B_k]\} \; \cap \; \{\cap_{l=1}^{n}[X_{j_l} \in B'_l]\})$

$= \prod_{k=1}^{m} P(X_{i_k} \in B_k) \prod_{l=1}^{n} P(X_{j_l} \in B'_l)$ by independence

$= P(\cap_{k=1}^{m}[X_{i_k} \in B_k]) \; P(\cap_{l=1}^{n}[X_{j_l} \in B'_l])$ by independence

(q) $\quad = P(C) \, P(D)$,

so that $\mathcal{C}$ and $\mathcal{D}$ are independent classes. Thus $\sigma[\mathcal{C}]$ and $\sigma[\mathcal{D}]$ are independent by proposition 1.1(b), as is required for (6). The extension to countably many disjoint sets of indices is done by induction using proposition 1.1(c), and is left to the exercises. [The corollary is immediate.] □

Exercise 1.2 Prove theorem 1.2(b).

Exercise 1.3 Prove proposition 1.1(c).

Criteria for Independence

Theorem 1.3 The rvs $(X_1, \ldots, X_n)$ are independent rvs if and only if

(7) $\quad F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = F_{X_1}(x_1) \cdots F_{X_n}(x_n) \qquad$ for all $x_1, \ldots, x_n$.

Proof. Clearly, independence implies that the joint df factors. For the converse we suppose that the joint df factors. Then for all $x_1, \ldots, x_n$ we have

$$P(X_1 \le x_1, \ldots, X_n \le x_n) = P(X_1 \le x_1) \cdots P(X_n \le x_n).$$

That is, the classes $\mathcal{C}_i \equiv \{[X_i \le x_i] : x_i \in R\}$ are independent, and they are π-systems, with $\sigma[\mathcal{C}_i] = \mathcal{F}(X_i)$. Independence of $X_1, \ldots, X_n$ then follows from proposition 1.1(c). □

Exercise 1.4 Rvs X, Y that take on only a countable number of values are independent if and only if $P([X = a_i][Y = b_j]) = P(X = a_i)P(Y = b_j)$ for all i, j.

Exercise 1.5 Show that rvs X, Y having a joint density $f(\cdot, \cdot)$ are independent if and only if the joint density factors to give $f(x, y) = f_X(x) f_Y(y)$ for a.e. x, y.

Remark 1.2 That rv's X, Y are independent if and only if their characteristic function factors appears as theorem 9.5.3 below. □

2 The Tail σ-Field

Definition 2.1 (The tail σ-field) Start with an arbitrary random element $\mathsf{X} \equiv (X_1, X_2, \ldots)$ from $(\Omega, \mathcal{A}, P)$ to $(R_\infty, \mathcal{B}_\infty)$. Then $\mathcal{T} \equiv \bigcap_{n=1}^{\infty} \mathcal{F}(X_n, X_{n+1}, \ldots)$ is called the *tail σ-field*, and any event $D \in \mathcal{T}$ is called a *tail event*.

Theorem 2.1 (Kolmogorov's 0-1 law) If $X_1, X_2, \ldots$ are independent rvs, then $P(D)$ equals 0 or 1 for all tail events D in the tail σ-field $\mathcal{T}$.

Proof. Fix a set $D \in \mathcal{T}$, and then let $D \in \mathcal{F}(\mathsf{X}) = \mathsf{X}^{-1}(\mathcal{B}_\infty)$. By the Halmos approximation lemma of exercise 1.2.3 and the introduction to section 2.5, there exists an integer n and a set D_n in the nth member of $\cup_m \mathcal{F}(X_1, \ldots, X_m) = \cup_m \mathsf{X}_m^{-1}(\mathcal{B}_m) =$ (a field) such that $P(D_n \Delta D) \to 0$. Thus both $P(D \cap D_n) \to P(D)$ and $P(D_n) \to P(D)$ occur. Happily, $D \in \mathcal{T} \subset \mathcal{F}(X_{n+1}, \ldots)$, so that D and $D_n \in \mathcal{F}(X_1, \ldots, X_n)$ are independent. Hence

$$P^2(D) \leftarrow P(D)\,P(D_n) = P(D \cap D_n) \to P(D),$$

yielding $P^2(D) = P(D)$. Thus $P(D) = 0$ or 1. □

Remark 2.1 (Sequences and series of independent rvs converge a.s., or almost never) Note that for any Borel sets $B_1, B_2, \ldots$ in $\mathcal{B}$,

(1) $$[X_n \in B_n \text{ i.o.}] \text{ equals } \overline{\lim}\,[X_n \in B_n] = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} [X_m \in B_m] \in \mathcal{T},$$

since $\bigcup_{m=n}^{\infty} [X_m \in B_m] \in \mathcal{F}(X_n, \ldots)$. Also,

$$\underline{\lim}\,[X_n \in B_n] = \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} [X_m \in B_m] = (\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} [X_m \in B_m^c])^c$$

(2) $$= (\overline{\lim}\,[X_n \in B_n^c])^c \in \mathcal{T}.$$

Note also that

$$[\omega : X_n(\omega) \to \text{ (some finite } X(\omega))]^c = [\omega : X_n(\omega) \not\to \text{ (some finite } X(\omega))]$$

(3) $$= \bigcup_{k=1}^{\infty} \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} [\omega : |X_m(\omega) - X_n(\omega)| > 1/k] \in (\bigcup_{k=1}^{\infty} \mathcal{T}) = \mathcal{T}.$$

Likewise, if $S_n = \sum_{i=1}^{n} X_i$, then

$$[\omega : S_n(\omega) \to \text{(some finite } S(\omega))]^c$$

(4) $$= \bigcup_{k=1}^{\infty} \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} [\omega : |\textstyle\sum_{i=n+1}^{m} X_i(\omega)| > 1/k] \in \mathcal{T}.$$

The following result has thus been established. □

Theorem 2.2 Sequences and series of independent rvs can only converge either a.s. or almost never.

The Symmetric σ-Field

Definition 2.2 (Symmetric sets) Let π denote any mapping of the integers onto themselves that (for some finite n) merely permutes the first n integers. Let $\mathsf{X} \equiv (X_1, X_2, \ldots)$ be $\mathcal{B}_\infty$-measurable, and set $\mathsf{X}_\pi \equiv (X_{\pi(1)}, X_{\pi(2)}, \ldots)$. Then $A \equiv \mathsf{X}^{-1}(B)$ for some $B \in \mathcal{B}_\infty$ is called a *symmetric set* if $A = \mathsf{X}_\pi^{-1}(B)$ for all such π. Let $\mathcal{S}$ denote the collection of all symmetric sets.

Exercise 2.1 (Hewitt–Savage 0-1 law)
Let $\mathsf{X} \equiv (X_1, X_2, \ldots)$ have iid coordinates X_k.
(a) Show that $P(A)$ equals 0 or 1 for every A in $\mathcal{S}$.
(b) Show that $\mathcal{S}$ is a σ-field, called the *symmetric σ-field*.
(c) Show that the tail σ-field $\mathcal{T}$ is a subset of the symmetric σ-field $\mathcal{S}$.
(d) Give an example where $\mathcal{T}$ is a proper subset of $\mathcal{S}$.
[Hint. Use the approximation lemma of exercise 1.2.3 as it was used above.]

3 Uncorrelated Random Variables

Recall from definition 3.4.1 that $X \cong (\mu, \sigma^2)$ denotes that X has mean μ and a variance σ^2 that is assumed to be finite.

Definition 3.1 (Correlation) If $X_1, \ldots, X_n$ have finite variances, then we call them *uncorrelated* if $\mathrm{Cov}[X_i, X_j] \equiv \mathrm{E}\{(X_i - \mathrm{E}X_i)(X_j - \mathrm{E}X_j)\} = 0$ for all $i \neq j$. Define the dimensionless quantity

$$\mathrm{Corr}[X_i, X_j] \equiv \frac{\mathrm{Cov}[X_i, X_j]}{\sqrt{\mathrm{Var}[X_i]\mathrm{Var}[X_j]}}$$

to be the *correlation* between X_i and X_j. If $\mathsf{X} \equiv (X_1, \ldots, X_n)'$, then the $n \times n$ *covariance matrix* of X is defined to be the matrix $\Sigma \equiv |[\sigma_{ij}]|$ whose (i,j)th element is $\sigma_{ij} \equiv \mathrm{Cov}[X_i, X_j]$.

Proposition 3.1 Independent rvs with finite variances are uncorrelated.

Proof. Now,

(a) $\quad \mathrm{Cov}[X, Y] = \mathrm{E}[(X - \mu_X)(Y - \mu_Y)] = \mathrm{E}(X - \mu_X)\mathrm{E}(Y - \mu_Y) = 0 \cdot 0 = 0,$

where

(b) $\quad 0 \leq \{\mathrm{Cov}[X, Y]\}^2 \leq \mathrm{Var}[X]\,\mathrm{Var}[Y] < \infty$

by Cauchy–Schwarz. □

Note from (b) (or recall from (3.4.14)) that

(1) $\quad |\,\mathrm{Corr}[X, Y]\,| \leq 1 \qquad$ for any X and Y having finite variances.

Proposition 3.2 If $(X_1, \ldots, X_n)$ are uncorrelated and $X_i \cong (\mu_i, \sigma_i^2)$, then

(2) $$\sum_1^n a_i X_i \cong \Big(\sum_1^n a_i\, \mu_i, \sum_1^n a_i^2\, \sigma_i^2 \Big).$$

In particular, suppose $X_1, \ldots, X_n$ are uncorrelated (μ, σ^2). Then

(3) $$\bar{X}_n \equiv \frac{1}{n} \sum_1^n X_i \cong \Big(\mu, \frac{\sigma^2}{n}\Big), \qquad \text{while} \qquad \sqrt{n}(\bar{X}_n - \mu)/\sigma \cong (0, 1),$$

provided that $0 < \sigma < \infty$. Moreover

$$\mathrm{Cov}\Big[\sum_{i=1}^m a_i\, X_i, \sum_{j=1}^n b_j Y_j\Big] = \mathrm{Cov}\Big[\sum_{i=1}^m a_i\,(X_i - \mu_{X_i}), \sum_{j=1}^n b_j\,(Y_j - \mu_{Y_j})\Big]$$

(4) $$= \sum_{i=1}^m \sum_{j=1}^n a_i\, b_j\, \mathrm{Cov}[X_i, Y_j]\,.$$

Note that if

(5) $\quad Y = AX, \qquad \text{then} \qquad \Sigma_Y = A\,\Sigma_X A'.$

Proof. This is trivial. □

4 Basic Properties of Conditional Expectation

The Lebesgue integral is a widely applicable tool that extended the value of the Reimann approach. It allows more general "heavy duty results." So too, we now need to extend and rigorize our elementary approach to conditional expectation, in a way that keeps the useful results intact. (Illustrations follow the definitions.)

Definition 4.1 (Conditional expectation) Let $(\Omega, \mathcal{A}, P)$ denote a probability space. Let $\mathcal{D}$ denote a sub σ-field of $\mathcal{A}$. Let Y be a rv on $(\Omega, \mathcal{A}, P)$ for which $\mathrm{E}|Y| < \infty$. By $\mathrm{E}(Y|\mathcal{D})(\cdot)$ we mean any $\mathcal{D}$-measurable function on Ω such that

$$(1) \qquad \int_D \mathrm{E}(Y|\mathcal{D})(\omega)\, dP(\omega) = \int_D Y(\omega)\, dP(\omega) \qquad \text{for all } D \in \mathcal{D}.$$

Such a function exists and is unique a.e. P, as is seen below; we call this the *conditional expectation of* Y *given* $\mathcal{D}$. If X is another rv on $(\Omega, \mathcal{A}, P)$, then

$$(2) \qquad \mathrm{E}(Y|X)(\omega) \equiv \mathrm{E}(Y|\mathcal{F}(X))(\omega);$$

we recall that $\mathcal{F}(X) \equiv X^{-1}(\mathcal{B})$ for the Borel subsets $\mathcal{B}$ of the real line R.

Justification of definition 4.1. Let $\mathrm{E}|Y| < \infty$. Define a signed measure ν on $\mathcal{D}$ by

$$(a) \qquad \nu(D) \equiv \int_D Y\, dP \qquad \text{for all } D \in \mathcal{D}.$$

Now, ν is a signed measure on $(\Omega, \mathcal{D})$ by example 4.1.1, and the restriction of P to $\mathcal{D}$ (denoted by $P|\mathcal{D}$) is another signed measure on $(\Omega, \mathcal{D})$. Moreover, that $\nu \ll P|\mathcal{D}$ is trivial. Thus the Radon–Nikodym theorem guarantees, uniquely a.e. $P|\mathcal{D}$, a $\mathcal{D}$-measurable function h such that (recall exercise 3.2.3 for the second equality)

$$(b) \qquad \nu(D) = \int_D h\, d(P|\mathcal{D}) = \int_D h\, dP \qquad \text{for all } D \in \mathcal{D}.$$

Now, being $\mathcal{D}$-measurable and unique a.s. $P|\mathcal{D}$ implies that the function h is unique a.s. P. Define $\mathrm{E}(Y|\mathcal{D}) \equiv h$. Radon–Nikodym derivatives are only unique a.e., and any function that works is called a *version* of the Radon–Nikodym derivative. □

Proposition 4.1 Suppose that Z is a rv on $(\Omega, \mathcal{A})$ that is $\mathcal{F}(X)$-measurable. Then there exists a measurable function g on $(R, \mathcal{B})$ such that $Z = g(X)$.

Proof. This is just proposition 2.2.5 again. □

Notation 4.1 Since $\mathrm{E}(Y|X) = \mathrm{E}(Y|\mathcal{F}(X))$ is $\mathcal{F}(X)$-measurable, the previous proposition shows that $h \equiv \mathrm{E}(Y|X) = g(X)$ for some measurable function g on $(R, \mathcal{B})$. The theorem of the unconscious statistician gives $\int_{X^{-1}(B)} g(X)\, dP = \int_B g\, dP_X$, where we have written the general set $D \in \mathcal{F}(X)$ as $D = X^{-1}(B)$ for some $B \in \mathcal{B}$. Thus we may define $\mathrm{E}(Y|X = x) = g(x)$ to be a $\mathcal{B}$-measurable function on R for which

$$(3) \qquad \int_B \mathrm{E}(Y \mid X = x)\, dP_X(x) = \int_{X^{-1}(B)} Y(\omega)\, dP(\omega) \qquad \text{for all } B \in \mathcal{B}.$$

This function $\mathrm{E}(Y \mid X = x)$ exists and is unique a.s. P_X, as above. In summary:

$$(4) \qquad \text{If } \ g(x) \equiv \mathrm{E}(Y \mid X = x), \quad \text{then } \ h(\omega) \equiv \mathrm{E}(Y \mid X)(\omega) = g(X(\omega)). \qquad \square$$

Definition 4.2 (Conditional probability) Since $P(A) = \mathrm{E}1_A$ for standard probability, define the *conditional probability of A given $\mathcal{D}$*, denoted by $P(A|\mathcal{D})$, by

$$(5) \qquad P(A|\mathcal{D}) \equiv \mathrm{E}(1_A|\mathcal{D}).$$

Equivalently, $P(A|\mathcal{D})$ is a $\mathcal{D}$-measurable function on Ω satisfying

$$(6) \qquad P(A \cap D) = \int_D P(A|\mathcal{D})\, dP \qquad \text{for all } D \in \mathcal{D},$$

and it exists and is unique a.s. P. Also,

$$P(A|X) \equiv P(A|\mathcal{F}(X)).$$

Thus $P(A|X)(\omega) = g(X(\omega))$, where $g(x) \equiv P(A|X = x)$ is a $\mathcal{B}$-measurable function satisfying

$$(7) \qquad P(A \cap X^{-1}(B)) = \int_B P(A|X = x)\, dP_X(x) \qquad \text{for all } B \in \mathcal{B}.$$

This function exists and is unique a.s. P_X.

Discussion 4.1 (Discrete case; elementary treatment) Given that the event B has occurred (with $P(B) > 0$), how likely is it now for the event A to occur. The classic elementary approach defines the *conditional probability* of A *given* B by $P(A|B) \equiv P(AB)/P(B)$. Thus we have taken a revisualized view of things, while regarding B as the updated sample space. For any event B, only that portion AB of the event A is relevant (as B was known to have occurred). Thus all that matters is the probabilistic size $P(AB)$ of AB relative to the probabilistic size $P(B)$ of B. The resulting $P(\cdot|B)$ is a probability distribution over $\mathcal{A}_B \equiv \{AB : A \in \mathcal{A}\}$.

For discrete rvs X and Y with mass function $p(\cdot,\cdot)$ this leads to

$$p_{Y|X=x}(y) \equiv p(x,y)/\, p_X(x), \qquad \text{for each } x \text{ for which } p_X(x) \neq 0,$$

for the conditional mass function. It is then natural to define

$$\mathrm{E}(\psi(Y)|X = x) \equiv \textstyle\sum_{\text{all } y} \psi(y)\, p_{Y|X=x}(y), \quad \text{for each } x \text{ with } p_X(x) \neq 0$$

when $\mathrm{E}|\psi(Y)| < \infty$. It is then elementary to show that $\mathrm{E}(Y) = \mathrm{E}(\,\mathrm{E}(\psi(Y)|X))$. $\square$

Exercise 4.1 ("Discrete" conditional probability; general case) Suppose $\Omega = \sum_i D_i$, and then define $\mathcal{D} = \sigma[D_1, D_2, \ldots]$. Show that (whether the summation is finite or countable) the different expressions needed on the different sets D_i of the partition $\mathcal{D}$ can be combined together via

$$(8) \qquad P(A|\mathcal{D}) = \sum_i \frac{P(AD_i)}{P(D_i)}\, 1_{D_i}\,,$$

where (just for definiteness) $\frac{P(AD_i)}{P(D_i)} \equiv P(A)$ if $P(D_i) = 0$. For general $Y \in \mathcal{L}_1$ show that the function $\mathrm{E}(Y|\mathcal{D})$ takes the form

$$(9) \qquad \mathrm{E}(Y|\mathcal{D}) = \sum_i \left\{ \frac{1}{P(D_i)} \int_{D_i} Y\, dP \right\} 1_{D_i}\,,$$

with the term in braces defined to be 0 if $P(D_i) = 0$ (just for definiteness). (We note that the standard elementary approach to conditional probability is, in the discrete case, embedded within (8) and (9)—but it sits there "sideways.".)

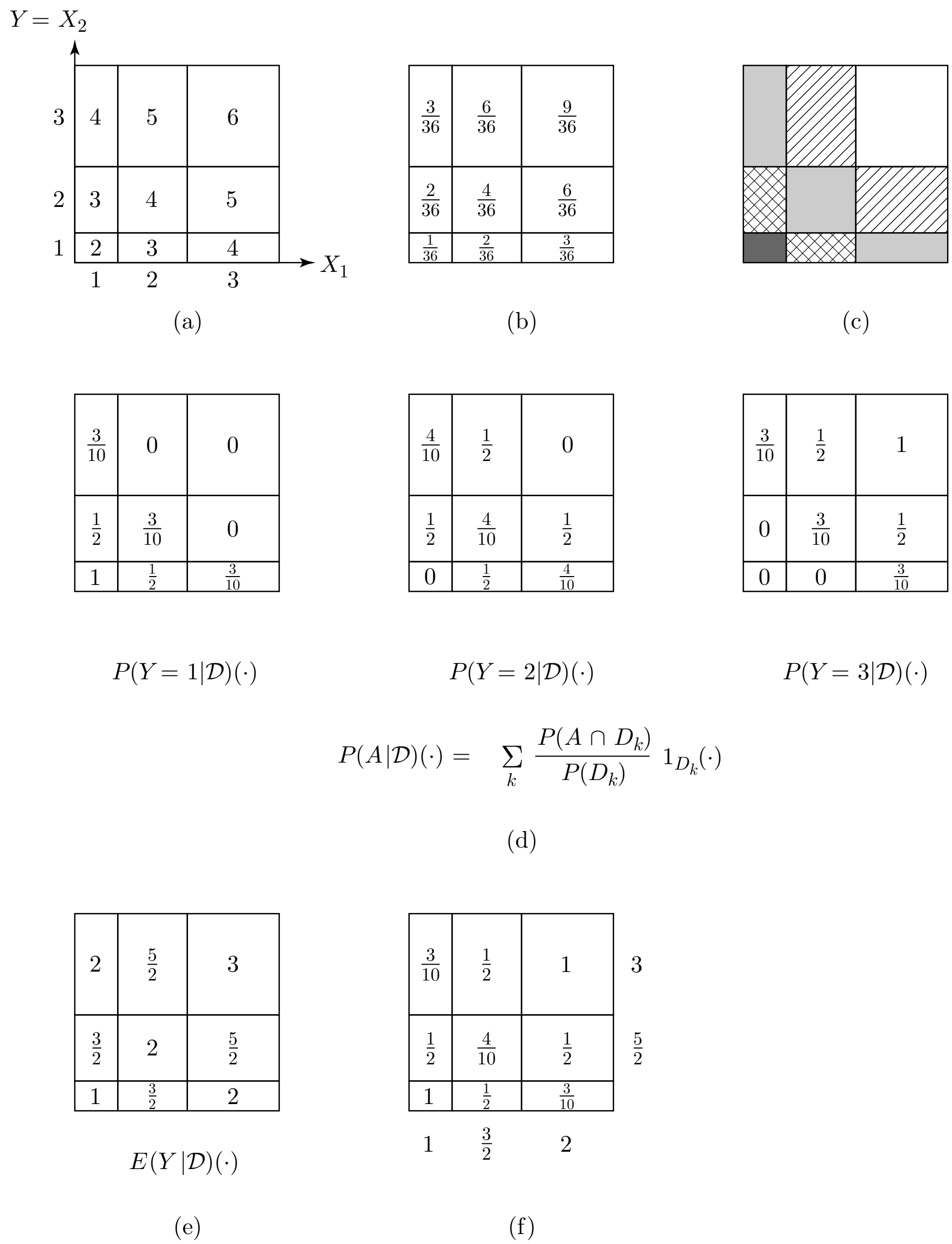


Figure 4.1 Conditional probability and conditional expectation.

Remark 4.1 It will be seen that

(10) $0 \le P(A|\mathcal{D}) \le 1$ a.s. P,

(11) $P(\sum_1^\infty A_i|\mathcal{D}) = \sum_1^\infty P(A_i\,|\mathcal{D})$ a.s. P,

(12) $P(\emptyset|\mathcal{D}) = 0$ a.s. P,

(13) $A_1 \subset A_2$ implies $P(A_1|\mathcal{D}) \le P(A_2|\mathcal{D})$ a.s. P.

[To see these, just apply parts (16)(monotonicity) and (17)(MCT) of theorem 4.1 appearing below.] These properties remind us of a probability distribution. □

Example 4.1 Let an urn consist of six balls identical except for the numbers $1, 2, 2, 3, 3, 3$. Let X_1 and X_2 represent a sample of size two drawn with replacement, and set $Y = X_2$ and $S = X_1 + X_2$. Consider figure 4.1 below. In the figure (a) we see the sample space Ω of (X_1, X_2) with the values of S superimposed, while the figure (b) superimposes the probability function on the same representation of Ω. In the figure (c) we picture the five "diagonal sets" that generate $\mathcal{D} \equiv S^{-1}(\mathcal{B})$. The three-part figure (d) depicts $P(Y = i\,|\mathcal{D})(\cdot)$ as a $\mathcal{D}$-measurable function on Ω for each of the three choices $[Y = 1]$, $[Y = 2]$, and $[Y = 3]$ for A, while the figure (e) depicts $\mathrm{E}(Y|\mathcal{D})(\omega)$ as a $\mathcal{D}$-measurable function. [Had we used the elementary definition of $P(Y = \cdot\,|S = k)$ as a function of y for each fixed k, then the conditional distributions would have been those shown along the five diagonals in the figure (f), while $\mathrm{E}(Y|S = k)$ is shown at the end of each diagonal.] □

Discussion 4.2 (Continuous case; an elementary treatment) For discrete rvs, the conditional distribution is specified by

$$g(y|x) \equiv P(Y = y\,|X = x) \equiv P(Y = y \text{ and } X = x)/P(X = x).$$

(This is in line with discussion 4.1.) One "natural approximation" of this approach for continuous rvs considers

$$g(y|x) = \lim_{h\to 0} \int_{x-h}^{x+h} f_{X,Y}(r, y)\,dr \,/ \int_{x-h}^{x+h} f_X(r)\,dr\ .$$

But making this approach rigorous fails without sufficient smoothness, and leads to a tedious and limited theory. So elementary texts just suggest the even more blatant and "less rigorous" imitation of the discrete result via

$$g(y|x)\,\Delta y \doteq \frac{f_{X,Y}(x, y)\,\Delta x\Delta y}{f_X(x)\,\Delta x} \doteq \frac{f_{X,Y}(x, y)\,\Delta y}{f_X(x)}\ .$$

Discussion 4.3 suggests that the general approach of this section should ultimately lead to this same elementary result in the case when densities do exist.

Moreover, if $(x(t), y(t))$, $a \le t \le b$, parametrizes a smooth curve (imagine a circle about the origin, or a line of slope 135°), it is definition 4.2 that leads rigorously to formulas of the type

$$\frac{f(x(t), y(t))\sqrt{(dx/dt)^2 + (dy/dt)^2}}{\int_a^b f(x(t'), y(t'))\sqrt{(dx/dt')^2 + (dy/dt')^2}\,dt'} \qquad \text{for } a \le t \le b$$

for the conditional density at the point t given that one is on the curve. □

Discussion 4.3 (Continuous case; general treatment) Let us consider the current approach to conditional probability. We will illustrate it in a special case. Let $A \in \mathcal{B}_2$ denote a two-dimensional Borel set. Let $T \equiv T(\mathsf{X}) \equiv (X_1^2 + X_2^2)^{1/2}$, so that $T = t$ defines (in the plane $\Omega = R_2$) the circle $C_t \equiv \{(x_1, x_2) : x_1^2 + x_2^2 = t^2\}$. Let $B \in \mathcal{B}_1$ denote a one-dimensional Borel set of t's, and then let $D \equiv T^{-1}(B) = \cup\{C_t : t \in B\}$. Requirements (6) and (7) (in a manner similar to exercise 4.1, but from a different point of view than used in discussion 4.1) become

$$P(AD) = P(A \cap (\cup\{C_t : t \in B\})) = \int_{\cup\{C_t : t \in B\}} P(A|T)(\mathsf{x})\, dP_{\mathsf{X}}(\mathsf{x})$$

$$\equiv \int_{\cup\{C_t : t \in B\}} h_A(\mathsf{x})\, dP_{\mathsf{X}}(\mathsf{x}) = \int_B g_A(t)\, dP_T(t) \equiv \int_B P(A|T = t)\, dP_T(t).$$

So if $g_A(\cdot)$ is given a value at t indicating the probabilistic proportion of $A \cap C_t$ that belongs to A (or $h(\mathsf{x})$ is given this same value at all $\mathsf{x} \in C_t$), then the above equation ought to be satisfied. (When densities exist, such a value would seem to be $g_A(t) = \int_{C_t} 1_A(\mathsf{x})\, p_{\mathsf{X}}(\mathsf{x})\, d\mathsf{x} / \int_{C_t} p_{\mathsf{X}}(\mathsf{x})\, d\mathsf{x}$, while $h_A(\mathsf{x}) = g_A(T(\mathsf{x}))$ would be assigned this same value at each $\mathsf{x} \in C_t$.) Requirements (1) and (2) become

$$\int_{\cup\{C_t : t \in B\}} Y\, dP = \int_{\cup\{C_t : t \in B\}} h(\mathsf{x})\, dP(\mathsf{x}) = \int_B g(t)\, dP_T(t)$$

(When densities exist, then $\mathrm{E}(Y|T = t) = g(t) = \int_{C_t} Y(\mathsf{x})\, p(\mathsf{x})\, d\mathsf{x} / \int_{C_t} p(\mathsf{x})\, d\mathsf{x}$ seems appropriate, with $h(\mathsf{x})$ getting the same value for all x in C_t.) □

Exercise 4.2 (A) (i) Mimic discussion 4.2 in case $T \equiv X_1 + X_2$, instead.
(ii) Make up another interesting example.
(B) (iii) Repeat example 4.1 and the accompanying figure, but now in the context of sampling without replacement.
(iv) Make up another interesting example.

Exercise 4.3 Let Y be a rv on some $(\Omega, \mathcal{A}, P)$ that takes on the eight values $1, \ldots, 8$ with probabilities $1/32, 2/32, 3/32, 4/32, 15/32, 4/32, 1/32, 2/32$, respectively. Let $\mathcal{C} \equiv \mathcal{F}(Y)$, and let $C_i \equiv [Y = i]$ and $p_i \equiv P(C_i)$ for $1 \le i \le 8$. Let $\mathcal{D} \equiv \sigma[\{C_1+C_5, C_2+C_6, C_3+C_7, C_4+C_8\}]$, $\mathcal{E} \equiv \sigma[\{C_1+C_2+C_5+C_6, C_3+C_4+C_7+C_8\}]$, and $\mathcal{F} \equiv \{\Omega, \emptyset\}$.
(a) Represent Ω as a 2×4 rectangle having eight 1×1 cells representing $C_1, \ldots, C_4$ in the first row and $C_5, \ldots, C_8$ in the second row. Enter the appropriate values of $Y(\omega)$ and p_i in each cell, forming a table. Evaluate $\mathrm{E}(Y)$.
(b) Evaluate $\mathrm{E}(Y|\mathcal{D})$. Present this function in a similar table. Evaluate $\mathrm{E}(\mathrm{E}(Y|\mathcal{D}))$.
(c) Evaluate $\mathrm{E}(Y|\mathcal{E})$. Present this function in a similar table. Evaluate $\mathrm{E}(\mathrm{E}(Y|\mathcal{E}))$.
(d) Evaluate $\mathrm{E}(Y|\mathcal{F})$. Present this function in a similar table. Evaluate $\mathrm{E}(\mathrm{E}(Y|\mathcal{F}))$.

For the most part, the use of regular conditional probabilities (as defined in the next section) can be avoided by appeal to the following theorem. Just as with the Reimann integral, we want to use the old method for most examples, but need the new method to justify heavy duty mathematical arguments. The next theorem shows that this is possible, and it shows how to do it. The next section shows that the old thinking is still possible as well. Just knowing that is enough to go only so, it is optional to read the rest of this chapter.

Theorem 4.1 (Properties of conditional expectation) Let X, Y, Y_n be integrable rvs on $(\Omega, \mathcal{A}, P)$. Let $\mathcal{D}$ be a sub σ-field of $\mathcal{A}$. Let g be measurable. Then for *any* versions of the conditional expectations, the following hold:

(14) (Linearity) $\mathrm{E}(aX + bY|\mathcal{D}) = a\,\mathrm{E}(X|\mathcal{D}) + b\,\mathrm{E}(Y|\mathcal{D})$ a.s. P (or, a.s. $P|\mathcal{D}$)).

(15) $\mathrm{E}Y = \mathrm{E}\,[\,\mathrm{E}(Y|\mathcal{D})]$.

(16) (Monotonicity) $X \le Y$ a.s. P implies $\mathrm{E}(X|\mathcal{D}) \le \mathrm{E}(Y|\mathcal{D})$ a.s. P.

(17) (MCT) If $0 \le Y_n \nearrow Y$ a.s. P, then $\mathrm{E}(Y_n|\mathcal{D}) \nearrow \mathrm{E}(Y|\mathcal{D})$ a.s. P.

(18) (Fatou) If $0 \le Y_n$ a.s. P, then $\mathrm{E}(\,\underline{\lim}\, Y_n|\mathcal{D}) \le \underline{\lim}\, \mathrm{E}(Y_n|\mathcal{D})$ a.s. P.

(19) (DCT) If all $|Y_n| \le X$ and $Y_n \to_{a.s.} Y$, then $\mathrm{E}(Y_n|\mathcal{D}) \to_{a.s.} \mathrm{E}(Y|\mathcal{D})$.

(20) If Y is $\mathcal{D}$-measurable and $XY \in \mathcal{L}_1(P)$, then $\mathrm{E}(XY|\mathcal{D}) =_{a.s.} Y\mathrm{E}(X|\mathcal{D})$.

(21) If $\mathcal{F}(Y)$ and $\mathcal{D}$ are independent, then $\mathrm{E}(Y|\mathcal{D}) = \mathrm{E}Y$ a.s. P.

(22) (Stepwise smoothing). If $\mathcal{D} \subset \mathcal{E} \subset \mathcal{A}$, then $\mathrm{E}[\mathrm{E}(Y|\mathcal{E})|\mathcal{D}] = \mathrm{E}(Y|\mathcal{D})$ a.s. P.

(23) If $\mathcal{F}(Y, X_1)$ is independent of $\mathcal{F}(X_2)$, then $\mathrm{E}(Y|X_1, X_2) = \mathrm{E}(Y|X_1)$ a.s. P.

(24) C_r, Hölder, Liapunov, Minkowski, and Jensen inequalities hold for $\mathrm{E}(\cdot|\mathcal{D})$.

Jensen: $g(\mathrm{E}(Y|\mathcal{D})) \le_{a.s.} \mathrm{E}[g(Y)|\mathcal{D}]$ for g convex with $g(Y)$ integrable.

(25) Let $r \ge 1$. If $Y_n \to_{\mathcal{L}_r} Y$, then $\mathrm{E}(Y_n|\mathcal{D}) \to_{\mathcal{L}_r} \mathrm{E}(Y|\mathcal{D})$. In fact,

$\mathrm{E}|\,\mathrm{E}(X|\mathcal{D}) - \mathrm{E}(Y|\mathcal{D})|^r \le \mathrm{E}|X - Y|^r$.

(26) $h_{\mathcal{D}}(\cdot)$ is a determination of $\mathrm{E}(Y|\mathcal{D})$ if and only if

$\mathrm{E}(XY) = \mathrm{E}(Xh_{\mathcal{D}})$ for all $\mathcal{D}$-measurable rvs X.

(27) If $P(D) = 0$ or 1 for all $D \in \mathcal{D}$, then $\mathrm{E}(Y|\mathcal{D}) = \mathrm{E}Y$ a.s. P.

Proof. We first prove (14). Now, by linearity of expectation,

$$\int_D [\,a\,\mathrm{E}(X|\mathcal{D}) + b\,\mathrm{E}(Y|\mathcal{D})]\,dP = a\int_D \mathrm{E}(X|\mathcal{D})\,dP + b\int_D \mathrm{E}(Y|\mathcal{D})\,dP$$

(a) $$= a\int_D X\,dP + b\int_D Y\,dP \qquad \text{by definition of } \mathrm{E}(X|\mathcal{D}), \text{ etc.}$$

$$= \int_D [\,a\,X + b\,Y\,]\,dP \qquad \text{for all } D \in \mathcal{D}, \text{ as required.}$$

To prove (15), simply note that

(b) $$\mathrm{E}\,Y = \int_\Omega Y\,dP = \int_\Omega \mathrm{E}(Y|\mathcal{D})\,dP = \mathrm{E}\,[\,\mathrm{E}(Y|\mathcal{D})\,].$$

For (16), use (14) for the first step of

$$\int_D \{E(Y|\mathcal{D}) - E(X|\mathcal{D})\}\, dP = \int_D E(Y - X|\mathcal{D})\, dP$$

(c) $$= \int_D (Y - X)\, dP \equiv \nu(D) \ge 0 \qquad \text{is a measure } \nu(\cdot).$$

Then $E(Y|\mathcal{D}) - E(X|\mathcal{D})$ is a Radon–Nikodym derivative of $\nu(\cdot)$, and so is ≥ 0 a.s.

Statement (17) follows easily from (16), since we have

(d) $$E(Y_n|\mathcal{D}) \le E(Y_{n+1}|\mathcal{D}) \le E(Y|\mathcal{D}) \qquad \text{a.s., for all } n.$$

Thus $\lim_n E(Y_n|\mathcal{D})$ exists a.s., and

(e) $$\int_D \lim_n E(Y_n|\mathcal{D})\, dP = \lim_n \int_D E(Y_n|\mathcal{D})\, dP \qquad \text{by the MCT}$$

$$= \lim_n \int_D Y_n\, dP = \int_D Y\, dP \qquad \text{by the MCT}$$

$$= \int_D E(Y|\mathcal{D})\, dP;$$

and we can appeal to the uniqueness of Radon–Nikodym derivatives, or apply exercise 3.2.2. Now we use (16) and (17) to prove (18). Thus

$$E(\underline{\lim}\, Y_n|\mathcal{D}) = E(\lim_n \inf_{k\ge n} Y_k|\mathcal{D}) \qquad \text{by the definition of } \underline{\lim}$$

(f) $$= \lim_n E(\inf_{k\ge n} Y_k|\mathcal{D}) \qquad \text{a.s., by the MCT of (17)}$$

$$\le \lim_n \inf_{k\ge n} E(Y_k|\mathcal{D}) \qquad \text{by the monotonicity of (16)}$$

$$= \underline{\lim}\, E(Y_n|\mathcal{D}) \qquad \text{by the definition of } \underline{\lim}.$$

To prove (19), apply the Fatou of (18) to $Y_n + X$ to get

$$E(Y|\mathcal{D}) + E(X|\mathcal{D}) = E(X + Y|\mathcal{D}) = E(\underline{\lim}\,(X + Y_n)|\mathcal{D})$$

(g) $$\le \underline{\lim}\, E(X + Y_n|\mathcal{D}) = \underline{\lim}\, E(Y_n|\mathcal{D}) + E(X|\mathcal{D}).$$

Canceling the a.e. finite $E(X|\mathcal{D})$ from both ends of (g) gives

$$E(Y|\mathcal{D}) \le \underline{\lim}\, E(Y_n|\mathcal{D}) \le \overline{\lim}\, E(Y_n|\mathcal{D});$$

(h) $$\le E(Y|\mathcal{D}) \qquad \text{by applying the Fatou of (18) again, to } X - Y_n.$$

To prove (20) we proceed through indicator, simple, nonnegative, and then general functions, and each time we apply exercise 3.2.2 at the final step.

Case 1: $Y = 1_{D^*}$. Then

$$\int_D Y\, E(X|\mathcal{D})\, dP = \int_D 1_{D^*}\, E(X|\mathcal{D})\, dP = \int_{D\cap D^*} E(X|\mathcal{D})\, dP$$

$$= \int_{D\cap D^*} X\, dP = \int_D 1_{D^*}\, X\, dP = \int_D Y\, X\, dP = \int_D E(YX|\mathcal{D})\, dP.$$

Case 2: $Y = \sum_1^n a_i 1_{D_i}$. Then

$$\int_D Y\,\mathrm{E}(X|\mathcal{D})\,dP = \sum_1^n a_i \int_D 1_{D_i}\,\mathrm{E}(X|\mathcal{D})\,dP$$

$$= \sum_1^n a_i \int_D 1_{D_i} X\,dP \qquad \text{by case 1}$$

$$= \int_D Y\,X\,dP = \int_D \mathrm{E}(Y X\,|\mathcal{D})\,dP.$$

Case 3: $Y \geq 0$. Let simple functions $Y_n \nearrow Y$ where $Y_n \geq 0$. Suppose first that $X \geq 0$. Then we have

$$\int_D \mathrm{E}(Y X|\mathcal{D})\,dP = \int_D \mathrm{E}(\lim_n Y_n X|\mathcal{D})\,dP$$

$$= \int_D \lim_n \mathrm{E}(Y_n X|\mathcal{D})\,dP \qquad \text{by the MCT of (17)}$$

$$= \int_D \lim_n Y_n\,\mathrm{E}(X|\mathcal{D})\,dP \qquad \text{by case 2}$$

$$= \int_D Y\,\mathrm{E}(X|\mathcal{D})\,dP \qquad \text{by the MCT.}$$

For general X, use $X = X^+ - X^-$ and the linearity of (14).
Case 4: General Y. Just write $Y = Y^+ - Y^-$.

To prove (21), simply note that for each $D \in \mathcal{D}$ one has

$$\int_D \mathrm{E}(Y|\mathcal{D})\,dP = \int_D Y\,dP = \int 1_D\,Y\,dP = \mathrm{E}\,(1_D)\,\mathrm{E}(Y)$$

$$= P(D)\,\mathrm{E}(Y) = \int_D \mathrm{E}(Y)\,dP;$$

and apply exercise 3.2.2. Assertion (22) is proved by noting that

$$\int_D \mathrm{E}[\,\mathrm{E}(Y|\mathcal{E})\,|\mathcal{D}]\,dP = \int_D \mathrm{E}[\,Y|\mathcal{E}]\,dP = \int_D Y\,dP \qquad \text{since } D \in \mathcal{D} \subset \mathcal{E}$$

$$= \int_D \mathrm{E}(Y|\mathcal{D})\,dP.$$

The integrands of the two extreme terms must be equal a.s. by the exercise 3.2.2.

Consider (23). Now,

$$\mathcal{F}(X_1, X_2) = \sigma[D \equiv D_1 \cap D_2 \equiv X_1^{-1}(B_1) \cap X_2^{-1}(B_2) : B_1, B_2 \in \mathcal{B}]\,.$$

Let $D = D_1 \cap D_2$ be any one of the generators of $\mathcal{F}(X_1, X_2)$. Then

$$\nu_1(D) \equiv \int_D \mathrm{E}(Y|X_1, X_2)\,dP = \int_D Y\,dP = \int 1_{D_1} 1_{D_2} Y\,dP$$

$$= \int 1_{D_2} \int_{D_1} Y\,dP = \int 1_{D_2}\,dP \int_{D_1} \mathrm{E}(Y|X_1)\,dP = \int 1_{D_2} 1_{D_1}\,\mathrm{E}(Y|X_1)\,dP$$

$$= \int_D \mathrm{E}(Y|X_1)\,dP \equiv \nu_2(D).$$

Since ν_1 and ν_2 are measures on $\mathcal{F}(X_1, X_2)$ that agree on all sets in the $\bar{\pi}$-system consisting of all sets of the form $D = D_1 \cap D_2$, they agree on the σ-field $\mathcal{F}(X_1, X_2)$ by the Dynkin $\pi - \lambda$ theorem. Thus the integrands satisfy $\mathrm{E}(Y|X_1, X_2) = \mathrm{E}(Y|X_1)$ a.s.

We next prove (25), leaving most of (24) and (26) to the exercises. We have

$$\mathrm{E}\,|\mathrm{E}(Y_n|\mathcal{D}) - \mathrm{E}(Y|\mathcal{D})|^r = \mathrm{E}\,|\mathrm{E}(Y_n - Y|\mathcal{D})|^r$$

(i) $\le \mathrm{E}[\,\mathrm{E}(|Y_n - Y|^r|\mathcal{D})\,]$ by the conditional Jensen inequality of (24)

$= \mathrm{E}|Y_n - Y|^r$ by (15)

$\to 0.$

To prove (27), note that for all $D \in \mathcal{D}$ we have

(j) $$\int_D \mathrm{E}(Y|\mathcal{D})\,dP = \int_D Y\,dP = \left\{ \begin{array}{ll} \mathrm{E}(Y) & \text{if } P(D) = 1 \\ 0 & \text{if } P(D) = 0 \end{array} \right\} = \int_D \mathrm{E}(Y)\,dP.$$

(Durrett) We now turn to the Jensen inequality of (24). The result is trivial for linear g. Otherwise, we define

(k) $$C \equiv \{(c,d) : c,d \text{ are rational, and } \ell(x) \equiv cx + d \le g(x) \text{ for all } x \text{ in } I\},$$

and observe that

(l) $$g(x) = \sup_{\text{all } c,d \in C}\,(cx + d) \quad \text{for all } x \in I^o;$$

this follows from the supporting hyperplane result (below (3.4.3)). For any fixed $cx + d$ for which $(c,d) \in C$ with $cx + d \le g(x)$ on all (a,b) we have

(m) $$\mathrm{E}(g(X)|\mathcal{D}) \ge \mathrm{E}(cX + d\,|\mathcal{D}) = c\,\mathrm{E}(X|\mathcal{D}) + d \quad \text{a.s.} \qquad \text{by (16).}$$

Hence (as the union of a countable number of null sets is null), (l) and (m) give

(28) $$\mathrm{E}(g(X)|\mathcal{D}) \ge \sup_{\text{all } c,d \in C}\,\{c\,\mathrm{E}(X|\mathcal{D}) + d\} = g(\mathrm{E}(X|\mathcal{D})) \quad \text{a.s.}$$

since $(\inf I) < \mathrm{E}(X|\mathcal{D}) < (\sup I)$ a.s. (since $\mathrm{E}(X) \in I^0$ was asumed in (3.4.21)). □

Exercise 4.4 Prove (26) and the rest of (24), in theorem 4.1.

Exercise 4.5 (Dispersion inequality) Suppose that X and Y are independent rvs with $\mu_Y = 0$. Let $r \ge 1$. Show that $|X + Y|$ is more dispersed than X in that

(29) $$\mathrm{E}|X|^r \le \mathrm{E}|X + Y|^r \qquad (\text{or, } \mathrm{E}|X + \mu_Y|^r \le \mathrm{E}|X + Y|^r \text{ more generally}).$$

[Hint. Use Fubini on the induced distribution in $(R_2, \mathcal{B}_2)$ and then apply Jensen's inequality to $g_x(y) = |x + y|^r$ to the inner integral. Note also exercise 8.2.3 below.]

Exercise 4.6 (a) Let P denote the Uniform$(-1,1)$ distribution on the Borel subsets $\mathcal{B}$ of $\Omega = [-1,1]$. Let $W(\omega) \equiv |\omega|$, $X(\omega) \equiv \omega^2$, $Y(\omega) \equiv \omega^3$, and $Z(\omega) \equiv \omega^4$. Fix $A \in \mathcal{B}$. Show that versions of various conditional probabilities are given by

$$P(A|W)(\omega) = P(A|X)(\omega) = P(A|Z)(\omega) = \tfrac{1}{2}\,\{1_A(\omega) + 1_A(-\omega)\} \quad \text{on } \Omega,$$

while $P(A|Y)(\omega) = 1_A(\omega)$ on Ω.

Exercise 4.7 Determine $P(A|W)(\omega)$, $P(A|X)(\omega)$, $P(A|Y)(\omega)$, and $P(A|Z)(\omega)$ for W, X, Y, and Z as in exercise 4.6 when P has the density $1-|x|$ on $([-1,1],\mathcal{B})$.

Exercise 4.8 Determine $P(A|W)(\omega)$, $P(A|X)(\omega)$, and $P(A|Y)(\omega)$ for W, X, and Y as in exercise 4.6 when P has density $1-x$ on $(0,1]$ and $\frac{3}{2}\,(1-x^2)$ on $[-1,0]$.

5 Regular Conditional Probability

For fixed A, the function $P(A|\mathcal{D})(\cdot)$ is a $\mathcal{D}$-measurable function on Ω that is only a.s. unique. We wish that for each fixed ω the set function $P(\cdot|\mathcal{D})(\omega)$ were a probability measure. But for each disjoint sequence of A_i's there is a null set where (7.4.10)–(7.4.12) may fail, and there typically are uncountably many such sets. The union of all such null sets need not be null. In the most important special cases, though, we may assume that $P(\cdot|\mathcal{D})(\omega)$ behaves as we would like, where the nonuniqueness of $P(A|\mathcal{D})(\cdot)$ also provides the key, by allowing us to make whatever negligible changes are required. [For added useful generality, we will work on a sub σ-field $\tilde{\mathcal{A}}$ of the basic σ-field $\mathcal{A}$.]

Definition 5.1 We will call $P(A|\mathcal{D})(\omega)$ a *regular conditional probability* on a sub σ-field $\tilde{\mathcal{A}}$ of the σ-field $\mathcal{A}$, given the σ-field $\mathcal{D}$, if

(1) for each fixed $A \in \tilde{\mathcal{A}}$, the function $P(A|\mathcal{D})(\cdot)$ of ω satisfies definition 4.2,

(2) for each fixed ω, $P^{\omega}(\cdot|\mathcal{D}) \equiv P(\cdot|\mathcal{D})(\omega)$ is a probability measure on $\tilde{\mathcal{A}}$.

Exercise 5.1 Verify that the discrete conditional probability of exercise 7.4.1 is a regular conditional probability.

When a regular conditional probability exists, conditional expectation can be computed by integrating with respect to conditional probability, and we first show this general theorem 5.1. In theorem 5.2 and beyond we shall show specifically how to construct such conditional probabilities in some of the most important examples.

Theorem 5.1 Let $P(A|\mathcal{D})(\omega)$ be a regular conditional probability on $\tilde{\mathcal{A}}$, and let $Y \in \mathcal{L}_1(\Omega, \tilde{\mathcal{A}}, P)$. Then a version of the conditional expectation of Y given $\mathcal{D}$ is formed by setting

(3) $$\mathrm{E}\{Y|\mathcal{D}\}(\omega) = \int Y(\omega')\, dP^{\omega}(\omega'|\mathcal{D}), \qquad \text{for each fixed } \omega.$$

Proof. If $Y = 1_A$, then (3) follows from

(a) $\int Y\, dP^{\omega}(\cdot\,|\mathcal{D}) = \int 1_A\, dP^{\omega}(\cdot\,|\mathcal{D}) = \int_A dP^{\omega}(\omega'|\mathcal{D}) = P^{\omega}(A|\mathcal{D}) = P(A|\mathcal{D})(\omega)$

(b) $=_{a.s.} \mathrm{E}\{Y|\mathcal{D}\}(\omega)$, no matter which version of the latter is used,

with the various steps true by definition. Thus (3) is trivial for simple functions Y. If $Y \geq 0$ and Y_n are simple functions for which $Y_n \nearrow Y$, then for any version of the conditional expectation function $\mathrm{E}(Y|\mathcal{D})(\cdot)$ we have

$$\int Y(\omega')\, dP^{\omega}(\omega'|\mathcal{D}) = \lim \int Y_n(\omega')\, dP^{\omega}(\omega'|\mathcal{D})$$

(c) $=_{a.s.} \lim \mathrm{E}\{Y_n|\mathcal{D}\}(\omega) =_{a.s.} \mathrm{E}\{Y|\mathcal{D}\}(\omega)$

using the MCT (ordinary, and of (17)) in the first and last steps. Finally, let $Y = Y^+ - Y^-$. □

Regular conditional probabilities need not exist; the null sets on which things fail may have a nonnull union. However, if $Y : (\Omega, \mathcal{A}) \to (R, \mathcal{B})$ is a rv, then things necessarily work out on $(R, \mathcal{B})$, and this will be generalized to any "Borel space." We will now start from scratch with regular conditional probability, and will choose to regard it as a measure over the image σ-field.

Definition 5.2 (Borel space) If a 1-to-1 bimeasurable mapping ϕ from $(M, \mathcal{G})$ to a measurable subset B_o of $(R, \mathcal{B})$ exists, then $(M, \mathcal{G})$ is called a *Borel space.*

Exercise 5.2 (a) Show that $(R_n, \mathcal{B}_n)$ is a Borel space.
(b) Show that $(R_\infty, \mathcal{B}_\infty)$ is a Borel space.
(c) The spaces $(C, \mathcal{C})$ and $(D, \mathcal{D})$ to be encountered below are also Borel spaces.
(d) Let (M, d) be a complete and separable metric space having Borel sets $\mathcal{M}$, and let $M_0 \in \mathcal{M}$. Then $(M_0, M_0 \cap \mathcal{M})$ is a Borel space.
[This exercise is the only mathematically difficult part of the chapter that we have encountered so far.]

Definition 5.3 (Regular conditional distribution) Suppose that $Z : (\Omega, \mathcal{A}) \to (M, \mathcal{G})$. Let $\tilde{\mathcal{A}} \equiv Z^{-1}(\mathcal{G})$, and let $\mathcal{D}$ be a sub σ-field of $\mathcal{A}$. Then $P_Z(G|\mathcal{D})(\omega)$ will be called a *regular conditional distribution for* Z *given* $\mathcal{D}$ if

(4) for each fixed $G \in \mathcal{G}$,
the function $P_Z(G|\mathcal{D})(\cdot)$ is a version of $P(Z \in G|\mathcal{D})(\cdot)$ on Ω,

(5) for each fixed $\omega \in \Omega$,
the set function $P_Z(\cdot|\mathcal{D})(\omega)$ is a probability distribution on $(M, \mathcal{G})$;

and $P_Z^\omega(\cdot|\mathcal{D}) \equiv P_Z(\cdot|\mathcal{D})(\omega)$ will be used to denote this probability distribution.

Theorem 5.2 (Existence of a regular conditional distribution) Suppose $Z : (\Omega, \mathcal{A}) \to (M, \mathcal{G})$ with $(M, \mathcal{G})$ a Borel space. Then the existence of a regular conditional probability distribution $P_Z^\omega(G|\mathcal{D}) \equiv P_Z(G|\mathcal{D})(\omega)$ is guaranteed.

Proof. Case 1: Suppose first that $Z : (\Omega, \mathcal{A}) \to (R, \mathcal{B})$. Let $r_1, r_2, \ldots$ denote the set of rational numbers. Consider $P(Z \leq r_i|\mathcal{D})$ and note that except on a null set N, all of the following hold:

(a) $r_i \leq r_j$ implies $P(Z \leq r_i|\mathcal{D}) \leq P(Z \leq r_j|\mathcal{D})$.

(b) $\lim_{r_j \searrow r_i} P(Z \leq r_j|\mathcal{D}) = P(Z \leq r_i|\mathcal{D})$.

(c) $\lim_{r_j \nearrow \infty} P(Z \leq r_j|\mathcal{D}) = P(\Omega|\mathcal{D}) = 1$.

(d) $\lim_{r_j \searrow -\infty} P(Z \leq r_j|\mathcal{D}) = P(\emptyset|\mathcal{D}) = 0$.

Now define, for an arbitrary but fixed df F_0,

(e) $$F(z|\mathcal{D})(\omega) = \begin{cases} \lim_{r_j \searrow z} P(Z \le r_j|\mathcal{D})(\omega) & \text{if } \omega \notin N, \\ F_0(z) & \text{if } \omega \in N. \end{cases}$$

Then for every ω, the function $F(\cdot|\mathcal{D})(\omega)$ is a df. Also, (e) and the DCT of theorem 7.4.1 show that $F(z|\mathcal{D})(\cdot)$ is a version of $P(Z \le z|\mathcal{D})(\cdot)$.

Now extend $P(Z \le \cdot|\mathcal{D})(\omega)$ to a distribution [labeled $P_Z(B|\mathcal{D})(\omega)$] over all $B \in \mathcal{B}$ via the correspondence theorem. We now define

(f) $$\mathcal{M} \equiv \{C \in \mathcal{B} : P_Z(C|\mathcal{D}) \text{ is a version of } P(Z \in C|\mathcal{D})\}.$$

Now, $\mathcal{M}$ contains all $(a, b]$ and all $\sum_1^m (a_i, b_i]$, and $\mathcal{M}$ is closed under monotone limits. Thus $\mathcal{M} = \mathcal{B}$, by the minimal monotone class result of proposition 1.1.6, completing the proof in this case.

Case 2: Let $Y \equiv \phi(Z)$, so Y is a rv. Thus a regular conditional distribution $P_Y(B|\mathcal{D})$ exists by case 1. Then for $G \in \mathcal{G}$, define $P_Z(G|\mathcal{D}) \equiv P_Y(\phi^{-1}(G)|\mathcal{D})$. □

Example 5.1 (Elementary conditional densities) Suppose that

$$P((X, Y) \in B_2) = \int_{B_2}\int f(x, y)\, dx\, dy \qquad \text{for all } B_2 \in \mathcal{B}_2,$$

where $f \ge 0$ is measurable; and then $f(x, y)$ is called the *joint density* of X, Y (or, the Radon-Nikodym derivative $dP/d\lambda_2$). Let $B_2 \equiv B \times R$, for all $B \in \mathcal{B}$. We can conclude that $P_X \ll \lambda \equiv$ (Lebesgue measure), with

(6) $$\frac{dP_X(x)}{d\lambda} = f_X(x) \equiv \int_R f(x, y)\, dy\,;$$

we call $f_X(x)$ the *marginal density* of X. We first define

(7) $$g(y|x) \equiv \begin{cases} f(x, y)/f_X(x) & \text{for all } x \text{ with } f_X(x) \ne 0, \\ \text{an arbitrary density } f_0(y) & \text{for all } x \text{ with } f_X(x) = 0, \end{cases}$$

and then define

(8) $$P(Y \in A|X = x) = \int_A g(y|x)\, dy \qquad \text{for } A \in \mathcal{B}.$$

Call $g(y|x)$ the *conditional density* of Y *given* $X = x$, and this $P(Y \in A|X = x)$ will now be shown to be a regular conditional distribution (if modified appropriately on the set where $f_X(x) = 0$). Moreover, if $\mathrm{E}|h(Y)| < \infty$, then

(9) $$\mathrm{E}\{h(Y)|X = x\} = \int_{-\infty}^{\infty} h(y)\, g(y|x)\, dy \qquad \text{a.s. } P_X.$$

Thus (8) (also written as (10)) fulfills theorem 5.2, and (9) will be seen to fulfill theorem 5.3. (Note that this example also holds for vectors $x \in R_m$ and $y \in R_n$.) □

Proof. By Fubini's theorem,

(a) $P(X \in B) = \int\int_{B\times R} f(x,y)\,dx \times dy = \int_B [\int_R f(x,y)\,dy]\,dx = \int_B f_X(x)\,dx.$

Moreover, Fubini's theorem tells us that $f_X(x)$ is $\mathcal{B}$-measurable.

Let $S \equiv \{x : f_X(x) \neq 0\}$. We may assume that $(\Omega, \mathcal{A}) = (R_2, \mathcal{B}_2)$. Let $\tilde{\mathcal{A}} = Y^{-1}(\mathcal{B})$. We will verify that (7.4.6) holds. For $A \in \mathcal{B}$ and for $[Y \in A] \in \tilde{\mathcal{A}} \equiv R \times \mathcal{B}$, we note that for all $B \in \mathcal{B}$ we have (writing $f(x)$ for $f_X(x)$)

$$\int_B [\int_A g(y|x)\,dy]\,dP_X(x)$$

$$= \int_B [\int_A g(y|x)\,dy] f(x)\,dx = \int_{B\cap S} \left[\int_A \frac{f(x,y)}{f(x)}\,dy\right] f(x)\,dx$$

$$= \int_{B\cap S} [\int_A f(x,y)\,dy]\,dx = \int\int_{B\times A} f(x,y)\,(dx \times dy)$$

$$= P((X,Y) \in B \times A) = P([Y \in A] \cap X^{-1}(B))$$

(b) $$= \int_B P(Y \in A|X = x)\,dP_X(x).$$

Thus

(c) $$P(Y \in A|X = x) = \int_A g(y\,|\,x)\,dy \qquad \text{a.s. } P_X,$$

and so $g(y|x)$ works as a version. Now, for any fixed set $A \in \mathcal{B}$ we note that

(10) $$\int_A g(y|x)\,dy = 1_S(x) \times [\int_A f(x,y)\,dy \,/\, f(x)] + 1_{S^c}(x) \times \int_A f_0(y)\,dy$$

is a measurable function on $(R, \mathcal{B})$. It is clear that for each fixed x the function of (10) acts like a probability distribution. Thus (10) defines completely a regular conditional probability distribution.

Suppose that $\mathrm{E}|h(Y)| < \infty$. Then (9) holds since

$$\int_B [\int_R h(y)g(y|x)\,dy]\,dP_X(x) = \int_{B\cap S} [\int_R h(y)g(y|x)\,dy] f(x)\,dx$$

$$= \int_{B\cap S} [\int_R h(y)f(x,y)\,dy]\,dx = \int_B [\int_R h(y)f(x,y)\,dy]\,dx$$

$$= \int\int_{B\times R} h(y)f(x,y)\,(dx \times dy)$$

(d) $$= \int\int_{X^{-1}(B)} h(y)\,dP_{X,Y}(x,y) = \int_B \mathrm{E}(h(Y)|X = x)\,dP_X(x),$$ □

Theorem 5.3 (Conditional expectation exists as an expectation) Given a measurable mapping $Z : (\Omega, \mathcal{A}) \to (M, \mathcal{S})$, where $(M, \mathcal{S})$ is a Borel space, consider a transformation $\phi : (M, \mathcal{S}) \to (R, \mathcal{B})$ with $\mathrm{E}|\phi(Z)| < \infty$. Then a version of the conditional expectation of $\phi(Z)$ given $\mathcal{D}$ is formed by setting

(11) $$\mathrm{E}\{\phi(Z)|\mathcal{D}\}(\omega) = \int_M \phi(z)\,dP_Z^{\omega}(z|\mathcal{D}) \qquad \text{for all } \omega.$$

Proof. Apply theorem 5.1 to the regular conditional distribution of theorem 5.2. □

Theorem 5.4 (A most useful format for conditional expectation) Suppose that $X : (\Omega, \mathcal{A}, P) \to (M_1, \mathcal{G}_1)$ and $Y : (\Omega, \mathcal{A}, P) \to (M_2, \mathcal{G}_2)$ (with Borel space images). Then $(X, Y) : (\Omega, \mathcal{A}, P) \to (M_1 \times M_2, \mathcal{G}_1 \times \mathcal{G}_2)$. Also (as above)

(12) a regular conditional probability $P(A|X = x)$ exists,

for sets $A \in \tilde{A} \equiv Y^{-1}(\mathcal{G}_2) \subset \mathcal{A}$ and for $x \in M_1$. Let $\mathrm{E}\,|h(X, Y)| < \infty$. (a) Then

(13) $\mathrm{E}(h(X, Y)|X = x) = \int_{M_2} h(x, y) dP(y|X = x)$ a.s.

(b) If X and Y are independent, then

(14) $\mathrm{E}(h(X, Y)|X = x) = \mathrm{E}(h(x, Y))$ a.s.

Exercise 5.3 Prove theorem 5.4 above. (Give a separate trivial proof of (14).) Hint. Begin with indicator functions $h = 1_{G_1} 1_{G_2}$.

Example 5.2 (Sufficiency of the order statistics) Let $X_1, \ldots, X_n$ be iid with df F in the class $\mathcal{F}_c$ of all continuous dfs. Let $T(\mathsf{x}) \equiv (x_{n:1}, \ldots, x_{n:n})$ denote the vector of ordered values of x, and let $\mathcal{T} \equiv \{\mathsf{x} : x_{n:1} < \cdots < x_{n:n}\}$. Exercise 5.5 below asks the reader to verify that $P_F(T(\mathsf{X}) \in \mathcal{T}) = 1$ for all $F \in \mathcal{F}_c$. Let $\mathcal{X}$ denote those $\mathsf{x} \in R_n$ having distinct coordinates. Let $\mathcal{A}$ and $\mathcal{B}$ denote all Borel subsets of $\mathcal{X}$ and $\mathcal{T}$, respectively. Then $\mathcal{D} \equiv T^{-1}(\mathcal{B})$ denotes all symmetric subsets of $\mathcal{A}$ (in that $\mathsf{x} \in D \subset \mathcal{D}$ implies $\pi(\mathsf{x}) \in D$ for all $n!$ permutations $\pi(\mathsf{x})$ of x). Let $dP_F^{(n)}(\cdot)$ denote the n-fold product measure $dF(\cdot) \times \cdots \times dF(\cdot)$. Suppose $\phi(\cdot)$ is $P_F^{(n)}$-integrable. Then define

(15) $\phi_0(\mathsf{x}) \equiv \frac{1}{n!} \sum_{\text{all } n! \text{ permutations}} \phi(\pi(\mathsf{x})),$

which is a $\mathcal{D}$-measurable function. Since $P_F^{(n)}$ is symmetric, for any symmetric set $D \in \mathcal{D}$ we have

$$\int_D \phi(\mathsf{x})\, dP_F^{(n)}(\mathsf{x}) = \int_D \phi(\pi(\mathsf{x}))\, dP_F^{(n)}(\mathsf{x}) \qquad \text{for all } n! \text{ permutations } \pi(\cdot)$$

(16) $= \int_D \phi_0(\mathsf{x})\, dP_F^{(n)}(\mathsf{x})$ for every $D \in \mathcal{D}$.

But this means that

(17) $\mathrm{E}_F\{\phi(\mathsf{x})|T\}(\mathsf{x}) = \phi_0(\mathsf{x})$ a.s. $P_F^{(n)}(\cdot)$.

Now, for any $A \in \mathcal{A}$, the function $1_A(\cdot)$ is $P_F^{(n)}$-integrable for all $F \in \mathcal{F}_c$. (Thus conclusion (14) can be applied.) Fix $F \in \mathcal{F}_c$. For any fixed $A \in \mathcal{A}$ we have

$$P_F^{(n)}(A|T)(\mathsf{x}) = \mathrm{E}_F\{1_A(\mathsf{x})|T\}(\mathsf{x}) = \frac{1}{n!} \sum_{\text{all } \pi} 1_A(\pi(\mathsf{x}))$$

(18) $= \dfrac{(\text{the } \# \text{ of times } \pi(\mathsf{x}) \text{ is in } A)}{n!}.$

Note that the right-hand side of (15) does not depend on the particular $F \in \mathcal{F}_c$, and so T is said to be a *sufficient statistic* for the family of distributions $\mathcal{F}_c$. (Note discussion 4.2 once again.) □

Example 5.3 (Ranks) Consider the ranks $\mathsf{R}_n = (R_{n1}, \ldots, R_{nn})'$ and the antiranks $\mathsf{D}_n = (D_{n1}, \ldots, D_{nn})'$ in a sample from some $F \in \mathcal{F}_c$ (see the previous example, and (6.5.17)). Let $\prod_n$ denote the set of all $n!$ permutations of $(1, \ldots, n)$. Now, R_n takes on values in $\prod_n$ when $F \in \mathcal{F}_c$ (since ties occur with probability 0). By symmetry, for every $F \in \mathcal{F}_c$ we have

(19) $$P_F^{(n)}(\mathsf{R}_n = \mathsf{r}) = 1/n! \qquad \text{for each } \mathsf{r} \in \textstyle\prod_n .$$

Note that X is equivalent to (T, R_n), in the sense that each determines the other. Note also that

(20) T and R_n are independent rvs (for each fixed $F \in \mathcal{F}_c$),

in that for all $B \in \mathcal{B} \equiv$ (the Borel subsets of $\mathcal{T}$) and for all $\pi \in \prod_n$ we have

$$P_F^{(n)}([T \in B] \text{ and } [\mathsf{R}_n = \mathsf{r}])$$

$$= \{\tfrac{1}{n!}\} \times \textstyle\int_{T^{-1}(B)} n!\, dP_F^{(n)}(\mathsf{x}) = \{P_F^{(n)}(\mathsf{R}_n = \mathsf{r})\} \times P_F^{(n)}(T \in B)$$

(21) $$= P(\mathsf{R}_n = \mathsf{r}) \times P_F^{(n)}(T \in B),$$

since $P(\mathsf{R}_n = \mathsf{r})$ does not depend on the particular $F \in \mathcal{F}_c$. Since the ranks are independent of the sufficient statistic, they are called *ancillary* rvs. (Note that D_n is also equally likely distributed over $\prod_n$, and that it is distributed independent of the order statistics T.) □

Exercise 5.4 Suppose that the n observations are sampled from a continuous distribution F. Verify that with probability one all observations are distinct. [Hint. Use corollary 2 to theorem 5.1.3.]

Exercise 5.5 Suppose $X_1, \ldots, X_n$ are iid Bernoulli(p) rvs, for some $p \in (0,1)$. Let $T \equiv \sum_1^n X_k$ denote the total number of successes. Show that this rv T is sufficient for this family of probability distributions (that is, T is "sufficient for p").

Exercise 5.6 Let ξ_1 and ξ_2 be independent Uniform$(0,1)$ rvs. Let $\Theta \equiv 2\pi\xi_1$ and $Y \equiv -\log \xi_2$. Let $R \equiv (2Y)^{1/2}$. Now let $Z_1 \equiv R\cos\Theta$ and $Z_2 \equiv R\sin\Theta$. Determine the joint distribution of (Y, Θ) and of (Z_1, Z_2).

Chapter 8

WLLN, SLLN, LIL, and Series

0 Introduction

This is one of the classically important chapters of this text. The first three sections of it are devoted to developing the specific tools we will need. In the second section we also present Khinchin's weak law of large numbers (WLLN), which can be viewed as anticipating both of the classical laws of large numbers (LLNs). Both the classical weak law of large numbers (Feller's WLLN) and classical strong law of large numbers (Kolmogorov's SLLN) are presented in section 8.4, where appropriate negligibility of the summands is also emphasized. This section is the main focus of the chapter. Some applications of these LLNs are given in the following section 8.5. Then we branch out. The law of the iterated logarithm (LIL), the strong Markov property, and convergence of infinite series are treated in sections 8.6 – 8.8. The choice was made to be rather specific in section 8.4, with easy generalizations in section 8.8. The usual choice is to begin more generally, and then specialize. Martingales (mgs) are introduced briefly in section 8.9, both for limited use in chapter 12 and so that the inequalities in the following section 8.10 can be presented in appropriate generality.

1 Borel–Cantelli and Kronecker lemmas

The first three sections will develop the required tools, while applications will begin with the LLNs (the first of which appears in section 8.2). We use the notation

(1) $\quad [A_n \text{ i.o.}] = [\omega : \omega \in A_n \text{ infinitely often}] = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m = \overline{\lim}_n A_n.$

This concept is important in dealing with convergence of various random elements. The following lemmas exhibit a nice dichotomy relative to sequences of independent events.

Lemma 1.1 (Borel–Cantelli lemma) For any events A_n,

(2) $$\sum_{n=1}^{\infty} P(A_n) < \infty \qquad \text{implies} \qquad P(A_n \text{ i.o.}) = 0.$$

Lemma 1.2 (Second Borel–Cantelli lemma) For a sequence of independent events $A_1, A_2, \ldots$, we have the converse that

(3) $$\sum_{n=1}^{\infty} P(A_n) = \infty \qquad \text{implies} \qquad P(A_n \text{ i.o.}) = 1.$$

Thus independent events $A_1, A_2, \ldots$ have $P(A_n$ i.o.$)$ equal to 0 or 1 according as $\sum_1^{\infty} P(A_n)$ is finite or infinite.

Proof. We use proposition 1.1.2 freely. Now,

(a) $\quad P(A_n \text{ i.o.}) = P(\lim_n \bigcup_n^{\infty} A_m) = \lim_n P(\bigcup_n^{\infty} A_m) \le \lim_n \sum_n^{\infty} P(A_m) = 0$

whenever $\sum_1^{\infty} P(A_m) < \infty$. Also,

$$P([\overline{\lim}\, A_n]^c) = P(\bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m^c) = \lim_n P(\bigcap_{m=n}^{\infty} A_m^c)$$

$$= \lim_n \lim_N P(\bigcap_{m=n}^{N} A_m^c)$$

(b) $\quad = \lim_n \lim_N \prod_{m=n}^{N} [1 - P(A_m)] \qquad$ by independence

(c) $\quad \le \lim_n \lim_N \exp(-\sum_{m=n}^{N} P(A_m)) \qquad$ since $1 - x \le \exp(-x)$

(d) $\quad = \lim_n \exp(-\sum_{m=n}^{\infty} P(A_m)) = \lim_n \exp(-\infty) = \lim_n 0 = 0,$

using $\sum_1^{\infty} P(A_n) = \infty$. □

Remark 1.1 (Kolmogorov's 0-1 law) In theorem 7.2.1 we considered the tail σ-field $\mathcal{T} \equiv \cap_{n=1}^{\infty} \mathcal{F}(X_n, X_{n+1}, \ldots)$ of an arbitrary sequence of independent rvs $X_1, X_2, \ldots$. We learned that $P(D) = 0$ or 1 for all $D \in \mathcal{T}$. (Here, let $X_n \equiv 1_{A_n}$ and obtain the characterization via the finiteness of $\sum_1^{\infty} P(A_n)$ at the end of lemma 1.2. The tail event in question is $[X_n = 1 \text{ i.o.}]$.) □

Lemmas About Real Numbers

An important bridge going from the convergence of series to the convergence of averages is provided by Kronecker's lemma. (An alternative bridge is provided by the monotone inequality 8.10.1 (see exercise 8.4.10).)

Lemma 1.3 (Kronecker's lemma) (a) Let $b_n \ge 0$ and $\nearrow \infty$. For $x_1, x_2, \ldots$ real,

(4) $$\sum_{k=1}^{n} x_k \to (\text{some real } r) \qquad \text{implies} \qquad \frac{1}{b_n}\sum_{k=1}^{n} b_k\, x_k \to 0.$$

(b) So, $\sum_{k=1}^n x_k/k \to$ (some real r) implies $\frac{1}{n}\sum_{k=1}^n x_k \to 0$.

Proof. Let $s_k \equiv \sum_1^k x_j$ with $s_0 \equiv 0$ and $b_0 \equiv 0$. Summing by parts gives

(a) $$\tfrac{1}{b_n}\sum_1^n b_k x_k = \tfrac{1}{b_n}\sum_1^n b_k(s_k - s_{k-1}) = \tfrac{1}{b_n}\sum_0^{n-1}(b_k - b_{k+1})\, s_k + \tfrac{1}{b_n}\, b_n s_n$$

(b) $$= -\sum_1^n a_k\, s_{k-1} + s_n \qquad \text{where} \qquad a_k \equiv \frac{b_k - b_{k-1}}{b_n} \ge 0 \ \text{ with } \ \sum_1^n a_k = 1$$

(c) $$= -\sum_1^n a_k\,(s_{k-1} - r) + (s_n - r).$$

Since $|s_k - r| \le \epsilon$ for all $k \ge$ (some N_ϵ), we have

(d) $$\left|\tfrac{1}{b_n}\sum_1^n b_k\, x_k\right| \le \sum_1^{N_\epsilon} |a_k(s_{k-1} - r)| + \sum_{N_\epsilon+1}^n |a_k(s_{k-1} - r)| + |s_n - r|$$

(e) $$\le \frac{\sum_1^{N_\epsilon}(b_k - b_{k-1})|s_{k-1} - r|}{b_n} + \epsilon\,(\textstyle\sum_{N_\epsilon+1}^n a_k) + \epsilon \qquad \text{for } n \ge N_\epsilon$$

$$\le 3\,\epsilon \qquad \text{for } n \text{ sufficiently larger than } N_\epsilon.$$

[Since $\sum_1^n x_k \to r$, we must have $x_k \to 0$. Note that $\sum_1^n b_k x_k/b_n$ puts large weight only on the later terms.] □

Lemma 1.4 (Convergence of sums and products) Suppose $a \in [0,\infty]$, all constants $c_{nk} \ge 0$, and $m_n \equiv [\max_{1\le k\le n} c_{nk}] \to 0$. Then

(5) $$\prod_{k=1}^{n}(1 - c_{nk}) \to e^{-a} \qquad \text{if and only if} \qquad \sum_{k=1}^{n} c_{nk} \to a.$$

Proof. We will write $a = b \oplus c$ to mean that $|a - b| \le c$. For $m_n \le 1/2$,

(6) $$\log\prod_1^n (1-c_{nk}) = \sum_1^n \log(1-c_{nk}) = -\sum_1^n c_{nk} \oplus \sum_1^n c_{nk}^2 = -(1\oplus m_n)\sum_1^n c_{nk}$$

via an expansion of $\log(1+x)$. This yields the result, as $m_n \to 0$. □

Exercise 1.1 (Cesàro summability) If $s_n \equiv \sum_{k=1}^n x_k \to r$, then $\frac{1}{n}\sum_{k=1}^n s_k \to r$.

Exercise 1.2 Let all $a_n \geq 0$. Suppose $\sum_1^\infty a_n b_n < \infty$ holds whenever $\sum_1^\infty b_n^2 < \infty$ with all $b_n \geq 0$. Show that $\sum_1^\infty a_n^2 < \infty$.

Exercise 1.3 (Toeplitz) Let a_{nk} (for $1 \leq k \leq k_n$, with $k_n \to \infty$) be such that: (i) For every fixed k, we have $a_{nk} \to 0$. (ii) $\sum_{k=1}^{k_n} |a_{nk}| \leq c < \infty$, for every n. Let $x_n' \equiv \sum_{k=1}^{k_n} a_{nk} x_k$. Then

(a) $\quad x_n \to 0 \quad$ implies $\quad x_n' \to 0$.

If $\sum_{k=1}^{k_n} a_{nk} \to 1$, then

(b) $\quad x_n \to x \quad$ implies $\quad x_n' \to x$.

In particular, if $b_n \equiv \sum_{k=1}^{n} a_k \nearrow \infty$, then

(c) $\quad x_n \to x$ finite $\quad$ entails $\quad \sum_{k=1}^{n} a_k x_k / b_n \to x$.

[This exercise will not be employed anywhere in this text.]

Exercise 1.4 Show that

(7) $\quad 1 - x \leq e^{-x} \leq 1 - x/(1+x) \quad$ for all $x \geq 0$.

Exercise 1.5 Let $X_1, \ldots, X_n$ be independent rvs. (i) Show that

(8)
$$\begin{aligned} &\textstyle\sum_1^n P(|X_k| > x) / \left[1 + \sum_1^n P(|X_k| > x)\right] \\ &\quad \leq p_{\max}(x) \equiv P(\max_{1 \leq k \leq n} |X_n| > x) \leq \textstyle\sum P(|X_k| > x) \quad \text{for all } x \geq 0. \end{aligned}$$

(ii) So whenever $p_{\max}(x) \equiv P(\max_{1 \leq k \leq n} |X_n| > x) \leq \frac{1}{2}$, then

(9) $\quad \frac{1}{2} \sum_1^n P(|X_k| > x) \leq p_{\max} \leq \sum_1^n P(|X_k| > x)$.

2 Truncation, WLLN, and Review of Inequalities

Truncated rvs necessarily have moments, and this makes them easier to work with. But it is crucial not to lose anything in the truncation.

Definition 2.1 (Khinchin equivalence) Two sequences of rvs $X_1, X_2, \ldots$ and $Y_1, Y_2, \ldots$ for which $\sum_{n=1}^{\infty} P(X_n \neq Y_n) < \infty$ are called *Khinchin equivalent.*

Proposition 2.1 (i) Let $X_1, X_2, \ldots$ and $Y_1, Y_2, \ldots$ be Khinchin equivalent rvs.
(a) If $X_n \to_{a.s.}$ (some rv X), then $Y_n \to_{a.s.}$ (the same rv X).
(b) If $S_n \equiv \sum_1^n X_k \to_{a.s.}$ (some rv S), then $T_n \equiv \sum_1^n Y_k \to_{a.s.}$ (some rv T).
(c) If $S_n/b_n \to_{a.s.}$ (some rv U) and $b_n \to \infty$, then $T_n/b_n \to_{a.s.}$ (the same rv U).
(ii) Of less interest, $\to_p$ may replace $\to_{a.s.}$ in (a), (b), and (c).

Proof. The Borel–Cantelli lemma gives $P(X_n \neq Y_n \text{ i.o.}) = 0$; or

(p) $X_n(\omega) = Y_n(\omega)$ for all $n \geq$ (some $n(\omega)$) holds true for a.e. ω.

Thus the a.s. statements for X_n and S_n are trivial. Moreover, since $X_n(\omega) = Y_n(\omega)$ for all $n \geq$ (some fixed $n(\omega)$), we have

(q) $$\frac{S_n}{b_n} = \frac{S_{n(\omega)} + S_n - S_{n(\omega)}}{b_n} = \frac{S_{n(\omega)} + T_n - T_{n(\omega)}}{b_n} = \frac{S_{n(\omega)} - T_{n(\omega)}}{b_n} + \frac{T_n}{b_n}$$

(r) $$= o(1) + T_n/b_n$$

using $b_n \to \infty$.

Since a sequence (such as X_n, S_n or S_n/b_n) converges in probability if and only if each subsequence n' contains a further subsequence n'' on which the convergence is a.s., the in probability statements follow directly from the a.s. statements. □

Inequality 2.1 (Sandwiching E$|X|$) For any rv X we have

(1) $$\sum_{n=1}^{\infty} P(|X| \geq n) \leq \mathrm{E}|X| = \int_0^{\infty} P(|X| > x)\, dx \leq \sum_{n=0}^{\infty} P(|X| \geq n).$$

If X is a rv with values $0, 1, 2, \ldots$, then

(2) $$\mathrm{E}(X) = \sum_{n=1}^{\infty} P(X \geq n).$$

Proof. If $X \geq 0$, then $\mathrm{E}X = \int_0^{\infty} [1 - F(x)]\, dx$ by (6.4.11); consult figure 2.1. If $X \geq 0$ is integer valued, then

$$\mathrm{E}(X) = \textstyle\sum_{k=0}^{\infty} k\, P(X = k) = \sum_{k=1}^{\infty} \sum_{n=1}^{k} P(X = k)$$

$$= \textstyle\sum_{n=1}^{\infty} \sum_{k=n}^{\infty} P(X = k)$$

(a) $$= \textstyle\sum_{n=1}^{\infty} P(X \geq n).$$

For the greatest integer function $[\cdot]$, an arbitrary rv satisfies

(b) $[|X|] \leq |X| \leq [|X|] + 1$.

Moreover, (a) shows that

(c) $\quad \mathrm{E}[|X|] = \sum_{n=1}^{\infty} P([|X|] \geq n) = \sum_{n=1}^{\infty} P(|X| \geq n),$

while (consult figure 2.1 again)

(d) $\quad \mathrm{E}\{\, [|X|] + 1\,\} = \sum_{n=1}^{\infty} P(|X| \geq n) + 1 = \sum_{n=1}^{\infty} P(|X| \geq n) + P(|X| \geq 0).$ □

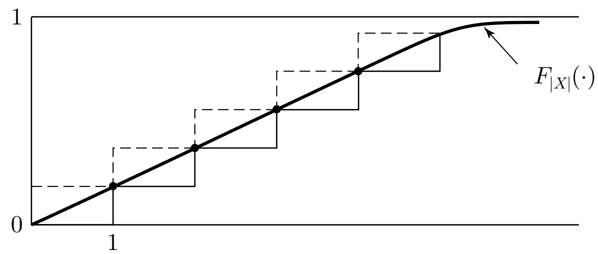


Figure 2.1 The moment $E|X| = \int_0^\infty [1 - F_{|X|}(x)]dx$ is sandwiched.

Example 2.1 (Ťruncating and W̃insorizing) Let $X_1, X_2, \ldots$ be iid as X. Let us truncate and Winsorize the rv X_n by defining

(3) $\quad \check{X}_n = X_n \times 1_{[|X_n|<n]}$ and $\tilde{X}_n = -n \times 1_{[X_n \leq -n]} + X_n \times 1_{[|X_n|<n]} + n \times 1_{[X_n \geq n]}.$

From (2) we see that

(4) $$\begin{array}{llll} \mathrm{E}|X| < \infty & \text{if and only if} & \int_0^\infty P(|X| > t)\, dt < \infty & \text{iff} \\ \sum_1^\infty P(X_n \neq \check{X}_n) < \infty & \text{if and only if} & \sum_1^\infty P(X_n \neq \tilde{X}_n) < \infty\,, & \end{array}$$

so that these $\check{X}_n$'s and $\tilde{X}_n$'s are Khinchin equivalent to the X_n's if and only if the absolute moment $\mathrm{E}|X| < \infty$. (Do not lose sight of this during the SLLN.) □

Proof. Using inequality 2.1, then iid, and then the Borel–Cantelli lemmas, we obtain that $\mathrm{E}|X| < \infty$ if and only if $\sum_1^\infty P(|X| \geq n) < \infty$ if and only if $\sum_1^\infty P(|X_n| \geq n) < \infty$ if and only if $P(|X_n| \geq n \text{ i.o.}) = 0$. This gives (4), as well as the additional fact that

(5) $\quad \mathrm{E}|X| < \infty \quad$ if and only if $\quad P(|X_n| \geq n \text{ i.o.}) = 0 \quad$ for X_n's iid as X.

This final fact (5) is a useful supplementary result. Recall (6.4.11). □

Exercise 2.1 (a) Show $\mathrm{E}X^2 = 2\int_0^\infty xP(|X| > x)\, dx = \sum_{k=1}^\infty (2k-1)P(|X| \geq k)$ for an integer valued rv $X \geq 0$. (Let $\tau(x) \equiv xP(|X| > x)$.) Thus, for any rv X,

(6) $\quad \mathrm{E}X^2 < \infty \quad \text{iff} \quad \int_0^\infty x\, P(|X| > x)\, dx < \infty \quad \text{iff} \quad \sum_{n=1}^\infty nP(|X| \geq n) < \infty.$

(b) Now let $X_1, X_2, \ldots$ be iid as X. Let $\check{X}_n$ or $\tilde{X}_n$ result when X_n is truncated or Winsorized outside either $[-\sqrt{n}, \sqrt{n}\,]$ (or, $(-\sqrt{n}, \sqrt{n}\,)$). Show

(7) $\mathrm{E}X^2 < \infty \quad$ if and only if
these new $\check{X}_n$'s and $\tilde{X}_n$'s are Khinchin equivalent to the X_n's.

Khinchin's WLLN

We begin with an easy result that illustrates our path rather clearly. As we improve our technique, we will be able to improve this result.

Theorem 2.1 (WLLN; Khinchin) Let $X_1, \ldots, X_n$be iid with mean μ. Then

(8) $\quad \bar{X}_n \to_p \mu\,.$

Proof. Truncate via $Y_{nk} \equiv X_k \times 1_{[-n \le X_k \le n]}$, with $\mu_n \equiv \mathrm{E}Y_{nk}$. Note that $\mu_n \to \mu$ by the DCT. Now, for any $\epsilon > 0$, Chebyshev's inequality gives

(a) $\quad P(\,|\bar{Y}_n - \mu_n| \ge \epsilon\,) \le \frac{1}{\epsilon^2\, n^2} \sum_{k=1}^n \mathrm{Var}[Y_{nk}] \le \frac{1}{\epsilon^2\, n^2} \sum_{k=1}^n \mathrm{E}(Y_{nk}^2)$

(b) $\quad = \frac{1}{\epsilon^2\, n^2} \sum_{k=1}^n \mathrm{E}\Big\{ Y_{nk}^2 \Big(1_{[|X_k| \le \epsilon^{3/2}\, \sqrt{n}\,]} \;+\; 1_{[\epsilon^{3/2}\, \sqrt{n} < X_{nk} \le n\,]} \Big) \Big\}$

(c) $\quad \le \frac{1}{\epsilon^2\, n^2} \sum_{k=1}^n \epsilon^3\, n \;+\; \frac{1}{\epsilon^2\, n^2} \sum_{k=1}^n \int_{[\,\epsilon^{3/2}\, \sqrt{n} < |x| \le n\,]} n\,|x|\, dF_X(x)$

(d) $\quad \le \epsilon \;+\; \frac{1}{\epsilon^2} \int_{[|x| > \epsilon^{3/2}\, \sqrt{n}\,]} |x|\, dF_X(x)$

(e) $\quad \le 2\,\epsilon \qquad$ for all $\epsilon \ge$ (some n_ϵ),

since $\mathrm{E}|X_1| < \infty$ implies that the integral in (d) goes to 0 as $n \to \infty$ (for any $\epsilon > 0$). Thus $\bar{Y}_n - \mu_n \to_p 0$. Since $\mu_n \to \mu$, this gives $\bar{Y}_n \to_p \mu$. Thus, by the Khinchin equivalence of example 2.1 and then proposition 2.1(c) we have $\bar{X}_n \to_p \mu$. □

Exercise 2.2 (More general WLLN) Suppose that $X_{n1}, \ldots, X_{nn}$ are independent. Truncate as before via $Y_{nk} \equiv X_{nk} \times 1_{[-n \le X_{nk} \le n]}$, and let $\mu_{nk} \equiv \mathrm{E}(Y_{nk})$.
(i) Set $\bar{\mu}_n \equiv \sum_1^n \mu_{nk}/n$. Show that the same proof as above gives

(a) $\quad \bar{Y}_n - \bar{\mu}_n \to_p 0, \qquad$ even in this non iid case,

provided that

(b) $$\frac{1}{n} \sum_1^n \mathrm{E}\{|X_{nk}|\;\; 1_{[\epsilon\sqrt{n} < |X_{nk}| \le n]}\} \to 0 \qquad \text{for all } \epsilon > 0.$$

(ii) Show that $\bar{X}_n - \mathrm{E}\bar{X}_n \to_p 0$ holds if all $\frac{1}{n}\sum_{k=1}^n \mathrm{E}\{|X_{nk}|\;\; 1_{[\epsilon\sqrt{n} < |X_{nk}|}\} \to 0$, or if the collection of rvs $\{X_{nk} : n \ge 1 \text{ and } 1 \le k \le n\}$ is uniformly integrable.

Remark 2.1 There are two natural ways to proceed to improve Khinchin's WLLN in the iid case. One way is to obtain the conclusion $\bar{X}_n \to_{a.s.} \mu$; and this is done in Kolmogorov's SLLN (theorem 8.4.2 below). Another way is to relax the assumption of a finite mean and center differently; and this is done in Feller's WLLN (theorem 8.4.1 below). [Other possibilities and other approaches will be outlined in the exercises of section 8.4.]

In section 8.3 we will develop a number of inequalities (so called "maximal inequalities") to help us to the stated goal. (At the end of this section the reader could go directly to section 8.4, and then go to section 8.3 for the inequalities as they are needed.) □

Review of General Inequalities from Measure Theory

As we have completed the transition from measure theory to probability theory, we take this opportunity to restate without comment a few of the most important inequalities presented earlier. (See theorem 2.1 for the Khinchin inequality below.)

Inequality 2.2 (Review) Let X and Y be rvs on a probability space $(\Omega, \mathcal{A}, P)$. Then:

(9) C_r-inequality: $\mathrm{E}|X+Y|^r \le C_r\{\mathrm{E}|X|^r + \mathrm{E}|Y|^r\}$ for $r > 0$, $C_r \equiv 2^{(r\vee 1)-1}$.

(10) Hölder: $\mathrm{E}|XY| \le (\mathrm{E}|X|^r)^{1/r}(\mathrm{E}|Y|^s)^{1/s}$ for $r > 1$, and $\frac{1}{r}+\frac{1}{s} = 1$.

(11) Liapunov: $(\mathrm{E}|X|^r)^{1/r}$ is $\nearrow$ in r, for $r \ge 0$.

(12) Markov: $P(|X| \ge \lambda) \le \mathrm{E}|X|^r/\lambda^r$ for all $\lambda > 0$, when $r > 0$.

(13) Dispersion : $\mathrm{E}|X|^r \le \mathrm{E}|X+Y|^r$ if independence, $\mu_Y = 0$, and $r \ge 1$.

(14) Jensen: $g(\mathrm{E}X) \le \mathrm{E}\,g(X)$ if g is convex on some $(a,b) \subset R$ having $P(X \in (a,b)) = 1$, and if $\mathrm{E}X$ is finite.

(15) Littlewood: $m_s^{t-r} \le m_r^{t-s}\, m_t^{s-r}$ for $0 \le r \le s \le t$, with $m_r \equiv \mathrm{E}|X|^r$.

(16) Khinchin: $P(\,|\bar{X}_n - \bar{\mu}_n| \ge \epsilon) \le \epsilon^2 + \frac{1}{\epsilon^2 n}\sum_1^n \int_{[\epsilon^2\sqrt{n}<|X_{nk}|]} |X_{nk}|\,dP$ for independent rvs X_{nk} having finite means μ_{nk}.

(17) Minkowski: $\mathrm{E}^{1/r}|X+Y|^r \le \mathrm{E}^{1/r}|X|^r + \mathrm{E}^{1/r}|Y|^r$ for all $r \ge 1$.

Definition 2.2 ("Big oh$_p$," and "little oh$_p$," $=_a$, $\sim$, and "at most" $\oplus$)
(a) We say that Z_n is *bounded in probability* [and write $Z_n = O_p(1)$] if for all $\epsilon > 0$ there exists a constant M_ϵ for which $P(|Z_n| \ge M_\epsilon) < \epsilon$. For a sequence a_n, we write $Z_n = O_p(a_n)$ if $Z_n/a_n = O_p(1)$; and we say that Z_n is *of order a_n, in probability.*
(b) If $Z_n \to_p 0$, we write $Z_n = o_p(1)$. We write $Z_n = o_p(a_n)$ if $Z_n/a_n \to_p 0$.
(c) This notation (without subscript p) was also used for sequences of real numbers z_n and a_n. For example, $z_n = o(a_n)$ if $z_n/a_n \to 0$. (Note that $o(a_n) = o_p(a_n)$.)
(d) Write $U_n =_a V_n$ if $U_n - V_n \to_p 0$; and call U_n and V_n *asymptotically equal.* (This is effectively a passage to the limit that still allows n to appear on the right-side.)
(e) We write $a_n \sim b_n$ if $a_n/b_n \to 1$.
(f) We write $a = b \oplus c$ if $|a-b| \le c$. (This can be used in the same fashion as $o_p(\cdot)$, but it allows one to keep track of an absolute bound on the difference. Especially, it allows inequalities to be strung together more effectively.)

Exercise 2.3 Let X and Y be independent rvs, and let $r > 0$. Then

(18) $\quad \mathrm{E}|X+Y|^r$ is finite if and only if $\mathrm{E}|X|^r$ and $\mathrm{E}|Y|^r$ are finite.

That is, $(X+Y) \in \mathcal{L}_r$ if and only if both $X \in \mathcal{L}_r$ and $Y \in \mathcal{L}_r$, for any $r > 0$.
Hint. Condition on $Y = y$. Or, note the symmetrization inequality 8.3.2 below.

Exercise 2.4 If $V_n = O_p(1)$ and $\gamma_n = o_p(1)$ are rvs on the same $(\Omega, \mathcal{A}, P)$, then $\gamma_n V_n \to_p 0$.

Exercise 2.5 Let $a_n \geq 0$ be $\nearrow$. Show that for any rv X we have

(19) $\quad \sum_1^\infty n\, P(a_{n-1} \leq |X| < a_n) = \sum_1^\infty P(\,|X| \geq a_{n-1})$.

3 Maximal Inequalities and Symmetrization

Sums of independent random variables play an important role in probability and statistics. Our goal initially in this section is to develop probability bounds for the maximum of the first n partial sums. Such inequalities are called *maximal inequalities*. The most famous of these is Kolmogorov's inequality. For symmetric rvs, Lévy's inequality is an extremely clean and powerful version of such a maximal inequality; it does not require the underlying rvs to have any moments. Neither does the Ottavani–Skorokhod inequality, which is true for arbitrary rvs, though it is not nearly as clean. (Recall (2.3.7) which shows that $S_n \to_{a.s.}$ (some rv S) if and only if $P(\max_{n\le m\le N}|S_m - S_n| \ge \epsilon) \le \epsilon$ for all $N \ge n \ge$ (some n_ϵ).)

Inequality 3.1 (Kolmogorov) Let $X_1, X_2, \ldots$ be independent, with $X_k \cong (0, \sigma_k^2)$. Let $S_k \equiv X_1 + \cdots + X_k$. Then

$$(1) \qquad P\left(\max_{1\le k\le n}|S_k| \ge \lambda\right) \le \mathrm{Var}[S_n]/\lambda^2 = \sum_{k=1}^n \sigma_k^2/\lambda^2 \qquad \text{for all } \lambda > 0.$$

[This contains Chebyshev's inequality that $P(|S_n| \ge \lambda) \le \mathrm{Var}[S_n]/\lambda^2$ for all $\lambda > 0$.]

Proof. Let $A_k \equiv [\max_{1\le j<k}|S_j| < \lambda \le |S_k|]$, so that $A \equiv \sum_1^n A_k = [\max_{1\le k\le n]}|S_k| \ge \lambda]$. Thus k is the first index for which $|S_k|$ exceeds λ; call k the *first passage time*. Then

$$\mathrm{Var}[S_n] = \int S_n^2\,dP \ge \int_A S_n^2\,dP = \sum_1^n \int_{A_k}[(S_n - S_k) + S_k]^2\,dP$$

$$= \sum_1^n \int[(S_n - S_k)^2 1_{A_k} + (S_n - S_k)2S_k 1_{A_k} + S_k^2 1_{A_k}]\,dP$$

$$\text{(a)} \qquad \ge \sum_1^n \left[\int 0\,dP + \mathrm{E}(S_n - S_k)\,\mathrm{E}(2S_k 1_{A_k}) + \int_{A_k} S_k^2\,dP\right] \qquad \text{by independence}$$

$$\text{(b)} \qquad \ge \sum_1^n \left[0 + 0\cdot(\text{a number}) + \int_{A_k}\lambda^2\,dP\right] = \sum_1^n \lambda^2 P(A_k) = \lambda^2 P(A). \qquad \Box$$

Definition 3.1 (Symmetric rvs) A rv X is called *symmetric* if $X \cong -X$. Note that this is equivalent to its df satisfying $F(-x) = 1 - F_-(x)$ for all $x \ge 0$. Suppose $X \cong X'$ are independent rvs; then $X^s \equiv X - X'$ is called the *symmetrization* of the initial rv X.

Definition 3.2 (Medians) Let X be an arbitrary rv. Then $m \equiv$ median(X) is any number for which $P(X \ge m) \ge \frac{1}{2}$ and $P(X \le m) \ge \frac{1}{2}$. [One median of the symmetrization X^s of any rv X is always 0. And (2) below shows that the tails of X^s behave roughly the same as do those of X.]

Inequality 3.2 (Symmetrization inequality) Let $X^s \equiv X - X'$ where $X \cong X'$ with X and X' independent. Let $r > 0$ and let a be any real number. Then both

$$(2) \qquad \begin{array}{l} 2^{-1}P(|X - \mathrm{median}(X)| \ge \lambda) \le P(|X^s| \ge \lambda) \le 2\,P(|X - a| \ge \lambda/2) \quad \text{and} \\ 2^{-1}\,\mathrm{E}|X - \mathrm{median}(X)|^r \le \mathrm{E}|X^s|^r \le 2^{1+r}\,\mathrm{E}|X - a|^r. \end{array}$$

We may replace $\ge$ by $>$ in the three events in the upper half of (2).

Proof. Let $m \equiv \text{median}(X)$. Now, the first inequality comes from

$$P(X^s \geq \lambda) = P[(X - m) - (X' - m) \geq \lambda]$$

(a) $$\geq P(X - m \geq \lambda)\, P(X' - m \leq 0) \;\geq\; P(X - m \geq \lambda)/2.$$

The second inequality holds, since for any real a,

$$P(|X^s| \geq \lambda) = P(|(X - a) - (X' - a)| \geq \lambda)$$

(b) $$\leq P(|X - a| \geq \lambda/2) + P(|X' - a| \geq \lambda/2) = 2P(|X - a| \geq \lambda/2).$$

Plug (2) into (6.4.13) for the moment inequalities. □

Inequality 3.3 (Lévy) Let $X_1, \ldots, X_n$ be independent and symmetric rvs. Let $S_n \equiv X_1 + \cdots + X_n$. Then both

(3) $$P\left(\max_{1\leq k\leq n} |S_k| \geq \lambda\right) \leq 2\,P(|S_n| \geq \lambda) \qquad \text{for all } \lambda > 0 \qquad \text{and}$$

(4) $$P\left(\max_{1\leq k\leq n} |X_k| \geq \lambda\right) \leq 2P(|S_n| \geq \lambda) \qquad \text{for all } \lambda > 0.$$

Thus, $2\,\mathrm{E}|S_n|^r \geq \{\,\mathrm{E}(\max_{1\leq k\leq n} |S_k|^r) \vee \mathrm{E}(\max_{1\leq k\leq n} |X_k|^r)\}$, for each $r > 0$.

Proof. Let $A_k \equiv [\max_{1\leq j<k} S_j < \lambda \leq S_k]$ for $1 \leq k \leq n$, so that k is the smallest index for which S_k exceeds λ. Then

(a) $$P(S_n \geq \lambda) = \sum_{k=1}^n P(A_k \cap [S_n \geq \lambda]) \geq \sum_1^n P(A_k \cap [S_n \geq S_k])$$

(5) $$= \sum_1^n P(A_k) P(S_n - S_k \geq 0)$$

by independence of $X_1, \ldots, X_k$ from $X_{k+1}, \ldots, X_n$

(b) $$\geq \sum_1^n P(A_k)/2 \qquad \text{by symmetry}$$

(c) $$= P(\max_{1\leq k\leq n} S_k \geq \lambda)/2.$$

Combine this with the symmetric result, and achieve the first claim.

Now let $A_k \equiv [\,\max_{1\leq j<k} |X_j| < \lambda \leq |X_k|\,]$ for $1 \leq k \leq n$. Fix k. Let $S_n^o \equiv 2X_k - S_n \cong S_n$, and note that $2\lambda \leq 2|X_k| \leq |S_n| + |S_n^o|$ on A_k. Moreover,

(d) $$P(A_k) \leq P(A_k \cap [|S_n| \geq \lambda]) + P(A_k \cap [|S_n^o| \geq \lambda]) = 2P(A_k \cap [|S_n| \geq \lambda]).$$

So summing on k gives $P(A) \leq 2P(A \cap [|S_n| \geq \lambda]) \leq 2P(|S_n| \geq \lambda)$.

See Feller(1966) proof: Let $M \equiv X_K$, where $K \equiv \min\{k : |X_k| = \max_{1\leq j\leq n} |X_j|\}$. Let $T \equiv S_n - X_K$. Then, for all four choices of $+$ or $-$ signs, the rvs $(\pm M, \pm T)$ have the same distribution. Then we require both

$$P(M \geq \lambda) \leq P(M \geq \lambda \text{ and } T \geq 0) + P(M \geq \lambda \text{ and } T \leq 0)$$

(e) $$= 2P(M \geq \lambda \text{ and } T \geq 0)$$

(f) $$\leq 2P(M + T \geq \lambda) = 2P(S_n \geq \lambda)$$

and the symmetric result. [See exercise 3.2 below for more on (4).] □

Remark 3.1 Kolmogorov's inequality is a moment inequality. Since the rv $S_n/\mathrm{StDev}[S_n] \cong (0,1)$ is often approximately normal $(2\pi)^{-1/2}\exp(-x^2/2)$ on the line, and since

$$P(|N(0,1)| \ge \lambda) = \int_{[|x|\ge\lambda]} \frac{1}{\sqrt{2\pi}}\exp(-x^2/2)\,dx \le \sqrt{\frac{2}{\pi}}\int_\lambda^\infty \frac{x}{\lambda}\exp(-x^2/2)\,dx$$

$$(6) \qquad \le \sqrt{\tfrac{2}{\pi}}\tfrac{1}{\lambda}\exp(-\lambda^2/2) \qquad \text{for all } \lambda > 0,$$

both Lévy's inequality and the Ottaviani–Skorokhod inequality to follow offer the hope of a much better bound. □

Inequality 3.4 Let $S_k \equiv X_1 + \cdots + X_k$ for independent rvs X_k.
(Ottaviani–Skorokhod) For all $0 < c < 1$ we have

$$(7) \qquad P\left(\max_{1\le k\le n}|S_k| \ge \lambda\right) \le \frac{P(|S_n| \ge c\lambda)}{[1-\max_{1\le k\le n}P(|S_n - S_k| > (1-c)\lambda)]} \quad \text{for } \lambda > 0$$

$$\le 2\,P(|S_n| \ge c\lambda) \qquad \text{for all } \lambda \ge \sqrt{2}\,\mathrm{StDev}[S_n]/(1-c).$$

(Etemadi) Alternatively,

$$(8) \qquad P(\max_{1\le k\le n}|S_k| \ge 4\lambda) \le 4\max_{1\le k\le n}P(|S_k| \ge \lambda) \qquad \text{for all } \lambda > 0.$$

Hence, $\mathrm{E}(\max_{1\le k\le n}|S_k|^r) \le 4^{1+r}\max_{1\le k\le n}\mathrm{E}|S_k|^r$ for each $r > 0$. (See (8.2.13).)

Proof. Let $A_k \equiv [S_1 < \lambda, \ldots, S_{k-1} < \lambda, S_k \ge \lambda]$, so that $\sum_{k=1}^n A_k = [\max_{1\le k\le n} S_k \ge \lambda]$. Thus k is the smallest index for which S_k exceeds λ. (This is now the third time we have used this same trick.) Note that

$$\text{(a)} \qquad a \equiv \min_{1\le k\le n}P(|S_n - S_k| \le (1-c)\lambda)$$

$$\text{(b)} \qquad = 1 - \max_{1\le k\le n}P(|S_n - S_k| > (1-c)\lambda)$$

$$\ge 1 - \max_{1\le k\le n}\mathrm{Var}[S_n - S_k]/[(1-c)\lambda]^2 \qquad \text{by Chebyshev's inequality}$$

$$\ge 1 - \mathrm{Var}[S_n]/[(1-c)\lambda]^2$$

$$\text{(c)} \qquad \ge \tfrac{1}{2} \qquad \text{if } \lambda \ge \sqrt{2}\,\mathrm{StDev}[S_n]/(1-c)$$

allows us to "improve" (7) to (8). Meanwhile, (7) comes from

$$\text{(d)} \qquad a \times P(\max_{1\le k\le n} S_k \ge \lambda) \le \sum_{k=1}^n P(|S_n - S_k| \le (1-c)\lambda)P(A_k)$$

$$\text{(e)} \qquad = \sum_{k=1}^n P(A_k \cap [|S_n - S_k| \le (1-c)\lambda]) \qquad \text{by independence}$$

$$\text{(f)} \qquad \le P(S_n \ge c\lambda).$$

Combining (f) and (b) with the analogous result for $-S_n$ completes the proof. □

Exercise 3.1 Prove Etemadi's inequality.

Exercise 3.2 Consider independent rvs $X_k^s \equiv X_k - X_k'$, for $1 \le k \le n$, with all X_k and X_k' independent, and with each $X_k' \cong X_k$. Let m_k denote a median of X_k, and let a denote any real number. Let $\lambda > 0$ and $r > 0$. Show that both:

$$\begin{aligned} &2^{-1} P(\max |X_k - m_k| \ge \lambda) \le P(\max |X_k^s| \ge \lambda) \le 2P(\max |X_k - a| \ge \lambda/2). \\ &2^{-1}\,\mathrm{E}(\max |X_k - m_k|^r) \le \mathrm{E}(\max |X_k^s|^r) \le 2^{1+r}\,\mathrm{E}(\max |X_k - a|^r). \end{aligned} \tag{9}$$

[Also, $2^{-1} P(\max |X_k^s| \ge \lambda) \le P(|S_n^s| \ge \lambda) \le 2P(|S_n - a| \ge \lambda/2)$ (for any real a), by inequality 3.3 and inequality 3.2.]

Inequalities for Rademacher RVs

Inequality 3.5 (Symmetrization; Giné–Zinn) Let $X_1, \ldots, X_n$ be iid rvs, and let $\epsilon_1, \ldots, \epsilon_n$ denote an independent sample of iid Radamacher rvs (that satisfy $P(\epsilon_k = \pm 1) = \frac{1}{2}$). Then

$$P\left(\frac{1}{\sqrt{n}}\left|\sum_{k=1}^n \epsilon_k X_k\right| > 2\lambda\right) \le \sup_{1\le m\le n} 2\,P\left(\left|\frac{1}{\sqrt{m}}\sum_{k=1}^m X_k\right| > \lambda\right) \quad \text{for all } \lambda > 0\,. \tag{10}$$

Proof. By conditioning on the Rademacher rvs we obtain

(a) $\quad P(n^{-1/2}\,|\sum_1^n \epsilon_k X_k| > 2\,\lambda)$

$\le P(n^{-1/2}\,|\sum_{k:\epsilon_k=1} \epsilon_k X_k| > \lambda) + P(n^{-1/2}\,|\sum_{k:\epsilon_k=-1} \epsilon_k X_k| > \lambda)$

$\le \mathrm{E}_\epsilon\, P(n^{-1/2}\,|\sum_{k:\epsilon_k=1} X_k| > \lambda) + \mathrm{E}_\epsilon\, P(n^{-1/2}\,|\sum_{k:\epsilon_k=-1} X_k| > \lambda)$

(b) $\quad \le 2\,\sup_{m\le n} P(n^{-1/2}\,|\sum_1^m X_k| > \lambda)$

$\le 2\,\sup_{m\le n} P(m^{-1/2}\,|\sum_1^m X_k| > \lambda)$

(c) $\quad \le 2\,\sup_{m\ge 1} P(m^{-1/2}\,|\sum_1^m X_k| > \lambda)$, as required. □

Exercise 3.3 (a) (Khinchin inequality) Suppose $\epsilon_1, \ldots, \epsilon_n$ are iid Rademacher rvs. Let $a_1, \ldots, a_n$ be real constants. Then

$$A_r\left(\textstyle\sum_1^n a_k^2\right)^{1/2} \le \left(\mathrm{E}\,|\textstyle\sum_1^n a_k \epsilon_k|^r\right)^{1/r} \le B_r\left(\textstyle\sum_1^n a_k^2\right)^{1/2}, \qquad \text{for each } r \ge 1, \tag{11}$$

for some constants A_r and B_r. Establish this for $r = 1$, with $A_1 = 1/\sqrt{3}$ and $B_1 = 1$. [Hint. Use Littlewood's inequality with r, s, t equal to $1, 2, 4$.]

(b) (Marcinkiewicz–Zygmund inequality) For $X_1, \ldots, X_n$ independent 0 mean rvs,

$$(\tfrac{1}{2}A_r)^r\,\mathrm{E}(\textstyle\sum_1^n X_k^2)^{r/2} \le \mathrm{E}|\textstyle\sum_1^n X_k|^r \le (2B_r)^r\,\mathrm{E}(\textstyle\sum_1^n X_k^2)^{r/2}, \quad \text{for each } r \ge 1. \tag{12}$$

Exercise 3.4 Let $X_1, \ldots, X_n$ be independent with 0 means, and independent of the iid Rademacher rvs $\epsilon_1, \ldots, \epsilon_n$. Let ϕ be $\nearrow$ and convex on R. Then

$$\mathrm{E}\phi(|\textstyle\sum_1^n \epsilon_k X_k|/2) \le \mathrm{E}\phi(|\textstyle\sum_1^n X_k|) \le \mathrm{E}\phi(2|\textstyle\sum_1^n \epsilon_k X_k|)\,.$$

[Hint. The left side is an average of terms like $\mathrm{E}\phi(|\sum_1^n e_k(X_k - \mathrm{E}X_k')|/2)$, for independent $X_k' \cong X_k$ and with each e_k equal to ± 1.]

Weak Negligibility, or Maximal Inequalities of Another Ilk

Discussion 3.1 (Weak negligibility) Let $Y_{n1}, \ldots, Y_{nn}$ be independent with dfs $F_{n1}, \ldots, F_{nn}$. Let $\theta > 0$ be given. For any $\epsilon > 0$, let $p_{nk}^{\epsilon} \equiv P(|Y_{nk}| > \epsilon)$. Now, the maximum $\text{Max}_n \equiv [\max_{1 \le k \le n} |Y_{nk}|]$ satisfies

(13) $$1 - \exp(-\textstyle\sum_1^n p_{nk}^{\epsilon}) \le 1 - \prod_1^n (1 - p_{nk}^{\epsilon}) = P(\text{Max}_n > \epsilon) \le \sum_1^n p_{nk}^{\epsilon}\,.$$

[The equality uses $\cap_1^n A_k^c = [\cup_1^n A_k]^c$, and the first bound follows from the inequality $1 - x \le \exp(-x)$.] This gives (so does exercise 8.1.5) the standard result that

(14) $$\text{Max}_n \to_p 0 \qquad \text{if and only if} \qquad n\bar{P}_n(\epsilon) \equiv \textstyle\sum_1^n p_{nk}^{\epsilon} \to 0 \quad \text{for all } \epsilon > 0.$$

Define $x_{\theta n}$ by requiring $[-x_{\theta n}, x_{\theta n}]$ to be the smallest interval that is both closed and symmetric to which $\bar{F}_n \equiv \sum_1^n F_{nk}/n$ assigns probability at least $1 - \theta/n$. Let $\bar{P}_n(x) \equiv \frac{1}{n} \sum_1^n P(|Y_{nk}| > x)$ denote the average tail probability, and then let $\overline{K}_n$ denote the qf of the df $1 - \bar{P}_n(\cdot)$. Note the quantile relationship $x_{\theta n} = \overline{K}_n(1-\theta/n)$. Since $\overline{K}_n(1-\theta/n) = \inf\{x : 1 - \bar{P}_n(x) \ge 1 - \theta/n\} = \inf\{x : \bar{P}_n(x) \le \theta/n\}$, we have

(15) $$n\bar{P}_n(\epsilon) \le \theta \qquad \text{if and only if} \qquad \overline{K}_n(1 - \theta/n) \le \epsilon\,.$$

Fix $0 < \epsilon \le 1$ and $0 < \theta \le 1$, and suppose that we are considering all n exceeding some $n_{\epsilon,\theta}$. Conclusions (14) and (15) give (the seemingly new emphasis)

(16) $$\text{Max}_n \to_p 0 \quad \text{if and only if} \quad x_{\theta n} = \overline{K}_n(1 - \theta/n) \to 0 \;\text{ for all } 0 < \theta \le 1. \square$$

Discussion 3.2 (Weak negligibility in the LLN context) Let $\nu_n > 0$ be constants. Applying the previous paragraph to the rvs $|Y_{nk}|/n\,\nu_n$ (whose average df has the $(1-\theta/n)$th quantile $x_{\theta n}/n\,\nu_n$) gives the equivalencies

(17) $$M_n/\nu_n \equiv [\max_k \tfrac{1}{n}\,|Y_{nk}|\,]/\,\nu_n \to_p 0\,,$$

(18) $$x_{\theta n}/n\,\nu_n \to 0 \qquad\qquad \text{for all } 0 < \theta \le 1\,,$$

(19) $$\textstyle\sum_1^n P(|Y_{nk}|/n\,\nu_n > \epsilon) \to 0 \qquad \text{for all } 0 < \epsilon \le 1\,.$$

Useful choices for ν_n are the *truncated absolute moment* $u_{1n} \equiv \int_{[\,|y| \le x_{1n}]} |y|\, d\bar{F}_n(y)$ and the *Winsorized absolute moment* $\tilde{u}_{1n} \equiv u_{1n} + x_{1n}\bar{P}_n(x_{1n})$. (Here x_{1n} means the quantile $x_{\theta n}$ with $\theta = 1$.) $\square$

Inequality 3.6 (Daniels' equality) With high probability there is an upper linear bound on the uniform empirical df $\mathbb{G}_n$. That is, for each $0 < \lambda < 1$,

(20) $$P(\mathbb{G}_n(t) \le t/\lambda \text{ for all } 0 \le t \le 1) = P(\xi_{n:k} \ge \lambda k/n \text{ for } 1 \le k \le n) = 1 - \lambda\,.$$

Proof. (Robbins) The vector of Uniform(0,1) order statistics $(\xi_{n:1}, \ldots \xi_{n:n})$ has joint density $n!$ on its domain $0 < t_1 < \cdots < t_n < 1$. Thus

(a) $$P(\mathbb{G}_n(t) \le t/\lambda \text{ for } 0 \le t \le 1) = P(\xi_{n:k} \ge \lambda k/n \text{ for } 1 \le k \le n)$$

(b) $$= \int_\lambda^1 \int_{\lambda(n-1)/n}^{t_n} \cdots \int_{\lambda 2/n}^{t_3} \int_{\lambda/n}^{t_2} n!\; dt_1 \cdots dt_n = \cdots =$$

(c) $$= n! \left[\frac{t^n}{n!} - \frac{\lambda t^{n-1}}{n!}\right]\Bigg|_\lambda^1 = 1 - \lambda\,. \qquad \square$$

Inequality 3.7 (Chang's inequality) With high probability there is a lower linear bound on the uniform empirical df $\mathbb{G}_n$. That is,

$$
\begin{aligned}
&P(\|I/\mathbb{G}_n^{-1}\|_{\xi_{n:1}}^{1} \le \lambda) = P(\mathbb{G}_n(t) \ge t/\lambda \text{ on all of } [\xi_{n:1}, 1]\,) \\
&\quad \ge 1 - 2\,\lambda^2\, e^{-\lambda} \quad \text{for all } \lambda \ge 1.
\end{aligned} \tag{21}
$$

(This provides a nice symmetrty with the previous inequality, though it will not be proven until chapter 12. It is stated now in the spirit of symmetry, completeness, and fun.)

4 The Classical Laws of Large Numbers, LLNs

It is now time to present versions of the laws of large numbers under minimal hypotheses. The *weak* law of large numbers (WLLN) will establish $\to_p$, while the *strong* law of large numbers (SLLN) will establish $\to_{a.e.}$ of a sample average $\bar{X}_n$.

Theorem 4.1 (WLLN; Feller) Let $X_{n1}, \ldots, X_{nn}$ be iid with df F and qf K, for each n. Let $\bar{X}_n \equiv (X_{n1} + \cdots + X_{nn})/n$. The following are equivalent:

(1) $\bar{X}_n - \mu_n \to_p 0$ for some choice of constants μ_n .

(2) $\tau(x) \equiv xP(|X| > x) \to 0$ (true, iff $\tau^{\pm}(x) \equiv xP(X^{\pm} > x) \to 0$).

(3) $t\,\{|F_+^{-1}(t)| + |F^{-1}(1-t)|\} \to 0$ (true, iff $t\,F_{|X|}^{-1}(1-t) \to 0$).

(4) $M_n \equiv [\,\frac{1}{n} \max_{1\le k\le n} |X_{nk}|\,] \to_p 0.$

When (1) holds, possible choices include $\mu_n \equiv \int_{[-n,n]} x\,dF(x)$, $\nu_n \equiv \int_{1/n}^{1-1/n} K(t)\,dt$, and $m_n \equiv \text{median}(\bar{X}_n)$. If $E|X| < \infty$, then (1) holds with $\mu_n = \mu \equiv \mathrm{E}X$.

Theorem 4.2 (SLLN; Kolmogorov) Let $X, X_1, X_2, \ldots$ be iid rvs. Then:

(5) $\mathrm{E}|X| < \infty$ implies $\bar{X}_n \to_{a.s.} \mu \equiv \mathrm{E}X$.

(6) $\mathrm{E}|X| = \infty$ implies $\overline{\lim}\,|\bar{X}_n| = \infty$ a.s.

(7) $\mathrm{E}|X| < \infty$ iff $\overline{\lim}\,|\bar{X}_n| < \infty$ a.s. iff $\bar{X}_n \to_{\mathcal{L}_1}$ (some rv).

(8) $\mathrm{E}|X| < \infty$ iff $M_n \equiv [\,\frac{1}{n} \max_{1\le k\le n} |X_k|\,] \to_{a.s.} 0$ iff $\frac{X_n}{n} \to_{a.s.} 0$ iff $M_n \to_{\mathcal{L}_1} 0$.

Conditions (4) and (8) show the sense in which these LLNs are tied to the size of the maximal summand. This is an important theme, do not lose sight of it. We now give a symmetric version of condition (4). (See also exercises 4.18– 4.21.)

Theorem 4.3 (The maximal summand) Let $X_{n1}, \ldots, X_{nn}, n \ge 1$, be iid row independent rvs with df F. We then let $X_{nk}^s \equiv X_{nk} - X_{nk}'$ denote their symmetrized versions. Fix $r > 0$. [Most important is $r = 1$.] Then

(9) $\tau(x) \equiv x^r P(|X| > x) \to 0$ iff $\tau^s(x) \equiv x^r P(|X^s| > x) \to 0$ iff

(10) $[\,\max_{1\le k\le n} \frac{1}{n}\,|X_{nk} - a|^r\,] \to_p 0$ for all/some a iff $[\,\max_{1\le k\le n} \frac{1}{n}\,|X_{nk}^s|^r\,] \to_p 0$.

Proof. We first consider the SLLN. Let $Y_n \equiv X_n \times 1_{[\,|X_n|<n]}$, for the iid X_n's.

Suppose $E|X| < \infty$. Using inequality 8.2.1 in the second step and iid in the third, we obtain

(a) $\mu = \mathrm{E}X$ is finite iff $\mathrm{E}|X| < \infty$ iff $\sum_{n=1}^{\infty} P(|X| \ge n) < \infty$

iff $\sum_{n=1}^{\infty} P(|X_n| \ge n) < \infty$ iff $\sum_{n=1}^{\infty} P(Y_n \ne X_n) < \infty$

(b) iff the X_n's and Y_n's are Khinchin equivalent rvs.

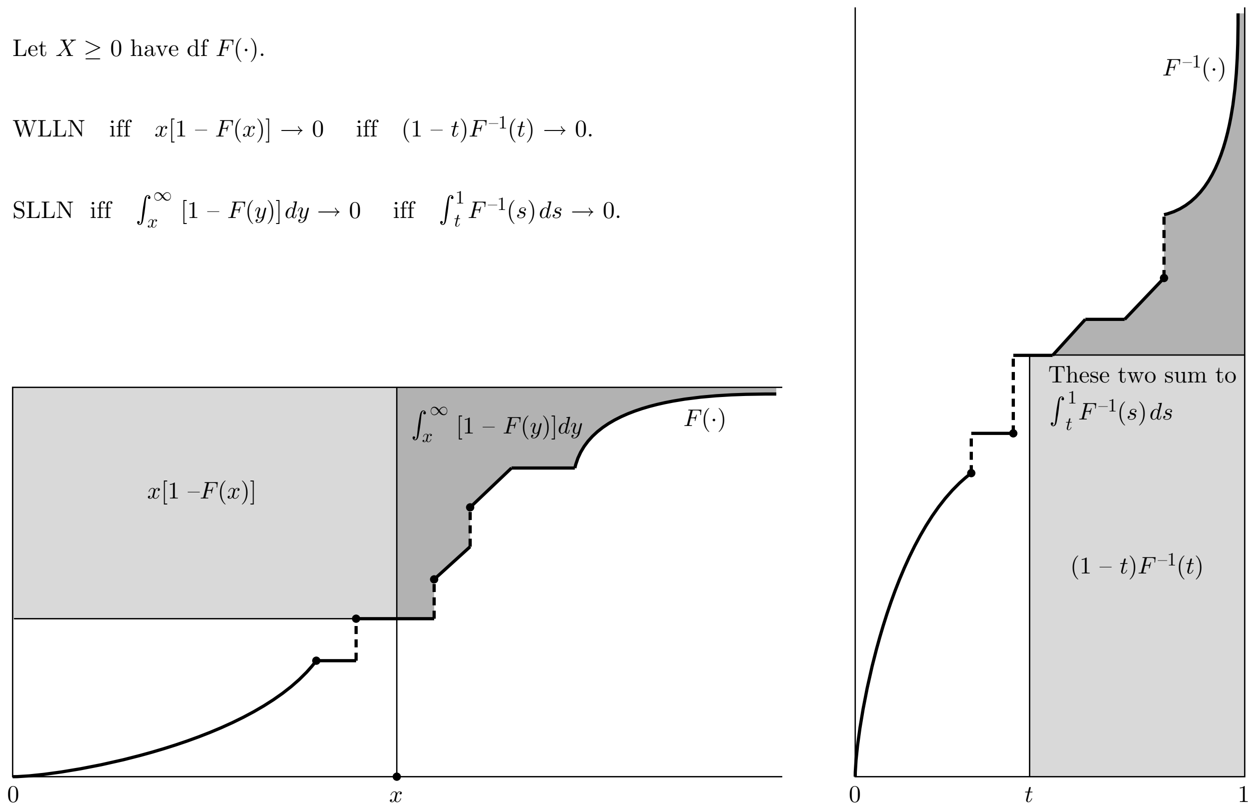


Figure 4.1 Conditions for the WLLN and for the SLLN. (Use $(a \vee b) \leq a+b \leq 2(a \vee b)$ for general X.)

Comment Recall that $\mathrm{E}|X| = \int_0^\infty P(|X| > y)\,dy$ and (see (6.3.12))

(*i*) $\int_{(x,\infty)} y\,dF_{|X|}(y) = xP(|X| > x) + \int_x^\infty P(|X| > y)\,dy$;

(*ii*) $\int_{[0,x]} y\,dF_{|X|}(y) = \int_0^x P(|X| > y)\,dy - xP(|X| > x)$.

so

Thus $\bar{X}_n \to_{a.s.} \mu$ wherever $\bar{Y}_n \to_{a.s.} \mu$, by proposition 8.2.1(c). So we now make the definitions $\mu_n \equiv \mathrm{E}Y_n$ and $\sigma_n^2 \equiv \mathrm{Var}[Y_n]$. Then we can write

(c) $\quad \bar{Y}_n = [\sum_1^n (Y_i - \mu_i)/n] + \bar{\mu}_n, \qquad$ where $\bar{\mu}_n \equiv (\mu_1 + \cdots + \mu_n)/n \to \mu$

by the Cesàro summability exercise 8.1.1 (since $\mu_n = \mathrm{E}Y_n = \int_{(-n,n)} x\,dF(x) \to \mu$ by the DCT with dominating function given by $|x|$). It thus suffices to show that $\sum_1^n (Y_i - \mu_i)/n \to_{a.s.} 0$. By Kronecker's lemma it thus suffices to show that

(d) $\quad Z_n \equiv \sum_{i=1}^n (Y_i - \mu_i)/i \to_{a.s.}$ (some rv Z).

But $Z_n \to_{a.s.}$ (some Z), by proposition 2.3.3, if for all $\epsilon > 0$ we have

$$p_{nN}^{\epsilon} \equiv P(\max_{n \le m \le N} |Z_m - Z_n| \ge \epsilon)$$

(e) $$\quad = P(\max_{n \le m \le N} |\textstyle\sum_{i=n+1}^m [(Y_i - \mu_i)/i]\,| \ge \epsilon) \to 0.$$

Kolmogorov's inequality yields (e) via

$$p_{nN}^{\epsilon} \le \epsilon^{-2} \textstyle\sum_{n+1}^N \mathrm{Var}[(Y_i - \mu_i)/i] = \epsilon^{-2} \sum_{n+1}^N \sigma_i^2/i^2$$

(f) $$\quad \le \epsilon^{-2} \textstyle\sum_{n+1}^\infty \sigma_i^2/i^2 \qquad \text{for all } N$$

$$\to 0 \qquad \text{as } n \to \infty,$$

provided that

(g) $\quad \sum_{n=1}^\infty \sigma_n^2/n^2 = \sum_{n=1}^\infty \mathrm{Var}[Y_n - \mu_n]/n^2 < \infty.$

Now, this last is seen to be true via the following Kolmogorov argument that

$$\textstyle\sum_1^\infty \mathrm{Var}[Y_n - \mu_n]/n^2 \le \sum_1^\infty \mathrm{E}Y_n^2/n^2 = \sum_1^\infty \int_{[|x|<n]} x^2\,dF(x)/n^2$$

$$= \textstyle\sum_1^\infty \sum_{k=1}^n \int_{[k-1 \le |x| < k]} x^2\,dF(x)/n^2$$

(h) $$\quad = \textstyle\sum_{k=1}^\infty \sum_{n=k}^\infty \frac{1}{n^2} \int_{[k-1 \le |x| < k]} x^2\,dF(x)$$

(i) $$\quad \le \textstyle\sum_{k=1}^\infty 2 \int_{[k-1 \le |x| < k]} x^2\,dF(x)/k$$

$$\text{since} \quad \textstyle\sum_{n=k}^\infty 1/n^2 \le \int_k^\infty (2/x^2)\,dx = 2/k$$

(j) $$\quad \le 2 \textstyle\sum_{k=1}^\infty \int_{[k-1 \le |x| < k]} |x|\,dF(x)$$

(k) $$\quad = 2\,\mathrm{E}|X| < \infty.$$

Thus we do have $Z_n \to_{a.s.}$ (some rv Z), and so $\bar{X}_n \to_{a.s.} \mu$.

Suppose $\mathrm{E}|X| = \infty$. Then the "sandwich" inequality 8.2.1 gives

(l) $\quad \sum_{n=0}^\infty P(|X_n| \ge nC) = \sum_{n=0}^\infty P(|X|/C \ge n) \ge \mathrm{E}|X/C| = \infty$ for all $C > 0$,

so that applying the second Borel–Cantelli lemma to (l) gives

(m) $\quad P(|X_n| \ge nC \text{ i.o.}) = 1 \qquad$ for all $C > 0 \quad$ (and hence for all large $C > 0$).

Since $S_n = S_{n-1} + X_n$, (m) implies (using the fact that $|S_n| < nC/2$ and $|X_n| \ge nC$ yields $|S_{n-1}| > nC/2 > (n-1)C/2$) that

(n) $\quad P(|S_n| \ge nC/2 \text{ i.o.}) = 1 \qquad$ for all $C > 0$.

That is, $\overline{\lim}\,|S_n|/n \ge C/2$ a.s. for all C. That is, $\overline{\lim}\,|S_n|/n = \infty$ a.s.

Thus (5)–(7) hold. [Apply Vitali with exercise 4.16 below for $\bar{X}_n \to_{\mathcal{L}_1} \mu$ in (7). If $\bar{X}_n \to_{\mathcal{L}_1}$ (some rv W), then $\bar{X}_n \to_p W$ and the averages $\bar{X}_n$ are u.i. by Vitali. Exercise 4.16 below then shows that EX is finite.]

Consider (8). Suppose $M_n \to_{a.s.} 0$. Then a.s. for all $n \geq$ (some n_ω) we have

(o) $\quad [\max_{1\le k\le n}|X_k|]/n < \epsilon, \quad$ and hence $\quad |X_n|/n < \epsilon.$

We merely repeat this last statement, writing

(p) $\quad A_n \equiv [\,|X_n|/n \ge \epsilon] \quad$ satisfies $\quad P(A_n \text{ i.o.}) = P(|X_n|/n \ge \epsilon \text{ i.o.}) = 0.$

Thus inequality 8.2.1 (by applying iid, and then the second Borel–Cantelli) gives

(q) $\quad \mathrm{E}|X|/\epsilon = \mathrm{E}|X/\epsilon| \le \sum_{n=0}^\infty P(|X/\epsilon| \ge n) = \sum_{n=0}^\infty P(|X_n|/n \ge \epsilon) < \infty.$

Conversely, suppose $\mathrm{E}|X| < \infty$. Then $S_n/n \to_{a.s.} \mu$ by the SLLN. Since

(r) $$\frac{X_n}{n} = \frac{S_n - n\mu}{n} - \frac{n-1}{n}\left[\frac{S_{n-1}-(n-1)\mu}{n-1}\right] + \frac{\mu}{n} \to_{a.s.} 0 - 1\cdot 0 + 0 = 0,$$

we have a.s. that

(s) $\quad |X_n|/n \le \epsilon \qquad$ for all $n \ge$ (some n_ω).

Thus for all n exceeding some even larger n'_ω we have

(11) $$\left[\max_{1\le k\le n}\frac{|X_k|}{n}\right] = \left[\max_{1\le k\le n}\frac{k}{n}\cdot\frac{|X_k|}{k}\right] \le \left[\max_{1\le k\le n_\omega}\frac{|X_k|}{n}\right] \vee \left[\max_{k\ge n_\omega}\left|\frac{|X_k|}{k}\right|\right]$$

(t) $\qquad \le n^{-1}$[a fixed number depending on ω] $+ \epsilon \le 2\epsilon \qquad$ using (s),

where we will have to increase the specification on n'_ω for (t). Thus $M_n \to_{a.s.} 0$.

Finally, note exercise 4.17 for $M_n \to_{\mathcal{L}_1} 0$ if and only if $\mathrm{E}|X| < \infty$.

From (6.4.11) we see (note figure 4.1) that

(12) $\quad \mathrm{E}|X| < \infty \quad$ iff $\quad \int_0^\infty P(|X| > x)\,dx < \infty \quad$ iff $\quad \int_0^1 |F^{-1}(t)|\,dt < \infty\,. \quad \square$

Remark 4.1 Suppose $X_1, \ldots, X_n$ are independent, with $X_i \cong (0, \sigma_i^2)$. Then

$$S_n \equiv X_1 + \cdots + X_n \cong (0, \textstyle\sum_1^n \sigma_i^2), \quad \text{while} \quad \bar{X}_n \equiv S_n/n \cong (0, \textstyle\sum_1^n \sigma_i^2/n^2).$$

Chebyshev's inequality and Kolmogorov's inequality give, respectively,

(13)
$(a)\ \ P(|S_n| \ge \lambda) \le \mathrm{Var}[S_n]/\lambda^2 = \sum_{i=1}^n \sigma_i^2/\lambda^2 \qquad$ for all $\lambda > 0$,

$(b)\ \ P(\max_{1\le k\le n}|S_k| \ge \lambda) \le \mathrm{Var}[S_n]/\lambda^2 \qquad$ for all $\lambda > 0$.

For $X_1, X_2, \ldots$ iid (μ, σ^2), the inequality (13)(a) gives $\bar{X}_n \to_p \mu$, by Chebyshev's inequality. But the WLLN conclusion $\bar{X}_n \to_p \mu$ should not require the variance σ^2 to be finite, as this cheap proof based on (13)(a) requires. Indeed, Khintchine's WLLN of theorem 8.2.1 didn't. Exercise 4.8 below outlines one very cheap proof of the SLLN using "only" the Borel–Cantelli lemma, and exercise 4.9 outlines a slightly improved version that also uses Kolmogorov's inequality. Kolmogorov's proof of the full SLLN made the key step of incorporating truncation. Exercise 4.10 describes an elementary way to avoid use of Kronecker's lemma. $\square$

Proof. Consider the WLLN. Suppose (2) holds. Define $Y_{nk} \equiv X_{nk} \times 1_{[|X_{nk}| \le n]}$ and $\mu_n \equiv \mathrm{E}Y_{nk} = \int_{[-n,n]} x\,dF(x)$. Now (using integration by parts for (b)),

$$P(|\bar{X}_n - \mu_n| \ge \epsilon\,) = P(|\bar{X}_n - \mathrm{E}\bar{Y}_n| \ge \epsilon\,)$$

(14) $$\le P(|\bar{Y}_n - \mathrm{E}\bar{Y}_n| \ge \epsilon\,) + \textstyle\sum_1^n P(Y_{nk} \ne X_k)$$

(15) $$\le \tfrac{1}{\epsilon^2}\mathrm{Var}[\bar{Y}_n] + nP(|X| > n) \le \tfrac{1}{n\epsilon^2}\mathrm{E}\,Y_{n1}^2 + \tau(n) \textbf{ (Truncation inequality)}$$

(a) $$= -\tfrac{1}{n\epsilon^2}\int_{[0,n]} x^2\,dP(|X|>x) + \tau(n)$$

(b) $$= \tfrac{1}{\epsilon^2}\left[-\tfrac{1}{n}\,x^2 P(|X|>x)\Big|_{0-}^{n+} + \tfrac{1}{n}\int_{[0,n]} P(|X|>x)\,2x\,dx\right] + \tau(n)$$

$$= \tfrac{1}{\epsilon^2}\left[-\tau(n) + 0 + 2\,n^{-1}\int_{[0,n]}\tau(x)\,dx\right] + \tau(n)$$

(16) $$\le \frac{2}{n\epsilon^2}\int_{[0,n]}\tau(x)\,dx = \frac{2}{n\epsilon^2}\int_{[0,n]} x\,P(|X|>x)\,dx \quad \text{(compare (6.4.18))}$$

for $0 < \epsilon \le 1$. Note that $\tau(x) \le x$, and choose $M > 0$ so large that $\tau(x) < \epsilon^3/4$ for $x > M$. Applying these to (16) gives

(c) $$\tfrac{2}{n\epsilon^2}\int_0^n \tau(x)\,dx \le \tfrac{2}{n\epsilon^2}\{\int_0^M x\,dx + \int_M^n \tfrac{\epsilon^3}{4}\,dx\} \le \tfrac{M^2}{n\epsilon^2} + \tfrac{\epsilon}{2} \le \epsilon$$

for all $n \ge 2M^2/\epsilon^3$. Combining (16) and (c), it follows that

(d) $$P(|\bar{X}_n - \mu_n| \ge \epsilon\,) \le \epsilon \qquad \text{for } n \ge (\text{some } N_\epsilon)\,.$$

Thus $\bar{X}_n - \mu_n \to_p 0$. We also have $\bar{X}_n - \mathrm{median}(\bar{X}_n) \to_p 0$, since the symmetrization inequality 8.3.2 gives

(e) $$P(|\bar{X}_n - \mathrm{median}(\bar{X}_n)| \ge \epsilon) \le 4P(|\bar{X}_n - \mu_n| \ge \epsilon/2) \to 0.$$

[The acceptability of the third exhibited choice for the centering constant is left to exercise 4.1 below.] In any case, we have shown that (2) implies (1).

The equivalence of (2) and (3) follows. Note figure 4.1, bearing in mind that $(a \vee b) \le a + b \le 2(a \vee b)$ for the definitions $a \equiv \tau^+(x) \equiv x\,P(X^+ > x)$ and $b \equiv \tau^-(x) \equiv x\,P(X^- > x)$. Figure 4.1 thus shows that $\tau^-(x) \to 0$ holds if and only if $t\,|F^{-1}(t)| \to 0$, that $\tau^+(x) \to 0$ holds if and only if $t\,|F^{-1}(1-t)| \to 0$, and that $\tau(x) \to 0$ holds if and only if $t\,F_{|X|}^{-1}(1-t) \to 0$.

Consider the equivalence of (4) and (2). We know from (8.3.14) that $M_n \to_p 0$ if and only if $nP(|X| > \epsilon n) \to 0$ for all $\epsilon > 0$, that is, if and only if $\tau(\epsilon n) = \epsilon n P(|X| > \epsilon n) \to 0$ for all $\epsilon > 0$. Thus $M_n \to_p 0$ if and only if $\tau(x) \to 0$ as $x \to \infty$.

We still neeed to show that (1) implies (2), but we'll wait a paragraph for this.

Consider next theorem 4.3. We will only provide a proof with $r = 1$ (we may just replace $|X|$ by $|Y| \equiv |X|^r$, after raising $|X|$ to the power r). Now, $\tau_X(x) \to 0$ implies $\tau^s(x) \to 0$ by the right-hand side of inequality 8.3.2 with $a = 0$, while the left-hand side then gives $\tau_{X-\mathrm{med}}(x) \to 0$. The equivalence of (4) and (2) then gives $\max|X_{nk} - \mathrm{med}|/n \to_p 0$, which trivially gives $\max|X_{nk}|/n \to_p 0$, which gives $\tau_X(x) \to 0$ by the equivalence of (4) and (2). This completes theorem 4.3.

Finally, we prove that (1) implies (2). Suppose that there exist some constants μ_n such that $S_n/n - \mu_n = \bar{X}_n - \mu_n \to_p 0$. Let $S_n^s = S_n - S_n'$, where $S_n' \equiv X_{n1}' + \cdots + X_{nn}'$ with $X_{nk}' \cong X_{nk}$ and with X_{nk}'s and X_{nk}''s independent. Thus $S_n^s/n \to_p 0$. Then $M_n^s \equiv \max_{1\le k\le n}|X_{nk}^s|/n \to_p 0$ by the (8.3.4) Lévy inequality. Thus $\tau^s(x) \to 0$ by theorem 4.3, and hence $\tau(x) \to 0$ by theorem 4.3. □

Exercise 4.1 Verify that the choice $\nu_{\theta,n} \equiv \int_{\theta/n}^{1-\theta/n} K(t)\,dt$ (for any $0 < \theta \le 1$) also works in the WLLN as a centering constant in (1).

We have just seen that *good inequalities lead to good theorems*! In sections 8.8 and 12.11 we will add to our growing collection of good inequalities. Some will be used in this text, and some will not. But the author thinks it important to illustrate these possibilities.

Exercise 4.2* When $\mathrm{E}|X| = \infty$, the SLLN above showed that $\lim_n |\bar{X}_n| = \infty$ a.s. Show the following stronger result. If $X_1, X_2, \ldots$ are iid with $\mathrm{E}|X| = \infty$, then

$$\overline{\lim}_{n\to\infty}|\bar{X}_n - c_n| =_{a.s.} \infty$$

for every sequence of constants c_n. (Note exercise 4.23 below.)

Exercise 4.3 (Erickson) (a) If $\mathrm{E}X^- < \infty$ but $\mathrm{E}X^+ = \infty$, then $\lim S_n/n =_{a.s.} +\infty$. (b) (Kesten)* If both $\mathrm{E}X^+ = \infty$ and $\mathrm{E}X^- = \infty$, then either $S_n/n \to_{as} \infty$, $S_n/n \to_{as} -\infty$, or both $\overline{\lim}\, S_n/n = \infty$ and $\underline{\lim}\, S_n/n = -\infty$.

Exercise 4.4 (Marcinkiewicz–Zygmund) Let $X_1, X_2, \ldots$ be iid. Let $0 < r < 2$. Establish the equivalence

$$(17) \qquad \mathrm{E}|X|^r < \infty \qquad \text{if and only if} \qquad \frac{1}{n^{1/r}}\sum_{k=1}^{n}(X_k - c) \to_{a.s.} 0 \quad \text{for some } c.$$

If so, then $c = \mathrm{E}X$ when $1 \le r < 2$, while c is arbitrary (so $c = 0$ works) when $0 < r < 1$. [Hint. Truncate via $Y_n \equiv X_n \times 1_{[|X_n| < n^{1/r}]}$ in a SLLN type proof.]

Exercise 4.5*(Feller) Let $X_1, X_2, \ldots$ be iid with $\mathrm{E}|X| = \infty$. If $a_n/n \uparrow$, then

$$(18) \qquad \overline{\lim}\,|S_n|/a_n = \begin{cases} = 0 & \text{a.s.,} \\ = \infty & \text{a.s.,} \end{cases} \quad \text{according as} \quad \sum_{n=1}^{\infty} P(|X_n| \ge a_n) = \begin{cases} < \infty, \\ = \infty. \end{cases}$$

[Note that $P(|X_n| \ge a_n \text{ i.o.})$ equals 0 or 1 as $\sum_1^\infty P(|X_n| \ge a_n)$ is finite or infinite.]

Exercise 4.6 Clarify the overlap between (17) and (18).

Exercise 4.7 (Random sample size) (a) Let $X_1, X_2, \ldots$ be iid with $\tau(x) \to 0$ as $x \to \infty$. Let $N_n \ge 0$ be any integer-valued rv satisfying $N_n/n \to_p c \in (0, \infty)$. Then

$$(19) \qquad S_{N_n}/N_n - \mu_n \to_p 0, \qquad \text{for } \mu_n \equiv \int_{[-n,n]} x\,dF(x).$$

(b) Suppose $X_1, X_2, \ldots$ are iid and $\mu \equiv \mathrm{E}X$ is finite. Let $N_n \ge 0$ be any positive integer-valued rv satisfying $N_n/n \to_{a.s.} c \in (0, \infty)$. Then

$$(20) \qquad S_{N_n}/N_n \to_{a.s.} \mu.$$

Exercise 4.8 (A weak SLLN) (a) For $X_{n1}, \ldots, X_{nn}$ independent (or, uncorrelated) with $X_{nk} \cong (0, \sigma_{nk}^2)$ and all $\sigma_{nk}^2 \le$ (some M) $< \infty$, we have $\bar{X}_n \to_{a.s.} 0$.
Hint. Show that $P(|S_n| \ge n\epsilon) \le M/(n\epsilon^2)$, so that $P(|\bar{X}_{n^2}| = |S_{n^2}/n^2| > 0 \text{ i.o.}) = 0$. Then show that the "block maximum"

$$\Delta_n \equiv \max_{n^2<k<(n+1)^2} |S_k - S_{n^2}|$$

has $\mathrm{E}\Delta_n^2 \le 2n\mathrm{E}\{|S_{(n+1)^2-1} - S_{n^2}|^2\} \le 4n^2 M$, so that $P(\Delta_n/n^2 > \epsilon \text{ i.o.}) = 0$.
(b) Use Kolmogorov's inequality to obtain $\mathrm{E}\Delta_n^2 \le 2nM$, under independence.

Exercise 4.9 Let $X_{n1}, \ldots, X_{nn}$ be row independent rvs (here, merely uncorrelated is much harder to consider) with means 0 and having all $\mathrm{E}X_{nk}^4 \le$ (some M) $< \infty$.
(a) (Cantelli's inequality) Verify that $\bar{X}_n \equiv S_n/n \equiv (X_{n1} + \cdots + X_{nn})/n$ satisfies

$$P(\,|S_n| \ge \lambda) \le 3Mn^2/\lambda^4 \qquad \text{for all } \lambda > 0.$$

(b) (A very weak SLLN) Show that under these circumstances $\bar{X}_n \to_{a.s.} 0$.

Exercise 4.10 (Alternative proof of the SLLN) Apply either the Hájek–Rényi inequality (inequality 8.10.3) or the monotone inequality (inequality 8.10.1) as a replacement for the use of the Kronecker lemma in the SLLN proof.

Exercise 4.11 (St. Petersburg paradox) Let $X_1, X_2, \ldots$ be iid rvs for which $P(X = 2^m) = 1/2^m$ for $m \ge 1$. Show that $S_n/a_n - 1 = (S_n - b_n)/a_n \to_p 0$ for $b_n \equiv n \log_{\text{Base } 2} n$ and $a_n \equiv n \log_{\text{Base } 2} n$ also. Hint. Let $Y_{nk} \equiv X_k 1_{[X_k \le n \log_{\text{Base } 2} n]}$. (While $S_n/a_n \to_p 1$ was just shown, it can also be shown that $S_n/a_n \to_{a.s.} \infty$.)

Exercise 4.12* (Spitzer) Let $X, X_1, X_2, \ldots$ be iid. Establish the following claim.

(21) $$\mathrm{E}(X) = 0 \quad \text{for } X \in \mathcal{L}_1 \qquad \text{iff} \qquad \sum_{n=1}^{\infty} \frac{1}{n} P(|S_n| \ge n\,\epsilon) < \infty \quad \text{for all } \epsilon > 0.$$

Exercise 4.13 If $X_1, X_2, \ldots$ are iid Exponential(1), then $\overline{\lim}\, X_n/\log n = 1$ a.s. and $X_{n:n}/\log n \to 1$ a.s.

Exercise 4.14 If $X_1, X_2, \ldots$ are iid N(0,1), then $X_{n:n}/\sqrt{2 \log n} \to_p 1$.

Exercise 4.15 (a) Does the WLLN hold for the Cauchy distribution?
(b) Does the WLLN hold if $P(|X| > x) = e/[2x \log x]$ for $x \ge e$, X symmetric?
(c) Make up one more example of each of these two types.

Exercise 4.16 (Uniform integrability of sample averages) Let $X_1, X_2, \ldots$ be iid, and let $\bar{X}_n \equiv (X_1 + \cdots + X_n)/n$. Then the rvs $\{X_n : n \ge 1\}$ are uniformly integrable if and only if the rvs $\{\bar{X}_n : n \ge 1\}$ are uniformly integrable. (Relate this to the SLLN result in (7).) (We only need independence for u.i. X_k's to yield u.i. $\bar{X}_n$'s.)

Exercise 4.17 (a) Let row independent rvs $X_{n1}, \ldots, X_{nn}$ be iid with the df $F(\cdot)$. Let F have finite mean $\mu \equiv \mathrm{E}X$. We know $M_n \equiv [\max_{1 \le k \le n} |X_{nk}|/n] \to_p 0$ by the WLLN. Trivially, $\mathrm{E}M_n \le \mathrm{E}|X|$. Show that

(22) $$\mathrm{E}M_n = \mathrm{E}[\max_{1 \le k \le n} \tfrac{1}{n} |X_{nk}|] \to 0 \qquad (\text{that is, } M_n \to_{\mathcal{L}_1} 0).$$

(b) Let $X_1, X_2, \ldots$ be iid. Show that $\mathrm{E}|X| < \infty$ if and only if $M_n \to_{\mathcal{L}_1} 0$.

Exercise 4.18 (Negligibility for $r = 1$, a.s.) (i) Let $X, X_1, X_2, \ldots$ be iid rvs. Let $r > 0$ (with $r = 1$ the most important case). Prove that the following are equivalent:

(23) $\mathrm{E}|X|^r < \infty$.

(24) $M_{rn} \equiv [\frac{1}{n} \max_{1 \le k \le n} |X_k|^r] \to_{a.s.} 0$.

(25) $\mathrm{E}M_{rn} \to 0$.

(ii) Since $\mathrm{E}|X|^r < \infty$ if and only if the symmetrized rv $X - X'$ has $\mathrm{E}|X - X'|^r < \infty$ (by exercise 8.2.3), we can add three analogous equivalences for iid symmetric rvs distributed as $X - X'$.

Exercise 4.19* (Maller–Resnick; Kesten) For a sequence of iid rvs $X, X_1, X_2, \ldots$ let $\bar{X}_n \equiv (X_1 + \cdots + X_n)/n$ and let $M_{1n} \equiv [\frac{1}{n} \max_{1 \le k \le n} |X_k|]$. Then (difficult)

(26) $M_{1n}/|\bar{X}_n| \to_{a.s.} 0$ if and only if $0 < |\mathrm{E}X| < \infty$.

Exercise 4.20 (Negligibility for $r = 2$, a.s.) (i) Let $X, X_1, X_2, \ldots$ be iid rvs (that are not identically equal to 0). Let $r > 0$ (with $r = 2$ the most important case). Prove that the following are equivalent (you should use the difficult (26) for(30)):

(27) $\mathrm{E}|X|^r < \infty$.

(28) $M_{rn} \equiv [\frac{1}{n} \max_{1 \le k \le n} |X_k|^r] \to_{a.s.} 0$.

(29) $\mathrm{E}M_{rn} \to 0$.

(30) $M_{rn}/[\frac{1}{n} \sum_1^n |X_k|^r] \to_{a.s.} 0$.

(ii) When $r = 2$, we may add the equivalent condition (by (6.6.6))

(31) $\mathcal{D}_n^2 \equiv [\frac{1}{n} \max_{1 \le k \le n} (X_k - \bar{X}_n)^2]/[\frac{1}{n} \sum_{k=1}^n (X_k - \bar{X}_n)^2] \to_{a.s.} 0$.

(iii) Again (by (8.2.18)), $\mathrm{E}|X|^r < \infty$ if and only if $\mathrm{E}|X - X'|^r < \infty$.

Exercise 4.21 (Neglibility, in probability) Let $X_{n1}, \ldots, X_{nn}$ be iid F, for $n \ge 1$. Let $X \cong F$. Let $r > 0$. Prove that the following are equivalent:

(32) $y\, P(|X|^r > y) \to 0$ as $y \to \infty$.

(33) $x^r\, P(|X| > x) \to 0$ as $x \to \infty$.

(34) $x^r\, P(|X - X'| > x) \to 0$ as $x \to \infty$, here X and X' are iid F.

(35) $x^r\, P(|X - (\text{some } `a')| > x) \to 0$. Then any '$a$' works; so med($X$) works.

(36) $M_{rn} \equiv [\frac{1}{n} \max_{1 \le k \le n} |X_{nk}|^r] \to_p 0$. (Any $|X_{nk} - a|^r$ may replace $|X_{nk}|^r$.)

(37) $\mathrm{E}M_{rn}^\alpha = \mathrm{E}[\frac{1}{n^\alpha} \max_{1 \le k \le n} |X_{nk}|^{r\alpha}] \to 0$ for all $0 < \alpha < 1$.

In case $r > 1$ (and especially for $r = 2$) add to this list the equivalent condition

(38) $\mathrm{E}M_{rn}^{1/r} = \mathrm{E}[\frac{1}{n^{1/r}} \max_{1 \le k \le n} |X_{nk}|] \to 0$.

Because (34) is on the list, the iid X_{nk}'s may be replaced by iid symmetrized X_{nk}^s's in (32), (33), and (36)–(38). Moreover

(39) $\mathrm{E}|X|^p < \infty$ for all $0 < p < r$ whenever (33) holds.

The remaining problems in this subsectiion are mainly quite substantial. They are here for "flavor." Some also make good exercises for section 8.8. (Some of the martingale inequalities found in section 8.10 should prove useful (here and below).)

Exercise 4.22* (Kesten) Let $X_1, X_2, \ldots$ be iid as $X \geq 0$, with $\mathrm{E}|X| = \infty$. Then

(40) $\quad \overline{\lim}\, \frac{X_n}{S_{n-1}} = \infty$ a.s.

Exercise 4.23* (Chow–Robbins) Let $X_1, X_2, \ldots$ be iid as X, with $\mathrm{E}|X| = \infty$. Let $b_n > 0$ denote *any* sequence. Then

(41) $\quad$ either $\quad \underline{\lim}\, \frac{|S_n|}{b_n} = 0$ a.s. $\quad$ or $\quad \overline{\lim}\, \frac{|S_n|}{b_n} = \infty$ a.s.

Exercise 4.24* Suppose $\sum_1^\infty \mathrm{E}|X_n|^r < \infty$ for some $r > 0$. Show that $X_n \to_{a.s.} 0$.

Exercise 4.25* (Hsu–Robbins; Edrös) Let $X, X_1, X_2, \ldots$ be iid rvs. Then

(42) $\quad \mathrm{E}X = 0$ and $\mathrm{Var}[X] < \infty$ $\quad$ if and only if $\quad \sum_{n=1}^\infty P(|S_n| > \epsilon\, n) < \infty$ for all $\epsilon > 0$.

Exercise 4.26* Let $X_1, X_2, \ldots$ be independent with 0 means. Let $r \geq 1$. Then

(43) $\quad S_n/n \to_{a.s.\text{ and }\mathcal{L}_{2r}} 0 \quad$ whenever $\quad \sum_{n=1}^\infty \mathrm{E}|X_n|^{2r}/n^{r+1} < \infty$.

Exercise 4.27* Let $X, X_1, X_2, \ldots$ be iid. Let $\mathrm{Log}\, x \equiv 1 \vee \log x$. (a) Show that

(44) $\quad \mathrm{E}(|X|\ \mathrm{Log}^+(|X|)) < \infty$ $\quad$ if and only if
$\quad \mathrm{E}\{\sup_{n\geq 1}(|X_n|/n)\} < \infty$ $\quad$ if and only if $\quad \mathrm{E}\{\sup_{n\geq 1}(|S_n|/n)\} < \infty$.

(b) Show that for each $r > 1$,

(45) $\quad \mathrm{E}|X|^r < \infty \quad$ if and only if $\quad \mathrm{E}\{\sup_{n\geq 1}(|S_n|^r/n^r)\} < \infty$.

Exercise 4.28* (Stone) Let $X, X_1, X_2, \ldots$ be iid nondegenerate rvs with 0 means. Let $S_n \equiv X_1 + \cdots + X_n$. Then

(46) $\quad \overline{\lim}_n\, S_n/\sqrt{n} =_{a.s.} +\infty \quad$ and $\quad \underline{\lim}_n\, S_n/\sqrt{n} =_{a.s.} -\infty$.

Generalizations of the LLNs

Our results allow simple generalizations of both the WLLN and SLLN.

Theorem 4.4 (General WLLN and SLLN) Let $X_1, X_2, \ldots$ be independent. Then

(47) $\quad \sum_1^n \sigma_k^2/b_n^2 \to 0 \quad$ implies $\quad \sum_1^n (X_k - \mu_k)/b_n \to_p 0$,

(48) $\quad \sum_1^\infty \sigma_k^2/b_k^2 < \infty$ with $b_n \nearrow \infty \quad$ implies $\quad \sum_1^n (X_k - \mu_k)/b_n \to_{a.s.} 0$.

Proof. The first claim is immediate from Chebyshev's inequality. Also, (f) in the SLLN proof shows that

(49) $\quad \sum_1^\infty \sigma_k^2/b_k^2 < \infty \quad$ implies $\quad \sum_1^n (X_k - \mu_k)/b_k \to_{a.s.}$ (some rv S).

Then Kronecker's lemma gives $\sum_1^n (X_k - \mu_k)/b_n \to_{a.s.} 0$. (This result is often the starting point for a development of the SLLN.) □

Exercise 4.29* (More general WLLN) Let $X_{n1}, \ldots, X_{nn}$ be independent, and set $S_n \equiv X_{n1} + \cdots + X_{nn}$. Truncate via $Y_{nk} \equiv X_{nk} \times 1_{[\,|X_{nk}| \le b_n]}$, for some $b_n > 0$ having $b_n \nearrow \infty$. Let μ_{nk} and σ^2_{nk} denote the mean and variance of Y_{nk}. Then

(50) $\quad S_n/b_n \to_p 0$

if and only if we have all three of

(51) $\quad \sum_1^n P(|X_{nk}| > b_n) \to 0, \qquad \sum_1^n \mu_{nk}/b_n \to 0, \qquad \text{and} \qquad \sum_1^n \sigma^2_{nk}/b^2_n \to 0.$

[The converse is a substantial problem.]

5 Applications of the Laws of Large Numbers

Let $X_1, X_2, \ldots$ be iid F. Let $\mathbb{F}_n$ denote the empirical df of $X_1, \ldots, X_n$, given by

(1) $$\mathbb{F}_n(x) \equiv \mathbb{F}_n(x,\omega) \equiv \frac{1}{n}\sum_{k=1}^{n} 1_{(-\infty,x]}(X_k(\omega)) = \frac{1}{n}\sum_{k=1}^{n} 1_{[X_k \le x]} \quad \text{for all real } x.$$

Theorem 5.1 (Glivenko–Cantelli) We have

(2) $$\|\mathbb{F}_n - F\| \equiv \sup_{-\infty<x<\infty} |\mathbb{F}_n(x) - F(x)| \to_{a.s.} 0 \qquad \text{as } n \to \infty.$$

[This is a uniform SLLN for the random function $\mathbb{F}_n(\cdot)$.]

Proof. Let $X_k = F^{-1}(\dot{\xi}_k)$ for $k \ge 1$ be iid F, with the $\dot{\xi}_k$'s iid Uniform$(0,1)$. Let $\mathbb{G}_n$ denote the empirical df of the first n of these $\dot{\xi}_k$'s, and let $\mathbb{F}_n$ denote the empirical df of the first n of these X_k's. Let I denote the identity function. Then

(3) $$(\mathbb{F}_n - F) = [\mathbb{G}_n(F) - I(F)] \qquad \text{on } (-\infty,\infty) \qquad \text{for every } \omega$$

by (6.3.4). Thus by theorem 5.3.3, it will suffice to prove the result in the special case of uniform empirical df's $\mathbb{G}_n$'s. (Recall the remark in bold above (6.4.3) that the representation of X as $F^{-1}(\xi)$ allows alternative ways to approach problems. Moreover, using the $\dot{\xi}_k$'s of (6.3.8) gives us back the original X_k's.)

Now, $\mathbb{G}_n(k/M) - k/M \to_{a.s.} 0$ as $n \to \infty$ for $0 < k \le M$ by the SLLN applied to the iid Bernoulli(k/M) rv's $1_{[0,k/M]}(\dot{\xi}_i)$. We now assume that M is so large that $1/M < \epsilon$. Then for $(k-1)/M \le t \le k/M$, with $1 \le k \le M$, we have both

(a) $$\mathbb{G}_n(t) - t \le \mathbb{G}_n\left(\tfrac{k}{M}\right) - \tfrac{k-1}{M} \le \mathbb{G}_n\left(\tfrac{k}{M}\right) - \tfrac{k}{M} + \tfrac{1}{M} \qquad \text{and}$$

(b) $$\mathbb{G}_n(t) - t \ge \mathbb{G}_n\left(\tfrac{k-1}{M}\right) - \tfrac{k}{M} \ge \mathbb{G}_n\left(\tfrac{k-1}{M}\right) - \tfrac{k-1}{M} - \tfrac{1}{M}.$$

These combine to give

(c) $$\sup_{0\le t\le 1} |\mathbb{G}_n(t) - t| \le \left[\max_{0\le k\le M} \left|\mathbb{G}_n\left(\tfrac{k}{M}\right) - \tfrac{k}{M}\right|\right] + \tfrac{1}{M}$$

(d) $$\to_{a.s.} 0 + 1/M < \epsilon.$$

Since $\epsilon > 0$ is arbitrary, we have shown that $\sup_{0\le t\le 1} |\mathbb{G}_n(t) - t| \to_{a.s.} 0$. That is,

(4) $$\|\mathbb{F}_n - F\| = \|\mathbb{G}_n(F) - F\| \le \|\mathbb{G}_n - I\| \to_{a.s.} 0\,,$$

as claimed. □

Exercise 5.1 Let $\xi_{n1}, \ldots, \xi_{nn}$ denote any row independent Uniform(0,1) rvs, and let all $X_{nk} = F^{-1}(\xi_{nk})$ for a fixed df F. Let $\mathbb{F}_n$ and $\mathbb{G}_n$ denote the empirical dfs of the nth rows of these two arrays. Show that (2) still holds.

Example 5.1 (Weierstrass approximation theorem) If f is continuous on $[0,1]$, then there exist polynomials B_n such that $\|B_n - f\| = \sup_{0\le t\le 1} |B_n(t) - f(t)| \to 0$ as $n \to \infty$.

Proof. (Bernstein) Define the *Bernoulli polynomials*

(5) $$B_n(t) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\binom{n}{k} t^k (1-t)^{n-k} \qquad \text{for } 0 \le t \le 1$$

(a) $$= \mathrm{E} f(T/n) \qquad \text{where } T \cong \text{Binomial}(n,t).$$

Since f is continuous, f is bounded by some M, and f is uniformly continuous on $[0,1]$ having $|f(x) - f(y)| < \epsilon$ whenever $|x - y| < \delta_\epsilon$. Then

$$|f(t) - B_n(t)| = \left|\sum_{k=0}^{n} [f(t) - f(k/n)]\tbinom{n}{k} t^k (1-t)^{n-k}\right|$$

$$\le |\textstyle\sum_{\{k:|k/n-t|<\delta_\epsilon\}} \text{ same } | + |\textstyle\sum_{\{k:|k/n-t|\ge\delta_\epsilon\}} \text{ same } |$$

(b) $$< \epsilon + 2M\ P(|T/n - t| \ge \delta_\epsilon) \qquad \text{by uniform continuity of } f$$

$$\le \epsilon + 2M\ t(1-t)/n\delta_\epsilon^2 \qquad \text{for all } t \quad \text{(by Chebyshev)}$$

(c) $$\le \epsilon + 2M/4n\delta_\epsilon^2 \le 2\epsilon \qquad \text{for } n \ge \text{ some } N_\epsilon, \qquad \text{for all } 0 \le t \le 1.$$

As the choice of N_ϵ does not depend on t, the convergence is uniform. Note that this is just an application of a weak form of the WLLN (that is, of the Chebyshev inequality). □

Example 5.2 (Borel's normal numbers) A number x in $[0,1]$ is called *normal to base* d if when expanded to base d, the fraction of each of the digits $0,\ldots,d-1$ converges to $1/d$. The number is *normal* if it is normal to base d for each $d > 1$. We are able to conclude that

a.e. number in $[0,1]$ is normal with respect to Lebesgue measure λ.

[Comment: $\frac{1}{3} = 0.010101\ldots$ in base 2 is normal in base 2, but $\frac{1}{3} = 0.1000\ldots$ in base 3 is not normal in base 3.] [This was a historically important example, which spurred some of the original development.]

Proof. Let $(\Omega, \mathcal{A}, P) = ([0,1], \mathcal{B} \cap [0,1], \lambda)$. Let

(a) $$\omega = \textstyle\sum_{n=1}^{\infty} \beta_n(\omega)/d^n \qquad \text{define rvs } \beta_1, \beta_2, \ldots .$$

Note that the β_n's are iid discrete uniform on $0,1,\ldots,d-1$. Thus, letting $\eta_{nk} = 0$ or 1 according as $\beta_n = k$ or $\beta_n \ne k$, we have

(b) $$\lambda(A_{d,k}) \equiv \lambda(\{\omega : n^{-1}\textstyle\sum_{j=1}^{n} \eta_{jk} \to 1/d\}) = 1$$

by the SLLN. Thus $A_d \equiv \cap_{k=0}^{d-1} A_{d,k}$ has $\lambda(A_d) = 1$; that is, a.e. ω in $[0,1]$ is normal to base d. Then trivially, $A \equiv \cap_{d=1}^{\infty} A_d$ has $\lambda(A) = 1$. And so, a.e. ω in $[0,1]$ is normal. □

Example 5.3 (SLLN for random sample size) Let N_n be positive-integer valued rvs for which $N_n/n \to_{a.s.} c \in (0, \infty)$, and let $X_1, X_2, \ldots$ be iid with mean μ.
(a) Then

(6) $$S_{N_n}/n \to_{a.s.} \mu \cdot c \qquad \text{as } n \to \infty.$$

(b) If $X_1, X_2, \ldots$ are iid Bernoulli(p) and $N_n(\omega) \equiv \min\{k : S_k(\omega) = n\}$, then the waiting times N_n satisfy $N_n/n \to_{a.s.} 1/p$.

Proof. (a) Now, $S_n/n \to_{a.s} \mu$ by the SLLN, and thus $N_n \to_{a.s.} \infty$ implies $S_{N_n}/N_n \to_{a.s.} \mu$. Thus

$$S_{N_n}/n = (S_{N_n}/N_n)\,(N_n/n) \to_{a.s.} \mu \cdot c,$$

using $N_n \to_{a.s.} \infty$ by $c > 0$.
(b) We also have (since $\mu = p$)

$$1 = S_{N_n}/n = (S_{N_n}/N_n)\,(N_n/n), \quad \text{so} \quad N_n/n = 1/(S_{N_n}/N_n) \to_{a.s.} 1/p,$$

completing the proof. Note that we could also view N_n as the sum of n iid Geometric(p) rvs, and then apply the SLLN. □

Exercise 5.2 (Monte Carlo estimation) Let $h : [0, 1] \to [0, 1]$ be continuous.
(i) Let $X_k \equiv 1_{[h(\xi_k) \geq \Theta_k]}$, where $\xi_1, \xi_2, \ldots \Theta_1, \Theta_2, \ldots$ are iid Uniform(0, 1) rvs. Show that this sample average is a strongly consistent estimator of the integral; that is, show that $\bar{X}_n \to_{a.s.} \int_0^1 h(t)\,dt$.
(ii) Let $Y_k \equiv h(\xi_k)$. Show that $\bar{Y}_n \to_{a.s.} \int_0^1 h(t)\,dt$.
(iii) Evaluate Var[$\bar{X}_n$] and Var[$\bar{Y}_n$], and compare them.

6 Law of the Iterated Logarithm

Theorem 6.1 (LIL; Hartman–Wintner; Strassen) Let $X_1, X_2, \ldots$ be iid rvs. Consider the partial sums $S_n \equiv X_1 + \cdots + X_n$.
(a) If $\mathrm{E}X = 0$ and $\sigma^2 \equiv \mathrm{Var}[X] < \infty$, then

$$(1) \qquad \limsup_{n\to\infty} \frac{S_n}{\sqrt{2n\log\log n}} = \sigma \text{ a.s.,} \qquad \text{while} \qquad \liminf_{n\to\infty} \frac{S_n}{\sqrt{2n\log\log n}} = -\sigma \text{ a.s.}$$

(b) In fact,

$$(2) \qquad \frac{S_n}{\sqrt{2n\log\log n}} \rightsquigarrow_{a.s.} [-\sigma, \sigma].$$

[That is, for a.e. ω the limit set of $S_n/\sqrt{2n\log\log n}$ is exactly $[-\sigma, \sigma]$].
(c) Conversely, if

$$(3) \qquad \limsup_{n\to\infty} \frac{|S_n|}{\sqrt{2n\log\log n}} < \infty \text{ a.s.,} \qquad \text{then} \qquad \mathrm{E}X = 0 \text{ and } \sigma^2 < \infty.$$

Theorem 6.2 (The other LIL; Chung) If $X_1, X_2, \ldots$ are iid $(0, \sigma^2)$, then

$$(4) \qquad \liminf_{n\to\infty} \max_{1\le k\le n} \sqrt{2\log\log n}\, \frac{|S_k|}{\sqrt{n}\,\sigma} = \pi/2 \text{ a.s.}$$

[We state this for fun only, as it has seen little application.]

Versions of both theorems are also known for cases other than iid. The classical proof of theorem 6.1 in full generality begins with truncation, and then carefully uses exponential bounds for bounded rvs. A more modern proof relies upon Skorokhod embedding of the partial sum process in Brownian motion. This general proof is outlined in the straightforward exercise 12.8.2, after embedding is introduced. But the proof below for the special case of normal rvs contains several of the techniques used in the classical proof of the general case (and in other related problems). And it is also a crucial component of the general case in exercise 12.8.2.

Proposition 6.1 Let $Z_1, Z_2, \ldots$ be iid $N(0,1)$ rvs. Let $S_n \equiv Z_1 + \cdots + Z_n$ and $b_n \equiv \sqrt{2\log\log n}$. Then $\limsup_{n\to\infty} S_n/\sqrt{n}\, b_n = 1$ a.s.

Proof. Let $\epsilon > 0$. We will use the exponential bound

$$(5) \qquad \exp[-(1+\epsilon)\lambda^2/2] \le P(S_n/\sqrt{n} \ge \lambda) \le \exp[-(1-\epsilon)\lambda^2/2] \qquad \text{for all } \lambda > \lambda_\epsilon$$

(for some λ_ϵ) [see Mills' ratio exercise 6.1 below], and the Lévy maximal inequality

$$(a) \qquad P(\max_{1\le k\le n} S_k \ge \lambda) \le 2P(S_n \ge \lambda) \qquad \text{for all } \lambda > 0.$$

Let $n_k \equiv [a^k]$ for $a > 1$; a sufficiently *small* a will be specified below. Now,

$A_k \equiv \bigcup_{n_{k-1} \le m \le n_k} [S_m \ge \sqrt{m}(1+2\epsilon)b_m]$

(b) $\qquad \subset \left[\max_{n_{k-1} \le m \le n_k} S_m \ge (1+2\epsilon)\sqrt{\frac{n_{k-1}}{n_k}}\, b_{n_{k-1}}\sqrt{n_k}\right],$

since $\sqrt{n}$ is $\nearrow$ and b_n is $\nearrow$; so that for k sufficiently large,

(c) $\qquad P(A_k) \le 2P\left(S_{n_k}/\sqrt{n_k} \;\ge\; (1+2\epsilon)\sqrt{\frac{n_{k-1}}{n_k}}\, b_{n_{k-1}}\right) \qquad$ by (a)

$\qquad \le 2\exp\left(-\tfrac{1}{2}(1-\epsilon)(1+2\epsilon)^2\, \frac{1-\epsilon}{a}\, 2\log k\right) \qquad$ by (5)

$\qquad \le 2\exp(-(1+\epsilon)\log k) = 2/k^{1+\epsilon} \qquad$ for a sufficiently close to 1

(d) $\qquad = \;$ (kth term of a convergent series).

Thus $P(A_k \text{ i.o.}) = 0$ by Borel–Cantelli. Since $\epsilon > 0$ is arbitrary, we thus have

(e) $\qquad \limsup_{n\to\infty} \frac{S_n}{\sqrt{n}\, b_n} \le 1$ a.s.

Since (e) is true, on any subsequence $n_k \to \infty$ we can claim that

(f) $\qquad P(A_k \text{ i.o.}) = 0, \qquad$ including when $n_k \equiv [a^k]$ for a huge.

We must now show that the $\overline{\lim}$ in (e) is also ≥ 1 a.s. We will still use $n_k \equiv [a^k]$, but a will be specified sufficiently *large* below. We write $S_{n_k} = S_{n_{k-1}} + (S_{n_k} - S_{n_{k-1}})$, so that

(g) $\qquad \frac{S_{n_k}}{\sqrt{n_k}\, b_{n_k}} = \sqrt{\frac{n_{k-1}}{n_k}}\left(\frac{b_{n_{k-1}}}{b_{n_k}}\right)\frac{S_{n_{k-1}}}{\sqrt{n_{k-1}}\, b_{n_{k-1}}} + \frac{S_{n_k} - S_{n_{k-1}}}{\sqrt{n_k}\, b_{n_k}}$

(h) $\qquad \sim \frac{1}{\sqrt{a}} \cdot 1 \cdot \frac{S_{n_{k-1}}}{\sqrt{n_{k-1}}\, b_{n_{k-1}}} + \frac{S_{n_k} - S_{n_{k-1}}}{\sqrt{n_k}\, b_{n_k}}.$

Now, the independent events

(i) $\qquad B_k \equiv [S_{n_k} - S_{n_{k-1}} \ge (1-2\epsilon)\sqrt{n_k}\, b_{n_k}] = \left[\frac{S_{n_k} - S_{n_{k-1}}}{\sqrt{n_k - n_{k-1}}} \ge \frac{(1-2\epsilon)\sqrt{n_k}\, b_{n_k}}{\sqrt{n_k - n_{k-1}}}\right]$

have

(j) $\qquad P(B_k) \ge \exp\left(-\tfrac{1}{2}(1+\epsilon)(1-2\epsilon)^2\, \frac{n_k}{n_k - n_{k-1}}\, b_{n_k}^2\right) \qquad$ by (5)

$\qquad \ge \exp\left(-\tfrac{1}{2}(1+\epsilon)(1-2\epsilon)^2\, \frac{(1+\epsilon)a}{a-1}\, 2\log k\right)$

$\qquad \ge \exp\left(-(1-\epsilon)\log k\right) \qquad$ for a sufficiently large

(k) $\qquad = 1/k^{1-\epsilon} = \;$ (kth term of a series with an infinite sum), so that $P(B_k \text{ i.o.}) = 1$ by the second Borel–Cantelli lemma. But $P(A_k \text{ i.o.}) = 0$ and $P(B_k \text{ i.o.}) = 1$ means that

$$\text{(l)} \qquad P(A_k^c \cap B_k \text{ i.o.}) = 1.$$

Moreover, on $A_k^c \cap B_k$ we have, using (h), (i), and the symmetric version of (f),

$$\text{(m)} \qquad \frac{S_{n_k}}{\sqrt{n_k}\, b_{n_k}} \geq -\frac{(1+2\epsilon)(1+\epsilon)}{\sqrt{a}} + (1-2\epsilon) \geq (1-3\epsilon)$$

for the constant a specified sufficiently large. Thus, even focusing only on the subsequence n_k in (f) with this large $a \equiv a_\epsilon$, since $\epsilon > 0$ was arbitrary,

$$\text{(n)} \qquad \limsup_{k\to\infty} \frac{S_{n_k}}{\sqrt{n_k}\, b_{n_k}} \geq 1 \text{ a.s.}$$

Combining (e) and (n) gives the proposition. □

Exercise 6.1 (Mills' ratio) Show that for all $\lambda > 0$

$$\text{(6)} \qquad \frac{\lambda}{\lambda^2+1}\frac{1}{\sqrt{2\pi}}\exp(-\lambda^2/2) < P(N(0,1) > \lambda) < \frac{1}{\lambda}\frac{1}{\sqrt{2\pi}}\exp(-\lambda^2/2),$$

which can be rewritten as

$$\text{(7)} \qquad \frac{\lambda}{\lambda^2+1}\,\phi(\lambda) < 1 - \Phi(\lambda) < \tfrac{1}{\lambda}\,\phi(\lambda)$$

where ϕ and Φ denote the standard normal $N(0,1)$ density and df, respectively. Show that (5) follows from this. This is the end of this exercise.
(∗) For a standardized rv Z_n, one might then *hope* that as $\lambda_n \to \infty$

$$\text{(8)} \qquad \exp(-(1+\epsilon_n)\,\lambda_n^2/2) \leq P(Z_n \geq \lambda_n) \leq \exp(-(1-\epsilon_n)\,\lambda_n^2/2),$$

as was applied in (5). [This clean exponential bound for normal rvs was the key to the simple LIL proof in proposition 6.1. The classic Hartman–Wintner proof uses truncation to achieve a reasonable facsimile of this in other cases.]
(∗) (Ito–McKean) It is even true that for all $\lambda > 0$ there are the tighter bounds

$$\text{(9)} \qquad \tfrac{2}{\sqrt{\lambda^2+4}+\lambda}\,\phi(\lambda) < 1 - \Phi(\lambda) < \tfrac{2}{\sqrt{\lambda^2+2}+\lambda}\,\phi(\lambda).$$

Exercise 6.2 In place of (c) in the LIL proof of proposition 6.1, use Mills' ratio to bound $P(A_n) \equiv P(S_n/\sqrt{n} \geq (1+2\epsilon)\sqrt{2\log n}\,)$. Use that bound directly to show that $\limsup |S_n|/(\sqrt{n}\sqrt{2\log n}\,) \leq 1$ a.s. [This "poor" result will show the value of using the "block of indices" in the definition of A_k in the proof we gave.]

Exercise 6.3 Suppose arbitrary events A_n and B_n satisfy $P(A_n \text{ i.o.}) = 0$ and $P(B_n \text{ i.o.}) = 1$. Show that $P(A_n^c \cap B_n \text{ i.o.}) = 1$ (as in (l) above).

Summary Suppose $X, X_1, X_2, \ldots$ are iid $(\mu, 1)$. Then:

$$\frac{\sum_{k=1}^n (X_k - \mu)}{n} \to_{a.s.} 0 \qquad \text{by the SLLN.} \tag{10}$$

$$\frac{\sum_{k=1}^n (X_k - \mu)}{n^{1/r}} \to_{a.s.} 0 \qquad \text{for all } 1 \le r < 2, \text{ by Marcinkiewicz–Zygmund.} \tag{11}$$

$$\frac{\sum_{k=1}^n (X_k - \mu)}{\sqrt{2n \log \log n}} \rightsquigarrow_{a.s.} [-1, 1] \qquad \text{by the LIL.} \tag{12}$$

Suppose we go all the way to $\sqrt{n}$ in the denominator. Then the classical CLT gives

$$\frac{\sum_{k=1}^n (X_k - \mu)}{\sqrt{n}} \to_d N(0, 1) \qquad \text{by the CLT,} \tag{13}$$

even though we have divergence to $\pm\infty$ for a.e. ω (by the LIL). □

Exercise 6.4 (rth mean convergence theorem) Let $X, X_1, X_2, \ldots$ be iid, and consider the partial sums $S_n \equiv X_1 + \cdots + X_n$. Let $0 < r < 2$ (and suppose $\mathrm{E}X = 0$ in case $1 \le r < 2$). The following are equivalent:

(a) $\mathrm{E}|X|^r < \infty$.

(b) $S_n / n^{1/r} \to_{a.s.} 0$.

(c) $\mathrm{E}|S_n|^r = o(n)$.

(d) $\mathrm{E}(\max_{1 \le k \le n} |S_k|^r) = o(n)$.

Hint. For (a) and (b) imply (c), use the Hoffmann–Jorgensen (8.10.14) below.

7 Strong Markov Property for Sums of IID RVs

Let $X_1, X_2, \ldots$ be iid and let $S_n \equiv X_1 + \cdots + X_n$. Let $\mathsf{S} \equiv (S_1, S_2, \ldots)$.

Definition 7.1 The integer valued rv N is a *stopping time* for the sequence of rvs $S_1, S_2, \ldots$ if $[N = k] \in \mathcal{F}(S_1, \ldots, S_k)$ for all $k \ge 1$. It is elementary that

(1) $\quad \mathcal{F}_N \equiv \mathcal{F}(S_k : k \le N)$

(2) $\quad \equiv \{A \in \mathcal{F}(\mathsf{S}) : A \cap [N = k] \in \mathcal{F}(S_1, \ldots, S_k) \text{ for all } k \ge 1\} = (\text{a } \sigma\text{-field}),$

since it is clearly closed under complements and countable intersections. (Clearly, $[N = k]$ can be replaced by $[N \le k]$ in the definition of $\mathcal{F}_N$ in (2).)

Proposition 7.1 Both N and S_N are $\mathcal{F}_N$-measurable.

Proof. Now, to show that $[N \le m] \in \mathcal{F}_N$ we consider $[N \le m] \cap [N = k]$ equals $[N = k]$ or $\emptyset$, both of which are in $\mathcal{F}(\mathsf{S})$; this implies $[N \le m] \in \mathcal{F}_N$. Likewise,

(a) $\quad [S_N \le x] \cap [N = k] = [S_k \le x] \cap [N = k] \in \mathcal{F}(S_1, \ldots, S_k),$

implying that $[S_N \le x] \in \mathcal{F}_N$. □

Theorem 7.1 (The strong Markov property) If N is a stopping time, then the increments continuing from the random time

(3) $\quad \tilde{S}_k \equiv S_{N+k} - S_N, \qquad k \ge 1,$

have the same distribution on $(R_\infty, \mathcal{B}_\infty)$ as does $S_k,\ k \ge 1$. Moreover, defining $\tilde{S} \equiv (\tilde{S}_1, \tilde{S}_2, \ldots)$,

(4) $\quad \mathcal{F}(\tilde{S}) \equiv \mathcal{F}(\tilde{S}_1, \tilde{S}_2, \ldots)$ is independent of $\mathcal{F}_N$ (hence of N and S_N).

Proof. Let $B \in \mathcal{B}_\infty$ and $A \in \mathcal{F}_N$. Now,

(a) $\quad P([\tilde{S} \in B] \cap A) = \sum_{n=1}^\infty P([\tilde{S} \in B] \cap A \cap [N = n])$

$\quad = \sum_{n=1}^\infty P([(S_{n+1} - S_n, S_{n+2} - S_n, \ldots) \in B] \cap (A \cap [N = n]))$

$\quad\quad$ with $A \cap [N = n] \in \mathcal{F}(S_1, \ldots, S_n)$

$\quad = \sum_{n=1}^\infty P([(S_{n+1} - S_n, S_{n+2} - S_n, \ldots) \in B])\, P(A \cap [N = n])$

$\quad = P(\mathsf{S} \in B) \sum_{n=1}^\infty P(A \cap [N = n])$

(b) $\quad = P(\mathsf{S} \in B)\, P(A).$

Set $A = \Omega$ in (b) to conclude that $\tilde{S} \cong \mathsf{S}$. Then use $P(\tilde{S} \in B) = P(\mathsf{S} \in B)$ to rewrite (b) as

(c) $\quad P([\tilde{S} \in B] \cap A) = P(\tilde{S} \in B)\, P(A),$

which is the statement of independence. □

Exercise 7.1 (Manipulating stopping times) Let N_1 and N_2 denote stopping times relative to an $\nearrow$ sequence of σ-fields $\mathcal{A}_1 \subset \mathcal{A}_2 \subset \cdots$. Show that $N_1 \wedge N_2$, $N_1 \vee N_2$, $N_1 + N_2$, and $N_o \equiv i$ are all stopping times.

Definition 7.2 Define *waiting times* for return to the origin by

$$(5)\qquad \begin{array}{lll} W_1 \equiv \min\{n : S_n = 0\} & \text{with } W_1 = +\infty & \text{if the set is empty,} \\ \cdot & \cdot & \\ \cdot & \cdot & \\ \cdot & \cdot & \\ W_k \equiv \min\{n > W_{k-1} : S_n = 0\} & \text{with } W_k = +\infty & \text{if the set is empty.} \end{array}$$

Then define $T_k \equiv W_k - W_{k-1}$, with $W_0 \equiv 0$, to be the *interarrival times* for return to the origin.

Proposition 7.2 If $P(S_n = 0 \text{ i.o.}) = 1$, then $T_1, T_2, \ldots$ are well-defined rvs and are, in fact, iid.

Proof. Clearly, each W_k is always an extended-valued rv, and the condition $P(S_n = 0 \text{ i.o.}) = 1$ guarantees that $W_k(\omega)$ is well-defined for all $k \geq 1$ for a.e. ω.

Now, $T_1 = W_1$ is clearly a stopping time. Thus, by the strong Markov property, T_1 is independent of the rv $\tilde{S}^{(1)} \equiv \tilde{S}$ with kth coordinate $\tilde{S}_k^{(1)} \equiv \tilde{S}_k \equiv S_{T_1+k} - S_{T_1}$ and $\tilde{S}^{(1)} \equiv \tilde{S} \cong \vec{S}$. Thus T_2 is independent of the rv $\tilde{S}^{(2)}$ with kth coordinate $\tilde{S}_k^{(2)} \equiv \tilde{S}_{T_2+k}^{(1)} - \tilde{S}_{T_2}^{(1)} = S_{T_1+T_2+k} - S_{T_1+T_2}$ and $\tilde{S}^{(2)} \cong \tilde{S}^{(1)} \cong \vec{S}$. Continue with $\tilde{S}^{(3)}$, etc. [Note the relationship to interarrival times of a Bernoulli process.] $\square$

Exercise 7.2 (Wald's identity) (a) Suppose $X_1, X_2, \ldots$ are iid with mean μ, and N is a stopping time with finite mean. Show that $S_n \equiv X_1 + \cdots + X_n$ satisfies

$$(6)\qquad \mathrm{E}S_N = \mu \mathrm{E}N.$$

(b) Suppose each X_k equals 1 or -1 with probability p or $1-p$ for some $0 < p < 1$. Then define the rv $N \equiv \min\{n : S_n \text{ equals } -a \text{ or } b\}$, where a and b are strictly positive integers. Show that N is a stopping time that is a.s. finite. Then evaluate the mean $\mathrm{E}N$. [Hint. $[N \geq k] \in \mathcal{F}(S_1, \ldots, S_{k-1})$, and is thus independent of X_k, while $S_N = \sum_{k=1}^{\infty} X_k 1_{[N \geq k]}$.]

8 Convergence of Series of Independent RVs

In section 8.4 we proved the SLLN after recasting it (via Kronecker's lemma) as a theorem about a.s. convergence of infinite series. In this section we consider the convergence of infinite series directly. Since the convergence set of a series is a tail event (recall remark 7.2.1), convergence can happen only with probability 0 or 1. Moreover, the first theorem below seems to both limit the possibilities and broaden the possible approaches to them. All proofs are given at the end of this section.

Theorem 8.1 Let $X_1, X_2, \ldots$ be independent. Then, for some rv S, we have

(1) $$S_n \equiv \sum_{k=1}^n X_k \to_{a.s.} S \quad \text{iff} \quad \sum_{k=1}^n X_k \to_p S \quad \text{iff} \quad \sum_{k=1}^n X_k \to_d S.$$

[We will show the first equivalence now, and leave the second until section 10.2.]

Theorem 8.2 (The 2-series theorem) Let $X_1, X_2, \ldots$ be independent rvs for which $X_k \cong (\mu_k, \sigma_k^2)$. Let $S_n \equiv \sum_{k=1}^n X_k$ and $S_{0,n} \equiv \sum_{k=1}^n (X_k - \mu_k)$. **(a)** Then

(2) $$\sum_{k=1}^n \mu_k \to \mu \quad \text{and} \quad \sum_{k=1}^\infty \sigma_k^2 < \infty \quad \text{imply} \quad S_n \equiv \sum_{k=1}^n X_k \to_{a.s.} (\text{some rv } S).$$

Of course, in this situation $S_{0,n} \to_{a.s.} S_0 \equiv S - \mu$. Moreover,

(3) $$\mathrm{E}S = \mu \equiv \textstyle\sum_{k=1}^\infty \mu_k \qquad \mathrm{Var}[S] = \sigma^2 \equiv \sum_{k=1}^\infty \sigma_k^2, \qquad \text{and} \qquad S_n \to_{\mathcal{L}_2} S.$$

(b) Further, if all $|X_k| \le$ (some c), then (including converses) both:

(4) $$S_{0,n} \equiv \sum_{k=1}^n (X_k - \mu_k) \to_{a.s.} (\text{some rv } S_0) \quad \text{if and only if} \quad \sum_{k=1}^\infty \sigma_k^2 < \infty.$$

(5) $$S_n \equiv \sum_{k=1}^n X_k \to_{a.s.} (\text{some rv } S) \quad \text{iff} \quad \sum_{k=1}^n \mu_k \to \mu \ \text{ and } \ \sum_{k=1}^\infty \sigma_k^2 < \infty.$$

If a series is to converge, the size of its individual terms must be approaching zero. Thus the rvs must be effectively bounded. Thus truncation should be particularly effective for demonstrating the convergence of series.

Theorem 8.3 (The 3-series theorem) Let $X_1, X_2, \ldots$ be independent rvs.
(a) Define $X_k^{(c)}$ to be the trimmed X_k that equals X_k or 0 as $|X_k| \le c$ or as $|X_k| > c$. Then the series

(6) $$S_n \equiv \sum_{k=1}^n X_k \to_{a.s.} (\text{some rv } S)$$

if and only if for some $c > 0$ the following three series *all* converge:

(7) $$\mathrm{I}_c \equiv \sum_{k=1}^\infty P(|X_k| > c), \qquad \mathrm{II}_c \equiv \sum_{k=1}^\infty \mathrm{Var}[X_k^{(c)}], \qquad \mathrm{III}_c \equiv \sum_{k=1}^\infty \mathrm{E}X_k^{(c)}.$$

(b) The condition (7) holds for some $c > 0$ if and only if it holds for all $c > 0$.
(c) If either I_c, II_c, or III_c diverges for any $c > 0$, then $\sum_{k=1}^n X_k$ diverges a.s.

Example 8.1 Suppose $X_1, X_2, \ldots$ are independent and are uniformly bounded. They are assumed to be independent of the iid Rademacher rvs $\epsilon_1, \ldots, \epsilon_n$. Then

(8) $\sum_{k=1}^n \epsilon_k X_k \to_{a.s.}$ (some rv S) if and only if $\sum_{k=1}^\infty \sigma_k^2 < \infty$.

Moreover, $S \cong (0, \sum_{k=1}^\infty \sigma_k^2)$. [This is immediate from the 2-series theorem.] □

Exercise 8.1 Suppose $X_1, X_2, \ldots$ are iid with $P(X_k = 0) = P(X_k = 2) = \frac{1}{2}$. Show that $\sum_{k=1}^n X_k/3^k \to_{a.s.}$ (some S), and determine the mean, variance, and the name of the df F_S of S. Also determine the characteristic function of S (at some point after chapter 9).

Exercise 8.2 (a) Show that $\sum_{k=1}^n a^k X_k \to_{a.s.}$ (some S) when $X_1, X_2, \ldots$ are independent with $X_k \cong \text{Uniform}(-k, k)$ for $k \geq 1$, and where $0 < a < 1$.
(b) Evaluate the mean and the variance (give a simple expression) of S.

Exercise 8.3 Let $X_1, X_2, \ldots$ be arbitrary rvs with all $X_k \geq 0$ a.s. Let $c > 0$ be arbitrary. Then $\sum_{k=1}^\infty \mathrm{E}(X_k \wedge c) < \infty$ implies that $\sum_{k=1}^n X_k \to_{a.s.}$ (some rv S). The converse holds for independent rvs.

Exercise 8.4 (a) Let $Z_1, Z_2, \ldots$ be iid $N(0,1)$ rvs. Show that

$$\sum_{k=1}^\infty [Z_{2k-1}^2 + Z_{2k}^2]/2^k \to_{a.s.} \text{(some rv)},$$
$$\sum_{n=1}^\infty \left[\sum_{k=1}^n Z_k/2^{n+k}\right] \to_{a.s.} \text{(some rv)},$$

and determine (if possible) the mean, variance, and distribution of the limiting rvs.
(b) Let $Y_1, Y_2, \ldots$ be iid Cauchy$(0,1)$ rvs. Does $\sum_{k=1}^\infty Y_k/2^k \to_{a.s.}$(some rv)? If so, what is the distribution of the limit?

Proofs

Proof. Consider theorem 8.1. Now, $\to_{a.s.}$ always implies $\to_p$. So we turn to the converse. Suppose $S_n \to_p S$ (which is equivalent to $S_m - S_n \to_p 0$). To establish $S_n \to_{a.s.}$, it is enough to verify (2.3.7) that for all $\epsilon > 0$ and $\theta > 0$ we have

(9) $P(\max_{n \leq m \leq N} |S_m - S_n| \geq \epsilon) < \theta$ for all $n \geq$ (some $n_{\epsilon,\theta}$).

But Ottaviani–Skorokhod's inequality 8.3.4 gives

(a) $P(\max_{n \leq m \leq N} |S_m - S_n| \geq \epsilon) = P(\max_{n < m \leq N} |\sum_{n+1}^m X_k| \geq \epsilon)$

(b) $\leq P(|\sum_{n+1}^N X_k| \geq \epsilon/2)/[1 - \max_{n<m\leq N} P(|\sum_{n+1}^m X_k| > \epsilon/2)]$

(c) $\leq P(|S_N - S_n| \geq \epsilon/2)/[1 - \max_{n \leq m \leq N} P(|S_m - S_n| > \epsilon/2)]$

(d) $= o(1)/[1 - o(1)] < \theta$ for all $n, N \geq$ (some $n_{\epsilon,\theta}$),

using $S_N - S_n \to_p 0$ for (d). Thus (9) holds, and $S_n \to_{a.s.}$ (some rv S'). The a.s. limit S' equals S a.s. by proposition 2.3.4. □

Proof. Consider theorem 8.2, part (a): We first verify (2). By theorem 8.1, to establish that $S_{0,n} \to_{a.s.}$ (some S) we need only show that $S_{0,m} - S_{0,n} \to_p 0$. But this follows immediately from Chebyshev's inequality, since

$$P(|S_{0,m} - S_{0,n}| \ge \epsilon) \le \mathrm{Var}[S_{0,m} - S_{0,n}]/\epsilon^2 \le \textstyle\sum_{n+1}^{\infty} \sigma_k^2/\epsilon^2 < \epsilon$$

for all sufficiently large n. Thus (2) holds, since $S_n = S_{0,n} + \mu_n$.

We next verify (3) that $\mathrm{Var}[S] = \mathrm{Var}[S_0] = \sigma^2$ and $\mathrm{E}S = \mu$. Fatou gives

(e) $$\mathrm{E}(S_0^2) = \mathrm{E}(\lim S_{0,n}^2) = \mathrm{E}(\underline{\lim}\, S_{0,n}^2) \le \underline{\lim}\, \mathrm{E}(S_{0,n}^2) = \underline{\lim} \textstyle\sum_1^n \sigma_k^2 = \sigma^2,$$

and $\mathrm{E}(S_0^2) \ge \sigma^2$ since

(f) $$\mathrm{E}(S_0^2) = \mathrm{E}\{(\textstyle\sum_1^n (X_k - \mu_k))^2\} + \mathrm{E}\{(\sum_{n+1}^{\infty} (X_k - \mu_k))^2\} \ge \mathrm{E}(S_{0,n}^2) \to \sigma^2$$

(as the two rvs are independent, the first has mean 0, and both have finite variance (as follows from (e)). Thus $\mathrm{E}(S_0^2) = \sigma^2$. Inasmuch as both $S_{0,n} \to_{a.s.} S_0$ and $\mathrm{E}(S_{0,n}^2) \to \mathrm{E}(S_0^2)$, the Vitali theorem gives $S_{0,n} \to_{\mathcal{L}_2} S_0$. Then exercise 3.5.1b and Vitali show that $\mathrm{E}(S_0) = \lim \mathrm{E}(S_{0,n}) = \lim 0 = 0$. As $S = S_0 + \mu$, we have $\mathrm{E}S = \mathrm{E}(S_0 + \mu) = \mu$ and $\mathrm{Var}[S] = \mathrm{Var}[S_0] = \mathrm{E}(S_0^2) = \sigma^2 = \sum_1^{\infty} \sigma_k^2$. □

The proof of part (b) of the 2-series theorem above will require a converse of Kolmogorov's inequality that is valid for bounded rvs.

Inequality 8.1 (Kolmogorov's other inequality) Consider independent zero-mean rvs X_k, and set $S_k \equiv X_1 + \cdots + X_k$ for $1 \le k \le n$. Suppose $|X_k| \le (\text{some } M) < \infty$ for all k. Then

(10) $$P(\max_{1 \le k \le n} |S_k| \le \lambda) \le (\lambda + M)^2 / \textstyle\sum_{k=1}^n \sigma_k^2 \qquad \text{for all } \lambda > 0.$$

Proof. Let $A_k \equiv [\max_{1 \le j < k} |S_j| \le \lambda < |S_k|]$. Let $M_n \equiv [\max_{1 \le k \le n} |S_k|]$. We give another first passage argument. Thus

(a) $$\mathrm{E}\left\{S_n^2 1_{\sum_1^n A_k}\right\} = \textstyle\sum_1^n \mathrm{E}\{S_n^2 1_{A_k}\} = \sum_1^n \mathrm{E}\{[S_k + (S_n - S_k)]^2 1_{A_k}\}$$

$$= \textstyle\sum_1^n \{\mathrm{E}(S_k^2 1_{A_k}) + 2 \cdot 0 + P(A_k)\, \mathrm{E}(S_n - S_k)^2\} \qquad \text{by independence}$$

(b) $$\le (\lambda + M)^2 \textstyle\sum_1^n P(A_k) + \sum_1^n P(A_k) \sum_{j=k+1}^n \sigma_j^2$$

(c) $$\le \{(\lambda + M)^2 + \mathrm{Var}[S_n]\}(1 - P(M_n \le \lambda)),$$

where in step (b) we take advantage of $|S_k| \le |S_{k-1}| + |X_k| \le \lambda + M$ on A_k. We also note that

(d) $$\mathrm{E}\left\{S_n^2 1_{\sum_1^n A_k}\right\} = \mathrm{E}S_n^2 - \mathrm{E}\{S_n^2 1_{[M_n \le \lambda]}\} \quad \ge \mathrm{Var}[S_n] - \lambda^2 P(M_n \le \lambda).$$

using $|S_n| \le \lambda$ on the event $[M_n \le \lambda]$ to obtain (d). Combining (c) and (d) and doing algebra gives

(e) $$P(M_n \le \lambda) = \frac{(\lambda + M)^2}{(\lambda + M)^2 + \mathrm{Var}[S_n] - \lambda^2} \quad \le \frac{(\lambda + M)^2}{\mathrm{Var}[S_n]}.$$ □

Proof. Consider Theorem 8.2, part(b): Consider first the forward half of (4). Since $S_{0,n} \to_{a.s.}$ (some rv S_0), for some sufficiently large λ we have

(a) $\quad 0 < P(\sup_n |S_{0,n}| \le \lambda)$

$\quad = \lim_N P(\max_{1\le n\le N} |S_{0,n}| \le \lambda)$ since measures are monotone

(b) $\quad \le \lim_N (\lambda + c)^2 / \sum_1^N \sigma_k^2$ by Kolmogorov's other inequality 8.1

(c) $\quad = (\lambda + c)^2 / \sum_1^\infty \sigma_k^2\,.$

Then (c) implies that $\sum_1^\infty \sigma_k^2 < \infty$. Conversely, $\sum_1^\infty \sigma_k^2 < \infty$ implies by (2) that $S_{0,n} \to_{a.s.}$ (some rv S_0). So, both halves of (4) hold.

Consider (5). Again, (2) gives the converse half. Consider the forward half. Suppose that $S_n \to_{a.s.} S$. The plan is first to symmetrize, so that we can use (4) to prove (5). Let X_n''s be independent, and independent of the X_n's with $X_n' \cong X_n$; then $X_n^s \equiv X_n - X_n'$ denotes the symmetrized rv. Since $\to_{a.s.}$ depends only on the finite-dimensional distributions, the given fact that $S_n \to_{a.s.} S$ implies that the rv $S_n' \equiv \sum_1^n X_k' \to_{a.s.}$ (some rv S') $\cong S$. We can thus claim that

(d) $\quad S_n^s \equiv \sum_1^n X_k^s \to_{a.s.} S^s \equiv S - S'.$

Now, $|X_n^s| \le 2c$; thus (d) and (4) imply that $\sum_1^\infty \mathrm{Var}[X_n^s] < \infty$. Thus

(e) $\quad \sum_1^\infty \sigma_n^2 = \sum_1^\infty \mathrm{Var}[X_n^s]/2 < \infty.$

Now, (e) and (2) imply that $\sum_1^n (X_k - \mu_k) \to_{a.s.}$ (some rv S_0) with mean 0. Thus

(f) $\quad \sum_1^n \mu_k = [\,\sum_1^n X_k\,] - [\,\sum_1^n (X_k - \mu_k)\,] \to_{a.s.} S - S_0.$

Thus $S = S_0 + \mu$ with $\mu \equiv \sum_1^\infty \mu_k$ convergent, and the forward half of (5) holds. □

Proof. Consider the 3-series theorem. Consider (a) and (b) in its statement: Suppose that the 3 series converge for at least one value of c. Then II and III imply that $\sum_1^n X_k^{(c)} \to_{a.s.}$ by (2). Thus $\sum_1^n X_k \to_{a.s.}$ by proposition 8.2.1, since I $< \infty$ implies that $X_1, X_2, \ldots$ and $X_1^{(c)}, X_2^{(c)}, \ldots$ are Khinchin equivalent sequences.

Suppose that $\sum_1^n X_k \to_{a.s.}$. Then for all $c > 0$ we have $P(|X_n| > c \text{ i.o.}) = 0$, so that $I < \infty$ holds for all $c > 0$ by the second Borel–Cantelli lemma. Thus $\sum_1^n X_k^{(c)} \to_{a.s.}$ for all c, since I $< \infty$ implies that $X_1^{(c)}, X_2^{(c)}, \ldots$ and $X_1, X_2, \ldots$ are Khinchine equivalent sequences. Thus II $< \infty$ and III $< \infty$ for all c by the 2-series theorem result (4).

Consider (c). Kolmogorov's 0-1 law shows that S_n either converges a.s. or else diverges a.s.; and it is not convergent if one of the three series fails to converge. □

$\mathcal{L}_2$-Convergence of Infinite Series, and A.S. Convergence

Exercise 8.5 ($\mathcal{L}_2$-convergence of series) Let $X_1, X_2, \ldots$ be independent rvs in $\mathcal{L}_2$, where X_k has mean μ_k and variance σ_k^2. Then the sum $S_n \equiv X_1 + \cdots + X_n$ has mean $m_n \equiv \sum_{k=1}^n \mu_k$ and variance $v_n^2 \equiv \sum_{k=1}^n \sigma_k^2$. Show that

$$(11) \qquad S_n \to_{\mathcal{L}_2} (\text{some rv } S) \text{ if and only if } m_n \to (\text{some } \mu) \text{ and } v_n^2 \to (\text{some } \sigma^2).$$

If $S_n \to_{\mathcal{L}_2} S$, then $\mathrm{E}S = \mu$ and $\mathrm{Var}[S] = \sigma^2$.

Exercise 8.6 (Chow–Teicher) Let $X_1, X_2, \ldots$ be iid with finite mean. Suppose the series of real numbers $\sum_1^n a_k$ converges, where the $|a_k|$ are uniformly bounded. Show that $\sum_1^n a_k X_k \to_{a.s.}$ (some rv S).

Other Generalizations of the LLNs

Exercise 8.7 The following (with $r = 1$) can be compared to theorem 8.4.4. If $X_1, X_2, \ldots$ are independent with 0 means, then

$$(12) \qquad \sum_1^\infty \mathrm{E}|X_n|^{2r}/n^{r+1} < \infty \text{ for some } r \geq 1 \qquad \text{implies} \qquad S_n/n \to_{a.s.} 0.$$

Exercise 8.8 (Chung) Here is an even more general variation on theorem 8.4.4. Suppose that $\phi > 0$ is even and continuous, and either $\phi(x)/x \nearrow$ but $\phi(x)/x^2 \searrow$ or else $\phi(x) \nearrow$ but $\phi(x)/x \searrow$. Let $b_n \nearrow \infty$. Let $X_1, X_2, \ldots$ be independent with 0 means. Then

$$(13) \qquad \begin{array}{ll} \sum_{n=1}^\infty \mathrm{E}\phi(X_n)/\phi(b_n) < \infty & \text{implies both} \\ \sum_{n=1}^\infty X_n/b_n \to_{a.s.} (\text{some rv}) & \text{and} \quad \sum_1^n X_k/b_n \to_{a.s.} 0. \end{array}$$

The WLLN is taken up again in sections 10.1 and 10.2, after the characteristic function tool has been introduced in chapter 9.

9 Martingales

Definition 9.1 (Martingales) (a) Consider the sequence of rvs $S_1, S_2, \ldots$ defined on a probability space $(\Omega, \mathcal{A}, P)$ and adapted to an $\nearrow$sequence of σ-fields $\mathcal{A}_1 \subset \mathcal{A}_2 \subset \cdots$. Call it a *martingale* (abbreviated *mg*) if $\mathrm{E}|S_k| < \infty$ for all k, and

(1) $\quad \mathrm{E}(S_k|\mathcal{A}_i) =_{a.s.} S_i \qquad$ for all $i \le k$ in the index set.

If $(S_k, \mathcal{A}_k), k \ge 1$ is a mg, then the increments $X_k \equiv S_k - S_{k-1}$ are called the *martingale differences*.

(b) Let I denote a subinterval of the extended real line $\bar{R}$. A collection $\{S_t : t \in I\}$ of rvs on some $(\Omega, \mathcal{A}, P)$ that is adapted to an $\nearrow$family of σ-fields $\{\mathcal{A}_t : t \in I\}$ is called a *martingale* if $\mathrm{E}|S_t| < \infty$ for all $t \in I$, and

(2) $\quad \mathrm{E}(S_t|\mathcal{A}_r) =_{a.s.} S_r \qquad$ for all $r \le t$ in I.

(c) If "=" is replaced by "$\ge$" in either of (1) or (2), then either of $\{S_k : k \ge 1\}$ or $\{S_t : t \in I\}$ is called a *submartingale* (or *submg*).

Example 9.1 (The prototypical example) Let $X_1, \ldots, X_n$ denote independent rvs with 0 means, and set $S_k \equiv X_1 + \cdots + X_k$ and $\mathcal{A}_k \equiv \sigma[X_1, \ldots, X_k]$ for $1 \le k \le n$. Then the sequence of partial sums satisfies

(3) $\quad (S_k, \mathcal{A}_k),\ 1 \le k \le n,\ $ is a mg,

while (provided X_k also has finite variance σ_k^2)

(4) $\quad (S_k^2, \mathcal{A}_k),\ 1 \le k \le n,\ $ is a submg.

The first claim is trivial, and the second holds, since

$$\mathrm{E}(S_k^2|\mathcal{A}_i) = \mathrm{E}\{S_i^2 + 2S_i(S_k - S_i) + (S_k - S_i)^2|\mathcal{A}_i\}$$

(a) $$\ge S_i^2 + 2S_i\mathrm{E}\{S_k - S_i|\mathcal{A}_i\} + 0 = S_i^2 + 0 + 0 = S_i^2,$$

using (7.4.20) and (7.4.16). □

Exercise 9.1 (Equivalence) (a) Show that $(S_t, \mathcal{A}_t), t \in I$, is a martingale if and only if all $\mathrm{E}|S_t| < \infty$ and

(5) $$\int_{A_r} S_t\,dP = \int_{A_r} S_r\,dP \qquad \text{for all } A_r \in \mathcal{A}_r \text{ and all } r \le t \text{ with } r, t \in I.$$

(b) For a submartingale, just replace "=" by "$\ge$" in (5).

Notation 9.1 We will use the following notational system:

(6) $$\begin{cases} \text{mg} & \text{and} & = & \text{for a martingale.} \\ \text{submg} & \text{and} & \ge & \text{for a submartingale.} \\ \text{s-mg} & \text{and} & \geqq & \text{for a s-mg (mg or submg, as the case may be).} \end{cases}$$

Thus $(S_t, \mathcal{A}_t),\ t \in I$, is a s-mg if and only if all $\mathrm{E}\,|S_t| < \infty$ and

(7) $$\int_{A_r} S_t\,dP \geqq \int_{A_r} S_r\,dP \quad \text{for all } A_r \in \mathcal{A}_r, \text{ and for all } r \le t \text{ with } r, t \in I.$$ □

Exercise 9.2 Turn $(S_k^2, \mathcal{A}_k),\ 1 \le k \le n$ in (4) by centering it appropriately.

10 Maximal Inequalities, Some with $\nearrow$ Boundaries

Inequality 10.1 (Monotone inequality) For arbitrary rvs $X_1, \ldots, X_n$ and for constants $0 < b_1 \le \cdots \le b_n$ we let $S_k \equiv X_1 + \cdots + X_k$ and obtain

$$(1) \qquad \left(\max_{1\le k\le n} \frac{|S_k|}{b_k} \right) \le 2 \max_{1\le k\le n} \left| \sum_{i=1}^{k} \frac{X_i}{b_i} \right| .$$

If all $X_i \ge 0$, then we may replace 2 by 1. [This also holds in higher dimensions, when properly formulated. See Shorack and Smythe(1976).]

Proof. Define $b_0 = 0$, $X_0 = 0$, $Y_j = X_j/b_j$, and $T_k = \sum_{j=1}^{k} Y_j$. Then

$$(a) \qquad S_k = \sum_{j=1}^{k} b_j \Delta T_j = \sum_{j=1}^{k} \Delta T_j \sum_{i=1}^{j} \Delta b_i = \sum_{i=1}^{k} T_{ik} \Delta b_i \,,$$

where $\Delta b_j \equiv b_j - b_{j-1}$, $\Delta T_j \equiv T_j - T_{j-1}$, and $T_{ik} \equiv \sum_{j=i}^{k} Y_j$. As $\sum_{i=1}^{k}(\Delta b_i/b_k) = 1$ with each $\Delta b_i/b_k \ge 0$, we have

$$(b) \qquad \left(\max_{1\le k\le n} |S_k|/b_k \right) \le \left(\max_{1\le k\le n} \{ \textstyle\sum_{i=1}^{k} |T_{ik}| \, (\Delta b_i/b_k) \} \right)$$

$$(c) \qquad \le \max_{1\le k\le n} \left(\max_{1\le i\le k} |T_{ik}| \right) \quad \text{since an average does not exceed the maximum}$$

$$(d) \qquad \le 2 \left(\max_{1\le k\le n} |T_k| \right) .$$

Note that 1 can replace 2 in step (d) if all $X_i \ge 0$. □

Martingale Maximal Inequalities

Inequality 10.2 (Doob) Let $(S_k, \mathcal{A}_k), 1 \le k \le n$, be a submg and define the maximum $M_n \equiv \max_{1\le k\le n} S_k$. Then

$$(2) \qquad \lambda P(M_n \ge \lambda) \le \int_{[M_n \ge \lambda]} S_n \, dP \le \mathrm{E} S_n^+ \le \mathrm{E}|S_n| \qquad \text{for all } \lambda > 0,$$

$$(3) \qquad P(M_n \ge \lambda) \le \inf_{r>0} \mathrm{E}(e^{rS_n})/e^{r\lambda} \qquad \text{for all } \lambda > 0.$$

If $(S_k, \mathcal{A}_k)$, $1 \le k \le n$ is a zero-mean mg with all of the variances $\mathrm{E}\, S_k^2 < \infty$, then $(S_k^2, \mathcal{A}_k)$, $1 \le k \le n$ is a submg. This allows the maximum to be bounded by

$$(4) \qquad P(M_n \ge \lambda) \le \mathrm{Var}[S_n]/\lambda^2 \qquad \text{for all } \lambda > 0.$$

[This last is Kolmogorov's inequality, valid for zero-mean mgs.]

Proof. Since $\mathrm{E}(S_n|\mathcal{A}_k) \geq S_k$ a.s. by the definition of a submg, we have

(a) $\int_{A_k} S_n\, dP = \int_{A_k} \mathrm{E}(S_n|\mathcal{A}_k)\, dP \geq \int_{A_k} S_k\, dP \qquad$ for all $A_k \in \mathcal{A}_k$

by (7.4.1) in the definition of conditional expectation. Now let

(b) $A_k \equiv [\max_{1\leq j<k} S_j < \lambda \leq S_k]$,

so that k is the first index for which S_k is $\geq \lambda$. Then

$$\lambda P(M_n \geq \lambda) = \lambda \textstyle\sum_1^n P(A_k) \leq \sum_1^n \int_{A_k} S_k\, dP$$

(c) $\quad \leq \sum_1^n \int_{A_k} S_n\, dP \qquad$ using (a)

(d) $\quad = \int_{[M_n \geq \lambda]} S_n\, dP \leq \int_{[M_n\geq\lambda]} S_n^+\, dP \leq \int S_n^+\, dP \leq \int |S_n|\, dP$,

as claimed. In a s-mg context, these are called "first passage time" proofs. To this end, set τ equal to k on A_k for $1 \leq k \leq n$, and set τ equal to $n+1$ on $(\sum_1^n A_k)^c$. Then τ is the first passage time to the level λ.

That $\{(\exp(rS_k), \mathcal{A}_k), 1 \leq k \leq n\}$ is also a submg for any $r > 0$ follows from Jensen's inequality for conditional expectation with an $\nearrow g(\cdot)$ via

(e) $\mathrm{E}\,(e^{rS_k} \,|\, \mathcal{A}_j) \equiv \mathrm{E}\,(g_r(S_k) \,|\, \mathcal{A}_j) \geq_{a.s.} g_r(\,\mathrm{E}\,(S_k \,|\, \mathcal{A}_j)) \geq g_r(S_j) = e^{rS_j}$.

Applying Doob's first inequality (2) to (e) gives (3). [This is often sharper than (2), though it requires the existence of the moment generating function $\mathrm{E}\exp(rS_n)$.] When $(S_k, \mathcal{A}_k)$ is a mg, then $(S_k^2, \mathcal{A}_k)$ is also a submg (by another application of the same Jensen's inequality), so applying (2) to the latter submg gives (4). □

Inequality 10.3 (Hájek–Rényi) Let $(S_k, \mathcal{A}_k), 1 \leq k \leq N$, be a mg with all $\mathrm{E}S_k = 0$. Let $X_k \equiv S_k - S_{k-1}$ have variance σ_k^2. Let $0 < b_1 \leq \cdots \leq b_N$. Then

(5) $$P\left(\max_{n\leq k\leq N} |S_k|/b_k \geq \lambda\right) \leq \frac{4}{\lambda^2}\left\{\sum_{k=1}^n \sigma_k^2/b_n^2 + \sum_{k=n+1}^N \sigma_k^2/b_k^2\right\} \text{ for all } \lambda > 0.$$

Proof. (We give the proof for independent rvs.) The monotone inequality bounds the maximum partial sum via

(6) $$\left(\max_{n\leq k\leq N} |S_k|/b_k\right) \leq 2\left(\max_{n\leq k\leq N}\left|\frac{S_n}{b_n} + \sum_{i=n+1}^k \frac{X_i}{b_i}\right|\right).$$

Applying Kolmogorov's inequality (4) to (6) gives

(a) $$P\left(\max_{n\leq k\leq N} |S_k|/b_k \geq \lambda\right) \leq (\lambda/2)^{-2}\left\{\mathrm{Var}[S_n/b_n] + \textstyle\sum_{k=n+1}^N \mathrm{Var}[X_k]/b_k^2\right\}$$

(b) $$\quad = (4/\lambda^2)\left\{\textstyle\sum_1^n \sigma_k^2/b_n^2 + \sum_{n+1}^N \sigma_k^2/b_k^2\right\}.$$

(A more complicated proof can eliminate the factor 4.) □

Exercise 10.1 To complete the proof of the Hájek–Rényi inequality for mgs, one can show that $T_k \equiv S_n/b_n + \sum_{n+1}^k X_i/b_i$ is such that $(T_k, \mathcal{A}_k), n \le k \le N$, is also a mg, and that $\mathrm{Var}[T_N]$ is equal to the right-hand side of (b). Do it.

Inequality 10.4 (Birnbaum–Marshall) Let $(S(t), \mathcal{A}(t)), 0 \le t \le \theta$, be a mg having $S(0) = 0$, $\mathrm{E}S(t) = 0$, and $\nu(t) = \mathrm{E}S^2(t)$ finite and continuous on $[0, \theta]$. Suppose that paths of S are right (or left) continuous. Let $q(\cdot) > 0$ on $(0, \theta]$ be ↗ and right (or left) continuous. Then

$$(7) \qquad P(\|S/q\|_0^\theta > \lambda) \le 4\lambda^{-2} \int_0^\theta [q(t)]^{-2}\, d\nu(t) \qquad \text{for all } \lambda > 0.$$

Proof. Because of right (or left) continuity and $S(0) = 0$, we have

$$\begin{aligned}
\text{(a)} \qquad P(\|S/q\|_0^\theta \le \lambda) &= P\left(\max_{0 \le i \le 2^n} |S(\theta i/2^n)|/q(\theta i/2^n) \le \lambda \text{ for all } n \ge 1 \right) \\
&= \lim P\left(\max_{0 \le i \le 2^n} |S(\theta i/2^n)|/q(\theta i/2^n) \le \lambda \right) \qquad \text{by proposition 1.2.2} \\
&\ge \lim \{1 - 4\lambda^{-2} \textstyle\sum_1^{2^n} \mathrm{E}[S^2(\theta i/2^n) - S^2(\theta(i-1)/2^n)]/q^2(\theta i/2^n)\} \quad \text{by (5)} \\
&= 1 - 4\lambda^{-2} \lim \textstyle\sum_1^{2^n} q^{-2}(\theta i/2^n)\, [\nu(\theta i/2^n) - \nu(\theta(i-1)/2^n)] \\
\text{(b)} \qquad &\to 1 - 4\lambda^{-2} \textstyle\int_0^\theta [q(t)]^{-2}\, d\nu(t) \qquad \text{using the MCT.}
\end{aligned}$$

□

Inequality 10.5 (Doob's $\mathcal{L}_r$-inequality) (i) Let $(S_k, \mathcal{A}_k)$, for $1 \le k \le n$, be a submg. Consider $M_n \equiv \max_{1 \le k \le n} S_k^+$. Let $r > 1$. Then

$$(8) \qquad \mathrm{E}\, M_n^r \le (\tfrac{r}{r-1})^r\, \mathrm{E}\{(S_n^+)^r\}\,.$$

(ii) Let $(S_k, \mathcal{A}_k)$, $1 \le k \le n$, be a mg. Let $M_n \equiv \max_{1 \le k \le n} |S_k|$. Let $r > 1$. Then

$$(9) \qquad \mathrm{E}\, M_n^r \le (\tfrac{r}{r-1})^r\, \mathrm{E}\{|S_n|^r\}\,.$$

Proof. Now, $(S_k^+, \mathcal{A}_k)$, for $1 \le k \le n$, is also a submg, by the conditional version of Jensen's inequality. (Or, refer to (13.1.7).) [Refer to (13.1.6) for case (ii).] Thus in case (i) we have

$$\begin{aligned}
\text{(a)} \qquad \mathrm{E}\, M_n^r &= \textstyle\int_0^\infty r\lambda^{r-1} P(M_n > \lambda)\, d\lambda \qquad \text{by (6.4.13)} \\
\text{(b)} \qquad &\le \textstyle\int_0^\infty r\lambda^{r-1}\, \lambda^{-1}\, \mathrm{E}\{S_n^+ 1_{[M_n \ge \lambda]}\}\, d\lambda \qquad \text{by Doob's inequality 10.2} \\
\text{(c)} \qquad &= \mathrm{E}\{S_n^+ \textstyle\int_0^{M_n} r\lambda^{r-2}\, d\lambda\} \qquad \text{by Fubini} \\
&= \mathrm{E}\{\, S_n^+\, (\tfrac{r}{r-1})\, M_n^{r-1}\, \} \\
\text{(d)} \qquad &\le (\tfrac{r}{r-1})\, (\mathrm{E}\{(S_n^+)^r\})^{1/r}\, (\mathrm{E}\{M_n^r\})^{(r-1)/r} \qquad \text{by Hölder's inequality,}
\end{aligned}$$

where $r^{-1} + s^{-1} = 1$ implies that $s = r/(r-1)$. So

$$\text{(e)} \qquad (\mathrm{E}\, M_n^r)^{1-(r-1)/r} \le (\tfrac{r}{r-1})\, (\mathrm{E}(S_n^+)^r)^{1/r}\,,$$

which gives the results. (Just change S_n^+ to $|S_n|$ for case (ii).) □

Hoffman–Jorgensen Inequalities

The following inequalities show that "in probability" control of the overall sum and of the maximal summand actually gives control of moments of sums of independent rvs.

Inequality 10.6 (Hoffmann–Jorgensen, probability form). Let $X_1, \ldots, X_n$ be independent rvs, and let $S_k \equiv X_1 + \cdots + X_k$ for $1 \le k \le n$. Let $\lambda, \eta > 0$. Then

$$(10)\qquad P(\max_{1\le k\le n} |S_k| > 3\lambda + \eta) \le \{P(\max_{1\le k\le n} |S_k| > \lambda)\}^2 + P(\max_{1\le i\le n} |X_i| > \eta)\,.$$

If the X_i's are also symmetric, then both

$$(11)\qquad P(\max_{1\le k\le n} |S_k| > 3\lambda + \eta) \le \{2P(|S_n| > \lambda)\}^2 + P(\max_{1\le i\le n} |X_i| > \eta) \qquad \text{and}$$

$$(12)\qquad P(|S_n| > 2\lambda + \eta) \le \{2P(|S_n| > \lambda)\}^2 + P(\max_{1\le i\le n} |X_i| > \eta)\,.$$

Inequality 10.7 (Hoffmann–Jorgensen, moment form). Let the rvs $X_1, \ldots, X_n$ be independent, and let $S_k \equiv X_1 + \cdots + X_k$ for $1 \le k \le n$. Suppose that each $X_i \in \mathcal{L}_r(P)$ for some $r > 0$. Then

$$(13)\qquad E\left(\max_{1\le k\le n} |S_k|^r\right) \le 2\,(4t_0)^r + 2\cdot 4^4\,\mathrm{E}\left(\max_{1\le i\le n} |X_i|^r\right),$$

where $t_0 \equiv \inf\{t > 0 : P(\max_{1\le k\le n} |S_k| > t) \le 1/(2\cdot 4^r)\}$.
If the X_i's are also symmetric, then

$$(14)\qquad \mathrm{E}|S_n|^r \le 2\,(3t_0)^r + 2\cdot 3^r\,\mathrm{E}\left(\max_{1\le i\le n} |X_i|^r\right),$$

where $t_0 \equiv \inf\{t > 0 : P(|S_n| > t) \le 1/(8\cdot 3^r)\}$.

Proof. Consider inequality 10.6. Let $\tau \equiv \inf\{k \le n : |S_k| > \lambda\}$. Then $[\tau = k]$ depends only on $X_1, \ldots, X_k$, and $[\max_{k\le n} |S_k| > \lambda] = \sum_{k=1}^n [\tau = k]$. On $[\tau = k]$, $|S_j| \le \lambda]$ if $j < k$, and for $j \ge k$,

$$(a)\qquad |S_j| = |S_j - S_k + X_k + S_{k-1}| \le \lambda + |X_k| + |S_j - S_k|\,;$$

hence

$$(b)\qquad \max_{1\le j\le n}|S_j| \le \lambda + \max_{1\le i\le n}|X_i| + \max_{k<j\le n}|S_j - S_k|\,.$$

Therefore, by independence,

$$\begin{aligned} &P(\tau = k, \max_{1\le k\le n}|S_k| > 3\lambda + \eta) \\ (c)\qquad &\le P(\tau = k, \max_{1\le i\le n}|X_i| > \eta) + P(\tau = k)\,P(\max_{k<j\le n}|S_j - S_k| > 2\lambda)\,. \end{aligned}$$

But $\max_{k<j\le n} |S_j - S_k| \le 2\max_{1\le k\le n} |S_k|$, and hence summing over k on both sides yields

$$(d)\qquad P(\max_{k\le n}|S_k| > 3\lambda + \eta) \le P(\max_{i\le n}|X_i| > \eta) + \{P(\max_{k\le n}|S_k| > \lambda)^2\}\,.$$

The second inequality follows from the first by Lévy's inequality 8.3.3.

For the symmetric case, first note that

(e) $|S_n| \le |S_{k-1}| + |X_k| + |S_n - S_k|\,,$

so that

$$P(\tau = k, |S_n| > 2\lambda + \eta)$$

(f) $$\le P(\tau = k,\, \max_{1\le i\le n} |X_i| > \eta) + P(\tau = k)\, P(|S_n - S_k| > \lambda)\,;$$

and hence summing over k then yields

$$P(|S_n| > 2\lambda + \eta)$$

(g) $$\le P(\max_{i\le n} |X_i| > \eta) + P(\max_{k\le n} |S_k| > \lambda)\, P(\max_{k\le n} |S_n - S_k| > \lambda)\,.$$

The third inequality again follows from Lévy's inequality. □

Proof. Consider inequality 10.7. Here is the proof of (14); the proof of (13) is similar. Let $u > t_0$. Then, using (12) for (i),

(h) $\mathrm{E}|S_n|^r = 3^r(\int_0^u + \int_u^\infty)\, P(\,|S_n| > 3t)\, d(t^r)$ by (6.4.13).

(i) $$\le (3u)^r + 4\cdot 3^r \int_u^\infty P(\,|S_n| > t)^2\, d(t^r) + 3^r \int_u^\infty P(\max_{1\le i\le n} |X_i| > t) d(t^r)$$

(j) $$\le (3u)^r + 4\cdot 3^r P(\,|S_n| > u) \int_u^\infty P(\,|S_n| > t)\, d(t^r) + 3^r\, \mathrm{E}(\max_{1\le i\le n} |X_i|^r)\,.$$

Since $4\cdot 3^r P(|S_n| > u) \le \frac{1}{2}$ by our choice of u, applying (6.4.13) again (to (j)) gives

(k) $\mathrm{E}|S_n|^r \le (3u)^r + \frac{1}{2}\,\mathrm{E}|S_n|^r + 3^r\, \mathrm{E}(\max_{1\le i\le n} |X_i|^r)\,.$

Simple algebra now gives (14). □

Exercise 10.2 Provide the details in the case of (13)

Chapter 9

Characteristic Functions and Determining Classes

1 Classical Convergence in Distribution

Definition 1.1 (Sub-dfs) (a) Suppose we have rvs $X_n \cong F_n$ and X. We now wish to allow the possibility that X is an extended rv. In this case, we assume that H is a *sub-df* (we will not use the notation F in this context), and we will write $X \cong H$. The interpretation in the case of an extended rv X is that $H(-\infty) = P(X = -\infty)$, $H(x) = P(-\infty \le X \le x)$ for all $-\infty < x < \infty$, and $1 - H(+\infty) = P(X = +\infty)$. The set C_H of all points at which H is continuous is called the *continuity set* of H.
(b) If $F_n(x) \to H(x)$ as $n \to \infty$ at each $x \in C_H$ of a sub-df H, then we say that X_n (or F_n) *converges in sub-df* to X (or H), and we write $X_n \to_{sd} X$ (or $F_n \to_{sd} H$) as $n \to \infty$. [What has happened in the case of sub-df convergence is that amounts $H(-\infty)$ and $1 - H(+\infty)$ of mass have escaped to $-\infty$ and $+\infty$, respectively.]
(c) We have agreed that F_n, F, etc. denote a bona fide df, while H_n, H, etc. may denote a sub-df. Thus $F_n \to_d F$ (with letter F rather than letter H) will still imply that the limit is necessarily a bona fide df. [The next definition provides a condition that guarantees (in a totally obvious way, on R at least) that any possible limit is a bona fide df.]

Definition 1.2 (Tightness) A family $\mathcal{P}$ of distributions P on R is called *tight* if for each $\epsilon > 0$ there is a compact set (which for one-dimensional rvs is just a closed and bounded set) K_ϵ with

$$(1) \qquad P(K_\epsilon) = P(X \in K_\epsilon) \ge 1 - \epsilon \qquad \text{for all dfs } P \in \mathcal{P}.$$

Theorem 1.1 (Helly–Bray) If $F_n \to_d F$ and g is bounded and continuous a.s. F, then the expectations satisfy

$$(2) \qquad \int g\, dF_n = \mathrm{E}g(X_n) \to \mathrm{E}g(X) = \int g\, dF.$$

Conversely, if (2) holds for all bounded continuous g, then $F_n \to_d F$.
[Thus $F_n \to_d F$ if and only if $\int g\, dF_n \to \int g\, dF$ for all bounded and continuous g.]

Theorem 1.2 (Continuous mapping theorem; Mann–Wald) Suppose that $X_n \to X$ and suppose that g is continuous a.s. F. Then $g(X_n) \to_d g(X)$.

How do we establish that $F_n \to_d F$? We have the necessary and sufficient condition of the Helly–Bray theorem 1.1 (presented earlier as theorem 3.5.1). (We should now recall our definition of the determining class used in the context of the proof of theorem 3.5.1.) We can also show convergence in distribution of more complicated functions of rvs via Mann–Wald's continuous mapping theorem 1.2 (presented earlier as theorem 3.5.2); an example is given by

$$Z_n \to_d Z \quad \text{implies that} \quad Z_n^2 \to_d Z^2 \cong \chi_1^2$$

where $g(x) = x^2$. The concept of tightness was introduced above to guarantee that any possible limit is necessarily a bona fide df. This becomes more important in light of the next theorem.

Theorem 1.3 (Helly's selection theorem) Let $F_1, F_2, \ldots$ be any sequence of dfs. There necessarily exists a subsequence $F_{n'}$ and a sub-df H for which $F_{n'} \to_{sd} H$. If the subsequence of dfs is tight, then the limit is necessarily a bona fide df.

Corollary 1 Let $F_1, F_2, \ldots$ be any sequence of dfs. Let H be a fixed sub-df. Suppose every sd-convergent subsequence $\{F_{n'}\}$ satisfies $F_{n'} \to_{sd}$ (this same H). Then the whole sequence satisfies $F_n \to_{sd} H$. (Here is an alternative phrasing. Suppose every subsequence n' contains a further subsequence n'' for which $F_{n''}$ converges in distribution to this *one fixed* sub-df H. Then the whole sequence satisfies $F_n \to_{sd} H$.)

Proof. Let $r_1, r_2, \ldots$ denote a sequence which is dense in R. Using Bolzano–Weierstrass, choose a subsequence n_{1j} such that $F_{n_{1j}}(r_1) \to$ (some a_1). A further subsequence n_{2j} also satisfies $F_{n_{2j}}(r_2) \to$ (some a_2). Continue in this fashion. The diagonal subsequence n_{jj} converges to a_i at r_i for all $i \ge 1$. [This *Cantor diagonalization technique* is important. Learn it!] Define H_o on the r_i's via $H_o(r_i) = a_i$. Now define H on all real values via

(a) $\qquad H(x) = \inf\{H_o(r_i) : r_i > x\};$

this H is clearly is $\nearrow$ and takes values in $[0,1]$. We must now verify that H is also right-continuous, and that $F_{n_{jj}} \to_{sd} H$. That is, the diagonal subsequence, which we will now refer to as n', is such that $F_{n'} = F_{n'_j} = F_{n_{jj}} \to_{sd} H$.

The monotonicity of H_o trivially gives $\inf_{y \searrow x} H(y) \ge H(x)$. Meanwhile,

(b) $\qquad H(x) > H_o(r_{\kappa_\epsilon}) - \epsilon \qquad$ for some $x < r_{\kappa_\epsilon}$ that is sufficiently close to x

(c) $\qquad \ge H(y) - \epsilon \qquad$ for any $x < y < r_{\kappa_\epsilon}$

yields $\inf_{y \searrow x} H(y) \le H(x)$. Hence $\inf_{y \searrow x} H(y) = H(x)$, and H is right continuous.

We next show that $F_{n'_j}(x) = F_{n_{jj}}(x) \to H(x)$ for any $x \in \mathcal{C}_H$. Well,

(d) $\qquad F_{n'_j}(r_k) \le F_{n'_j}(x) \le F_{n'_j}(r_\ell) \qquad$ for all $r_k < x < r_\ell$.

Passing to the limit on j gives

(e) $\quad H_o(r_k) \le \underline{\lim}\, F_{n'_j}(x) \le \overline{\lim}\, F_{n'_j}(x) \le H_o(r_\ell) \qquad$ for all $r_k < x < r_\ell$.

Now let $r_k \nearrow x$ and $r_\ell \searrow x$ in (e) to get

(f) $\quad \lim_j F_{n'_j}(x) = H(x) \qquad$ for all $x \in \mathcal{C}_H \qquad$ (that is, $F_{n'_j} \to_{sd} H$).

Consider the corollary. Fact: Any bounded sequence of real numbers contains a convergent subsequence; and the whole original sequence converges if and only if all subsequential limit points are the same. Or, if every subsequence $a_{n'}$ contains a further subsequence $a_{n''}$ that converges to the one fixed number a_o, then we have $a_n \to a_o$. We effectively showed above that every subsequence $F_{n'}$ contains a further subsequence $F_{n''}$ for which

$$F_{n''}(x) \to (\text{the same } H(x)) \qquad \text{for each fixed } x \in C_H.$$

Thus the whole sequence has $F_n(x) \to H(x)$ for each $x \in C_H$. So, $F_n \to_{sd} H$. $\quad\square$

Exercise 1.1 (Convergence of expectations and moments)
(a) Suppose $F_n \to_{sd} H$ and that both $F_{n-}(a) \to H_-(a)$ and $F_n(b) \to H(b)$ for some constants $-\infty < a < b < \infty$ in C_H having $H(a) < H(b)$. Then

(3) $\quad \int_{[a,b]} g\, dF_n \to \int_{[a,b]} g\, dH \quad$ for all $g \in C_{[a,b]} \equiv \{g : g \text{ is continuous on } [a,b]\}$.

Moreover, if $F_n \to_{s.d.} H$, then

(4) $\quad \int g\, dF_n \to \int g\, dH \qquad$ for all $g \in C_0$,

where $C_0 \equiv \{g : g$ is continuous on R and $g(x) \to 0$ as $|x| \to \infty\}$.
(b) Suppose $F_n \to_d F$ and g is continuous on the line. Suppose $|g(x)|/\psi(x) \to 0$ as $|x| \to \infty$, where $\psi \ge 0$ has $\int \psi\, dF_n \le K < \infty$ for all n. Then $\int g\, dF_n \to \int g\, dF$.
(c) If $\mathrm{E}|X_n|^{r_0} <$ (some M) $< \infty$ for all large n, then $F_n \to_d F$ implies that

(5) $\quad \mathrm{E}|X_n|^r \to \mathrm{E}|X|^r \quad$ and $\quad \mathrm{E}X_n^k \to \mathrm{E}X^k \quad$ for $0 < r < r_0$ and $0 < k < r_0$.

(d) Let g be continuous. If $F_n \to_{sd} H$, then $\liminf \int |g|\, dF_n \ge \int |g|\, dH$.
[Actually, g continuous a.s. H suffices in (a), (b), and (d) above.]

Exercise 1.2 (Pólya's lemma) If $F_n \to_d F$ for a continuous df F, then

(6) $\quad \|F_n - F\| \to 0.$

Thus if $F_n \to_d F$ with F continuous and $x_n \to x$, then $F_n(x_n) \to F(x)$.

Exercise 1.3 (Verifying tightness) Suppose $X_n \cong F_n$. Show that $\{F_n : n \ge 1\}$ is tight if either

(a) $\quad \overline{\lim}\, \mathrm{E}|X_n|^r < \infty \qquad$ for some $r > 0$, $\qquad$ or

(b) $\quad F_n \to_d F$.

Equivalent Definitions of Convergence in Distribution

The condition $F_n(x) \to F(x)$ can be rewritten as $P_n((-\infty, x]) \to P((-\infty, x])$, and as $\mathrm{E}1_{(-\infty,x]}(X_n) \to \mathrm{E}1_{(-\infty,x]}(X)$. Thus $\to_d$ is reduced to computing expectations of the particularly simple function $1_{(-\infty,x]}$; but these simple functions have the disadvantage of being discontinuous.

Definition 1.3 (Closure, interior, and boundary) The *closure* of B is defined to be $\bar{B} \equiv \cap\{C : B \subset C \text{ and } C \text{ is closed}\}$, while $B^0 \equiv \cup\{U : U \subset B \text{ and } U \text{ is open}\}$ is called the *interior* of B. These have the property that $\bar{B}$ is the smallest closed set containing B, while B^0 is the largest open set contained within B. The *boundary* of B is defined to be $\partial B \equiv \bar{B} \setminus B^0$. A set B is called a P-*continuity set* if $P(\partial B) = 0$. (These definitions are valid on a general metric space, not just on R.)

Theorem 1.4 ($\to_d$ equivalencies) Let $F, F_1, F_2, \ldots$ be the dfs associated with the probability distributions $P_1, P_2, \ldots$. Let C_b denote all bounded, continuous functions g on R, and then let C_{bu} denote all bounded and uniformly continuous functions g on R. The following are equivalent:

(7) $\quad F_n \to_d F.$

(8) $\quad F_n(x) \to F(x) \quad$ for all x in a dense set.

(9) $\quad \mathrm{E}g(X_n) = \int g\,dF_n \to \int g\,dF = \mathrm{E}g(X) \quad$ for all g in C_b.

(10) $\quad \int g\,dF_n \to \int g\,dF \quad$ for all g in C_{bu}.

(11) $\quad \overline{\lim}\, P_n(B) \le P(B) \quad$ for all closed sets B.

(12) $\quad \underline{\lim}\, P_n(B) \ge P(B) \quad$ for all open sets B.

(13) $\quad \lim P_n(B) = P(B) \quad$ for all P-continuity sets B.

(14) $\quad \lim P_n(I) = P(I) \quad$ for all (even unbounded) P-continuity intervals I.

(15) $\quad L(F_n, F) \to 0 \quad$ for the Lévy metric L (see below).

Exercise 1.4 That (7)–(10) are equivalent is either trivial, or done previously. Cite the various reasons. Then show that (11)–(15) are also equivalent to $\to_d$.

Exercise 1.5 (Lévy's metric) For any dfs F and G define (the 45° distance between F and G)

(16) $\quad L(F, G) \equiv \inf\{\epsilon > 0 : F(x - \epsilon) - \epsilon \le G(x) \le F(x + \epsilon) + \epsilon \quad \text{for all } x\}.$

Show that L is a metric and that the set of all dfs under L forms a complete and separable metric space. Also show that $F_n \to_d F$ is equivalent to $L(F_n, F) \to 0$.

Convergence of Types

Definition 1.4 (Type) When $Y \cong (X - b)/a$ for some $a \neq 0$, we say that X and Y are of the same *type.* [Suppose that $X_n \to_d X$ where X is not degenerate. Then if $a_n \to a \neq 0$ and $b_n \to b$, we know from Slutsky's theorem that $(X_n - b_n)/a_n \to_d Y \cong (X - b)/a$.]

Theorem 1.5 (Convergence of types) Suppose $(X_n - b_n)/a_n \to_d X \cong F$, and $(X_n - \beta_n)/\alpha_n \to_d Y \cong G$, where $a_n > 0$, $\alpha_n > 0$, and both X and Y are nondegenerate. Then there exists $a > 0$ and a real b such that

$$(17) \qquad a_n/\alpha_n \to (\text{some positive } a) \qquad \text{and} \qquad (\beta_n - b_n)/a_n \to (\text{some real } b)$$

and $Y \cong (X - b)/a$ (or, equivalently, $G(x) = F(ax + b)$ for all x).

Remark 1.1 The classical CLT implies that if $X_1, X_2, \ldots$ are iid $(0, \sigma^2)$, then $S_n/\sqrt{n} \to_d N(0,1)$. The above theorem tells us that no matter how we normalize S_n, the only possible nondegenerate limits in distribution are normal distributions. Moreover, if $S_n/a_n \to_d$ (some rv), the limiting distribution can be nondegenerate only if $a_n/\sqrt{n} \to$ (some constant) $\in (0, \infty)$.

Exercise 1.6 (Proof of the convergence of types theorem) Prove theorem 1.5 on the convergence of types.
[Hint. Start with continuity points $x < x'$ of the df G and then continuity points y, y' of the df F for which $F(y) < G(x) \leq G(x') < F(y')$. Then for all n large enough one will have $a_n y + b_n \leq \alpha_n\, x + \beta_n \leq \alpha_n\, x' + \beta_n \leq a_n\, y' + b_n$.]

Higher Dimensions

If $X, X_1, X_2, \ldots$ are k-dimensional random vectors with dfs $F, F_1, F_2, \ldots$, then we say that X_n *converges in distribution* to X if

$$(18) \qquad F_n(x) \to F(x) \qquad \text{for all } x \in C_F,$$

just as in one dimension.

The Helly–Bray theorem, the Mann–Wald theorem, Helly's selection theorem, and Polya's lemma all hold in k dimensions; generalizations of the other results also hold. Moreover, if X_n' denotes the first j coordinates of X_n, with $1 \leq j < k$, then $X_n \to_d X$ implies $X_n' \to_d X'$.

Exercise 1.7 Prove the k-dimensional Helly–Bray theorem (along the lines of exercise 3.5.2) using Helly's selection theorem and Pólya's lemma. Prove that $X_n \to_d X$ implies $X_n' \to_d X'$. After reading section 9.2, prove the k-dimensional version of the Mann–Wald theorem.

Exercise 1.8 Prove that theorem 1.4 holds in k dimensions.

See also theorem 9.5.2 and theorem 10.1.3 below.

2 Determining Classes of Functions

We can approximate the functions $1_{(-\infty,z]}(\cdot)$ to an arbitrary degree of accuracy within various classes of particularly smooth functions. Within these classes of functions we do not have to worry about the continuity of the limiting measure at z, and this will make these classes more convenient. Indeed, the specialized class $\mathcal{H}^0$ below is of this type.

Definition 2.1 (Determining class) A collection $\mathcal{G}$ of bounded and continuous functions g is called a *determining class* if for any choice of dfs $\tilde{F}$ and F, the requirement that $\int g d\tilde{F} = \int g\, dF$ for all $g \in \mathcal{G}$ implies $\tilde{F} = F$.

Definition 2.2 (Various classes of smooth functions) (i) Let C (let C_b) [let C_{bu}] denote the class of continuous (bounded and continuous) [bounded and also uniformly continuous] functions on R. Let $C_b^{(k)}$ (let $C_b^{(\infty)}$) denote the subclasses with k (with all) derivatives bounded and continuous.
(ii) An extra c on these classes will indicate that all functions vanish outside some compact subset of R.
(iii) Let C_0 denote the subclass of C that converge to 0 as $|x| \to \infty$.
(iv) Let $\mathcal{H}^0$ denote the class of all $h_{z,\epsilon}$ with z real and $\epsilon > 0$; here $h_{z,\epsilon}(x)$ equals 1, is linear, equals 0 according as x is in $(-\infty, z]$, is in $[z, z+\epsilon]$, is in $[z+\epsilon, \infty)$ (this class was introduced in the proof of the Helly–Bray theorem 3.5.1).
(v) Let $\mathcal{G}^0$ denote the class of all continuous functions $g_{a,b,\epsilon}$ with $a < b$ and $\epsilon > 0$; here $g_{a,b,\epsilon}(x)$ equals 0, is linear, equals 1 according as x is in $(-\infty, a-\epsilon] \cup [b+\epsilon, \infty)$, is in $[a-\epsilon, a] \cup [b, b+\epsilon]$, is in $[a, b]$.

Theorem 2.1 (Criteria for $\to_d$; a kinder and gentler Helly–Bray)
(i) Let $F_1, F_2, \ldots$ be tight. Let $\mathcal{G}$ be a determining class.
(a) If $\int g\, dF_n \to$ (some $\#_g$) for each $g \in \mathcal{G}$, then $F_n \to_d F$. Further, $\#_g = \int g\, dF$.
(b) Conversely: If $F_n \to_d F$, then $\int g\, dF_n \to \int g\, dF$ for each $g \in \mathcal{G}$.
(ii) Each of the various classes C_0, C_b, C_{bu}, $C_b^{(k)}$ with $k \ge 1$, $C_b^{(\infty)}$, $\mathcal{H}^0$, and $\mathcal{G}^0$ is a determining class.
(iii) So, too, if we add an extra subscript c to the various C-classes in (ii). (That is, we require they take on the value 0 outside some compact subset of R.)
[For some proofs in the literature, functions g with sharp corners are unhandy.]

Exercise 2.1 Prove the previous theorem.

Exercise 2.2 (Higher dimensions) Show that the natural extension of each of the results of this section to R_k is valid.

Exercise 2.3 Exhibit at least one more determining class.

Moments as a Determining Class for a Moment Unique Limit

Theorem 2.2 (CLT via moments; Fréchet–Shohat) (a) Suppose F is the unique df having the specific finite moment values $\mu_k = \int x^k \, dF(x)$, for all integers $k \geq 1$. Then $F_n \to_d F$ whenever

(1) $\mu_{nk} \equiv \int x^k \, dF_n(x) \to \mu_k \equiv \int x^k \, dF(x)$ for all $k \geq 1$.

(b) Any *normal* df is determined by its moments.

Proof. Let n' denote an arbitrary subsequence. By the Helly selection theorem we have $F_{n''} \to_{sd} H$ for some further subsequence n'' and some sub-df H. However, $\overline{\lim}\, \mathrm{E}|X_n|^2 < \infty$, so that $\{F_n : n \geq 1\}$ is tight by Markov's inequality. Thus H is a bona fide df, and $F_{n''} \to_d H$. Also, for all $k \geq 1$

$\int x^k \, dF(x) = \lim \int x^k \, dF_{n''}(x)$ by hypotheses

(a) $= \int x^k \, dH(x)$ by exercise 9.1.1(c).

Thus $\int x^k \, dH(x) = \int x^k \, dF(x)$ for all $k \geq 1$; and since only F has these moments, we conclude that $H = F$. Thus $F_{n''} \to_d F$. Moreover, $F_{n''} \to_d$ (this same F) on any such convergent subsequence n''. Thus $F_n \to_d F$, by the corollary to theorem 9.1.3. See exercise 9.2.6 below for part (b) of the theorem. □

In general, moments do not determine a distribution uniquely; thus $\{x^k : k \geq 1\}$ is *not* a determining class. This is shown by the following exercise.

Exercise 2.4* (Moments need not determine the df; Heyde) Suppose that the rv $\log X \cong N(0,1)$; thus

$f_X(x) = x^{-1} e^{-(\log x)^2/2} / \sqrt{2\pi}$ for $x > 0$.

For each $-1 \leq a \leq 1$, let Y_a have the density function

$f_a(y) = f_X(y)[1 + a \sin(2\pi \log y)]$ for $y > 0$.

Show that X and each Y_a have exactly the same moments. [Knowing that these particular distributions have this property is not worth much; it is knowing that some dfs have this property that matters.]

Though we have just seen that moments do not necessarily determine a df, it is often true that a given df F is the *unique* df having its particular moments (name them $\{\mu_k : k \geq 1\}$). Here is an "exercise" giving various sufficient conditions.

Exercise 2.5 (When moments *do* determine a df) Suppose either of the following conditions hold:

(a) $\overline{\lim}\, |\mu_k|^{1/k}/k < \infty$.

(b) $\sum_1^\infty \mu_{2k} t^{2k}/(2k)! < \infty$ in some interval of t values.

Then at most one df F can possess the moment values $\mu_k = \int x^k \, dF(x)$. [Wait to prove this until it appears again as part of exercise 9.6.1

Comment. A condition for convergence due to Carleman $\sum_1^\infty \mu_{2k}^{-1/2k} = \infty$ has often been claimed to be necessary and sufficient. It is not. See Stoyanov (1977; p. 113).

Exercise 2.6 Show that the $N(0,1)$ distribution is uniquely determined by its moments.

Summary The methods of this section that establish $\to_d$ by verifying the moment condition that $\mathrm{E}g(X_n) \to \mathrm{E}g(X)$ for all functions g in a given determining class $\mathcal{G}$ can be extended from the present setting of the real line to more general settings; note Chapter 15. This chapter now turns to the development of results associated with the particular determining class $\mathcal{G} \equiv \{g_t(\cdot) \equiv e^{it\cdot} : t \in R\}$. The resulting function is called the *characteristic function* of the rv X. The rest of Chapter 9 includes a specialized study of the characteristic function. Chapter 10 will apply this characteristic function tool to the CLT. □

3 Characteristic Functions, with Basic Results

Elementary Facts

Definition 3.1 (Characteristic function) Let X be an arbitrary rv, and let F denote its df. The *characteristic function* of X (abbreviated *chf*) is defined (for all $t \in R$) by

$$\begin{aligned} &\phi(t) \equiv \phi_X(t) \equiv Ee^{itX} = \int_{-\infty}^{\infty} e^{itx}\, d\mu_F(x) = \int_{-\infty}^{\infty} e^{itx}\, dF(x) \\ &\quad \equiv \int_{-\infty}^{\infty} \cos(tx)\, dF(x) + i \int_{-\infty}^{\infty} \sin(tx)\, dF(x). \end{aligned} \tag{1}$$

With dF replaced by $h\, d\mu$, we call this the *Fourier transform* of the signed measure $h\, d\mu$. (We note that the chf $\phi_X(t)$ exists for $-\infty < t < \infty$ for *all* rvs X, since $|e^{itX(\omega)}| \le 1$ for all t and all ω.)

Proposition 3.1 (Elementary properties) Let ϕ denote an arbitrary chf.

(a) $\phi(0) = 1 \quad \text{and} \quad |\phi(t)| \le 1$ for all $t \in R$.

(b) $\phi_{aX+b}(t) = e^{itb}\phi_X(at)$ for all $t \in R$.

(c) $\phi_{\sum_1^n X_i}(\cdot) = \prod_{i=1}^n \phi_{X_i}(\cdot)$ when $X_1, \ldots, X_n$ are independent.

(d) $\bar{\phi}_X(t) = \phi_X(-t) = \phi_{-X}(t) = \mathrm{E} \cos tX - i\, \mathrm{E} \sin tX$ for all $t \in R$.

(e) ϕ is real-valued if and only if $X \cong -X$.

(f) $|\phi(\cdot)|^2$ is a chf [of the rv $X^s \equiv X - X'$, with X and X' iid with chf ϕ].

(g) $\phi(\cdot)$ is uniformly continuous on R.

Proof. Now, (a), (b), (c), and (d) are trivial. (e) If $X \cong -X$, then $\phi_X = \bar{\phi}_X$; so ϕ_X is real. If ϕ is real, then $\phi_X = \bar{\phi}_X = \phi_{-X}$; so $X \cong -X$ by the uniqueness theorem below. (f) If X and X' are independent with characteristic function ϕ, then $\phi_{X-X'} = \phi_X \phi_{-X} = \phi\bar{\phi} = |\phi|^2$. For (g), we note that for all t,

$$|\phi(t+h) - \phi(t)| = |\int [\exp^{i(t+h)x} - e^{itx}]\, dF(x)|$$

$$\le \int |e^{itx}|\, |e^{ihx} - 1|\, dF(x) \le \int |e^{ihx} - 1|\, dF(x) \to 0 \tag{a}$$

as $h \to 0$, by the DCT with dominating function 2.

The converse of (g) is false. Let $X_1 \equiv X_2$ and X_3 be two iid Cauchy$(0,1)$ rvs. We will see below that $\phi_{\text{Cauchy}}(t) = \exp(-|t|)$, giving $\phi_{2X_1}(t) = \phi_{X_1+X_2}(t) = \phi_{X_1+X_3}(t)$ for all t. □

Motivation 3.1 (Proving the CLT via chfs) In this chapter we present an alternative method for establishing $F_n \to_d F$. It is based on the fact (to be demonstrated below) that the complex exponential functions $e^{it\cdot}$ on R, indexed by $t \in R$, form a limit determining class. Saying this another way, the chf ϕ determines the distribution P, or the df F (or the density f, if there is one). Thus (as is shown in the continuity theorem below) we can establish that $F_n \to_d F$ by showing that $\phi_n(\cdot) \to \phi(\cdot)$ on R. Indeed, using just the elementary properties listed above, it is trivial to give an informal "proof" of the classical CLT. Thus, we begin by expanding the chf of one rv X as

(a) $$\phi_{(X-\mu)/\sqrt{n}}(t) = \phi_{X-\mu}(t/\sqrt{n}\,) = \mathrm{E}e^{it(X-\mu)/\sqrt{n}}$$

(b) $$= \mathrm{E}\{1 + \frac{it}{\sqrt{n}}(X-\mu) + \frac{(it)^2}{n}(X-\mu)^2/2 + o(t^2/n)\}$$

$$= 1 + \frac{it}{\sqrt{n}}\mathrm{E}(X-\mu) + \frac{(it)^2}{n}\mathrm{E}(X-\mu)^2/2 + o(t^2/n)$$

(c) $$= 1 + 0 - t^2\sigma^2/2n + o(t^2/n) = 1 - t^2[\sigma^2 + o(1)]/2n.$$

(In section 9.6 we will make such expansions rigorous, and in section 9.7 we will estimate more carefully the size of the errors that were made.)

Then the standardized sum of the iid rvs $X_1, \ldots, X_n$ is

(d) $$Z_n \equiv \sqrt{n}(\bar{X}_n - \mu) = \textstyle\sum_1^n (X_k - \mu)/\sqrt{n}\,,$$

and it has chf

(e) $$\phi_{Z_n}(t) = \textstyle\prod_{k=1}^n \phi_{(X_k-\mu)/\sqrt{n}}(t) = [\phi_{(X-\mu)/\sqrt{n}}(t)]^n$$

(f) $$= \left\{1 - \frac{t^2[\sigma^2 + o(1)]}{2n}\right\}^n \to e^{-t^2\sigma^2/2}$$

(g) $$= \phi_{N(0,\sigma^2)}(t) \qquad \text{as will be shown below.}$$

Since $\phi_{Z_n}(\cdot) \to \phi_Z(\cdot)$ on R, where $Z \cong N(0,1)$, the uniqueness theorem and the continuity theorem combine to guarantee that $Z_n \to_d Z$. In principle, this is a rather elementary way to prove the CLT.

Think of it this way. To have all the information on the distribution of X, we must know $P(X \in B)$ for all $B \in \mathcal{B}$. We have seen that the df F also contains all this information, but it is presented in a different format; a statistician may well regard this F format as the "tabular probability calculating format." When a density f exists, it also contains all the information about P; but it is again presented in a different format, which the statistician may regard as the "distribution visualization format." We will see that the chf presents all the information about P too. It is just one more format, which we may well come to regard as the "theorem proving format." □

Some Important Characteristic Functions

Distribution	**Density**	**Chf**
Binomial(n,p)	$\binom{n}{k}p^k(1-p)^{n-k}$; for $0 \le k \le n$	$[1+p(e^{it}-1)]^n$
Poisson(λ)	$e^{-\lambda}\lambda^k/k!$; for $k \ge 0$	$\exp(\lambda(e^{it}-1))$
GeometricF(p)	pq^k; for $k \ge 0$	$p(1-qe^{it})^{-1}$
Normal(μ,σ^2)	$e^{-(x-\mu)^2/2\sigma^2}/\sqrt{2\pi}\sigma$ on R	$\exp(it\mu - \sigma^2t^2/2)$
Exponential(θ)	$e^{-x/\theta}/\theta$ on R^+	$(1-it\theta)^{-1}$
Chisquare(n)	$x^{(n/2)-1}e^{-x/2}/[2^{n/2}\Gamma(n/2)]$	$(1-2it)^{-n/2}$
Gamma(r,θ)	$x^{r-1}e^{-x/\theta}/[\theta^r\Gamma(r)]$ on R^+	$(1-it\theta)^{-r}$
Uniform$(0,1)$	$1_{[0,1]}(x)$	$[\exp(it)-1]/it$
Double Exp(θ)	$e^{-\|x\|/\theta}/2\theta$	$1/(1+\theta^2t^2)$
Cauchy(0,1)	$1/[\pi(1+x^2)]$	$e^{-\|t\|}$
de la Vallée Poussin	$(1-\cos x)/(\pi x^2)$ on R	$[1-\|t\|]\times 1_{[-1,1]}(t)$
Triangular(0,1)	$[1-\|x\|]\times 1_{[-1,1]}(x)$	$2(1-\cos t)/(t^2)$ on R

Table 1.1

Review of Some Useful Complex Analysis

A function f is called *analytic* on a *region* (a connected open subset of the complex plane) if it has a derivative at each point of the region; if it does, then it necessarily has derivatives of all orders at each point in the region. If z_0 is an isolated singularity of f and $f(z) = \sum_{n=0}^{\infty} a_n(z-z_0)^n + \sum_{n=1}^{m} b_n(z-z_0)^{-n}$, then $k \equiv$ (the *residue* of f at z_0) $= b_1$. Thus if f has a pole of order m at z_0 (that is, $b_n = 0$ for $n > m$ in the expansion above), then $g(z) \equiv (z-z_0)^m f(z) = b_m + \cdots + b_1(z-z_0)^{m-1} + \sum_0^{\infty} a_n(z-z_0)^{m+n}$ has $b_1 = g^{(m-1)}(z_0)/(m-1)!$. Thus

$$(2) \qquad b_1 = k = (\text{residue of } f \text{ at } z_0) \qquad \{= \lim_{z\to z_0}(z-z_0)f(z) \text{ for a simple pole at } z_0\}.$$

We also note that a *smooth arc* is described via equations $x = \phi(t)$ and $y = \psi(t)$ for $a \le t \le b$ when ϕ' and ψ' are continuous and not simultaneously zero. A *contour* is a continuous chain of a finite number of smooth arcs that do not cross the same point twice. *Closed* means that the starting and ending points are identical. (See Ahlfors (1953, pp. 102, 123) for what follows.)

Lemma 3.1 (Residue theorem) If f is analytic on a region containing a closed contour C, except for a finite number of singularities $z_1, \ldots, z_n$ interior to C at which f has residues $k_1, \ldots, k_n$, then (for counterclockwise integration over C)

$$(3) \qquad \int_C f(z)\,dz = 2\pi i \sum_{j=1}^{n} k_j \qquad \begin{cases} = 0 & \text{if } f \text{ is analytic,} \\ = 2\pi i\,(z-z_o)f(z_0) & \text{for one simple pole at } z_o. \end{cases}$$

Lemma 3.2 Let f and g be functions analytic in a regions Ω. Suppose that $f(z) = g(z)$ for all z on a set S that has an accumulation point in Ω. We then have the equality $f(z) = g(z)$ for all $z \in \Omega$. (That is, f is determined on Ω by its values on S. So if there is a Taylor series representation $f(z) = \sum_{j=o}^{\infty} a_n(z - z_o)^j$ valid on some disk interior to Ω, then the coefficientsd $a_1, a_2, \ldots$ determine f on all of Ω.)

Evaluating Various Characteristic Functions

Example 3.1 (Derivation of the Cauchy(0, 1) chf) Let C denote the upper semicircle centered at the origin with radius R parametrized counterclockwise; and let A (for arc) denote C without its base. Let $t > 0$. The Cauchy chf is approached via

(a) $$\int_C \frac{e^{itz}}{\pi(1+z^2)}\, dz \equiv \int_C f(z)\, dz = 2\pi i \cdot (z - z_o) f(z_o) \qquad \text{(with } z_o = i\text{)}$$

(b) $$= 2\pi i \cdot (z - i)\, \frac{e^{itz}}{\pi(1+iz)(1-iz)}\bigg|_{z=i} = e^{-t} \qquad \text{for } t > 0.$$

It further holds that

(c) $$\int_C \frac{e^{itz}}{\pi(1+z^2)}\, dz = \int_{-R}^{R} \frac{e^{itz}}{\pi(1+x^2)}\, dx + \int_A \frac{e^{itz}}{\pi(1+z^2)}\, dz$$

(d) $$\to \int_{-\infty}^{\infty} \frac{e^{itx}}{\pi(1+x^2)}\, dx \;+\; 0 \;=\; \phi(t) \qquad \text{as } R \to \infty,$$

since the second integral in (b) is bounded in absolute value by

$$\tfrac{1}{\pi} \int_A \tfrac{1}{R^2-1}\, dz = \tfrac{1}{\pi}\, \tfrac{1}{R^2-1}\, \pi R \to 0 \qquad \text{as } R \to \infty.$$

Since the Cauchy is symmetric, $\phi(-t) = \phi(t) = \exp(-|t|)$; or, integrate the contour clockwise when $t < 0$. The tabular entry has been verified. That is,

(4) $\phi(t) = \exp(-|t|\,)$, for all t, gives the Cauchy(0,1) chf. □

Example 3.2 (Derivation of the $N(0, 1)$ chf) Let X be $N(0, 1)$. Then

$$\phi(t) = \int_{-\infty}^{\infty} e^{itx}\, \tfrac{1}{\sqrt{2\pi}}\, e^{-x^2/2}\, dx.$$

Let us instead think of ϕ as a function of a complex variable z. That is,

(a) $$\phi(z) = \int_{-\infty}^{\infty} e^{izx}\, \tfrac{1}{\sqrt{2\pi}}\, e^{-x^2/2}\, dx.$$

Let us define a second function ψ on the complex plane by

(b) $$\psi(z) \equiv e^{-z^2/2}\,.$$

Now ϕ and ψ are analytic on the whole complex plane. Let us now consider the purely imaginary line $z = iy$. On this line it is clear that

$$\psi(iy) = e^{y^2/2},$$

and since elementary calculations show that

$$\phi(iy) = \int_{-\infty}^{\infty} e^{-yx}\, \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2} = e^{y^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\, e^{-(x+y)^2/2}\, dx = e^{y^2/2},$$

we have $\psi = \phi$ on the line $z = iy$. Thus lemma 3.2 implies that $\psi(z) = \phi(z)$ for all z in the plane. Thus $\phi(t) = \psi(t)$ for all real $z = t$. That is,

(5) $\qquad \phi(t) = \exp(-t^2/2), \qquad$ for all real t, $\qquad$ gives the $N(0,1)$ chf.

(A similar approach works for the gamma distribution in exercise 3.3 below.) □

Exercise 3.1 Derive the $N(0,1)$ chf via the residue theorem. Then extend to $N(\mu, \sigma^2)$. [Hint. Let C denote a closed rectangle of height t with base $[-R, R]$ on the x-axis.]

Exercise 3.2 (a) Derive the Poisson(λ) chf (by summing power series).
(b) Derive the GeometricT(p) chf.
(c) Derive the Bernoulli(p), Binomial(n, p), and NegBiT(m, p) chfs.

Exercise 3.3 (a) Derive the Gamma(r, θ) chf. [Hint. Note example 3.2.]
(b) Derive the Exponential(θ) and Chisquare(n), and Double Exponential(θ) chfs.

Exercise 3.4 Derive the Logistic$(0, 1)$ chf.
Hint. Use lemma 3.2 approach.

Exercise 3.5 Show that the real part of a chf (or Re $\phi(\cdot)$) is itself a chf.

Exercise 3.6 Let ϕ be a chf. Show that $\frac{1}{c}\int_0^c \phi(tu)\, du$ is a chf.

4 Uniqueness and Inversion

For the chf to be a useful tool, there must be a 1-to-1 correspondence between dfs and chfs. The fact that this is so is called the *uniqueness theorem.* We give a simple proof of the uniqueness theorem at the end of this subsection. But the simple proof does not establish an *inversion formula* that expresses the df as a function of the chf. In order to establish an inversion formula, we will need some notation, and an inversion formula useful for other purposes will require a hypothesis on the chf that is strong enough to allow some useful simplification.

Let U denote a rv with continuous density $f_U(\cdot)$, and let W denote a rv with a bounded and continuous density $f_W(\cdot)$ and with chf $\phi_W(\cdot)$; and suppose we are lucky enough to determine a *complementary pair* that (for some constant c) satisfy the relationship

(1) $\quad f_U(t) = c\,\phi_W(-t) \qquad$ for all real t. **(Complementary pair)**

We give three examples of such pairs. Let $Z \cong N(0,1)$, $T \cong$ Triangular$(0,1)$, and let D have the de la Vallée Poussin density. Then examples of (1) are

(2) $\quad U = Z \quad$ and $\quad W = Z, \quad$ with $c = 1/\sqrt{2\pi}$,

(3) $\quad U = T \quad$ and $\quad W = D, \quad$ with $c = 1$,

(4) $\quad U = D \quad$ and $\quad W = T, \quad$ with $c = 1/2\pi$.

(The Cauchy$(0,1)$ and the Double Exponential$(0,1)$ then lead to two additional complementary pairs.) (The beauty of this is that we can nearly eliminate the use of complex analysis.) (In all such examples we have $2\pi c f_W(0) = 1$.)

An arbitrary rv X, having df $F_X(\cdot)$ and chf $\phi_X(\cdot)$, may not have a density. Let us recall from the convolution formula (A.2.2) that (if U has a density) a slightly perturbed version X_a of X is smoother than X, in that

(5) $\quad X_a \equiv X + aU$ always has a density $f_a(\cdot)$; and $X_a \to_d X$ as $a \to 0$

by Slutsky's theorem, since $a\,U \to_p 0$ as $a \to 0$. Thus $F(\cdot) = \lim F_a(\cdot)$ at each point in the continuity set C_F of F. This is the key to the approach we will follow to establish an inversion formula.

Theorem 4.1 (Uniqueness theorem) Every df on the line has a unique chf.

Theorem 4.2 (Inversion formula) If an arbitrary rv X has df $F_X(\cdot)$ and chf $\phi_X(\cdot)$, we can always write

(6) $$F_X(r_2) - F_X(r_1) = \lim_{a\to 0} \int_{r_1}^{r_2} f_a(y)\,dy \qquad \text{for all } r_1 < r_2 \text{ in } C_{F_X},$$

where the density $f_a(\cdot)$ of the rv $X_a \equiv X + aU$ of (5) [with U as in (1)] is given by

(7) $$f_a(y) = \int_{-\infty}^{\infty} e^{-iyv}\,\phi_X(v) c f_W(av)\,dv \qquad \text{for all } y \in R.$$

Theorem 4.3 (Inversion formula for densities) If a rv X has a chf $\phi_X(\cdot)$ that satisfies the integrability condition

$$(8) \qquad \int_{-\infty}^{\infty} |\phi_X(t)|\, dt < \infty,$$

then X has a uniformly continuous density $f_X(\cdot)$ given by

$$(9) \qquad f_X(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\, \phi_X(t)\, dt.$$

Remark 4.1 The uniqueness theorem can be restated as follows: The set of complex exponentials $\mathcal{G} \equiv \{e^{itx} \text{ for } x \in R : t \in R\}$ is a determining class. This is so because knowing all values of $\phi_X(t) = \mathrm{E}e^{itX}$ allows the df F to be determined, via the inversion formula. □

Proof. From the convolution formula (A.2.2) and $X_a \equiv X + aU$ we have

$$f_a(y) = \textstyle\int_{-\infty}^{\infty} \frac{1}{a}\, f_U(\frac{y-x}{a})\, dF_X(x)$$

$$(a) \qquad = c\textstyle\int_{-\infty}^{\infty} \frac{1}{a}\, \phi_W(\frac{x-y}{a})\, dF_X(x) \qquad \text{by (1)}$$

$$= (c/a)\textstyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{i(x-y)w/a}\, f_W(w)\, dw\, dF_X(x)$$

$$= (c/a)\textstyle\int_{-\infty}^{\infty} e^{-iyw/a}\, f_W(w) \int_{-\infty}^{\infty} e^{i(w/a)x}\, dF_X(x)\, dw \qquad \text{by Fubini}$$

$$(b) \qquad = (c/a)\textstyle\int_{-\infty}^{\infty} e^{-iyw/a}\, \phi_X(w/a)\, f_W(w)\, dw = c\int_{-\infty}^{\infty} e^{-iyv}\, \phi_X(v)\, f_W(av)\, dv.$$

Since $X_a \to_d X$, at continuity points $r_1 < r_2$ of F we have (with $X_a \cong F_a(\cdot)$)

$$(c) \qquad F_X(r_2) - F_X(r_1) = \lim_{a\to 0}\{F_a(r_2) - F_a(r_1)\} = \lim_{a\to 0} \textstyle\int_{r_1}^{r_2} f_a(y)\, dy.$$

This establishes theorems 4.1 and 4.2.

The particular formula given in (c) might look useless, but the mere fact that one can recover F_X from ϕ_X via *some* formula is enough to establish the important property of uniqueness. (See exercise 4.3 for some utility for (7).) We now turn to theorem 4.3, in which we have added a hypothesis that allows the previous formula to be manipulated into a simple and useful form.

Suppose that (8) holds, so that applying the DCT to (b) (using a constant times $|\phi_X(\cdot)|$ as a dominating function) gives [recall the hypothesis on the $f_W(\cdot)$ of (1)] as $a \to 0$ that

$$(d) \qquad f_a(y) \to f(y) \equiv [c f_W(0)] \textstyle\int_{-\infty}^{\infty} e^{-iyv} \phi_X(v)\, dv,$$

since f_W is bounded and is continuous at 0. Note that uniform continuity of f follows from the bound

$$|f(y+h) - f(y)| = [c f_W(0)]\; |\textstyle\int_{-\infty}^{\infty} [e^{-i(y+h)v} - e^{-iyv}]\phi_X(v)\, dv|$$

$$(e) \qquad \le [c f_W(0)] \textstyle\int_{-\infty}^{\infty} |e^{-ihv} - 1|\, |\phi_X(v)|\, dv \to 0 \qquad \text{as } h \to 0,$$

by applying the DCT (with dominating function $2c\,\|f_W\|\,|\phi_X(\cdot)|$. The uniform convergence of f_a to f on any finite interval involves only an $|f_W(0)-f_W(av)|$ term under the integral sign. That f really is the density of F_X follows from applying this uniform convergence in (c) to obtain

(f) $\qquad F_X(r_2)-F_X(r_1)=\int_{r_1}^{r_2} f(y)\,dy\,.$

The conclusion (9) holds since specifying $U=W=Z$ gives

(g) $\qquad [cf_W(0)]=1/(2\pi) \qquad\qquad$ (as it always must). □

Esseen's inequality 9.7.1 below provides an important extension of theorem 4.2 by showing that if two chfs are sufficiently close over most of their domain, then the corresponding dfs will be uniformly close over their entire domain.

Exercise 4.1 Show that setting $W=Z$ in line (c) of the previous proof leads, for any rv X, to the alternative inversion formula

(10) $$F_X(r_2)-F_X(r_1)=\lim_{a\to 0}\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{-itr_2}-e^{-itr_1}}{-it}\,\phi_X(t)\,e^{-a^2t^2/2}\,dt$$

at all continuity points $r_1<r_2$ of $F_X(\cdot)$. [This is one possible alternative to (6).]

Exercise 4.2 Derive the chf of the Triangular$(0,1)$ density on the interval $[-1,1]$ (perhaps, add two appropriate uniform rvs). Then use theorem 4.3 to derive the chf of the de la Vallée Poussin density, while simultaneously verifying that the non-negative and real integrable function $(1-\cos x)/(\pi x^2)$ really is a density. Following section 9.6, determine $\mathrm{E}|X|$ when X has the de la Vallée Poussin density.

Exercise 4.3 (Kernel density estimator) Since the rv X having df $F_X(\cdot)$ and chf $\phi_X(\cdot)$ may not have a density, we choose instead to estimate the density $f_a(\cdot)$ of (5) and (7) using

(11) $$\hat{f}_a(x)\equiv c\int_{-\infty}^{\infty} e^{-itx}\,\hat{\phi}_X(t)f_W(at)\,dt$$

[where $f_U(\cdot)=c\phi_W(-\cdot)$, and where we now insist that $\mu_U=0$ and σ_U^2 is finite] with the *empirical chf* $\hat{\phi}_X(\cdot)$ defined by

(12) $$\hat{\phi}_X(t)\equiv\int_{-\infty}^{\infty} e^{itx}\,d\,\mathbb{F}_n(x)=\frac{1}{n}\sum_{j=1}^{n} e^{itX_j}\qquad\text{for }-\infty<t<\infty.$$

(a) Verify that $\hat{f}_a(\cdot)$ is actually a kernel density estimator, meaning that it can be expressed as

(13) $$\hat{f}_a(x)=\frac{1}{a}\int_{-\infty}^{\infty} f_U\left(\frac{x-y}{a}\right)d\,\mathbb{F}_n(y)=\frac{1}{n}\sum_{j=1}^{n}\frac{1}{a}f_U\left(\frac{x-X_j}{a}\right).$$

[This has statistical meaning, since we are averaging densities centered at each of the observations.]

(b) Show that $\hat{f}_a(x)$ is always *unbiased* (in that it has mean $f_a(x)$) and has a finite variance we can calculate; thus for all $x \in R$ we can show that

$$\text{(14)} \qquad \mathrm{E}\hat{f}_a(x) = f_a(x)\,,$$

$$\text{(15)} \qquad \mathrm{Var}[\hat{f}_a(x)] = \frac{1}{n}\left\{\frac{1}{a^2}\int_{-\infty}^{\infty} f_U^2\left(\frac{x-y}{a}\right)\,dF(y) - [f_a(x)]^2\right\}.$$

(c) Supposing that $F_X(\cdot)$ has a density $f(\cdot) \in C_b^{(2)}$, determine the order of the mean squared error

$$\text{(16)} \qquad \mathrm{MSE}\{\hat{f}_a(x)\} \equiv \mathrm{Bias}^2\{\hat{f}_a(x)\} + \mathrm{Var}[\hat{f}_a(x)] \equiv \{\mathrm{E}(\hat{f}_a(x)) - f(x)\}^2 + \mathrm{Var}[\hat{f}_a(x)]$$

of $\hat{f}_a(x)$, viewed as an estimator of $f(x)$. (It is intended that you rewrite (16) by expanding $f_a(x)$ in a Taylor series in "a" (valid for $f(\cdot) \in C_b^{(2)}$), and then analyze the magnitude of (16) for values of "a" near 0. It might also be useful to relabel f_U by ψ now so that your work refers to any kernel density estimator, right from the beginning. This will avoid "starting over" in part (f).) Show that this MSE expression is of order $n^{-4/5}$ for $f(\cdot) \in C_b^{(2)}$ when a is of order $n^{-1/5}$, and that this is the minimal attainable order.

(d) Note that the choice $U = Z$ (or $U = T$) leads to an $\hat{f}_a(\cdot)$ that is the sum of n normal (or triangular) densities that are centered at the n data points and that have a scale parameter directly proportional to a.

(e) Obtain an expression for $\lim_{a\to 0}\, a^{4/5}$ MSE $\{\hat{f}_a(x)\}$ in terms of $f(x)$, $f'(x)$, and $f''(x)$ when $a = n^{-1/5}$ (and obtain it for both of the choices $U = Z$ and $U = T$).

(f) We could also motivate the idea of a kernel density estimator based on (13) alone. How much of what we have done still carries over for a general kernel? What properties should a good kernel exhibit? What can you prove in this more general setting? (Now, for sure, replace f_U by a function labeled ψ. A simple sentence that specifies the requirements on ψ should suffice.)

Exercise 4.4 Use the table of chfs above to show in what sense the sums of independent Binomial, Poisson, NegBiT, Normal, Cauchy, Chisquare and Gamma rvs have distributions that again belong to the same family. (Recall section A.2, noting that chfs have allowed the complicated operation of convolution of dfs or densities to be replaced by the simple operation of multiplication of chfs.)

5 The Continuity Theorem

Theorem 5.1 (Continuity theorem for chfs; Cramér–Lévy) (i) If $\phi_n \to \phi$ where ϕ is continuous at 0, then ϕ is the chf of a bona fide df F and $F_n \to_d F$.
(ii) $F_n \to_d F$ implies $\phi_n \to \phi$ uniformly on any finite interval $|t| \le T$.

Inequality 5.1 (Chf bound on the tails of a df) For any df F we have

(1) $\quad P(|X| \ge \lambda) \le 7\lambda \int_0^{1/\lambda} [1 - \text{Real}\,\phi(t)]\,dt \qquad$ for all $\lambda > 0$.

Proof. Now,

(a) $\quad \lambda \int_0^{1/\lambda} [1 - \text{Real}\ \phi(t)]\,dt = \lambda \int_0^{1/\lambda} \int_{-\infty}^{\infty} [1 - \cos(tx)]\,dF(x)\,dt$

$$= \int_{-\infty}^{\infty} \lambda \int_0^{1/\lambda} [1 - \cos(tx)]\,dt\,dF(x)$$

$$= \int_{-\infty}^{\infty} \{\lambda t[1 - \tfrac{\sin(xt)}{xt}]\}\Big|_0^{1/\lambda}\,dF(x)$$

$$= \int_{-\infty}^{\infty} [1 - \tfrac{\sin(x/\lambda)}{(x/\lambda)}]\,dF(x)$$

$$\ge \int_{[|x|/\lambda \ge 1]} [1 - \sin(x/\lambda)/(x/\lambda)]\,dF(x)$$

(b) $\quad = \inf_{[|y|\ge 1]} [1 - \sin(y)/y]\,P(|X| \ge \lambda) = [1 - \sin(1)]\,P(|X| \ge \lambda)$

(c) $\quad = (.1585\ldots)P(|X| \ge \lambda) \ge P(|X| \ge \lambda)/7,$

as claimed. (It may be interesting to compare this to the Chebyshev inequality.) [This idea will be carried further in (10.5.9) and (10.5.10).] □

Proof. Consider theorem 5.1. (i) The uniqueness theorem for chfs shows that the collection $\mathcal{G}$ of complex exponential functions form a determining class, and the expectations of these are hypothesized to converge. It thus suffices (by the kinder and gentler Helly–Bray theorem (theorem 9.2.1(i)(a)) below) to show that $\{F_n : n \ge 1\}$ is tight. Now,

(a) $\quad \overline{\lim}_{n\to\infty} P(|X_n| \ge \lambda) \le \overline{\lim}_{n\to\infty} 7\lambda \int_0^{1/\lambda} [1 - \text{Real}\ \phi_n(t)]\,dt$

$$= 7\lambda \int_0^{1/\lambda} [1 - \text{Real}\ \phi(t)]\,dt$$

by the DCT, with dominating function 2

(b) $\quad \to 0 \qquad$ as $\lambda \to \infty$,

so that $\{F_n : n \ge 1\}$ is tight.

(ii) Now replacing $X_n \to_d X$ by versions $Y_n \to_{a.s.} Y$ (and using Skorokhod's construction) gives for $|t| \le T$ that

(c) $\quad |\phi_n(t) - \phi(t)| \le \int |e^{itY_n} - e^{itY}|\,dP$

$$\le \int |e^{it(Y_n - Y)} - 1|\,dP$$

$$\le \int \sup_{|t|\le T} |e^{it(Y_n - Y)} - 1|\,dP$$

(d) $\quad \to 0 \qquad$ as $\qquad \sup\{|it(Y_n - Y)| : |t| \le T\} \le T\,|Y_n - Y| \to 0$

by the DCT, with dominating function 2. □

Higher Dimensions

If $X_1, \ldots, X_k$ are rvs on $(\Omega, \mathcal{A}, P)$, then the $\mathcal{B}_k$-$\mathcal{A}$-mapping $\mathsf{X} \equiv (X_1, \ldots, X_k)'$ from Ω to R_k induces a measure P_{X} on $(R_k, \mathcal{B}_k)$. The characteristic function of X is

(2) $\quad \phi_{\mathsf{X}}(\mathsf{t}) \equiv \mathrm{E}\, e^{i\,\mathsf{t}'\mathsf{X}} = \mathrm{E}\, e^{i\,[t_1 X_1 + \cdots + t_k X_k]} \qquad$ for $\mathsf{t} \equiv (t_1, \ldots, t_k)' \in R_k$.

Without further explanation, we state simply that the uniqueness theorem (that $\{g_{\mathsf{t}} \equiv \exp(i\mathsf{t}'\mathsf{x})$ for all $\mathsf{x} \in R_n : \mathsf{t} \in R_n\}$ is a determining class) and the Cramér–Lévy continuity theorem still hold, based on minor modifications of the previous proof. We also remark that all equivalences of $\to_d$ in theorem 9.1.1 are still valid. But we now take up an extremely useful approach to showing convergence in distribution in higher dimensions.

The characteristic function of the one-dimensional linear combination $\vec{\lambda}'\mathsf{X}$ is

(3) $\quad \phi_{\vec{\lambda}'\mathsf{X}}(t) = \mathrm{E}\, e^{i\,[t\lambda_1 X_1 + \cdots + t\lambda_k X_k]} \qquad$ for $t \in R$.

Comparison of this with (2) shows that knowing the joint chf $\phi_{\mathsf{X}}(\mathsf{t})$ for all $\mathsf{t} \in R_k$ is equivalent to knowing the one-dimensional chf $\phi_{\vec{\lambda}'\mathsf{X}}(t)$ for all $t \in R$ and $\vec{\lambda} \in R_k$ for which $|\vec{\lambda}| = 1$. This immediately yields the following useful result.

Theorem 5.2 (Cramér–Wold device) If $\mathsf{X}_n \equiv (X_{n1}, \ldots, X_{nk})'$ satisfy

(4) $\quad \phi_{\vec{\lambda}'\mathsf{X}_n}(t) \to \phi_{\vec{\lambda}'\mathsf{X}}(t) \qquad$ for all $t \in R \qquad$ and for each $\vec{\lambda} \in R_k$,

then $\mathsf{X}_n \to_d \mathsf{X}$. (It suffices to show (4) for all unit vectors $\vec{\lambda}$ in R_k.) [In fact, we only require that $\vec{\lambda}'\mathsf{X}_n \to_d \vec{\lambda}'\mathsf{X}$ for all such $\vec{\lambda}$ (no matter what method we use to show it), as such a result implies (4).]

Theorem 5.3 The rvs $X_1, \ldots, X_k$ are independent if and only if the joint chfs satisfy $\phi_{\mathsf{X}}(t_1, \ldots, t_k) = \prod_1^k \phi_{X_i}(t_i)$.

Exercise 5.1* Prove the claims made below (2) for the n-dimensional chf ϕ_{X}.

Exercise 5.2 Prove theorem 5.3.

6 Elementary Complex and Fourier Analysis

Lemma 6.1 (Taylor expansions of $\log(1+z)$ and e^z) [Note that $\log z$ is a many-valued function of a complex $z = re^{i\theta}$; any of $(\log r) + i[\theta + 2\pi m]$ for $m = 0, \pm 1, \pm 2, \ldots$ will work for $\log z$. However, when we write $\log z = \log r + i\theta$, we will always suppose that $-\pi < \theta \le \pi$. Moreover, we denote this unique determination by Log z; this is the *principal branch*.] The Taylor series expansion of Log $(1+z)$ gives

$$\begin{aligned}(1)\qquad &|\text{Log } (1+z) - \textstyle\sum_{k=1}^{m-1}(-1)^{k-1}z^k/k| = |\sum_{k=m}^{\infty}(-1)^{k-1}z^k/k\,| \\ &\le \frac{|z|^m}{m}(1+|z|+|z|^2+\cdots) \le \frac{|z|^m}{m(1-|z|)}\end{aligned}$$

for $|z| < 1$. Thus

$$(2)\qquad |\text{Log } (1+z) - z| \le |z|^2/(2(1-\theta)) \qquad \text{for} \quad |z| \le \theta < 1.$$

From another Taylor series expansion we have for all z that

$$(3)\qquad |e^z - \textstyle\sum_{k=0}^{m-1} z^k/k!\,| = |\sum_{k=m}^{\infty} z^k/k!\,| \le |z|^m \sum_{j=0}^{\infty} \dfrac{|z|^j}{j!}\dfrac{j!}{(j+m)!} \le \dfrac{|z|^m e^{|z|}}{m!}\,.$$

Lemma 6.2 (Taylor expansion of e^{it}) Let $m \ge 0$ and $0 \le \delta \le 1$ (and set the constant $K_{0,0} = 2$, below). Then for all real t we have

$$(4)\qquad \left| e^{it} - \sum_{k=0}^{m} \frac{(it)^k}{k!} \right| \le \frac{\delta 2^{1-\delta}}{(m+\delta)\cdots(2+\delta)(1+\delta)(0+\delta)}\ |t|^{m+\delta} \equiv K_{m,\delta}\ |t|^{m+\delta}\,.$$

Proof. The proof is by induction. For $m = 0$ we have both $|e^{it}-1| \le 2 \le 2\,|t/2|^{\delta}$ for $|t/2| \ge 1$, and (since $\int_0^t ie^{is}\,ds = \int_0^t (i\cos s - \sin s)\,ds = e^{it} - 1$)

$$(a)\qquad |e^{it} - 1| \le |\textstyle\int_0^t ie^{is} ds| \le \int_0^{|t|} ds = |t| \le 2\,|t/2|^{\delta} \qquad \text{for } |t/2| \le 1;$$

so that (4) holds for $m = 0$. We now assume that (4) holds for $m-1$, and we will verify that it thus holds for m. We again use $e^{it} - 1 = i\int_0^t e^{is}\,ds$ and further note that $i\sum_{k=0}^{m-1}\int_0^t [(is)^k/k!]\,ds = \sum_{k=0}^{m-1}(i^{k+1}/k!)\int_0^t s^k\,ds = \sum_1^m (it)^k/k!$ to obtain

$$\begin{aligned} &\left|e^{it} - \textstyle\sum_{k=0}^{m} (it)^k/k!\,\right| = \left| i\int_0^t \left[e^{is} - \sum_{k=0}^{m-1} (is)^k/k! \right] ds \right| \\ (b)\qquad &\le K_{m-1,\delta} \textstyle\int_0^{|t|} s^{m-1+\delta}\,ds \qquad \text{by the induction step} \\ (c)\qquad &= K_{m,\delta}\ |t|^{m+\delta}. \end{aligned}$$

[See Chow and Teicher (1997).] (The next inequality is immediate.) □

Inequality 6.1 (Moment expansion inequality) Suppose $\mathrm{E}|X|^{m+\delta} < \infty$ for some $m \ge 0$ and $0 \le \delta \le 1$. Then

$$(5)\qquad \left| \phi(t) - \sum_{k=0}^{m} \frac{(it)^k}{k!}\mathrm{E}X^k \right| \le K_{m,\delta}\ |t|^{m+\delta}\ \mathrm{E}|X|^{m+\delta} \qquad \text{for all } t.$$

Some Alternative Tools

Lemma 6.3 (The first product lemma) For all $n \geq 1$, let complex $\beta_{n1}, \ldots, \beta_{nn}$ satisfy the following conditions:

(a) $\beta_n \equiv \sum_1^n \beta_{nk} \to \beta$ as $n \to \infty$.

(b) $\delta_n \equiv [\max_{1 \leq k \leq n} |\beta_{nk}|] \to 0$.

(c) $M_n \equiv \sum_{k=1}^n |\beta_{nk}|$ satisfies $\delta_n M_n \to 0$.

Then (compare this with the stronger Lemma 8.1.4, which requires all $\beta_{nk} \geq 0$)

(6) $$\prod_{k=1}^{n} (1 + \beta_{nk}) \to e^{\beta} \qquad \text{as } n \to \infty.$$

Proof. When $0 < \delta_n \leq \frac{1}{2}$ (and we are on the principal branch), (2) gives

(p) $|\sum_{k=1}^n \text{Log}\,(1 + \beta_{nk}) - \sum_{k=1}^n \beta_{nk}| \leq \sum_{k=1}^n |\beta_{nk}|^2 \leq \delta_n M_n \to 0.$

Thus

(q) $\sum_{k=1}^n \text{Log}\,(1 + \beta_{nk}) \to \beta$ as $n \to \infty$.

Moreover, (q) shows that

(r) $\prod_{k=1}^n (1 + \beta_{nk}) = \exp(\text{Log}\,(\prod_{k=1}^n (1 + \beta_{nk}))) = \exp(\sum_{k=1}^n \text{Log}\,(1 + \beta_{nk}))$
$\to \exp(\beta)$,

and this gives (6). [See Chung (1974).] (Recall lemma 8.1.4.) □

Lemma 6.4 (The second product lemma) If $z_1, \ldots, z_n$ and $w_1, \ldots, w_n$ denote complex numbers with modulus at most 1, then

(7) $$\left| \prod_{k=1}^n z_k - \prod_{k=1}^n w_k \right| \leq \sum_{k=1}^n |z_k - w_k| \, .$$

Proof. This is trivial for $n = 1$. We will use induction. Now,

(a) $|\prod_{k=1}^n z_k - \prod_{k=1}^n w_k| \leq |z_n| \left| \prod_{k=1}^{n-1} z_k - \prod_{k=1}^{n-1} w_k \right| + |z_n - w_n| \left| \prod_{k=1}^{n-1} w_k \right|$

(b) $\leq \left| \prod_{k=1}^{n-1} z_k - \prod_{k=1}^{n-1} w_k \right| + |z_n - w_n| \cdot \prod_{k=1}^{n-1} 1 \leq \sum_{k=1}^{n-1} |z_k - w_k| + |z_n - w_n|$

by the induction step. [See most newer texts.] □

Inequality 6.2 (Moment expansions of chfs) Suppose $0 < E|X|^m < \infty$ for some $m \geq 0$. Then (for some $0 \leq g(t) \leq 1$) the chf ϕ of X satisfies

(8) $$\left| \phi(t) - \sum_{k=0}^m \frac{(it)^k}{k!} \, \text{E}X^k \right| \leq \tfrac{3}{m!} \, |t|^m \text{E}|X|^m g(t) \qquad \text{where } g(t) \to 0 \text{ as } t \to 0.$$

Proof. Use the real expansions for sin and cos to obtain

(a) $e^{itx} = \cos(tx) + i\sin(tx) = \sum_{k=0}^{m-1} \frac{(itx)^k}{k!} + \frac{(itx)^m}{m!}\,[\cos(\theta_1 tx) + i\,\sin(\theta_2 tx)]$

(b) $\quad = \sum_{k=0}^{m} \frac{(itx)^k}{k!} + \frac{(itx)^m}{m!}\,[\cos(\theta_1 tx) + i\,\sin(\theta_2 tx) - 1].$

Here, we have some θ_1, θ_2 with $0 \le |\theta_1| \vee |\theta_2| \le 1$. Then (8) follows from (b) via

(9) $\lim_{t\to 0} \mathrm{E}|X^m\,[\cos(\theta_1 tX) - 1 + i\,\sin(\theta_2 tX)]| = 0,$

by the DCT with dominating function $3|X|^m$. [See Breiman (1968).] □

Inequality 6.3 (Summary of useful facts) Let $X \cong (0, \sigma^2)$. Result (8) then gives the highly useful

$$|\,\phi(t) - (1 - \tfrac{1}{2}\,\sigma^2\,t^2)\,| \le \tfrac{3}{2}\,\sigma^2\,t^2\,g(t) \qquad \text{where } g(t) \to 0 \text{ as } t \to 0,$$

with $0 \le g(t) \le 1$. Applying this and (5) gives (since $K_{2,1} = \frac{1}{6}$, and since $K_{1,1} = \frac{1}{2}$ allows $\frac{1}{2}\,\sigma^2\,t^2$ to replace $\frac{3}{2}\ \sigma^2\,t^2 g(t)$)

(10) $\left|\phi(t) - (1 - \tfrac{1}{2}\,\sigma^2 t^2)\right| \le \tfrac{1}{2}\,\sigma^2\,t^2 \wedge \tfrac{1}{6}\,\mathrm{E}|X|^3\,|t|^3 \qquad \text{for all } t \in R.$

Exercise 6.1 (Distributions determined by their moments)
(a) Suppose that $\mathrm{E}|X|^n < \infty$. Then the nth derivative $\phi^{(n)}(\cdot)$ is a continuous function given by $\phi^{(n)}(t) = i^n \mathrm{E}(X^n e^{itX})$, so that $\mathrm{E}X^n = i^{-n}\phi^{(n)}(0)$.
(b) The series $\psi(t) = \sum_0^\infty (it)^k \mathrm{E}(X^k)/k!$ has radius of convergence $R \equiv 1/(eL)$, where $L = \overline{\lim}_k\,|\mu_k|^{1/k}/k = \overline{\lim}_k\,(\mathrm{E}|X|^k)^{1/k}/k = \overline{\lim}\,\mu_{2k}^{1/2k}/(2k)$; use the root test.
(c) The series in (b) has the same $R > 0$ if $\sum_{k=0}^\infty \mu_{2k}\,t^{2k}/(2k)! < \infty$ for some $t > 0$.
(d) The series (b) converges for $|t| < r$ if and only if $\mathrm{E}\exp(t\,|X|) < \infty$ for $|t| < r$ if and only if $\mathrm{E}\exp(t\,X) < \infty$ for $|t| < r$
(e) If the radius of convergence in (b) is strictly positive, then the distribution having the stated moments is uniquely determined by its moments μ_k.
(f) Show that the Normal$(0,1)$ distribution is uniquely determined by its moments.
(g) Show that any Gamma$(r,1)$ distribution is uniquely determined by its moments.
(h) Show that it is valid to expand the mgs of Normal$(0,1)$ and Gamma$(r,1)$ to compute their moments. Do it.

Exercise 6.2 (a) If $\phi''(0)$ is finite, then σ^2 is finite. Prove this.
(b) In fact, if $\phi^{(2k)}(0)$ is finite, then $\mathrm{E}X^{2k} < \infty$. Prove this.
$(*)$ Appeal to Exercise 4.3.2.

Exercise 6.3 (Bounds on $(1 - x/n)^n$) (i) Use $(1 + t) \le e^t \le 1/(1 - t)$ for $0 \le t < 1$ at $t = x/n$ to show that

$$0 \le e^{-x} - (1 - \tfrac{x}{n})^n \le \tfrac{x^2}{n}\,e^{-x} \qquad \text{for } 0 \le x < n\,.$$

(ii) (Hall and Wellner) Show that

$$2\,e^{-2} \le n\,\sup_{x\ge 0}\,\left|e^{-x} - (1 - x/n)^n\,1_{[0,n]}(x)\right| \le (2 + n^{-1})e^{-2} \ \text{ for all } n \ge 1.$$

Results from Fourier Analysis

On some occasions we will need to know the behavior of $\phi(t)$ for $|t|$ large.

Lemma 6.5 (Riemann–Lebesgue lemma) If $\int_{-\infty}^{\infty} |g(x)|dx < \infty$, then

$$(11) \qquad \int_{-\infty}^{\infty} e^{itx} g(x)dx \to 0 \qquad \text{as } t \to \infty.$$

Proof. Now, $\Psi \equiv \{\psi \equiv \sum_1^m c_i 1_{(a_i,b_i]} : a_i, b_i, c_i \in R \text{ and } m \geq 1\}$ is dense in $\mathcal{L}_1$ by theorem 3.5.8; that is, if $\int_{-\infty}^{\infty} |g(x)|dx < \infty$, then there exists $\psi \in \Psi$ such that $\int_{-\infty}^{\infty} |g - \psi|\, dx < \epsilon$. Thus $\gamma(t) \equiv |\int_{-\infty}^{\infty} e^{itx} g(x)dx|$ satisfies

$$\gamma(t) \leq \textstyle\int_{-\infty}^{\infty} |e^{itx}|\, |g(x) - \psi(x)|\, dx + |\int_{-\infty}^{\infty} e^{itx} \psi(x)\, dx|$$

$$(a) \qquad \leq \epsilon + \textstyle\sum_1^m |c_i|\, |\int_{a_i}^{b_i} e^{itx}\, dx|.$$

It thus suffices to show that for any a, b in R we have

$$(12) \qquad \int_a^b e^{itx}\, dx \to 0 \qquad \text{as} \quad |t| \to \infty.$$

A quick picture of sines and cosines oscillating very fast (and canceling out over the interval) shows that (12) is trivial. (Or write $e^{itx} = \cos(tx) + i\sin(tx)$ and compute the integrals.) □

Lemma 6.6 (Tail behavior of chfs)
(i) If F has density f with respect to Lebesgue measure, then $|\phi(t)| \to 0$ as $|t| \to \infty$.
(ii) If F has $n+1$ integrable derivatives $f, f', \ldots, f^{(n)}$ on R, then

$$(13) \qquad |t|^n\, |\phi(t)| \to 0 \qquad \text{as} \quad |t| \to \infty.$$

Proof. The fact that $|\phi(t)| \to 0$ as $|t| \to \infty$ follows from the Reimann–Lebesgue lemma, since f is integrable. Since f is absolutely continuous and a density, it follows that that $f(x) \to 0$ as $|x| \to \infty$. Then use

$$\phi(t) = \textstyle\int e^{itx} f(x)\, dx = \int f(x)\, d(e^{itx}/it)$$

$$(a) \qquad = (e^{itx}/it)\, f(x)\big|_{-\infty}^{\infty} - \textstyle\int e^{itx}\, f'(x)\, dx/(it)$$

$$(b) \qquad = -\textstyle\int e^{itx}\, f'(x)\, dx/(it) \qquad \text{with} \qquad f'(\cdot) \in \mathcal{L}_1,$$

using $f(x) \to 0$ as $|x| \to \infty$ in going from (a) to (b). (Note exercise 6.4 below.) Applying the Riemann–Lebesgue lemma to (b) gives $|t|\, |\phi(t)| \to 0$ as $|t| \to \infty$. Keep on integrating by parts and applying the Riemann–Lebesgue lemma. □

Exercise 6.4 Verify lemma 6.2(ii) when $n = 1$.

Other Alternative Tools

Then chf always exists, so it can always be used. However, if $X \geq 0$ or if X is integer valued, *then* Laplace transforms or probability generating functions offer more elementary tools.

Exercise 6.5* (Laplace transform) Let $\mathcal{F}^+$ denote the class of all dfs F having $F_-(0) = 0$. For any df $F \in \mathcal{F}^+$ we define the *Laplace transform* L of F by

(14) $$L(\lambda) = \mathrm{E}\, e^{-\lambda X} = \int_0^\infty e^{-\lambda x}\, dF(x) \qquad \text{for} \quad \lambda \geq 0.$$

(a) Establish an analogue of proposition 3.1(a), (b), (c), and (g).
(b) (Uniqueness) Show that each df in $\mathcal{F}^+$ has a unique Laplace transform.
(c) (Continuity) Let $X_n \cong F_n \in \mathcal{F}^+$. If $L_n(\lambda) \to$ (some $L(\lambda)$) for all $\lambda \geq 0$ with $L(\cdot)$ right continuous at 0, then L is the Laplace transform of a df $F \in \mathcal{F}^+$ for which the convergence in distribution $F_n \to_d F$ holds.
(d) Establish analogues of inequality 6.1 on moment expansions.

Exercise 6.6* (Probability generating function) Let $\mathcal{F}^I$ denote the class of all dfs F assigning mass 1 to the integers $0, 1, 2, \ldots$. For any df $F \in \mathcal{F}^I$ we define the *probability generating function* g of F by

(15) $$g(z) = \mathrm{E}\, z^X = \sum_{k=0}^\infty p_k\, z^k \qquad \text{for all complex } z \text{ having } |z| \leq 1.$$

(a) Establish an analogue of proposition 3.1.
(b) (Uniqueness) Show that each df F in $\mathcal{F}^I$ has a unique generating function.
(c) (Continuity) Let $X_n \cong F_n \in \mathcal{F}^I$. If $g_n(z) \to$ (some $g(z)$) for all $|z| \leq 1$ with $g(\cdot)$ continuous at 1, then g is the generating function of a df F in $\mathcal{F}^I$ for which $F_n \to_d F$.

Exercise 6.7 (Cumulant generating function) The *cumulant generating function* $\psi_X(\cdot)$ of a rv X is defined by

(16) $$\psi_X(t) \equiv \mathrm{Log}\ \phi_X(t) = \mathrm{Log}\ \mathrm{E}(e^{itX}),$$

and is necessarily finite for t-values in some neighborhood of the origin.
(a) Temporarily suppose that all moments of X are finite. Let $\mu_k \equiv \mathrm{E}(X - \mu)^k$ denote the k-th central moment, for $k \geq 1$. Then when $\mu = \mathrm{E}X = 0$ and with $\sigma^2 \equiv \mu_2$, we have the formal expansion

$$\phi_X(t) = 1 - t^2\sigma^2/2 + (it)^3\mu_3/3! + (it)^4\mu_4/4! + \cdots \equiv 1 + z\,.$$

Verify that further *formal* calculations based on this yield

$$\psi_X(t) = \mathrm{Log}\ \phi_X(t) = \mathrm{Log}\ (1 + z) = z - z^2/2 + z^3/3 + \cdots$$

$$= (it)^2\mu_2/2! + (it)^3\mu_3/3! + (it)^4(\mu_4 - 3\mu_2^2)/4! + \cdots$$

(17) $$= (it)^2\sigma^2/2! + (it)^3\mu_3/3! + t^4(\mu_4 - 3\sigma^4)/4! + \cdots$$

(18) $$\equiv \sum_{j=2}^\infty (it)^j \kappa_j/j!\,,$$

where κ_j is called the jth *cumulant* of X. Note that for independent rvs $X_1, \ldots, X_n$,

(19) $\quad$ (the jth cumulant of $\sum_{k=1}^n X_k$) $= \sum_{k=1}^n$(the jth cumulant of X_k),

which is nice. Then verify that in the iid case, the third and fourth cumulants of the standardized rv $Z_n \equiv \sqrt{n}(\bar{X}_n - \mu)/\sigma$ are

(20) $\quad \gamma_1/\sqrt{n} \equiv (\mu_3/\sigma^3)/\sqrt{n} \qquad \text{and} \qquad \gamma_2/n \equiv (\mu_4/\sigma^4 - 3)/n,$

where γ_1 measures skewness and γ_2 measures tail heaviness. [This is particularly nice; it shows that the effect (on the distribution of $\bar{X}_n$) of skewness disappears at rate $1/\sqrt{n}$, while the effect of tail heaviness disappears at rate $1/n$.]
(b) Finally, if only $\mathrm{E}|X|^m < \infty$ for some $m \geq 1$ is known, show that in a sufficiently small neighborhood of the origin

(21) $\quad |\psi(t) - \sum_{j=2}^m \kappa_j (it)^j/j!\,| \leq c_m\, |t|^m\, \mathrm{E}|X|^m\, \delta_m(t),$

where $\delta_m(t) \searrow 0$ as $t \searrow 0$, and

(22) $\quad |\psi(t) - \sum_{j=2}^{m-1} \kappa_j (it)^j/j!\,| \leq \bar{c}_m\, |t|^m\, \mathrm{E}|X|^m$

for some universal constant $\bar{c}_m$. The exercise is to establish carefully that all of this is true.

7 Esseen's Lemma

Let G denote a fixed function having $G(-\infty) = 0$, $G(+\infty) = 1$, having derivative g on the real line for which $|g(\cdot)|$ is bounded by some constant M, having $\int_{-\infty}^{\infty} xg(x)\,dx = 0$, and then let $\psi(t) \equiv \int_{-\infty}^{\infty} e^{itx} g(x)\,dx$. Let F denote a general df having mean 0, and let ϕ denote its characteristic function. We wish to estimate $\|F - G\| = \sup_{-\infty<x<\infty} |F(x) - G(x)|$ in terms of the distance between ϕ and ψ. Roughly speaking, the next inequality says that if ϕ and ψ are sufficiently close over most of their domain, then $\|F - G\|$ will be small. [In the initial application of this Esseen's lemma, we will take G, g, ψ to be the $N(0,1)$ df, density, and chf. In this context, the constant of (1) is $24\|g\|/\pi = 24M/\pi = 24/(\sqrt{2\pi}\pi) = 3.047695\ldots$.]

Inequality 7.1 (Esseen's lemma) Let F and G be as above. For any $a > 0$ we have the uniform bound

$$(1)\qquad \|F - G\| \le \frac{1}{\pi}\int_{-a}^{a} \left|\frac{\phi(t) - \psi(t)}{t}\right| dt + \frac{24\,\|g\|}{\pi a}.$$

Proof. The key to the technique is to smooth by convolving F and G with the df H_a whose density h_a and characteristic function γ_a are given by

$$(2)\qquad h_a(x) = \frac{1 - \cos(ax)}{\pi a x^2} \quad \text{on } R \qquad \text{and} \qquad \gamma_a(t) = \begin{cases} 1 - |t|/a & \text{if } |t| \le a, \\ 0 & \text{if } |t| > a. \end{cases}$$

This h_a is the density of V/a, when V has the de la Vallée density. Let F_a and G_a denote the convolutions of F and G with H_a, for "a" large. We will now show that

$$(3)\qquad \|F - G\| \le 2\|F_a - G_a\| + 24\,\|g\|/(\pi a).$$

Let $\Delta \equiv F - G$. Now, $\Delta(x) = \Delta_+(x)$ and $\Delta_-(x)$ exist for all x; thus there exists x_o such that either $D \equiv \|F - G\| = |\Delta(x_o)|$ or $D = |\Delta_-(x_o)|$. Without loss of generality, we suppose that $D = |\Delta(x_o)|$ (just replace X, Y by $-X, -Y$ if not). Note figure 5.1. Without loss of generality, we act below as though $\Delta(x_o) > 0$, and we let $z_o > x_o$. (If $\Delta(x_o) < 0$, then let $z_o < x_o$). Now, since F is $\nearrow$ and g is bounded by M, we have

$$(a)\qquad \Delta(z_o - x) \ge D/2 + Mx \qquad \text{for } |x| \le \epsilon = \tfrac{D}{2M},$$

where $\epsilon \equiv D/2M$ and $z_o \equiv x_o + \epsilon$. Trivially (since D was the supremum),

$$(b)\qquad \Delta(z_o - x) \ge -D \qquad \text{for } |x| > \epsilon.$$

Thus, with $\Delta_a \equiv F_a - G_a$, using (a) and (b) gives

$$(c)\qquad \|F_a - G_a\| \ge \Delta_a(z_o) = \int_{-\infty}^{\infty} \Delta(z_o - x) h_a(x)\,dx \qquad \text{by the convolution formula}$$

$$(d)\qquad \ge \int_{-\epsilon}^{\epsilon} [D/2 + Mx]\,h_a(x)\,dx - D\int_{[|x|>\epsilon]} h_a(x)\,dx$$

$$= (D/2)[1 - \int_{[|x|>\epsilon]} h_a(x)\,dx] + M\cdot 0 - D\int_{[|x|>\epsilon]} h_a(x)\,dx$$

since $xh_a(x)$ is odd

$$= (D/2) - (3D/2)\int_{[|x|>\epsilon]} h_a(x)\,dx \geq (D/2) - (12M/\pi a)$$

(e) $$= \|F - G\|/2 - (12M/\pi a),$$

(which is (3)), since

(f) $$\int_{[|x|>\epsilon]} h_a(x)\,dx \leq 2\int_\epsilon^\infty (2/\pi a x^2)\,dx = 4/(\pi a \epsilon) = 8M/(\pi a D).$$

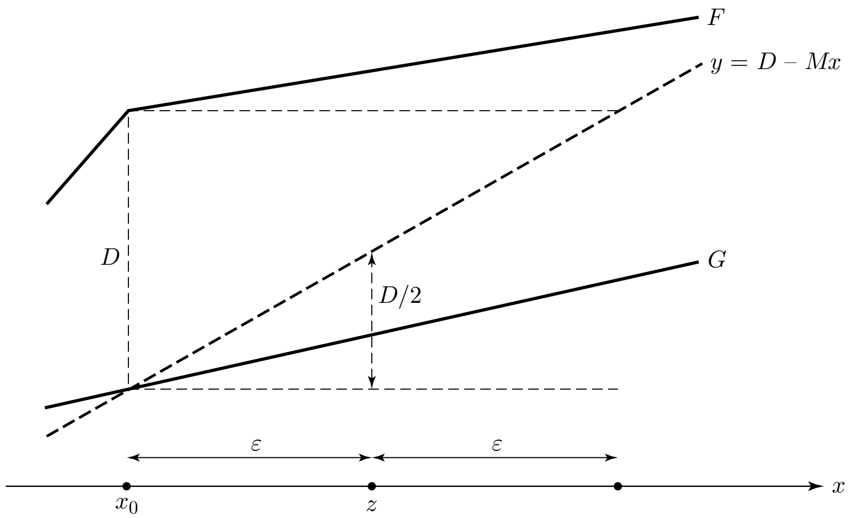


Figure 5.1 Bounds for Esseen's lemma.

We now bound $\|F_a - G_a\|$. By the Fourier inversion formula, F_a and G_a have bounded continuous "densities" that satisfy

(4) $$f_a(x) - g_a(x) = \int_{-a}^a e^{-itx}\,[\phi(t) - \psi(t)]\,\gamma_a(t)\,dt/(2\pi).$$

From this we suspect that

(5) $$\Delta_a(x) = \frac{1}{2\pi}\int_{-a}^a e^{-itx}\,\frac{\phi(t) - \psi(t)}{-it}\,\gamma_a(t)\,dt\,.$$

That the integrand is a continuous function that equals 0 at $t = 0$ (since F and G have 0 "means," inequality 9.6.1 gives this) makes the right-hand side well-defined, and we may differentiate under the integral sign by the DCT [with dominating function $\gamma_a(\cdot)$] to get the previous equation (4). Thus $\Delta_a(x)$ can differ from the right-hand side of (5) by at most a constant; but this constant is 0, since obviously $\Delta_a(x) \to 0$ as $|x| \to \infty$, while the right-hand side does the same by the Riemann–Lebesgue lemma. Equation (5) gives

(6) $$|\Delta_a(x)| \leq \frac{1}{2\pi}\int_{-a}^a \left|\frac{\phi(t) - \psi(t)}{t}\right| dt \qquad \text{for all } x.$$

Combining (3) and (6) gives (1). □

Corollary 1 (Stein) Suppose that instead of convolving F and G with the H_a of (2), we convolve with an arbitrary df H instead. In this situation we obtain

(7) $$\|F - G\| \le 2\,\|F * H - G * H\| + 8\,\|g\|\,\mathrm{E}|H^{-1}(\xi)|\,.$$

Proof. Picking up at line (d) of the previous proof (with $Y \cong H$), we obtain

(d) $$\|F * H - G * H\| \ge \int_{[-\epsilon,\epsilon]} [D/2 + My]\,dH(y) - D\,P(|Y| > \epsilon)$$

$$\ge (D/2)\,[1 - P(|Y| > \epsilon)] - M\,\mathrm{E}|Y| - D\,P(|Y| > \epsilon)$$

$$\ge (D/2) - (3D/2)\,P(|Y| > \epsilon) - M\,\mathrm{E}|Y|$$

(e) $$\ge (D/2) - 4M\,\mathrm{E}|Y|$$

using Markov's inequality and $\epsilon \equiv D/2M$ in the last step. □

8 Distributions on Grids

Definition 8.1 We say that a rv X is *distributed on a grid* if there exist real numbers a, d such that the probabilities $p_n \equiv P(X = a + nd)$ satisfy $\sum_{-\infty}^{\infty} p_n = 1$. We call d the *span* of the grid. The *maximal span* is $\sup\{|d| : |d| \text{ is a span}\}$.

Proposition 8.1 If $t_0 \neq 0$, the following are equivalent:
(a) $|\phi(t_0)| = 1$.
(b) $|\phi|$ has period t_0; that is, $|\phi(t + nt_0)| = |\phi(t)|$ for all n and t.
(c) The rv X is distributed on a grid of span $d = 2\pi/t_0$.

Proof. Suppose that (a) holds. Then $\phi(t_0) = e^{i\alpha}$ for some real α. That is, $\int e^{it_0 x} dF(x) = e^{i\alpha}$, or $\int e^{i(t_0 x - \alpha)} dF(x) = 1$. Taking real parts gives

(p) $\quad \int_{-\infty}^{\infty} [1 - \cos(t_0 x - \alpha)]\, dF(x) = 0.$

Since the integrand is nonnegative for all X, this means that

(q) $\quad 1 - \cos(t_0 x - \alpha) = 0 \qquad \text{a.s. } F;$

that is,

(r) $\quad t_0 X - \alpha \in \{2\pi m : m = 0, \pm 1, \pm 2, \ldots\} \qquad \text{a.s.}$

That is, $X \in \{\alpha/t_0 + (2\pi/t_0)m : m = 0, \pm 1, \pm 2, \ldots\}$ a.s.; so (c) holds.

Suppose (c) holds. Then (b) holds, since

$$|\phi(t + nt_0)| = |\textstyle\sum_{m=-\infty}^{\infty} p_m\, e^{i(t+nt_0)(a+dm)}|$$

$$= |e^{i(t+nt_0)a}|\, |\textstyle\sum_{m=-\infty}^{\infty} p_m\, e^{i(t+nt_0)dm}|$$

$$= |\textstyle\sum_{m=-\infty}^{\infty} p_m\, e^{i(t+2\pi n/d)\, dm}|$$

(s) $$= |\textstyle\sum_{m=-\infty}^{\infty} p_m\, e^{itdm}|\, |e^{i2\pi nm}| = |\textstyle\sum_{m=-\infty}^{\infty} p_m\, e^{itdm}\, e^{ita}| = |\phi(t)|.$$

Suppose that (b) holds. Then

(t) $\quad 1 = |\phi(0)| = |\phi(0 + t_0 n)| = |\phi(0 + t_0\, 1)| = |\phi(t_0)|,$

so that (a) holds. □

Corollary 2 If $a = 0$ in (c), then we may replace $|\phi|$ by ϕ in (a) and (b), and proposition 8.1 will still hold.

Proposition 8.2 One of the following possibilities must hold:
(d) $|\phi(t)| < 1$ for all $t \neq 0$.
(e) $|\phi(t)| < 1$ for $0 < t < 2\pi/d$ and $|\phi(2\pi/d)| = 1$. Thus, X has maximal span d.
(f) $|\phi(t)| = 1$ for all t. And so $\phi(t) = e^{iat}$ for all t and $P(X = a) = 1$, for some a.

Proof. Clearly, either (d), (e), or (f) holds, or else $|\phi(t_n)| = 1$ for some sequence $t_n \to 0$. In this latter case, $|\phi(mt_n)| = 1$ for all m, for each n by proposition 8.1. Since $\{mt_n : n \geq 1, m = 0, \pm 1, \pm 2, \ldots\}$ is dense in R and since ϕ, and thus $|\phi|$, is continuous, we must have case (f) again. It remains to establish the consequences of (e) and (f).

Consider (e). Proposition 8.1 shows that (e) holds if and only if both d is a span and no number exceeding d is a span.

In the case of (f), we have $|\phi(t_1)| = 1 = |\phi(t_2)|$ for some t_1 and t_2 having $t_1/t_2 =$ (an irrational number). But $|\phi(t_1)| = 1$ and $|\phi(t_2)| = 1$ imply that both $2\pi/t_1$ and $2\pi/t_2$ are spans. Thus if at least two points have positive mass, then the distance between them must equal $m_1 2\pi/t_1$ for some integer m_1 and it must equal $m_2 2\pi/t_2$ for some integer m_2. That is, $2\pi m_1/t_1 = 2\pi m_2/t_2$, or $t_1/t_2 = m_1/m_2 =$ (a rational number). This contradiction shows that there can be at most one mass point a. □

Exercise 8.1 (Inversion formula for distributions on a grid) Let X be distributed on a grid with $p_n = P(X = a + dn)$. Then $\phi(t) = \sum_{-\infty}^{\infty} p_n \, e^{it(a+dn)}$. Show that

$$(1) \qquad p_m = \frac{d}{2\pi} \int_{-\pi/d}^{\pi/d} \phi(t) \, e^{-it(a+dm)} \, dt.$$

9 Conditions for ϕ to Be a Characteristic Function

Example 9.1 Now, for $a > 0$,

$$F_a'(x) = [1 - \cos(ax)]/(\pi a x^2) \qquad \text{for } x \in R$$

is a de la Vallée Poussin density function with chf

$$\phi_a(t) = (1 - |t|/a)\, 1_{[-a,a]}(t) \qquad \text{for } t \in R.$$

Let F_a denote the df. Then

$$F \equiv \sum_1^n p_i\, F_{a_i} \qquad \text{with} \qquad p_i \ge 0, \quad \textstyle\sum_1^n p_i = 1, \quad \text{and} \quad 0 < a_1 < \cdots < a_n$$

is a df with characteristic function

$$\phi = \textstyle\sum_1^n p_i\, \phi_{a_i}.$$

Thus any even function $\phi \ge 0$ with $\phi(0) = 1$ whose graph on $[0, \infty)$ is a convex polygon is a chf. □

Proposition 9.1 (Pólya) Let $\phi \ge 0$ be an even function with $\phi(0) = 1$ whose graph on $[0, \infty)$ is convex and $\downarrow$. Then ϕ is a chf.

Proof. Pass to the limit in the obvious picture, using the continuity theorem to complete the proof. □

Bochner's theorem below gives necessary and sufficient conditions for a function to be a chf. We merely state it, as a background fact. Its proof can be found in a number of the standard texts.

Definition 9.1 A complex-valued function $\phi(\cdot)$ on R is *nonnegative definite* if for any finite set T and any complex-valued function $h(\cdot)$ we have

(1) $\qquad \sum_{s,t \in T} \phi(s - t) h(s) \bar{h}(t) \ge 0.$

Theorem 9.1 (Bochner) A complex-valued function $\phi(\cdot)$ is a chf if and only if it is nonnegative definite and continuous.

Chapter 10

CLTs via Characteristic Functions

0 Introduction

The classical CLT states that if $X_1, X_2, \ldots$ are iid (μ, σ^2), then

$$\sqrt{n}\,(\bar{X}_n - \mu) \to_d N(0, \sigma^2) \qquad \text{as } n \to \infty.$$

This chapter will also consider the following generalizations:

(i) Triangular arrays of row-independent non-iid rvs $X_{n1}, \ldots, X_{nn}$ for $n \geq 1$. Liapunov, Lindeberg–Feller, and the General CLT.

(ii) The speed of convergence of the dfs to the limiting df, via Berry–Esseen.

(iii) Necessary and sufficient conditions for iid rvs to satisfy the CLT.

(iv) Necessary and sufficient conditions for the weak and strong bootstrap.

(v) Convergence of the density functions as well (the local CLT.)

(vi) Random sample sizes, sample quantiles, and many other examples.

(vii) The multidimensional CLT.

(viii) Non-normal limits (with both the degenerate WLLN and the Poisson).

(ix) Note the summaries in Appendix D.

In chapter 11 we will consider situations that lead to both stable and infinitely divisible rvs as limits. Edgeworth and other approximations are also considered there. Chapter 13 includes a discussion of martingale CLTs. Chapter 15 has sections on trimmed means, asymptotic normality of L-statistics, and asymptotic normality of R-statistics (the latter includes a finite sampling CLT).

The proofs for (iii) and (iv) above require knowledge of Sections **??**.**??**–**??**.**??**. An inequality in Sections **??**.**??** is required for the proofs of the Chapter 15 examples.

1 Basic Limit Theorems

The goal of this section is to use a chf approach to present the classical central limit theorems for sums of iid random variables in R and in R_k. We also compare and contrast the central limit theorem with the Poisson limit theorem.

The Classical CLT

Theorem 1.1 (Classical CLT) For each $n \geq 1$, let $X_{n1}, \ldots, X_{nn}$ be iid $F(\mu, \sigma^2)$; this denotes that the df $F(\cdot)$ of the X_{nk}'s has mean μ and finite variance σ^2. Define the total $T_n \equiv X_{n1} + \cdots + X_{nn}$ and the average $\bar{X}_n \equiv T_n/n$. Then as $n \to \infty$,

(1) $$\sqrt{n}\,(\bar{X}_n - \mu) = \tfrac{1}{\sqrt{n}}\,(T_n - n\mu) = \tfrac{1}{\sqrt{n}} \textstyle\sum_{k=1}^n (X_{nk} - \mu) \to_d N(0, \sigma^2)\,.$$

Proof. Now, for fixed t we have (with $0 \leq g(t) \leq 3/2$ and $g(t) \to 0$ as $t \to 0$)

(a) $$\phi_{\sqrt{n}\,(\bar{X}_n-\mu)}(t) = \textstyle\prod_{k=1}^n \phi_{(X_{nk}-\mu)/\sqrt{n}}\,(t) = [\phi_{X_{nk}-\mu}(t/\sqrt{n}\,)]^n$$

$$= \left[1 - \frac{\sigma^2}{2}\left(\frac{t}{\sqrt{n}}\right)^2 + \left(\frac{t}{\sqrt{n}}\right)^2 \sigma^2\, g\left(\frac{t}{\sqrt{n}}\right)\right]^n \qquad \text{by inequality 9.6.2}$$

(b) $$= \left[1 - \frac{\sigma^2 t^2}{2\,n} + \frac{t^2}{n} \times \sigma^2\, g\left(\frac{t}{\sqrt{n}}\right)\right]^n .$$

The first product lemma 9.6.3 with $\theta = -\sigma^2 t^2/2$ trivially applies. Thus

(c) $$\phi_{\sqrt{n}\,(\bar{X}_n-\mu)}(t) \to e^{-\sigma^2 t^2/2} = \phi_{N(0,\sigma^2)}(t)\,,$$

using table 13.1.1. Thus $\sqrt{n}\,(\bar{X}_n - \mu) \to_d N(0, \sigma^2)$ by the Cramér–Lévy continuity theorem 9.5.1 and the uniqueness theorem 9.4.1.

Had we chosen to appeal to the second product lemma 9.6.4 instead, we would have instead claimed that

$$\left|\, \phi_{\sqrt{n}\,(\bar{X}_n-\mu)}(t) - (1 - \sigma^2 t^2/2n)^n \right|$$

(d) $$= \left|\, \textstyle\prod_{k=1}^n \phi_{(X_{nk}-\mu)/\sqrt{n}}\,(t) - \prod_{k=1}^n (1 - \sigma^2 t^2/2n) \right|$$

(e) $$\leq \textstyle\sum_{k=1}^n \left|\phi_{(X_{nk}-\mu)/\sqrt{n}}\,(t) - (1 - \sigma^2 t^2/2n)\right|$$

(f) $$\leq \textstyle\sum_{k=1}^n (t^2/n)\,\sigma^2\, g(t/\sqrt{n}\,) = t^2\,\sigma^2\, g(t/\sqrt{n}\,) \to 0.$$

But $(1 - \sigma^2 t^2/2n)^n \to \exp(-\sigma^2 t^2/2) = \phi_{N(0,\sigma^2)}(t)$, so the continuity theorem and the uniqueness theorem again complete the proof. □

Degenerate Limits

Exercise 1.1 (WLLN, or classical degenerate convergence theorem) For each $n \geq 1$, let $X_{n1}, \ldots, X_{nn}$ be iid with finite mean μ. Use chfs to show the WLLN result that $\bar{X}_n \to_p \mu$ as $n \to \infty$. Equivalently,

(2) $\bar{X}_n \to_d$ (the degenerate distribution with mass 1 at μ).

The Classical PLT

Theorem 1.2 (Classical Poisson limit theorem; the PLT) For each $n \geq 1$, suppose that $X_{n1}, \ldots, X_{nn}$ are independent Bernoulli(λ_{nk}) rvs for which the values of the parameters satisfy $\lambda_n \equiv \sum_{k=1}^n \lambda_{nk} \to \lambda \in (0, \infty)$ while $[\max_{1\leq k\leq n} \lambda_{nk}] \to 0$. (This is true if $\lambda_{n1} = \cdots = \lambda_{nn} = \lambda_n/n$ for all n, with $\lambda_n \to \lambda$). Then

(3) $\quad T_n \equiv X_{n1} + \cdots + X_{nn} \to_d \text{ Poisson}(\lambda) \qquad \text{as } n \to \infty.$

Proof. From table 13.1.1 we have $\phi_{X_{nk}}(t) = 1 + \lambda_{nk}(e^{it} - 1)$. Thus

(a) $\quad \phi_{T_n}(t) = \prod_{k=1}^n \phi_{X_{nk}}(t) = \prod_{k=1}^n [1 + \lambda_{nk}(e^{it} - 1)]$

(b) $\quad \to \exp(\lambda(e^{it} - 1))$ by the first product lemma 9.6.3

(c) $\quad = \phi_{\text{Poisson}(\lambda)}(t)$ by table 13.1.1.

Now apply the Cramér–Lévy continuity theorem and the uniqueness theorem. □

Exercise 1.2 (Poisson local limit theorem) Show that

(4) $\quad P(T_n = k) \to P(\text{ Poisson}(\lambda) = k) \qquad \text{as } n \to \infty, \text{ for } k = 0, 1, \ldots,$

when $\lambda_{n1} = \cdots = \lambda_{nn}$ in the PLT. Show that this implies

(5) $\quad d_{TV}(P_n, P) \equiv \sup\{|P_n(A) - P(A)| : A \in \mathcal{B}\} \to 0,$

where $T_n \cong P_n$ and Poisson(λ) $\cong P$. [Exercise 11.5.4 will improve this.]

Exercise 1.3 Show that if $T_\lambda \cong$ Poisson(λ), then $(T_\lambda - \lambda)/\sqrt{\lambda} \to_d N(0,1)$ as the parameter $\lambda \to \infty$.

A Comparison of Normal and Poisson Convergence

Exercise 1.4 (a) Suppose the hypotheses of the classical CLT hold. Show that

(6) $\quad M_n \equiv \left[\max_{1\leq k\leq n} \frac{1}{\sqrt{n}} |X_{nk} - \mu|\right] \to_p 0\,.$

(b) Suppose the hypotheses of the classical PLT hold. Show that

(7) $\quad M_n \equiv \left[\max_{1\leq k\leq n} |X_{nk}|\right] \to_d \text{Bernoulli}(1 - e^{-\lambda})\,.$

(∗) There is something fundamentally different regarding the negligibility of the corresponding terms in these two cases! The CLT involves summing many tiny pieces, but the PLT arises from very occasionally having a "large" piece.

Remark 1.1 Let $Y_{n1}, \ldots, Y_{nn}$ be independent. Let $p_{nk}^\epsilon \equiv P(|Y_{nk}| > \epsilon)$. Recall equation (8.3.14) for the conclusion

(8) $\quad M_n \equiv \left[\max_{1\leq k\leq n} |Y_{nk}|\right] \to_p 0 \quad \text{if and only if} \quad \sum_{k=1}^n p_{nk}^\epsilon \to 0 \quad \text{for all } \epsilon > 0.$

This was proved via the (8.3.13) inequality

(9) $\quad 1 - \exp(-\sum_{k=1}^n p_{nk}^\epsilon) \leq P(M_n \geq \epsilon) \leq \sum_{k=1}^n p_{nk}^\epsilon\,.$ □

The Multivariate CLT

Theorem 1.3 (Classical multivariate CLT) Let $\mathsf{X}_n \equiv (X_{n1}, \ldots, X_{nk})'$, $n \geq 1$, be a sequence of iid $(\vec{\mu}, \Sigma)$ random vectors. Then

(10) $$\frac{1}{n^{1/2}} \sum_{j=1}^{n} (\mathsf{X}_j - \vec{\mu}) \to_d N_k(\mathsf{O}, \Sigma) \qquad \text{as } n \to \infty.$$

Proof. For any $\vec{\lambda} \in R_k$ the rvs

(a) $Y_j \equiv \vec{\lambda}'(\mathsf{X}_j - \vec{\mu}\,) \cong (0, \vec{\lambda}'\Sigma\vec{\lambda}\,) \qquad$ are iid for $j = 1, \ldots, n$.

Thus the classic CLT gives

(b) $\sqrt{n}\,\bar{Y}_n \to_d N(0, \vec{\lambda}'\Sigma\vec{\lambda}\,)$.

That is, $\mathsf{Z}_n \equiv n^{-1/2} \sum_1^n (\mathsf{X}_j - \vec{\mu}\,)$ satisfies

(c) $\phi_{\vec{\lambda}'\mathsf{Z}_n}(t) = \phi_{\sqrt{n}\,\bar{Y}_n}(t) \to \exp(-\vec{\lambda}'\Sigma\vec{\lambda}\, t^2/2)$.

Now, if $\mathsf{Z} \cong N_k(\mathsf{O}, \Sigma)$, then $\vec{\lambda}'\mathsf{Z} \cong N(0, \vec{\lambda}'\Sigma\vec{\lambda}\,)$; and hence

(d) $\phi_{\vec{\lambda}'\mathsf{Z}}(t) = \exp(-\vec{\lambda}'\Sigma\vec{\lambda}\, t^2/2)$.

Thus (c) and (d) give $\phi_{\vec{\lambda}'\mathsf{Z}_n}(t) \to \phi_{\vec{\lambda}'\mathsf{Z}}(t)$ for all $t \in R$, for each $\vec{\lambda} \in R_k$. Thus the Cramér–Wold theorem (theorem 9.5.2) shows that $\mathsf{Z}_n \to_d \vec{Z}$. □

Exercise 1.5 (Empirical process; Doob) Let $\mathbb{U}_n \equiv \sqrt{n}\,[\mathbb{G}_n - I]$ be the uniform empirical process of sections 6.5 and 12.10, and let $\mathbb{U}$ denote the Brownian bridge of (A.4.13). Show that $\mathbb{U}_n \to_{fd} \mathbb{U}$ as $n \to \infty$; that is, show that for any set of points $0 < t_1 < \cdots < t_k < 1$ we have

$$(\mathbb{U}_n(t_1), \ldots, \mathbb{U}_n(t_k)) \to_d (\mathbb{U}(t_1), \ldots, \mathbb{U}(t_k)) \qquad \text{as } n \to \infty.$$

(Essentially, all results in chapter 12 derive from this example—via a suggestion of Doob(1949).)

Exercise 1.6 (Partial sum process of iid rvs) Let $\mathbb{S}_n$ denote the partial sum process of iid $(0,1)$ rvs (see (11) below) and let $\mathbb{S}$ denote Brownian motion (as in (A.4.12)). Show that $\mathbb{S}_n \to_{fd} \mathbb{S}$ as $n \to \infty$. [Hint. Set things up cumulating from the left, and then transform. Or note that the random element you must consider can be written in a form equivalent to something simpler. Or use the Cramér–Wold device. One of these methods is much simpler then the others.]

Exercise 1.7 (Partial sum process) Suppose that $X_{n1}, \ldots, X_{nn}$ are independent $(0, \sigma_{nk}^2)$ and satisfy Lindeberg's condition (10.2.11) below. Define $\mathbb{S}_n$ on $[0,1]$ via

(11) $$\mathbb{S}_n(t) = \textstyle\sum_{i=1}^{k} X_{ni}/s_{nn} \qquad \text{for } \frac{s_{nk}^2}{s_{nn}^2} \leq t < \frac{s_{n,k+1}^2}{s_{nn}^2}, \quad 0 \leq k \leq n,$$

with $s_{nk}^2 \equiv \sum_{i=1}^{k} \sigma_{ni}^2$ and $s_{n0}^2 \equiv 0$. Show that $\mathbb{S}_n \to_{fd} \mathbb{S}$, where $\mathbb{S}$ denotes Brownian motion. (Only attempt this problem following theorem 10.2.2.)

Example 1.1 (Chisquare goodness of fit statistic) Suppose $\Omega = \sum_{i=1}^{k} A_i$. Now let $X_1, \ldots, X_n$ be iid on $(\Omega, \mathcal{A})$ with all $p_i \equiv P(X \in A_i) > 0$. Let

$$(12) \qquad N_{ni} \equiv \sum_{j=1}^{n} 1_{A_i}(X_j) \equiv (\text{ the number of } X_j\text{'s that fall in } A_i) \quad \text{for } 1 \le i \le k.$$

(a) Now, $(Z_{1j}, \ldots, Z_{kj})'$, with $Z_{ij} \equiv (1_{A_i}(X_j) - p_i)/\sqrt{p_i}$, has mean vector $\mathsf{0}$ and covariance matrix $\Sigma = |[\sigma_{ii'}]|$ with $\sigma_{ii} = 1 - p_i$ and $\sigma_{ii'} = -\sqrt{p_i\, p_{i'}}$ for $i \ne i'$.
(b) Thus $\mathsf{W}_n \equiv \sum_1^n Z_j/\sqrt{n} \to_d \mathsf{W} \cong N_k(\mathsf{0}, \Sigma)$ as $n \to \infty$, by theorem 1.3.
(c) The usual chisquare goodness of fit statistic is

$$(13) \qquad Q_n(\mathsf{p}\,) \equiv \sum_{i=1}^{k} \frac{(N_{ni} - np_i)^2}{np_i} = \sum_{i=1}^{k} \frac{(\text{Observed}_i - \text{ Expected}_i)^2}{\text{Expected}_i}$$

$$= \mathsf{W}_n'\mathsf{W}_n \to_d \mathsf{W}'\mathsf{W} \qquad \text{by the Mann–Wald theorem}$$

$$(14) \qquad = (G\mathsf{W})'(G\mathsf{W}) \cong \text{Chisquare}(k-1)\,;$$

here G is $k \times k$ and orthogonal with first row $\sqrt{\mathsf{p}\,'}$, so that $G\,\Sigma\,G' = G[I - \sqrt{\mathsf{p}}\,\sqrt{\mathsf{p}\,'}\,]G' = I - (1, 0, \ldots, 0)'(1, 0, \ldots, 0)$. This has diagonal elements $(0, 1, 1, \ldots, 1)$ with all off-diagonal elements 0, and then $G\mathsf{W} \cong N(\mathsf{0}, G\Sigma G')$ (by (7.3.5) and (A.3.6)). We also use (A.1.29) for (16). [If a value of Expected is unknown, it should be replaced by an appropriate estimator $\widehat{\text{Expected}}$.] (See exercise 10.3.26 below.) (This statistic is just a quadratic form.) □

Exercise 1.8* [Independence in an $I \times J$ table] Suppose both $\Omega = \sum_{i=1}^{I} A_i$ and $\Omega = \sum_{j=1}^{J} B_j$ represent partitions of Ω.
(a) Let $p_{ij} \equiv P(A_i B_j) = p_{i\cdot} p_{\cdot j}$, where $p_{i\cdot} \equiv P(A_i)$ and $p_{\cdot j} \equiv P(B_j)$. Let

$$N_{ij} \equiv (\text{the number of iid observations } X_1, \ldots, X_n \text{ that fall in } A_i B_j).$$

Let $\hat{p}_{i\cdot} \equiv \sum_{j=1}^{J} N_{ij}/n$ and $\hat{p}_{\cdot j} \equiv \sum_{i=1}^{I} N_{ij}/n$. Show that

$$(15) \qquad Q_n^a \equiv \textstyle\sum_{i=1}^{I} \sum_{j=1}^{J} (N_{ij} - n\hat{p}_{i\cdot}\hat{p}_{\cdot j})^2/(n\hat{p}_{i\cdot}\hat{p}_{\cdot j}) \to_d \text{Chisquare}((I-1)(J-1)).$$

(b) Let $p_{i|j} \equiv P(A_i|B_j)$. Let $n \equiv n_{\cdot 1} + \cdots + n_{\cdot J}$. For each $1 \le j \le J$, let $N_{ij} \equiv$ (the number of iid $P(\cdot|B_j)$ observations $X_1^{(j)}, \ldots, X_{n_j}^{(j)}$ that fall in $A_i B_j$). Let $\hat{p}_{i|j} \equiv \sum_{j=1}^{J} N_{ij}/n_{\cdot j}$. Show that when $\sigma[A_1, \ldots, A_I]$ and $\sigma[B_1, \ldots, B_J]$ are independent, the chisquare statistic satisfies

$$(16) \qquad Q_n^b \equiv \textstyle\sum_{i=1}^{I} \sum_{j=1}^{J} (N_{ij} - n_j\,\hat{p}_{i|j})^2/(n_j\,\hat{p}_{i|j}) \to_d \text{Chisquare}((I-1)(J-1))$$

as $n_1 \wedge \cdots \wedge n_J \to \infty$.
$(*)$ Suppose that both sets of marginal totals $n_{1\cdot}, \ldots, n_{I\cdot}$ and $n_{\cdot 1}, \ldots, n_{\cdot J}$ are fixed, and that both sum to n. Suppose that n balls are assigned to the IJ cells at random without replacement, subject to the side conditions on the marginal totals stated above. Let N_{ij} denote the number assigned to the (i, j)-th cell. It holds that

$$(17) \qquad Q_n^c \equiv \textstyle\sum_{i=1}^{I} \sum_{j=1}^{J} (N_{ij} - n_{i\cdot} n_{\cdot j}/n)^2/(n_{i\cdot} n_{j\cdot}/n) \to_d \text{Chisquare}((I-1)(J-1))$$

as $(n_{1\cdot} \wedge \cdots \wedge n_{I\cdot}) \wedge (n_{\cdot 1} \wedge \cdots \wedge n_{\cdot J}) \to \infty$. [Suppose $I = 5$ different social groups are at work in broadcasting, where the sum of the I group sizes $n_{i\cdot}$ of our data is $n = 250$. The number whose salaries fall in each decile (thus $J = 10$) of the observed salaries is necessarily $n_{\cdot j} = n/J = 25$. The statistic in (17) can be used to test for independence of group and salary level.] □

Limiting Distributions of Extremes

Exercise 1.9 (a) Let $\xi_{n1}, \ldots, \xi_{nn}$ be iid Uniform$(0,1)$ rvs. Then the sample minimum $\xi_{n:n}$ satisfies $n\xi_{n:n} \to$ Exp(1).
(b) Now, $\xi_{n:n}$ is the sample maximum. Determine the joint asymptotic distribution of $n\xi_{n:1}$ and $n(1 - \xi_{n:n})$.

Exercise 1.10 (Special cases of Gnedenko's theorem) Let $X_{n:n}$ be the maximum of an iid sample $X_1, \ldots, X_n$ from $F(\cdot)$. Then:

(a) $$P(X_{n:n} - \log n \le y) \to \exp(-e^{-y}) \qquad \text{for all } y \in R,$$ $$\text{when } 1 - F(x) = e^{-x} \text{ for } x \ge 0.$$

(b) $$P(n^{1/b}\, X_{n:n} \le y) \to \exp(-|y|^b) \qquad \text{for all } y < 0,$$ $$\text{when } 1 - F(x) = |x|^b \quad \text{for } -1 \le x \le 0, \quad \text{with } b > 0.$$

(c) $$P(X_{n:n}/n^{1/a} \le y) \to \exp(-y^{-a}) \qquad \text{for all } y > 0,$$ $$\text{when } 1 - F(x) = 1/x^a \text{ for } x \ge 1, \qquad \text{with } a > 0.$$

[Distributions that are "suitably similar" to these prototypes yield the same limiting results, and these limits are the only possible limits.]

2 Variations on the Classical CLT

Notation 2.1 Let $X_{nk}, 1 \le k \le n$ for each $n \ge 1$, be row-independent rvs having means μ_{nk} and variances σ_{nk}^2, and let $\gamma_{nk} \equiv \mathrm{E}|X_{nk} - \mu_{nk}|^3 < \infty$ denote the third absolute central moments . Let

(1) $$\mathrm{sd}_n \equiv \sqrt{\textstyle\sum_{k=1}^n \sigma_{nk}^2} \qquad \text{and} \qquad \gamma_n \equiv \textstyle\sum_{k=1}^n \gamma_{nk}\,, \qquad \text{and let}$$

(2) $$Z_n \equiv \frac{1}{\mathrm{sd}_n} \sum_{k=1}^n [X_{nk} - \mu_{nk}]\,.$$

Let $\phi_{nk}(\cdot)$ denote the chf of $(X_{nk} - \mu_{nk})/\mathrm{sd}_n$. □

Theorem 2.1 (Rate of convergence in the CLT) Consider the rvs above. The df F_{Z_n} of the standardized Z_n is uniformly close to the $N(0,1)$ df Φ, in that

(3) $$\|F_{Z_n} - \Phi\| \le 13\, \gamma_n/\mathrm{sd}_n^3\,.$$

Corollary 1 (Liapunov CLT)

(4) $$Z_n \to_d N(0,1) \qquad \text{whenever} \qquad \gamma_n/\mathrm{sd}_n^3 \to 0.$$

Corollary 2 (Berry–Esseen for iid rvs) Let $X_{n1}, \ldots, X_{nn}$ be iid rvs with df $F(\mu, \sigma^2)$ having $\gamma \equiv \mathrm{E}|X - \mu|^3 < \infty$. Then

(5) $$\|F_{Z_n} - \Phi\| \le \frac{8\,\gamma}{\sigma^3 \sqrt{n}}\,.$$

Proof. A much simpler proof of (4) is asked for in exercise 2.4 below. Here we give a delicate proof of the rate of convergence to normality in (3) based on Esseen's lemma, with (4) as a (too difficult) corollary. Without loss, we assume that all $\mu_{nk} = 0$. Now, let $a \equiv \mathrm{sd}_n^3/\gamma_n$; and assume throughout that $a \ge 9$ (note that (3) is meaningless unless $a > 13$). (Recall that $a = b \oplus c$ means $|a - b| \le c$.) Note that

$$\left|\phi_{Z_n}(t) - e^{-t^2/2}\right| = \left|\textstyle\prod_{k=1}^n \phi_{nk}(t) - e^{-t^2/2}\right|$$

$$\le e^{-t^2/2}\left|e^{\left\{\sum_{k=1}^n \mathrm{Log}\ \phi_{nk}(t)\right\}+t^2/2} - 1\right|$$

(6) $$\equiv e^{-t^2/2}\,|e^z - 1| \le e^{-t^2/2}\,|z|\,e^{|z|} \qquad \text{by (9.6.3)}$$

for all z, where

(a) $$|z| = \left|\sum_{k=1}^n \mathrm{Log}\ \phi_{nk}(t) + t^2/2\right| = \left|\sum_{k=1}^n \left\{\mathrm{Log}\ \left(1 + [\phi_{nk}(t) - 1]\right) - \frac{i^2\, t^2 \sigma_{nk}^2}{2\ \mathrm{sd}_n^2}\right\}\right|$$

(7) $$\le \left|\sum_{k=1}^n \left\{[\phi_{nk}(t) - \left(1 + \frac{i^2 t^2 \sigma_{nk}^2}{2\,\mathrm{sd}_n^2}\right)] \oplus |\phi_{nk}(t) - 1|^2\right\}\right|$$

provided that $|\phi_{nk}(t) - 1| \le \frac{1}{2}$, using (9.6.2)

(b) $$\le \frac{|t|^3}{6}\,\frac{\gamma_n}{\mathrm{sd}_n^3} + \sum_{k=1}^{n}\left(\frac{|t|^{3/2}\mathrm{E}|X_{nk}|^{3/2}}{\mathrm{sd}_n^{3/2}}\,K_{1,1/2}\right)^2 \qquad \text{where } K_{1,1/2}^2 = \tfrac{8}{9}$$

using (9.6.4) [with $m = 2$ and $\delta = 1$, then with $m = 1$ and $\delta = \frac{1}{2}$]

(c) $$\le \tfrac{1}{2a}\,|t|^3\,\left(\tfrac{19}{9}\right) \qquad \text{with} \quad a \equiv \mathrm{sd}_n^3/\gamma_n$$

using the Liapunov inequality for $(\mathrm{E}|X_{nk}|^{3/2})^2 \le \gamma_{nk}$. But validity of (7) required that all $|\phi_{nk}(t) - 1| \le \frac{1}{2}$. However, (9.6.4) with $m = 1$ and $\delta = 1$ gives

(d) $$|\phi_{nk}(t) - 1| \le \tfrac{1}{2}\,t^2\,\sigma_{nk}^2/\mathrm{sd}_n^2 \le \tfrac{1}{2}\,a^{2/3}[\gamma_n/\mathrm{sd}_n^3]^{2/3} \qquad \text{on } |t| \le a^{1/3}$$

(e) $$\le \tfrac{1}{2} \qquad \text{on } |t| \le a^{1/3}.$$

Consider for a moment the Liapunov CLT of Corollary 1. For any fixed t, the bound on $|z|$ in (c) goes to 0 whenever $1/a = \gamma_n/\mathrm{sd}_n^3 \to 0$. Moreover, (e) always holds when $\gamma_n/\mathrm{sd}_n^3 \to 0$, since $(\max \sigma_{nk}^2/\mathrm{sd}_n^2)^{3/2} \le (\max \gamma_{nk}/\mathrm{sd}_n^3) \le (\gamma_n/\mathrm{sd}_n^3) \to 0$. Thus $\phi_n(t) \to \exp(-t^2/2)$ and $Z_n \to_d N(0,1)$ by (6), whenever $\gamma_n/\mathrm{sd}_n^3 \to 0$. That is, the Liapunov CLT corollary 1 holds. [This is already a good CLT!]

We now turn to theorem 2.1 itself. Since the bound of (c) gives

(f) $$|z| \le \tfrac{1}{2a}\,|t|^3\,\tfrac{19}{9} \le \tfrac{1}{4}\,t^2 \quad \text{when } |t| \le \tfrac{9}{38}\,a \qquad (\text{and as } a^{1/3} \le \tfrac{9}{38}\,a, \text{ if } a \ge 9),$$

we can claim from(6), (e) and (c) that

(g) $$|\phi_{Z_n}(t) - e^{-t^2/2}| \le \tfrac{19}{18a}\,|t|^3\,e^{-t^2/4} \le \tfrac{2}{a}\,|t|^3\,e^{-t^2/4} \quad \text{for } |t| \le a^{1/3} \text{ (when } a \ge 9).$$

(Having the bound in (g) only over the range $|t| \le a^{1/3}$ is not sufficient for what is too come; we extend it in the next paragraph.)

Now, $|\phi_n(t)|^2$ is the chf of the symmetrized rv $Z_n^s \equiv Z_n - Z_n'$ (and this rv has mean 0, variance 2, and third absolute moment bounded above by $8\gamma_n/\mathrm{sd}_n^3$ (via the C_r-inequality)). Thus

$$|\phi_{Z_n}(t)| \le [\,|\phi_n(t)|^2\,]^{1/2} \le [1 + 0 - \tfrac{2t^2}{2!} + \tfrac{|t|^3}{3!}\,8\gamma_n/\,\mathrm{sd}_n^3\,]^{1/2}$$

(h) $$\le \exp(-t^2\,[\tfrac{1}{2} - \tfrac{2|t|}{3}\,\gamma_n/\,\mathrm{sd}_n^3\,]) \qquad \text{using } 1 - x \le e^{-x}$$

(i) $$\le \exp(-t^2/4) \qquad \text{for} \quad |t| \le (3/8)\,(\mathrm{sd}_n^3/\gamma_n),$$

as was desired. This leads to

(j) $$|\phi_{Z_n}(t) - e^{-t^2/2}| \le 2\,e^{-t^2/4} \le \tfrac{2}{a}\,|t|^3\,e^{-t^2/4} \qquad \text{for } a^{1/3} \le |t| \le \tfrac{3}{8}a.$$

Key chf inequality Combining (6), (e), (f) and (h) gives (provided $a \ge 9$)

(8) $$\left|\phi_{Z_n}(t) - e^{-t^2/2}\right| \le (2\,|t|^3\,\gamma_n/\mathrm{sd}_n^3\,)\,e^{-t^2/4} \qquad \text{for } 0 \le |t| \le \tfrac{3}{8}\,\mathrm{sd}_n^3/\gamma_n = \tfrac{3}{8}\,a.$$

We apply Esseen's lemma to (8) and get (since we know the variance of a normal distribution)

(k) $$\|F_{Z_n} - \Phi\| \le \int_{-(3/8)a}^{(3/8)a} \frac{1}{\pi|t|}\frac{2|t|^3}{a} e^{-t^2/4}\, dt + \frac{3.04769}{(3/8)\,a}$$

$$\le \frac{1}{a}\left[\frac{2}{\pi}\int_{-\infty}^{\infty} t^2 e^{-t^2/4}\, dt + \frac{3.04769}{3/8}\right]$$

(l) $$= [8/\sqrt{\pi} + (8/3)\ 3.04769]/a \doteq 12.641/a \le 13/a.$$

In the iid case use $K_{1,1} = \frac{1}{2}$ and $\beta \equiv \mathrm{E}|X|^3/\sigma^3 = \gamma/\sigma^3 \ge 1$ in (b), and obtain

$$|z| \le \frac{|t|^3\beta}{6\sqrt{n}} + n\left(\frac{t^2\sigma^2}{2n\sigma^2}\right)^2 \le \frac{|t|^3\beta}{6\sqrt{n}} + \frac{t^4\beta^2}{4n}$$

(m) $$\le \frac{5}{12}\frac{\beta}{\sqrt{n}}\,|t|^3 \le \frac{5}{12}\,|t|^3 \qquad \text{for all } |t| \le \sqrt{n}/\beta,$$

with (e) necessarily valid. Thus (8) can be replaced in the iid case by

(9) $$|\phi_{Z_n}(t) - e^{-t^2/2}| \le \frac{5}{12}\,\frac{\gamma}{\sigma^3\sqrt{n}}\,|t|^3\, e^{-t^2/12} \qquad \text{on } 0 \le |t| \le \sqrt{n}\sigma^3/\gamma;$$

this yields $8\gamma/\sqrt{n}\,\sigma^3$ when the steps leading to (l) are repeated. □

Theorem 2.2 (Lindeberg–Feller) Let $X_{n1}, \ldots, X_{nn}$ be row independent, with $X_{nk} \cong (\mu_{nk}, \sigma_{nk}^2)$. Let $\mathrm{sd}_n^2 \equiv \sum_1^n \sigma_{nk}^2$. The following statements are equivalent:

(10) $$Z_n \to_d N(0,1) \qquad \text{and} \qquad [\max_{1\le k\le n} P(\,|X_{nk} - \mu_{nk}|/\mathrm{sd}_n > \epsilon)] \to 0\,.$$

(11) $$LF_n^\epsilon \equiv \sum_{k=1}^n \int_{[\,|x-\mu_{nk}|\ge\epsilon\,\mathrm{sd}_n]} \left[\frac{x-\mu_{nk}}{\mathrm{sd}_n}\right]^2 dF_{nk}(x) \to 0 \qquad \text{for all } \epsilon > 0\,.$$

[Condition (11) implies that $M_n \equiv [\max_{1\le k\le n}\ |X_{nk}-\mu_{nk}|/\mathrm{sd}_n] \to_p 0$, via (10.1.9).]

Proof. (Lindeberg) We prove the sufficiency here, with the necessity considered in the following separate proof. We note that the moment expansion inequality (found in inequality 9.6.1) gives bounds on $\beta_{nk}(t)$, where

(12) $$\phi_{nk}(t) \equiv 1 + \theta_{nk}(t) \equiv 1 - \frac{\sigma_{nk}^2}{\mathrm{sd}_n^2}\frac{t^2}{2} + \beta_{nk}(t)$$

defines $\theta_{nk}(t)$ and $\beta_{nk}(t)$. Moreover (in preparation for the product lemma)

(13) $$\phi_{Z_n}(t) = \textstyle\prod_{k=1}^n \phi_{nk}(t) = \prod_{k=1}^n\,[1+\theta_{nk}(t)] = \prod_{k=1}^n \left[1 - \frac{\sigma_{nk}^2}{\mathrm{sd}_n^2}\frac{t^2}{2} + \beta_{nk}(t)\right]$$

where

(14) $$\theta_n(t) \equiv \sum_{k=1}^n \theta_{nk}(t) = -\frac{t^2}{2}\sum_{k=1}^n \frac{\sigma_{nk}^2}{\mathrm{sd}_n^2} + \sum_{k=1}^n \beta_{nk}(t) = -t^2/2 + \sum_{k=1}^n \beta_{nk}(t)\,.$$

The inequality of (9.6.4) (compare this with (9.6.10), and with (7)) gives

(a) $$\left|\sum_{k=1}^{n} \beta_{nk}(t)\right| = \left|\sum_{k=1}^{n}\left[\phi_{nk}(t) - 1 - 0 + \frac{\sigma_{nk}^2}{\mathrm{sd}_n^2}\frac{t^2}{2}\right]\right|$$

(15) $$\le \sum_{k=1}^{n}\left|\int\left\{e^{it\,(x-\mu_{nk})/\mathrm{sd}_n} - \left[1 + \frac{it(x-\mu_{nk})}{\mathrm{sd}_n} + \frac{[it(x-\mu_{nk})]^2}{2\,\mathrm{sd}_n^2}\right]\right\} dF_{nk}(x)\right|$$

(b) $$\le \sum_{k=1}^{n}\int_{[|(x-\mu_{nk})|<\epsilon\,\mathrm{sd}_n]} \frac{1}{6}\left|\frac{it(x-\mu_{nk})}{\mathrm{sd}_n}\right|^3 dF_{nk}(x)$$

(c) $$+ \sum_{k=1}^{n}\int_{[|(x-\mu_{nk})|\ge\epsilon\,\mathrm{sd}_n]} \tfrac{1}{2}\left|\frac{it(x-\mu_{nk})}{\mathrm{sd}_n}\right|^2 dF_{nk}(x)$$

(16) $$\le \epsilon\,\frac{|t|^3}{6}\left[\frac{\sum_{k=1}^n \sigma_{nk}^2}{\mathrm{sd}_n^2}\right] + \tfrac{t^2}{2}\sum_{k=1}^{n}\int_{[|(x-\mu_{nk})|\ge\epsilon\,\mathrm{sd}_n]}\left[\frac{x-\mu_{nk}}{\mathrm{sd}_n}\right]^2 dF_{nk}(x)\,.$$

Thus normality holds, since the integral in (16) goes to 0 for all $\epsilon > 0$ by (11).

Note that $[\max_{1\le k\le n} |\theta_{nk}(t)|\,] \to 0$ as required by the product lemma, since we can use inequality 9.6.1 on the $\theta_{nk}(t)$ in (12) to claim that $|\theta_{nk}(t)| \le \frac{t^2}{2}\sigma_{nk}^2/\mathrm{sd}_n^2$, and then use (11) on the second term below to claim that

(d) $$\sigma_{nk}^2/\mathrm{sd}_n^2 \le [\textstyle\int_{[|x-\mu_{nk}|\le\epsilon\,\mathrm{sd}_n]} (x-\mu_{nk})^2\, dF_{nk}(x)/\mathrm{sd}_n^2$$

$$+\textstyle\int_{[|x-\mu_{nk}|>\epsilon\,\mathrm{sd}_n]} (x-\mu_{nk})^2\, dF_{nk}(x)/\mathrm{sd}_n^2$$

(e) $$\le \epsilon^2 + o(1) \le \epsilon, \qquad \text{for } n \ge (\text{some } n_\epsilon)\,.$$ □

Proof. (Feller) We proved sufficiency in the previous proof; we now turn to necessity. Suppose that condition (10) holds. Applying (9.6.2) [since the easy exercise 10.2.9 below applied to our uan rvs shows that the terms $z_{nk} = \phi_{nk}(t) - 1$ converge uniformly to 0 on any finite interval] gives

(17) $$|\,\textstyle\sum_{k=1}^n \mathrm{Log}\,\phi_{nk}(t) - \sum_{k=1}^n [\phi_{nk}(t) - 1]\,| \le \sum_{k=1}^n |\phi_{nk}(t) - 1|^2$$

(18) $$\le [\max_{1\le k\le n} |\phi_{nk}(t) - 1|] \times (t^2/2) \times [\,\textstyle\sum_{k=1}^n \sigma_{nk}^2/\mathrm{sd}_n^2] \qquad \text{by } (9.6.5)$$

(a) $$\le o(1) \times (t^2/2) \times 1 \to 0, \qquad \text{using (10) via exercise 2.1.}$$

We thus have (for any finite M)

(b) $$\mathrm{Log}\,\textstyle\prod_1^n \phi_{nk}(t) = \sum_1^n [\phi_{nk}(t) - 1] + o(1), \qquad \text{uniformly on any } |t| \le M\,.$$

But we also know that

(c) $\quad \text{Log} \prod_1^n \phi_{nk}(t) \to -t^2/2,$

since we have assumed asymptotic normality. [Recall that $a = b \oplus c$ means that $|a-b| \le c$.] Combining (b) and (c) shows that for every tiny $\epsilon > 0$ and every huge $M > 0$ we have

(19) $$-t^2/2 = \text{Real}(-t^2/2) = \text{Real}\{\textstyle\sum_1^n [\phi_{nk}(t) - 1]\} \oplus \epsilon \qquad \text{for } |t| \le M$$

for all large n; that is, for $n \ge$ (some $n_{\epsilon M}$) we have

(20) $$t^2/2 = \textstyle\sum_1^n \int [1 - \cos(t\,(x - \mu_{nk})/\text{sd}_n)]\, dF_{nk}(x) \oplus \epsilon \qquad \text{on } |t| \le M.$$

Define $y_k \equiv (x - \mu_{nk})$. We further define $I_{nk} \equiv [|x - \mu_{nk}| < \epsilon\, \text{sd}_n]$. Note that

(d) $\quad 0 \le 1 - \cos(ty/\text{sd}_n) \le (t^2 y^2 / 2\, \text{sd}_n^2).$

Thus for all $|t| \le M$ we have for all $n \ge n_{\epsilon M}$ that

(e) $$(t^2/2) \textstyle\sum_{k=1}^n \int_{I_{nk}^c} (y_k^2/\text{sd}_n^2)\, dF_{nk}(x) = (t^2/2)\, [1 - \sum_{k=1}^n \int_{I_{nk}} (y_k^2/\text{sd}_n^2)\, dF_{nk}(x)]$$

$$= (t^2/2) - \textstyle\sum_{k=1}^n \int_{I_{nk}} (t^2 y_k^2 / 2\, \text{sd}_n^2)\, dF_{nk}(x)$$

(f) $$\le (t^2/2) - \textstyle\sum_{k=1}^n \int_{I_{nk}} [1 - \cos(t y_k/\text{sd}_n)]\, dF_{nk}(x) \qquad \text{by (d)}$$

(21) $$= \textstyle\sum_{k=1}^n \int_{I_{nk}^c} [1 - \cos(t y_k/\text{sd}_n)]\, dF_{nk}(x) \oplus \epsilon \qquad \text{by (20)} \qquad \text{[the key step]}$$

(g) $$\le 2 \textstyle\sum_{k=1}^n \int_{I_{nk}^c} dF_{nk}(x) + \epsilon \qquad (= 2 \sum_{k=1}^n P(|\dfrac{X_{nk} - \mu_{nk}}{\text{sd}_n}| \ge \epsilon) + \epsilon = \frac{2}{\epsilon^2} + \epsilon)$$

(h) $$\le (2/\epsilon^2) \textstyle\sum_{k=1}^n \int_{I_{nk}^c} (y_k^2/\text{sd}_n^2)\, dF_{nk}(x) + \epsilon$$

(i) $$\le 2/\epsilon^2 + \epsilon\,.$$

Specifying $t^2 = M^2 = 4/(\epsilon^2 \times \theta)$ in (g) (for any fixed $0 < \theta < 1$) shows that for all $n \ge n_{\epsilon\theta}$ we have

(j) $$\text{sd}_n^{-2} \textstyle\sum_1^n \int_{I_{nk}^c} y_k^2\, dF_{nk}(x) \le (\frac{2}{\epsilon^2} + \epsilon)\, \frac{1}{2}\, \epsilon^2\, \theta \le 2\, \theta,$$

where $\theta > 0$ is arbitrary. Thus, the Lindeberg condition (11) holds. $\square$

Exercise 2.1 (Characterizations of "uan") The following are equivalent:

(22) $|X_{nk}|$'s are uan; that is, $[\max_{1\le k\le n} P(|X_{nk}|\ge\epsilon)]\to 0$ for all $\epsilon>0$.

(23) $[\max_{1\le k\le n} |\phi_{nk}(t)-1|\,]\to 0$ uniformly on every finite interval of t's.

(24) $\max_{1\le k\le n} \mathrm{E}(X_{nk}^2\wedge 1)=\max_{1\le k\le n}\int (x^2\wedge 1)\,dF_{nk}(x)\to 0$.

Exercise 2.2 Phrase a simple consequence of the Liapunov CLT that applies to uniformly bounded row independent rvs X_{nk}.

Exercise 2.3 Provide the steps leading to (9) referred to in the Berry-Esseen proof.

Remark 2.1 (Lindeberg's condition) (i) If Lindeberg's condition fails, it may still be true that

(a) $S_n/\mathrm{sd}_n \to_d N(0,a^2)$ with $a^2<1$ and $[\max_{1\le k\le n}\sigma_k^2/\mathrm{sd}_n^2]\to 0$.

Let $Y_1, Y_2,\ldots$ be iid $(0,1)$ rvs, so that $\sqrt{n}\;\bar{Y}_n\to_d N(0,1)$ by the CLT. Now let the rv's U_k be independent $(0,c^2)$ with U_k equal to $-ck, 0, ck$ with probabilities $1/(2k^2), 1-1/k^2, 1/(2k^2)$. Since $\sum_1^\infty P(|U_k|\ge\epsilon)=\sum_1^\infty k^{-2}<\infty$, the Borel–Cantelli lemma shows that for a.e. ω the rv sequence U_k satisfies $U_k\ne 0$ only finitely often. Thus $\sqrt{k}\,\bar{U}_k\to_{a.s.} 0$ follows. For $n\ge 1$ set $X_n\equiv Y_n+U_n$, and let $S_n\equiv X_1+\cdots+X_n$. Note that $\mathrm{sd}_n^2\equiv \mathrm{Var}[S_n]=(1+c^2)n$. So, by Slutsky's theorem,

$$
\begin{aligned}
S_n/\mathrm{sd}_n &= (\sqrt{n}\,\bar{Y}_n)/\sqrt{1+c^2}+(\sqrt{n}\,\bar{U}_n)/\sqrt{1+c^2}\\
&\to_d N(0,1)/\sqrt{1+c^2}+0\cong N(0,1/(1+c^2))\\
\text{(b)}\qquad &= N(0,a^2)\qquad \text{with } a^2=1/(1+c^2)<1.
\end{aligned}
$$

(One could also let $c=\sqrt{n}$, and have $a=0$ in (b).) Note that $[\max\sigma_{nk}^2/\mathrm{sd}_n^2]\to 0$. But, even so, Lindeberg's condition fails, since

$$
\begin{aligned}
LF_n^\epsilon &= \frac{a^2}{n}\sum_{k=1}^n \int_{[|x|\ge\epsilon\sqrt{n}/a\,]} x^2\,dF_{X_k}(x)\\
\text{(c)}\qquad &\sim \frac{a^2}{n}\sum_{\{k:ck\ge\epsilon\sqrt{n}/a\}}\frac{(kc)^2}{k^2}+o(1)\sim\frac{c^2}{1+c^2}\,\frac{1}{n}\sum_{\{k:ck\ge\epsilon\sqrt{n}/a\}}1\to\frac{c^2}{(1+c^2)}>0;
\end{aligned}
$$

the nonzero contribution shown in the last step is due to the U_k's, whereas we do already know that the contribution due to the Y_k's is $o(1)$. This example shows that it is possible to have $X_n\to_d X$ without having $\mathrm{Var}[X_n]\to\mathrm{Var}[X]$. Note that $\mathrm{Var}[N(0,1/(1+c^2)]=1/(1+c^2)<1=\lim 1=\underline{\lim}\mathrm{Var}[S_n/\mathrm{sd}_n]$ (via the Fatou lemma and Skorokhod's theorem).
(ii) Note that if $X_{n1}\cong N(0,pn)$, for some $0<p<1$, $X_{nk}\equiv 0$ for $2\le k\le [pn]$, and $X_{nk}\cong N(0,1)$ for $pn<k\le n$ for independent rvs X_{nk}, then $S_n/\mathrm{sd}_n\to_d N(0,1)$, while Lindeberg's condition fails and $[\max_{1\le k\le n}\,\sigma_{nk}^2/\mathrm{sd}_n^2]\to p>0$. □

Remark 2.2 It is known that the constant 8 in (5) can be replaced by 0.7975. It is also known in the iid case with $\mathrm{E}|X|^3 < \infty$ that the "limiting distribution measure" $d(F, \Phi) \equiv \lim_{n\to\infty} \sqrt{n}\, \|F_{Z_n} - \Phi\|$ exists, and that this measure achieves the bound $\sup_F (\sigma^3/\gamma)\, d(F, \Phi) = (\sqrt{10} + 3)/(6\sqrt{2\pi}\,) = 0.409$. This sup is achieved by $c\,[\text{Bernoulli}(a) - a]$, where $c = (\sqrt{10} - 3)/2$ and $a = (4 - \sqrt{10}\,)/2$. Thus the constant 0.7975 cannot be greatly improved. Many other improvements and refinements of the Berry–Esseen theorem are possible. The books by Bhattacharya and Rao (1976, pp. 110, 240) and Petrov(1977) both give many. We list three as "exercises" in exercise 2.15 below. □

Exercise 2.4 (Liapunov's $(2+\delta)$-CLT) Define $\gamma_{nk}^{\delta} = \mathrm{E}|X_{nk} - \mu_{nk}|^{2+\delta}$ for every value $0 < \delta \le 1$. Suppose we have *Liapunov's* $(2+\delta)$*-condition* that

(25) $\quad \sum_{k=1}^{n} \mathrm{E}|X_{nk} - \mu_{nk}|^{2+\delta} / \mathrm{sd}_n^{2+\delta} \to 0 \qquad \text{for some } 0 < \delta \le 1.$

Show that $\frac{1}{\mathrm{sd}_n} \sum_1^n (X_{nk} - \mu_{nk}) \to_d N(0,1)$. (Appeal first to (9.6.5). Alternatively, verify that all $LF_n^{\epsilon} \to 0$.)

Exercise 2.5 Construct an example with iid $X_1, X_2, \ldots$ for which the Lindeberg condition holds, but for which Liapunov's $(2+\delta)$-condition fails for each $0 < \delta \le 1$.

Exercise 2.6 (Liapunov-type WLLN) Let $X_{n1}, \ldots, X_{nn}$, $n \ge 1$, be a triangular array of row-independent rvs with 0 means. Then

$$\sum_{k=1}^{n} \mathrm{E}|X_{nk}|^{1+\delta} / n^{1+\delta} \to 0 \qquad \text{for some } 0 < \delta \le 1$$

implies that $\bar{X}_n \to_p 0$ as $n \to \infty$. (Or, mimic the WLLN proof.)

Exercise 2.7 (Pitman) For iid rvs, $\bar{X}_n \to_p a$ holds if and only if $\phi'(0) = ai$.

Exercise 2.8 (i) Show that Lindeberg's condition that all $LF_n^{\epsilon} \to 0$ implies *Feller's condition* (which is not strong enough to guarantee asymptotic normality) that

(26) $\quad [\,\max_{1\le k\le n} \sigma_{nk}^2\,] \,/\, \mathrm{sd}_n^2 \to 0.$

(ii) Let $X_{n1}, \ldots, X_{nn}$ be row independent Poisson(λ/n) rvs, with $\lambda > 0$. Discuss which of Lindeberg–Feller, Liapunov, and Feller conditions holds in this context.
(iii) Repeat part (ii) when $X_{n1}, \ldots, X_{nn}$ are row independent and all have the probability density $c\,x^{-3}(\log x)^{-2}$ on $x \ge e$ (for some constant $c > 0$).
(iv) Repeat part (ii) when $P(X_{nk} = a_k) = P(X_{nk} = -a_k) = 1/2$ for the row independent rvs. Discuss this for general $a_k \ge 0$, and present interesting examples.

Exercise 2.9 Let $X_{n1}, \ldots, X_{nn}$ be row independent, with $X_{nk} \cong (\mu_{nk}, \sigma_{nk}^2)$. Let $T_n \equiv Z_{n1} + \cdots + Z_{nn}$, and set $\mu_n \equiv \sum_1^n \mu_{nk} = \mathrm{E}\,T_n$ and $\mathrm{sd}_n^2 \equiv \sum_1^n \sigma_{nk}^2 = \mathrm{Var}[T_n]$. The following statements are equivalent:

(27) $\quad Z_n \equiv (T_n - \mu_n)/\,\mathrm{sd}_n \to_d N(0,1) \qquad \text{and} \qquad [\,\max_{1\le k\le n} \sigma_{nk}^2\,]/\,\mathrm{sd}_n^2 \to 0\,.$

(28) $\quad LF_n^{\epsilon} \equiv \sum_{k=1}^{n} \int_{[\,|x-\mu_{nk}|\ge\epsilon\,\mathrm{sd}_n]} \left[\frac{x-\mu_{nk}}{\mathrm{sd}_n}\right]^2 dF_{nk}(x) \to 0 \qquad \text{for all } \epsilon > 0\,.$

(Example 11.1.2 treats the current exercise by a different method.)

Exercise 2.10 Complete the proof of theorem 8.8.1 regarding the equivalence of $\to_d$, $\to_p$, and $\to_{a.s.}$ for sums of independent rvs.

Exercise 2.11 Formulate a WLLN in the spirit of the Lindeberg–Feller theorem.

Exercise 2.12 Establish the $(2+\delta)$-analogue of theorem 2.1. [Hint. Use both $2+\delta$ and $1+\delta/2$ moments in line (b) of the theorem 2.1 proof, via lemma 9.6.2.]

Exercise 2.13 Construct a second example that satisfies the key property of remark 2.1(i), that the limiting variance is not the limit of the variances.

Exercise 2.14 (Large deviations; Cramér)) Let $X_{n1}, \ldots, X_{nn}$ be iid F. Suppose $X \cong F$ has a moment generation function $M(t) \equiv \mathrm{E}\, e^{tX}$ that is finite in some neighborhood of the origin. Let $Z_n \equiv \sqrt{n}\,(\bar{X}_n - \mu)/\sigma$, and let $Z \cong N(0,1)$, Then

$$(29) \qquad P(Z_n > x_n)/P(Z > x_n) \to 1 \qquad \text{provided} \qquad x_n/n^{1/6} \to 0.$$

(This exercise is repeated again later as exercise 11.6.6.)

Exercise 2.15* **(a)** (Petrov) Suppose $X_{n1}, \ldots X_{nk}$ are row independent rvs for which $X_{nk} \cong (0, \sigma^2_{nk})$, and set $\sigma^2_n \equiv \sum_{k=1}^n \sigma^2_{nk}$ and $F_n(x) \equiv P(S_n/\sigma_n \le x)$. Then for some universal constant C we have

$$(30) \qquad \|F_n - \Phi\| \le C\left[\sigma_n^{-3} \sum_{k=1}^n \mathrm{E} X^3_{nk}\, 1_{[|X_{nk}| < \epsilon\sigma_n]} + \sigma_n^{-2} \sum_{k=1}^n \mathrm{E} X^2_{nk}\, 1_{[\,|X_{nk}| \ge \epsilon\sigma_n]}\right].$$

(b) (Petrov) If $E[X^2_{nk}\, g(X_{nk})] < \infty$ for $1 \le k \le n$ where

$$(31) \qquad g \ge 0 \text{ is even and } \nearrow \text{ for } x > 0, \text{ and } x/g(x) \text{ is } \nearrow \text{ for } x > 0,$$

then for some absolute constant C we have the very nice result

$$(32) \qquad \|F_n - \Phi\| \le C \sum_1^n \mathrm{E}[X^2_{nk}\, g(X_{nk})]/\sigma^2_n\, g(\sigma_n).$$

(c) (Nagaev) Bounds on $|F_n(x) - \Phi(x)|$ that decrease as $|x| \to \infty$ are given (in the iid case) in the expression

$$(33) \qquad |F_n(x) - \Phi(x)| \le C(\mathrm{E}|X|^3/\sigma^3)/(\sqrt{n}\,(1 + |x|^3)) \qquad \text{for all real } x.$$

(d) (Bernstein) Let $r > 2$. Consider row-independent rvs $X_{n1}, \ldots, X_{nn}$ for which we have $X_{nk} \cong (0, \sigma^2_{nk})$. Let $Z_n \equiv \sum_{k=1}^n (X_{nk} - \mu_{nk})/\sigma_n$.

(α) Let $\sum_1^n \mathrm{E}\,|X_{nk}|^r/\sigma_n^{r/2} \to 0$. Then $Z_n \to_d N(0,1)$ and $\mathrm{E}|Z_n|^r \to \mathrm{E}|N(0,1)|^r$.

(β) The converse holds if $[\max_{1 \le k \le n} \sigma^2_{nk}/\,\mathrm{sd}^2_n] \to 0$ is also true.

($*$) See Petrov (1975, pp. 118, 113, 125, 103) for (a), (b), (c), and (d).

Exercise 2.16 Beginning with (15), try to obtain the Berry–Esseen bound (but with a different constant) by appeal to the second product lemma.

3 Examples of Limiting Distributions

Example 3.1 (Delta method) (a) Suppose $c_n\,[W_n - a] \to_d V$ where $c_n \to \infty$, and suppose $g(\cdot)$ is differentiable at a (recall (4.3.6) and (4.3.12)). Then (as in the chain rule proof of calculus) immediately

$$(1)\qquad c_n\,[\,g(W_n) - g(a)] =_a \{g'(a)\} \cdot c_n\,[\,W_n - a] \to_d \{g'(a)\} \cdot V\,.$$

[Recall that $U_n =_a V_n$ means that $U_n - V_n \to_p 0$.]
(b) The obvious vector version of this has the conclusion

$$(2)\qquad c_n\,[\,g(\mathsf{W}_n\,) - g(\mathsf{a}\,)] =_a \{\,\nabla g(a)\} \cdot c_n\,[\,\mathsf{W}_n - \mathsf{a}\,]\,.$$ □

Example 3.2 (Asymptotic normality of the sample variance) Suppose the rvs $X_1, \ldots, X_n$ are iid (μ, σ^2) with $\mu_4 \equiv \mathrm{E}X^4 < \infty$ and $\sigma^2 > 0$. Then

$$(3)\qquad S_n^2 \equiv \frac{1}{n-1} \sum_{k=1}^{n} (X_k - \bar{X})^2 = (\text{the } \textit{sample variance}).$$

For a useful phrasing of conclusions, define

$$(4)\qquad \begin{aligned} &Z_k \equiv (X_k - \mu)/\sigma \cong (0,1)\,,\\ &Y_k \equiv Z_k^2 = [(X_k - \mu)/\sigma]^2 \cong (1, \tfrac{\mu_4}{\sigma^4} - 1) = (1,\, 2\,(1 + \gamma_2/2))\,, \end{aligned}$$

where $\gamma_2 \equiv (\mu_4 - 3\sigma^4)/\sigma^4 \equiv$ (the *kurtosis*) measures the tail heaviness of the distribution of X. We will show that as $n \to \infty$ both

$$(5)\qquad \sqrt{n}\,[S_n^2 - \sigma^2\,]\,\tfrac{1}{\sqrt{2}\,\sigma^2} =_a \tfrac{1}{\sqrt{2}}\,\sqrt{n}\,[\,\bar{Y}_n - 1] \to_d N(0, 1 + \gamma_2/2) \qquad \text{and}$$

$$(6)\qquad \sqrt{n}\,[S_n - \sigma]\,\tfrac{2}{\sigma} =_a \sqrt{n}\,[\,\bar{Y}_n - 1] \to_d \sqrt{2}\;N(0, 1 + \gamma_2/2)\,.$$ □

Proof. Now,

$$\tfrac{S_n^2}{\sigma^2} = \tfrac{1}{n-1} \textstyle\sum_{k=1}^n \tfrac{(X_k - \bar{X}_n)^2}{\sigma^2} = \tfrac{n}{n-1}\,\tfrac{1}{n} \textstyle\sum_{k=1}^n [(\tfrac{X_k - \mu}{\sigma}) - (\tfrac{\bar{X}_n - \mu}{\sigma})]^2$$

$$(a)\qquad = \tfrac{n}{n-1}\,\tfrac{1}{n} \textstyle\sum_{k=1}^n [Z_k - \bar{Z}_n]^2\,.$$

Then note from (a) that

$$(b)\qquad \frac{\sqrt{n}\,(S_n^2 - \sigma^2)}{\sqrt{2}\,\sigma^2} = \frac{n}{n-1}\frac{1}{\sqrt{2}}\left\{\sqrt{n}\,(\,\bar{Y}_n - 1) - \sqrt{n}\,\bar{Z}_n^2\right\} - \frac{\sqrt{n}}{\sqrt{2}\,(n-1)}$$

$$(c)\qquad =_a \tfrac{1}{\sqrt{2}}\,\sqrt{n}\,(\,\bar{Y}_n - 1) \to_d N(0, \mathrm{Var}[\,Y\,]/2).$$ □

Exercise 3.1 (a) Determine the joint limiting distribution of $\sqrt{n}\,(\bar{X}_n - \mu)$ and $\sqrt{n}\,(S_n - \sigma)$ in the iid case, where $S_n^2 \equiv [\sum_1^n (X_k - \bar{X}_n)^2/(n-1)]$. (Consider the representation of S_n in (6) as a normed sum of the rvs Y_k.) What condition on the moments is required for the result?
(b) Find the asymptotic distribution of the (appropriately normalized) *coefficient of variation* $S_n/\bar{X}_n$ in this iid case; that is, consider $\sqrt{n}\,(S_n/\bar{X}_n - \sigma/\mu)$. Obtain a useful representation by appealing to part (a). (Suppose now that all $X_k \geq 0$.)
(c) Note that (6) provides a stronger conclusion than just asymptotic normality, in that it forms a superb starting point for the further asymptotic work in (a) and (b). Note also (13) below.

Exercise 3.2 (Moments of $\bar{X}_n$ and S_n^2) Let $X_1, \ldots, X_n$ be iid. Note/show that (provided that μ or σ^2 is well-defined) $\bar{X}_n \cong (\mu, \sigma^2)$ and $\mathrm{E}S_n^2 = \sigma^2$. Show that (provided that μ_3 or μ_4 is well-defined):

(7) $\qquad \mathrm{E}(\bar{X}_n - \mu)^3 = \frac{\mu_3}{n^2}\,.$

(8) $\qquad \mathrm{E}(\bar{X}_n - \mu)^4 = \frac{3\,\sigma^4}{n^2} + \frac{\mu_4 - 3\,\sigma^4}{n^3}\,.$

(9) $\qquad \mathrm{Var}[S_n^2] = \frac{1}{n}\,\{\,\mu_4 - \frac{n-3}{n-1}\,\sigma^4\,\} \qquad$ and $\qquad \mathrm{Cov}[\bar{X}_n, S_n^2] = \frac{1}{n}\,\mu_3\,.$

Exercise 3.3 If $X_1, \ldots, X_n$ are iid (μ, σ^2), then $\sqrt{n}\,[\bar{X}_n^2 - \mu^2\,] \to_d 2\,\mu \times N(0, \sigma^2)$ (by the delta method). What is the asymptotic distribution of $n\bar{X}_n^2$ when $\mu = 0$?

Exercise 3.4 (Two sample problems) If $\sqrt{m}\,(S_m - \theta) \to_d N(0,1)$ as $m \to \infty$ and $\sqrt{n}\,(T_n - \theta) \to_d N(0,1)$ as $n \to \infty$ for independent rvs S_m and T_n, then $\sqrt{\frac{mn}{m+n}}\,(S_m - T_n) \to_d N(0,1)$ as $m \wedge n \to \infty$. [Hint: Suppose initially that $\lambda_{mn} \equiv m/(m+n) \to \lambda \in [0,1]$. Use Skorohod (or, use convolution or chfs) to extend it.] This is useful for the two-sample t-test and F-test.

Exercise 3.5 (Simple linear rank statistics) Let $T_N \equiv \frac{1}{\sqrt{N}} \sum_1^N c_i\, a_{\pi(i)}$, where $(\pi(1), \ldots, \pi(N))$ achieves each of the $N!$ permutations of $(1, \ldots, N)$ with probability $1/N!$. Here, the c_i and a_i are constants. Show that:

(10)
$$\mathrm{E}a_{\pi(i)} = \bar{a}_N\,, \quad \mathrm{Var}[a_{\pi(i)}] = \sigma_a^2 \equiv \frac{1}{N} \sum_1^N (a_i - \bar{a}_N)^2\,,$$
$$\mathrm{Cov}[\,a_{\pi(i)}\,, a_{\pi(j)}\,] = -\frac{1}{N-1}\,\sigma_a^2 \qquad \text{for all } i \neq j\,.$$

(11) $\qquad \frac{1}{\sqrt{N}}\,\mathrm{E}\,T_N = \bar{c}_N \cdot \bar{a}_N\,, \qquad \mathrm{Var}[T_N] = \frac{N}{N-1}\,\sigma_c^2\,\sigma_a^2\,.$

[Hint. $\mathrm{Var}[\sum_1^N a_{\pi(i)}\,] = 0$, as in (A.1.8).]

Example 3.3 (The median $\ddot{X}_n$) The population *median* of the distribution of a rv X's is any value θ satisfying $P(X \leq \theta) \geq \frac{1}{2}$ and $P(X \geq \theta) \geq \frac{1}{2}$. Let $X_1, \ldots, X_n$ be iid with df $F(\cdot - \theta)$, for some $\theta \in R$, where $F(0) = \frac{1}{2}$ and $F'(0) > 0$ exists and exceeds zero. (Thus $X_i \cong \theta + \epsilon_i$, for ϵ_i's that are iid $F(\cdot)$ with a unique median at 0.) The ordered values of the X_k's are denoted by $X_{n:1} \le \cdots \le X_{n:n}$, and are called the *order statistics*. The *sample median* $\ddot{X}_n$ is defined to be $X_{n:m}$ or *any* point in the interval $[X_{n:m}, X_{n:m+1}]$ according as n equals $2m+1$ or $2m$ is odd or even. Let X_n^l and X_n^r denote the left and right endpoints of the interval of possible sample medians (of course, $X_n^l = X_n^r = X_{n:m+1}$ if $n = 2m+1$ is odd). Let $\ddot{X}_n$ denote any sample median. (a) Then

$$(12) \qquad Z_{1n} \equiv \sqrt{n}\,[\ddot{X}_n - \theta] \to_d Z_1 \cong N(0, \tfrac{1}{4[F'(0)]^2}\,).$$

(b) If $F(\cdot) \cong (\mu, \sigma^2)$ also, then $Z_{2n} \equiv \sqrt{n}\,[\bar{X}_n - (\mu+\theta)] \to_d Z_2 \cong N(0, \sigma^2)$.
(c) In fact, the limiting normal distribution is given by $(Z_{1n}, Z_{2n}) \to_d (Z_1, Z_2)$, where the covariance of the limiting normal distribution is given by

$$(13) \qquad \mathrm{E}\{[\epsilon - \mu] \times [1_{[\epsilon>0]} - 1/2]/F'(0)\}\,. \qquad \square$$

Proof. By the event equality $[X_n^r - \theta \le y/\sqrt{n}\,] = [\,\sum_1^n 1_{[X_i - \theta \le y/\sqrt{n}\,]} > n/2]$, we have

$$P(\sqrt{n}\,[X_n^r - \theta] \le y) = P(X_n^r - \theta \le y/\sqrt{n}\,) = P(\textstyle\sum_1^n 1_{[X_i-\theta \le y/\sqrt{n}\,]} > n/2)$$

$$(14) \qquad = P(\tfrac{1}{n}\textstyle\sum_1^n 1_{[\epsilon_i \le y/\sqrt{n}\,]} > 1/2) \qquad \text{since } \epsilon_i \equiv X_i - \theta \cong F(\cdot)$$

$$= P(\tfrac{1}{\sqrt{n}}\textstyle\sum_1^n \{1_{[\epsilon_i > y/\sqrt{n}\,]} - \tfrac{1}{2}\} < 0)$$

$$(a) \qquad = P(W_n + A_n < 0)\,.$$

Here

$$(15) \qquad W_n \equiv \tfrac{1}{\sqrt{n}}\textstyle\sum_1^n \{1_{[\epsilon_i>0]} - P(\epsilon_i > 0)\} \to_d Z_1 \cong N(0, p(1-p))$$

with $p \equiv P(\epsilon_i > 0) = \frac{1}{2}$, and (as we will now show)

$$(16) \qquad A_n \equiv \tfrac{1}{\sqrt{n}}\textstyle\sum_1^n \{1_{[\epsilon_i > y/\sqrt{n}\,]} - 1_{[\epsilon_i>0]}\} \to_p -y\,F'(0)\,.$$

Note that all terms in the summation in A_n are of the same sign. Then

$$(b) \qquad A_n \to_p -y\,F'(0\pm) \qquad \text{according as } y > 0 \text{ or } y < 0,$$

since

$$\mathrm{E}A_n = \sqrt{n}\,[\,P(\epsilon > y/\sqrt{n}\,) - P(\epsilon > 0)] = -y\,[F(y/\sqrt{n}\,) - F(0)]/\,[y/\sqrt{n}\,]$$

$$(c) \qquad \to -y\,F'(0\pm) \qquad [\text{provided only that both } F'(0\pm) \text{ exist}]$$

and

$$(d) \qquad \mathrm{Var}[A_n] = [F(y/\sqrt{n}\,) - F(0)]\{1 - [F(y/\sqrt{n}\,) - F(0)]\} \to 0\,.$$

Thus $W_n + A_n \to_d Z_1 - yF'(0)$ via (15), (16), and Slutsky. By $\to_d$ we then have

$$P(\sqrt{n}\,[X_n^r - \theta] \le y) = P(W_n + A_n < 0)$$

$$(e) \qquad \to P(Z_1 - yF'(0) \le 0) = P(Z_1/F'(0) \le y) \text{ for each } y.$$

That is,

(f) $\sqrt{n}\,[X_n^r - \theta] \to_d Z_1/F'(0) \cong N(0, p(1-p)/\,[F'(0)]^2)\,.$

In like fashion, $[\sqrt{n}\,[X_n^l - \theta] \le y] = [\,\sum_1^n 1_{[\epsilon_i \le y/\sqrt{n}\,]} \ge n/2]$, so that

(g) $P(\sqrt{n}\,[X_n^l - \theta] \le y) = P(\,\sum_1^n 1_{[\epsilon_i \le y/\sqrt{n}\,]} \ge n/2) = P(W_n + A_n \le 0)\,.$

Thus the same argument as before gives

(h) $\sqrt{n}\,[X_n^l - \theta] \to_d Z_1/F'(0)\,.$

Now we squeeze the general $\ddot{X}_n$ in between, via

(i) $P(\sqrt{n}\,[X_n^r - \theta] \le y) \le P(\sqrt{n}\,[\ddot{X}_n - \theta] \le y) \le P(\sqrt{n}\,[X_n^l - \theta] \le y)\,,$

where both ends converge to $P(Z_1/F'(0) \le y)$. This completes the proof.

Summary It has been demonstrated that the events (note (15))

(17) $[\,\omega : \sqrt{n}\,(\ddot{X}_n(\omega) - \theta) \le y\,]$ and $[\,\omega : W_n(\omega) \le y\,F'(0)\,]$
differ by a probabilistically negligible amount.

For the joint result, apply (17) and the multivariate CLT to (W_n, Z_{2n}). □

Exercise 3.6 (Joint asymptotic normality of quantiles) For $0 < p < 1$, the pth *quantile* x_p of F is now defined as $x_p \equiv F^{-1}(p)$. (a) Show that if F has a derivative $F'(x_p) > 0$ at x_p, then

(18) $\sqrt{n}\,[\,X_{n:[np]} - x_p\,] \to_d N(0, p(1-p)/\,[F'(x_p)]^2\,)$ as $n \to \infty$.

(b) Establish joint normality for p_i and p_j quantiles, where the covariance matrix of the asymptotic distribution has (i,j)th entry

$$\sigma_{ij} \equiv [(p_i \wedge p_j) - p_i\,p_j]/\,[F'(x_{p_i})F'(x_{p_j})]\,.$$

Write out the analogue of (17), and use it.

Exercise 3.7 What happens when you try to apply (12) to:

$$\text{(a) } F'(x) = \exp(-|x|)/2\,? \quad \text{or} \quad \text{(b) } F'(x) = \tfrac{1}{2}\,1_{[-1,0)}(x) + \tfrac{1}{4}\,1_{[0,2]}(x)\,?$$

Show that $\sqrt{n}\,[\ddot{X}_n - \theta] \to_d$ (a rv) in both cases. (In case (b) it is not normal.)

Exercise 3.8 Verify (12) both for $n = 2m+1$ odd and for $n = 2m$ even.
Hint. Since $X_{n:m} \le \ddot{X}_n \le X_{n:m+1}$,

$$P(\,\sum_1^n 1_{[\epsilon_i \le y/\sqrt{n}\,]} > \tfrac{n}{2}\,]) \le P(\sqrt{n}\,(\ddot{X}_n - \theta) \le z) \le P(\,\sum_1^n 1_{[\epsilon_i \le y/\sqrt{n}\,]} \ge \tfrac{n}{2}\,])\,.$$

(The right side is an equality when n is odd.)

Exercise 3.9 Consider (with hypothesis as weak as possible) the asymptotic distribution of (appropriately normalized forms of) both

(19) $\quad \frac{1}{n}\sum_1^n |X_k - \bar{X}_n| \qquad \text{and} \qquad \frac{1}{n}\sum_1^n |X_k - \ddot{X}_n|$

for iid samples $X_1, \ldots, X_n$ from a df $F(\mu, \sigma^2)$ having median ν.

Exercise 3.10 Let $X_1, X_2, \ldots$ be independent with $X_k \cong \text{Uniform}(-k, k)$. Then establish that $S_n/\sigma_n \to_d N(0,1)$.

Exercise 3.11 Determine the limiting distribution of

$$\textstyle\sum_{k=1}^n (X_k - X_{2n+1-k}) / \left\{\sum_{k=1}^n (X_k - X_{2n+1-k})^2\right\}^{1/2},$$

where $X_1, X_2, \ldots$ are iid (μ, σ^2) rvs. Hint. Think "Slutsky."

Exercise 3.12 Determine the .95-quantile of the limiting distribution of

$$\textstyle\prod_{k=1}^n U_k^{-X_k/\sqrt{n}},$$

for independent rvs with $X_k \cong \text{Double Exponential}(0,1)$ and $U_k \cong \text{Uniform}(0,1)$.

Example 3.4 (Weighted sums of iid rvs) Suppose that rvs $X_{n1}, \ldots, X_{nn}$ are row independent and iid (μ, σ^2). Let $\mathsf{c}_n \equiv (c_{n1}, \ldots, c_{nn})'$ for $n \geq 1$, and set

$$\bar{c}_n \equiv \textstyle\sum_{k=1}^n c_{nk}/n \qquad \text{and} \qquad \sigma_{cn}^2 \equiv \sum_{k=1}^n (c_{nk} - \bar{c}_n)^2/n \equiv \text{SS}_{cc}/n\,.$$

Suppose we have the uan condition

$$(20) \qquad \mathcal{D}_c \equiv \mathcal{D}(\mathsf{c}_n) \equiv \frac{\max_{1\leq k\leq n} (c_{nk} - \bar{c}_n)^2}{\sum_{k=1}^n (c_{nk} - \bar{c}_n)^2} = \frac{\left[\max_{1\leq k\leq n} (c_{nk} - \bar{c}_n)^2/n\right]}{\sigma_{cn}^2} \to 0$$

as $n \to \infty$. Then

$$(21) \qquad \sum_{k=1}^n \frac{(c_{nk} - \bar{c}_n)}{\sqrt{n}\,\sigma_{cn}}\, \frac{X_{nk} - \mu}{\sigma} \to_d N(0,1).$$

[We need not center the c_{nk}'s if the X_{nk}'s have mean 0.]

Proof. Without loss of generality, set $\mu = 0$. Now, Lindeberg's condition holds, as we demonstrate via

$$\left|\sum_{k=1}^n \int_{[|c_{nk} - \bar{c}_n|\,|x| \geq \epsilon\,\sigma\sqrt{n}\,\sigma_{cn}\,]} \left[\frac{(c_{nk} - \bar{c}_n)}{\sigma\sqrt{n}\,\sigma_{cn}}\right]^2 x^2\, dF(x)\right|$$

$$\text{(a)} \qquad \leq \sigma^{-2} \cdot 1 \cdot \textstyle\int_{[|x| \geq \epsilon\sigma/\sqrt{\mathcal{D}_c}\,]} x^2\, dF(x) \to 0,$$

since $\mathcal{D}_c \to 0$ and $\int x^2\, dF(x) < \infty$. □

The preceding example is useful in regression situations, and in connection with the projection techniques of section A.3. See section 10.2.

Exercise 3.13 (Monte Carlo estimation) Let $h : [0,1] \to [0,1]$ be a measurable function, and let $\theta = \int_0^1 h(t)\,dt$. Let $X_1, Y_1, X_2, Y_2, \ldots$ be iid Uniform$(0,1)$ rvs. Define two different estimators of θ by

$$T_{1n} \equiv \sum_{k=1}^n h(X_k)/n \qquad \text{and} \qquad T_{2n} \equiv \sum_{k=1}^n 1_{[X_k \le h(Y_k)]}/n.$$

(a) Show that both T_{1n} and T_{2n} are unbiased estimators of θ, and determine which estimator has the smaller variance. Indicate how the variance of each estimator could be estimated.
(b) Determine the joint asymptotic distribution of appropriately normalized forms of T_{1n} and T_{2n}.

Exercise 3.14 (An analogue of the student-t statistic based on quartiles)
Let $X_1, \ldots, X_n$ be iid with df $F(\cdot)$. Let $m \equiv [n/4]$, for the greatest integer $[\cdot]$. Let

$$U_n \equiv X_{n:m}, \quad V_n \equiv \ddot{X}_n \equiv \text{(the median)}, \quad W_n \equiv X_{n:n+1-m}$$

denote the quartiles and median of the sample. Make appropriate assumptions regarding $F(\cdot)$.
(a) Determine the joint asymptotic distribution of

$$(\sqrt{n}\,[V_n - \nu],\ \sqrt{n}\,[W_n - U_n - \mu]\,)$$

for appropriately defined μ and ν.
(b) Simplify this if the X_i are symmetrically distributed about 0.
(c) Determine the asymptotic distribution under symmetry of the (student-t like) statistic (formed from three sample quantiles)

$$T_n \equiv \sqrt{n}\,[V_n - \nu]/\,[W_n - U_n].$$

Exercise 3.15 Let the X_k's be iid Cauchy$(0,1)$ in the previous exercise.
(d) Evaluate $F(x) = P(X \le x)$ for $x \in R$.
(e) Solve $p = F(x)$ for $x_p \equiv F^{-1}(p)$, when $0 < p < 1$.
(f) Express your answers to (b) and (c) of the previous exercise in the present context.

Exercise 3.16 (Poisson estimation) Let $X_1, \ldots, X_n$ be iid Poisson(θ).
(a) Reasonable estimators of θ include the sample mean $T_{1n} \equiv \bar{X}_n$, the sample variance $T_{2n} \equiv S_n^2$, and $T_{3n} \equiv \sum_1^n kX_k/\sum_1^n k$ (which puts more emphasis on the more recent observations). Evaluate lim Var$[T_{in}]$ for $i = 1, 2, 3$.
(b) Verify that $T_{4n} \equiv \bar{X}_n^2 - \bar{X}_n/n$ and $T_{5n} \equiv \overline{X_n^2} - \bar{X}_n$ are both unbiased estimators of θ^2. Evaluate lim Var$[T_{in}]$ for $i = 4, 5$.
(c) Determine the asymptotic distribution of $D_n \equiv \sqrt{n}\,[\,\bar{X}_n - S_n^2\,]/\,\bar{X}_n$ when the observations really do follow a Poisson distribution.
(d) What is the asymptotic distribution of D_n when the observations X_k actually follow a NegBiT(r, p) distribution?

Theorem 3.1 (Doeblin's CLT for a random number of rvs) Consider iid $(0, \sigma^2)$ rvs $X_1, X_2, \ldots$. Let $\{\nu_n\}_{n=1}^{\infty}$ be integer-valued rvs such that the proportion $\nu_n/n \to_p c \in (0, \infty)$ as $n \to \infty$. Let $T_n \equiv X_1 + \cdots + X_n$ denote the total. Then

$$(22) \qquad T_{\nu_n}/\sqrt{\nu_n} \to_d N(0, \sigma^2).$$

[Note that ν_n and $X_1, X_2, \ldots$ need *not* be independent.]

Proof. Now,

$$(23) \qquad \frac{T_{\nu_n}}{\sqrt{\nu_n}} = \sqrt{\frac{[cn]}{\nu_n}} \left\{ \frac{T_{[cn]}}{\sqrt{[cn]}} + \frac{T_{\nu_n} - T_{[cn]}}{\sqrt{[cn]}} \right\}.$$

Note that $T_{[cn]}/\sqrt{[cn]} \to_d N(0,1)$ and $[cn]/\nu_n = \frac{[cn]}{cn} \frac{c}{\nu_n/n} \to_p 1$. In the next paragraph we will show that

$$\text{(a)} \qquad (T_{\nu_n} - T_{[cn]})/\sqrt{[cn]} \to_p 0.$$

The theorem then follows from Slutsky's theorem.

We now let $A_n \equiv [|T_{\nu_n} - T_{[cn]}|/\sqrt{[cn]} > \epsilon]$, and note that

$$\text{(b)} \qquad P(A_n) = \textstyle\sum_{k=1}^{\infty} P(A_n \cap [\nu_n = k])$$

$$= \sum_{\{k: |k - [cn]| \le \epsilon^3 cn\}} P(A_n \cap [\nu_n = k]) + \sum_{\{k: |k - [cn]| > \epsilon^3 cn\}} P(A_n \cap [\nu_n = k])$$

$$\text{(c)} \qquad \equiv \textstyle\sum_1 + \sum_2.$$

Since $\nu_n/[cn] \to_p 1$, for n sufficiently large we have

$$\textstyle\sum_2 \le P(|\nu_n - [cn]| > \epsilon^3 cn) \le P\left(|\frac{\nu_n}{[cn]} - 1| > \epsilon^3 \frac{cn}{[cn]} \right)$$

$$\text{(d)} \qquad \le P\left(|\frac{\nu_n}{[cn]} - 1| > \epsilon^3 \right) < \epsilon.$$

Also, applying Kolmogorov's inequality twice,

$$\textstyle\sum_1 \le P\left(\max_{|k - [cn]| \le \epsilon^3 cn} |T_k - T_{[cn]}| > \epsilon\sqrt{[cn]} \right)$$

$$\le P\left(\max_{[cn] \le k \le \epsilon^3 cn} |T_k - T_{[cn]}| > \epsilon\sqrt{[cn]} \right)$$

$$+ P\left(\max_{[cn] - \epsilon^3 cn \le k \le [cn]} |T_k - T_{[cn]}| > \epsilon\sqrt{[cn]} \right)$$

$$\le 2 \sum_{k=[cn]+1}^{[cn]+\epsilon^3 cn} \mathrm{Var}[X_k]/\epsilon^2 [cn] \le 2\epsilon^3 cn\sigma^2/\epsilon^2 [cn]$$

$$\text{(e)} \qquad \le 4\,\sigma^2 \epsilon \qquad \text{for all } n \text{ sufficiently large.}$$

Combining (d) and (e) into (c) shows $P(A_n) \to 0$, as required. □

Exercise 3.17 Let V_n^2 now denote the sample variance. Show, in the context of Doeblin's CLT, that $T_{\nu_n}/V_{\nu_n} \to_d N(0,1)$ as $n \to \infty$.

Exercise 3.18 Prove a version of Doeblin's theorem for X_{nk}'s independent but not iid; assume the Lindeberg condition and $\nu_n/n \to_p c \in (0,\infty)$. [Revert to the Liapunov condition, if necessary.]

Exercise 3.19 (Sample correlation coefficient R_n; Cramér and Anderson) Let us suppose that

$$\begin{bmatrix} X_i \\ Y_i \end{bmatrix} \text{ are iid } \left[\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right] \qquad \text{for } 1 \le i \le n,$$

and that the Σ below has finite entries. Consider $\sqrt{n}\,[R_n - \rho]$, where R_n is the *sample correlation coefficient*. Thus $R_n \equiv \mathrm{SS}_{XY}/\{\,\mathrm{SS}_{XX}\,\mathrm{SS}_{YY}\,\}^{1/2}$ for the sums of squares $\mathrm{SS}_{XY} \equiv \sum_1^n (X_i - \bar{X}_n)(Y_i - \bar{Y}_n)$, etc.
(a) Reduce the case of general means, variances and covariances to this case.
(b) Note that

$$(24) \qquad \begin{bmatrix} \frac{1}{\sqrt{n}} \sum_1^n (X_i Y_i - \rho) \\ \frac{1}{\sqrt{n}} \sum_1^n (X_i^2 - 1) \\ \frac{1}{\sqrt{n}} \sum_1^n (Y_i^2 - 1) \end{bmatrix} \to_d \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \end{bmatrix} \cong N(0, \Sigma)$$

with

$$(25) \qquad \Sigma \equiv \begin{bmatrix} \mathrm{E}(X^2Y^2) - \rho^2 & \mathrm{E}(X^3Y) - \rho & \mathrm{E}(XY^3) - \rho \\ \mathrm{E}(X^3Y) - \rho & \mathrm{E}X^4 - 1 & \mathrm{E}(X^2Y^2) - 1 \\ \mathrm{E}(XY^3) - \rho & \mathrm{E}(X^2Y^2) - 1 & \mathrm{E}Y^4 - 1 \end{bmatrix}.$$

(c) Then show that $\sqrt{n}\,[R_n - \rho] \to_d Z_1 - \frac{\rho}{2} Z_2 - \frac{\rho}{2} Z_3 \cong N(0, \tau^2)$, and evaluate τ^2.
(d) Show that when X and Y are independent, then $\sqrt{n}\,[R_n - \rho] \to_d N(0,1)$.
(e) If the $(X_i, Y_i)'$ are jointly normal, show that

$$(26) \qquad \Sigma = \begin{bmatrix} 1 + \rho^2 & 2\rho & 2\rho \\ 2\rho & 2 & 2\rho^2 \\ 2\rho & 2\rho^2 & 2 \end{bmatrix}.$$

Then simplify the expression for τ^2 and obtain

$$(27) \qquad \sqrt{n}\,[R_n - \rho] \to_d N(0, (1 - \rho^2)^2).$$

(f) Show that $\sqrt{n}\,[g(R_n) - g(\rho)] \to_d N(0,1)$ for $g(t) \equiv \frac{1}{2} \log(\frac{1+t}{1-t})$.
(g) Approximating the distribution of $\sqrt{n-3}\,[g(R_n) - g(\rho) - \frac{\rho}{2(n-1)}]$ by $N(0,1)$ yields an excellent result.
(h) Show that $\mathrm{Cov}[X_i - \bar{X}_n, Y_i - \bar{Y}_n] = (1 - \frac{1}{n})\,\mathrm{Cov}[X_i, Y_i]$.

Exercise 3.20 (Extreme Value quantiles) Let X and $X_1, \ldots, X_n$ be iid with the *Weibull*(α, β) density $f(x) = (\beta x^{\beta-1}/\,\alpha^\beta)\exp(-(x/\alpha)^\beta)$ on $x \ge 0$. Now, $(X/\alpha)^\beta \cong$ Exponential(1), and thus $Y \equiv \log X$ satisfies

$$Y \cong \nu + \tau\, W \qquad \text{where } \nu \equiv \log\alpha \text{ and } \tau \equiv 1/\beta$$

and W has the *Extreme Value* density for minima given by $\exp(w-e^w)$ on $(-\infty,\infty)$. Let $Y_{n:1} \le \cdots \le Y_{n:n}$ denote the order statistics of the rvs $Y_k \equiv \log X_k$. First, let $0 < p_1 < p_2 < 1$, and then define $U_n \equiv Y_{n:[np_1]}$ and $V_n \equiv Y_{n:[np_2]}$. We seek values p_1 and p_2 such that

$$\begin{bmatrix} \sqrt{n}\,(V_n - \nu) \\ \sqrt{n}\,(V_n - U_n - \tau) \end{bmatrix} \to_d N(0, \Sigma).$$

(a) Let $0 < p < 1$. Evaluate $y_p \equiv F_Y^{-1}(p)$, $f_Y(y_p)$ and $p(1-p)/f_Y^2(y_p)$.
(b) Determine values of p_1 and p_2 that achieve the objective.
(c) Establish the claimed asymptotic normality, and evaluate Σ both symbolically and numerically.

Exercise 3.21 (Estimating a common normal mean) Consider independent rvs $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ from $N(\theta, \sigma^2)$ and $N(\theta, \tau^2)$. When $\gamma \equiv \sigma^2/\tau^2$ is known, the unbiased estimator of θ that has minimum variance (for all possible values of the parameters within this model) is known to be

$$\hat{\theta}_o \equiv \hat{\theta}_{o,mn} \equiv \tfrac{m\bar{X}_m + \gamma\cdot n\bar{Y}_n}{m+\gamma\cdot n}\,.$$

Define $\alpha \equiv \alpha_{mn} \equiv m/(m+\gamma\cdot n)$. Let $\hat{\alpha} \equiv \hat{\alpha}_{mn}(S_X^2, S_Y^2)$ depend only on the two sample variances $S_X^2 \equiv \sum_1^m (X_i - \bar{X}_m)^2/(m-1) = \mathrm{SS}_{XX}/(m-1)$ and also $S_Y^2 \equiv \sum_1^n (Y_j - \bar{Y}_n)^2/(n-1) = \mathrm{SS}_{YY}/(n-1)$, and suppose that $\hat{\alpha}$ is a rv with values in $[0,1]$. We hypothesize that

$$\hat{\alpha}^2/\,\alpha^2 \to_p 1 \qquad \text{as } m \wedge n \to \infty.$$

(All limits below are to be taken as $m \wedge n \to \infty$.) Then define

$$\hat{\theta} \;\equiv\; \hat{\theta}_{mn} \;=\; \hat{\alpha}\bar{X}_n + (1-\hat{\alpha})\bar{Y}_n\,,$$

$$v_o^2 \;\equiv v_{omn}^2 \equiv \tfrac{1}{m}\,\alpha^2\sigma^2 \;+ \tfrac{1}{n}\,(1-\alpha)^2\tau^2\,,$$

$$\hat{v}^2 \;\equiv \hat{v}_{mn}^2 \;\equiv \tfrac{1}{m}\,\alpha^2 S_X^2 + \tfrac{1}{n}\,(1-\alpha)^2 S_Y^2\,,$$

$$\hat{V}^2 \equiv \hat{V}_{mn}^2 \;\equiv \tfrac{1}{m}\,\hat{\alpha}^2 S_X^2 + \tfrac{1}{n}\,(1-\hat{\alpha})^2 S_Y^2\,,$$

$$\tilde{\alpha} \;\equiv\; \tilde{\alpha}_{mn} \;\equiv\; \tfrac{(m-1)}{(m-1)+(n-1)S_X^2/S_Y^2} = \mathrm{SS}_{YY}/(\mathrm{SS}_{XX} + \mathrm{SS}_{YY})\,.$$

Note that $\tilde{\gamma} \equiv \tilde{\gamma}_{mn} \equiv S_X^2/S_Y^2 \to_p \gamma$, and $\tilde{\alpha}^2/\,\alpha^2 \to_p 1$ is indeed true.
(a) Show that $\mathrm{E}\,\hat{\theta} = \theta$.
(b) Show that $(\hat{\theta}-\theta)/\,v_o \to_d N(0,1)$.
(c) Show that $|\hat{v}^2 - v_o^2|\,/\,v_o^2 \to_p 0$ and $|\hat{V}^2 - \hat{v}^2|\,/\,\hat{v}^2 \to_p 0$.
(d) Thus $(\hat{\theta}-\theta)/\,\hat{V} \to_d N(0,1)$.
(e) Evaluate $v^2 \equiv v_{mn}^2 \equiv \mathrm{Var}[\hat{\theta}]$ in terms of $\mathrm{E}\hat{\alpha}^2$ and $\mathrm{E}(1-\hat{\alpha})^2$.
(f) Determine the distribution of $\tilde{\alpha}$. Does $\tilde{\alpha}/\,\alpha \to_{\mathcal{L}_2} 1$?

Exercise 3.22 (Exponential estimation) Let $X_1, \dots, X_n$ be iid Exponential(θ). The minimum variance estimator of θ is known to be the sample mean $\bar{X}_n$. Another unbiased estimator of θ is $T_n \equiv \bar{G}_n / \Gamma^n(1 + 1/n)$, where $\bar{G}_n \equiv (\prod_1^n X_k)^{1/n}$ denotes the geometric mean of the observations. Evaluate the limiting ratio of the variances $\lim \operatorname{Var}[\bar{X}_n]/\operatorname{Var}[T_n]$.

Exercise 3.23 Let $X_1, \dots, X_n$ be iid Poisson(λ). Show the moment convergence $\mathrm{E}|\bar{X}_n - \lambda|^3 \to \mathrm{E}|N(0,1)|^3$.

Statistical Applications

Exercise 3.24 (Simple linear regression) Consider the simple linear regression model of (A.3.25); thus we are assuming that

$$(28) \qquad Y_{nk} = \gamma + \beta\, x_{nk} + \epsilon_k \equiv \alpha + \beta\,(x_{nk} - \bar{x}_n) + \epsilon_k \qquad \text{for iid rvs } \epsilon_k \cong (0, \sigma^2)$$

and for known constants x_{nk} for $1 \le k \le n$. The least squares estimators (LSEs) $\hat{\alpha}_n$ and $\hat{\beta}_n$ of α and β are defined to be those values of a and b that minimize the sum of squares $\sum_1^n [\,Y_{nk} - (a + b\,(x_{nk} - \bar{x}_n)\,]^2$.

(a) Show that the LSEs are given by

$$(29) \qquad \hat{\alpha}_n = \bar{Y}_n \qquad \text{and} \qquad \hat{\beta}_n = \frac{\sum_1^n (x_{nk} - \bar{x}_n)\, Y_{nk}}{\sum_1^n (x_{nk} - \bar{x}_n)^2} \equiv \sum_{k=1}^n d_{nk} Y_{nk}\,.$$

(b) Let $\mathrm{SS}_{xx} \equiv \sum_1^n (x_{nk} - \bar{x}_n)^2$, $\mathsf{d}_n = (d_{n1}, \dots, d_{nn})'$, $\mathsf{x}_n \equiv (x_{n1}, \dots, x_{nn})'$, and

$$(30) \qquad \mathcal{D}(\mathsf{x}_n) \equiv \Big[\max_{1 \le k \le n} |x_{nk} - \bar{x}_n|^2\Big] \Big/ \mathrm{SS}_{xx} = \Big[\max_{1 \le k \le n} d_{nk}\Big] = \mathcal{D}(\mathsf{d}_n)\,.$$

Use the Cramér–Wold device and the weighted sums of theorem 9.5.2 to show that

$$(31) \qquad \begin{bmatrix} \sqrt{n}\,[\hat{\alpha}_n - \alpha] \\ \sqrt{\mathrm{SS}_{xx}}\,[\hat{\beta}_n - \beta] \end{bmatrix} \to_d N(\mathsf{O}, \sigma^2\, I), \quad \text{provided that} \quad \mathcal{D}(\mathsf{d}_n) = \mathcal{D}(\mathsf{x}_n) \to 0$$

(recall (10.3.20)). [Note also that the LSE $\hat{\beta}_n$ of β is given by $\hat{\beta}_n = \mathrm{SS}_{xY}/\mathrm{SS}_{xx}$.]

Definition 3.1 (Noncentral distributions) (a) Let $X_1, \dots, X_m$ be independent, and suppose that $X_i \cong N(\theta_i, \sigma^2)$. Let $\theta^2 \equiv \sum_1^m \theta_i^2$, and define δ via $\delta^2 \equiv \theta^2/\sigma^2$. Show that the quadratic form

$$(32) \qquad U \equiv \textstyle\sum_1^m X_i^2/\sigma^2 \cong (Z_1 + \theta)^2 + \sum_{i=2}^m Z_i^2,$$

where $Z_1, \dots, Z_m$ are iid N(0,1) rvs. Denote this distribution by

$$(33) \qquad U \cong \chi^2_m(\delta^2/2),$$

and say that U is distributed as *noncentral chisquare* with m degrees of freedom and noncentrality parameter δ.

(b) Let $Y \cong N(\theta, 1)$, $U \cong \chi^2_m(\delta^2/2)$ and $V \cong \chi^2_n$ be independent rvs. We define the noncentral Student-$t_n(\theta)$ distribution and the noncentral Snedecor-$F_{m,n}(\delta^2/2)$ via

$$(34) \qquad T_n(\theta) \equiv \frac{Y}{V/n} \cong \text{Student-}t_n(\theta) \ \text{ and } \ \frac{n}{m}\,\frac{U}{V} = \frac{U/m}{V/n} \cong \text{Snedecor-}F_{m,n}(\delta^2/2)\,.$$

(Note that $T_n^2(\theta) \cong F_{1,n}(\theta^2/2)$.)

Proposition 3.1 (Form of the noncentral distributions) Consider the rvs U, V, and Y of the previous definition. Let $y > 0$. (a) The rv U of (32) satisfies

(35) $$P(\chi^2_m(\delta^2/2) > y) = \sum_{k=0}^{\infty} P(\text{Poisson}(\delta^2/2) = j) \times P(\chi^2_{m+2j} > y)\,.$$

Here, Poisson(λ) denotes a Poisson rv with mean λ, and χ^2_r denotes an ordinary chisquare rv with r degrees of freedom.
(b) It is thus trivial that

(36) $$P(F_{m,n}(\delta^2/2) > y) = \sum_{k=0}^{\infty} P(\text{Poisson}(\delta^2/2) = j) \times P(F_{m+2j,n} > y)\,.$$

(c) For $C_n \equiv 2^{(n+1)/2}\,\Gamma(n/2)\,\sqrt{\pi\, n}$ we have

$$P(T_n(\delta) > y)$$

(37) $$= \tfrac{1}{C_n} \int_y^{\infty} \int_0^{\infty} u^{(n-1)/2}\, e^{-u/2}\, \exp(-\tfrac{1}{2}\,(v(\tfrac{u}{n})^{1/2} - \delta)^2)\, du\, dv\,.$$

Exercise 3.25 Prove proposition 3.1.

Exercise 3.26 (Chisquare goodness of fit, again) (a) (Local alternatives) We suppose that the statistic $Q_n \equiv Q_n(\mathsf{p}_0)$ of (10.1.13) is computed, but that in reality the true parameter vector is now $\mathsf{p}_n \equiv \mathsf{p}_0 + \mathsf{a}/\sqrt{n}$ (with $\sum_1^k a_i = 0$, so that the coordinates p_{ni} add to 1). Let $\hat{p}_{ni} \equiv N_{ni}/n$ estimate p_{ni} for $1 \le i \le k$. Show that the vector

(38) $$\mathsf{W}_n^{k\times 1} \equiv |[\,\sqrt{n}\,(\hat{p}_{ni} - p_{0i})/\sqrt{p_{0i}}\,]| \to_d \mathsf{W} + \mathsf{d}\,,$$

where $\mathsf{W} \cong N(\mathsf{O}, I - \sqrt{\mathsf{p}}\sqrt{\mathsf{p}'})$ and $d_i \equiv a_i/\sqrt{p_{0i}}$ for $1 \le i \le k$. Thus

(39) $$Q_n = \mathsf{W}_n'\,\mathsf{W}_n \to_d Q \equiv (\mathsf{W} + \mathsf{d}\,)'(\mathsf{W} + \mathsf{d}\,) \cong \chi^2_{k-1}(\mathsf{d}\,'\mathsf{d}/2)\,.$$

(b) (Fixed alternatives) Suppose that $Q_n \equiv Q_n(\mathsf{p}_0)$ is computed, but a fixed p is true. Show that

(40) $$\tfrac{1}{n}\, Q_n \to_{a.s.} \sum_{i=1}^k (p_i - p_{0i})^2/\,p_{0i}\,.$$

Exercise 3.27 Suppose $\mathsf{X} \cong N(\vec{\theta}, \Sigma)$, with rank$(\Sigma) = r$. Show that

(41) $$\mathsf{X}'\,\Sigma^-\,\mathsf{X} = Y'Y \cong \chi^2_r(\vec{\theta}\,'\,\Sigma^-\,\vec{\theta}/2)\,, \qquad \text{where}$$

(42) $$Y \equiv \Sigma^{-1/2}X = (\Gamma D^{-1/2}\Gamma')X \cong N\left(\Sigma^{-1/2}\theta,\, \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}\right)\,.$$

4 Local Limit Theorems

Recall from Scheffé's theorem that if f_n and f are densities with respect to some dominating measure μ, then

(1) $$f_n(x) \to f(x) \quad \text{a.e. } \mu$$

implies that

(2) $$d_{TV}(P_n, P) \equiv \sup_{B \in \mathcal{B}} |P(X_n \in B) - P(X \in B)| \to 0 \qquad \text{as } n \to \infty.$$

Thus convergence of densities implies convergence in total variation distance, which is stronger than convergence in distribution. We will now establish (1) in a CLT context, for summands that are either suitably continuous or else are distributed on a grid.

Theorem 4.1 (Local limit theorem, continuous case) Let $X, X_1, X_2, \ldots$ be iid $(0, \sigma^2)$ with $\int_{-\infty}^{\infty} |\phi_X(t)|\, dt < \infty$. Then $S_n/\sqrt{n}$ has a density $f_n(\cdot)$ for which

(3) $$\sup_{-\infty < x < \infty} |f_n(x) - (1/\sigma) f_Z(x/\sigma)| \to 0 \qquad \text{as } n \to \infty,$$

for the $N(0,1)$ density $f_Z(\cdot)$.

Theorem 4.2 (Local limit theorem, discrete case) Let $X_1, X_2, \ldots$ be iid $(0, \sigma^2)$ rvs that take values on the grid $a \pm md$ for $m = 0, \pm 1, \pm 2, \ldots$. Now let $x^* = (na + md)/(\sigma\sqrt{n}\,)$, $m = 0, \pm 1, \pm 2, \ldots$, denote a possible value of $S_n/(\sigma\sqrt{n}\,)$, and let $p_n(x) \equiv P(S_n/(\sigma\sqrt{n}\,) = x)$. Then

(4) $$\sup_{-\infty < x^* < \infty} |(\sigma\sqrt{n}/d\,)\, p_n(x) - f_Z(x)| \to 0 \qquad \text{as } n \to \infty,$$

for the $N(0,1)$ density $f_Z(\cdot)$.

Example 4.1 Let $X_1, \ldots, X_n$ be iid Bernoulli(θ) rvs. Then (by (4))

(5) $$\sup_{0 \le m \le n} \sqrt{n} \left| P(S_n = m) - \frac{1}{\sqrt{n\theta(1-\theta)}}\, f_Z\left(\frac{m - n\theta}{\sqrt{n\theta(1-\theta)}}\right) \right| \to 0\,. \qquad \square$$

Exercise 4.1* Verify (5) by direct computation.

Exercise 4.2 Give an example where $X_n \to_d X$, but (2) fails.

Proof. Consider theorem 4.1. Without loss of generality we may suppose that $\sigma = 1$. Notice that

(a) $$\int |\phi_{S_n/\sqrt{n}}(t)|\, dt < \infty, \quad \text{since} \quad |\phi_{S_n/\sqrt{n}}(t)| = |\phi_X(t/\sqrt{n}\,)|^n \le |\phi_X(t/\sqrt{n}\,)|.$$

Thus the Fourier inversion formula of (9.4.9) gives

(b) $$f_n(y) \equiv f_{S_n/\sqrt{n}}(y) = (1/2\pi) \int_{-\infty}^{\infty} e^{-ity}\, \phi_{S_n/\sqrt{n}}(t)\, dt.$$

This same formula also holds for the distribution of a N(0, 1) rv Z. Thus

(c) $2\pi|f_n(x) - f_Z(x)| \le \int_{-\infty}^{\infty} |\phi_X^n(t/\sqrt{n}\,) - e^{-t^2/2}|\,dt$

$$= \left(\int_{[|t|\le a]} + \int_{[a<|t|<\delta\sqrt{n}\,]} + \int_{[|t|\ge\delta\sqrt{n}\,]}\right) |\phi_X^n(t/\sqrt{n}\,) - e^{-t^2/2}|\,dt$$

(d) $\equiv I_{1n} + I_{2n} + I_{3n}$.

We first specify $\delta > 0$ so small that $|\phi_X(t)| \le \exp(-t^2/4)$ for $|t| \le \delta$. This is possible, since $|\phi(t)| = |1 - 0 - t^2/2| + |o(t^2)| \le 1 - t^2/4 \le \exp(-t^2/4)$ in some neighborhood $|t| \le \delta$, by inequality 9.6.2.

Thus for a specified large enough (since $|\phi_X(t/\sqrt{n}\,)|^n \le e^{-t^2/4}$ for $|t| < \delta\sqrt{n}\,$) we have

(e) $I_{2n} \le \int_{[|t|>a]} 2e^{-t^2/4}\,dt < \epsilon.$

For this fixed large a we have

(f) $I_{1n} < \epsilon$ for $n \ge$ (some n_1),

since the Cramér–Lévy theorem implies that the convergence of these chfs is uniform on any $|t| \le a$.

Now, X is not distributed on a grid (we have a formula for the density). Thus $|\phi_X(t)| < 1$ for all $t \ne 0$, by proposition 9.8.2. Moreover, $|\phi_X(t)| \to 0$ as $|t| \to \infty$, by the Riemann–Lebesgue lemma, giving $|\phi_X(t)| <$ (some θ) < 1 for $|t| >$ (some λ). Thus, since $\theta < 1$,

$$I_{3n} < \theta^{n-1}\int_{-\infty}^{\infty} |\phi_X(t/\sqrt{n}\,)|\,dt + \int_{[|t|>\delta\sqrt{n}\,]} e^{-t^2/2}\,dt$$

$$= \sqrt{n}\,\theta^{n-1}\int_{-\infty}^{\infty} |\phi_X(t)|\,dt + 2(\delta\sqrt{n}\,)^{-1}e^{-n\delta^2/2}$$

(6) $= o(n^{-r})$ for any $r > 0$ whenever $\overline{\lim}_{|t|\to 0}\, |\phi_X(t)| < 1$.

(g) $< \epsilon$ for $n \ge$ (some n_2).

Combining (e), (f), and (g) into (d) establishes the claim made in the theorem. □

Proof. Consider theorem 4.2. By the inversion formula (9.8.1) given for distributions on grids,

(h) $$\frac{\sigma\sqrt{n}}{d}\,p_n(x) = \frac{\sigma\sqrt{n}}{d}\,\frac{d}{\sigma\sqrt{n}\,2\pi}\int_{-\pi\sigma\sqrt{n}/d}^{\pi\sigma\sqrt{n}/d} \phi_X^n(t/\sqrt{n}\,)e^{-itx}\,dt.$$

By the inversion formula (9.4.9) given for densities,

(i) $f_Z(x) = (1/2\pi)\int_{-\infty}^{\infty} e^{-t^2/2}e^{-itx}\,dt.$

Thus

$$\text{(j)} \qquad |(\sigma\sqrt{n}/d)\, p_n(x) - f_Z(x)|$$

$$\leq (1/2\pi) \int_{-\pi\sigma\sqrt{n}/d}^{\pi\sigma\sqrt{n}/d} |\phi_X^n(t/\sqrt{n}\,) - e^{-t^2/2}|\, dt$$

$$+(1/2\pi) \int_{[|t|>\pi\sigma\sqrt{n}/d\,]} e^{-t^2/2}\, dt$$

$$\text{(k)} \qquad = \left(\int_{[|t|\leq a]} + \int_{[a<|t|<\delta\sqrt{n}\,]} + \int_{[\delta\sqrt{n}<|t|<\pi\sigma\sqrt{n}/d\,]}\right) |\phi_X^n(t/\sqrt{n}\,) - e^{-t^2/2}|\, dt$$

$$+o(n^{-r})$$

$$\text{(l)} \qquad \equiv I_{1n} + I_{2n} + I_{3n} + o(n^{-r}).$$

The proof of theorem 4.1 applies, virtually verbatim; the only thing worthy of note is that $0 < \theta < 1$ now holds, since $\pi\sigma\sqrt{n}/d$ is only $\frac{1}{2}$ of the period. □

5 Normality Via $\tilde{\text{W}}$insorization and $\check{\text{T}}$runcation

Definition 5.1 (a) Call $X_{n1}, \ldots, X_{nn}$ *weakly negligible* (or, *strongly negligible*) if

(1) $\quad M_n \equiv [\max_{1\le k\le n} |X_{nk}|\,] \to_p 0 \qquad$ (or, if $\quad M_n \to_{a.s.} 0$).

(b) Call them *uniformly asymptotically negligible* (or *uan*) if

(2) $\quad \max_{1\le k\le n} P(\,|X_{nk}| \ge \epsilon\,) \to 0 \qquad$ for all $\epsilon > 0$.

The objective is to now investigate when some version of the CLT holds for uan pieces, as in (2). (That is, there are to be no monsters among the many, no giants among the peons, etc., and the law of the mob is to hold sway. These are cute descriptive phrases others have used to summarize the problem very cleverly.)

Theorem 5.1 (Asymptotic normality) Let $X_{n1}, \ldots, X_{nn}$ be independent rvs. Let $a > 0$ and b be arbitrary. Fix the truncation constant; let it be any $c > 0$. Let $\check{X}_{nk}$ be the $\check{\text{T}}$runcated rv that equals X_{nk} or 0 according as $|X_{nk}| < c$ or as $|X_{nk}| \ge c$. As above, let $M_n \equiv [\max_{1\le k\le n} |X_{nk}|\,]$. The following are equivalent:

(3) $\quad \sum_{k=1}^n X_{nk} \to_d N(b, a^2) \qquad$ for uan rvs $X_{n1}, \ldots, X_{nn}$.

(4) $\quad \sum_{k=1}^n P(\,|X_{nk}| \ge \epsilon) \to 0 \quad$ for all $\epsilon > 0 \quad$ (equivalently, $M_n \to_p 0$), while
$\quad \check{\mu}_n \equiv \sum_{k=1}^n \mathrm{E}\check{X}_{nk} \to b \qquad$ and $\qquad \check{\sigma}_n^2 \equiv \sum_{k=1}^n \mathrm{Var}[\check{X}_{nk}] \to a^2\,.$

If (4) holds for one $c > 0$, then it holds for each $c > 0$. [These results also hold with $\tilde{\text{W}}$insorized quantities $\tilde{X}_{nk}$, $\tilde{\mu}_n$, $\tilde{\sigma}_n^2$ replacing the $\check{\text{T}}$runcated quantities $\check{X}_{nk}$, $\check{\mu}_n$, $\check{\sigma}_n^2$.] (The deficiency in this result is that the norming of the *original* rvs fails to appear.)

Corollary 1 Suppose that $Y_n \equiv \sum_1^n X_{nk} \to_d$ (some rv Y), for uan X_{nk}'s. Then

(5) $\quad Y$ has a Normal distribution $\qquad$ iff $\qquad M_n \equiv [\max_{1\le k\le n} |X_{nk}|\,] \to_p 0.$

Corollary 2 Let the X_{nk} above be symmetric. Then (3) (with $b = 0$) holds iff

(6) $\quad \sum_1^n X_{nk}^2 \to_p a^2 \qquad$ for uan rvs $X_{n1}, \ldots, X_{nn}$.

Proof. Suppose (4). Then $P(\sum_1^n X_{nk} \ne \sum_1^n \check{X}_{nk}) \le \sum_1^n P(\,|X_{nk}| \ge c) \to 0$, so that $\sum_1^n X_{nk} =_a \sum_1^n \check{X}_{nk}$. Thus we need only show that the normalized rv $\check{Z}_n \equiv (\,\sum_1^n \check{X}_{nk} - \check{\mu}_n)/\check{\sigma}_n \to_d N(0,1)$. It suffices to verify that Lindeberg's $\check{L}F_n^\epsilon$ of (10.2.11) satisfies $\check{L}F_n^\epsilon \to 0$. We will do so presently. First, note that $M_n \to_p 0$, by (8.3.14). Thus $[\max|\check{X}_{nk}|\,] \le M_n \to_p 0$ (and $[\max|\tilde{X}_{nk}|\,] \le M_n \to_p 0$). Thus

(a) $\quad \check{m}_n \equiv \max|\,\check{\mu}_{nk}| \le \mathrm{E}\max_k |\check{X}_{nk}| \to 0,$

by the DCT with dominating function "c." Thus for $n \ge$ (some n_ϵ) we have

(b) $\quad \check{m}_n \le \epsilon\, a/8, \qquad$ and also $\qquad \check{\sigma}_n \ge a/2\,.$

Then (3) must hold, since the Lindeberg–Feller quantity $\check{L}F_n^\epsilon$ of (10.2.11) satisfies

(c) $\check{L}F_n^\epsilon \equiv \frac{1}{\check{\sigma}_n^2} \sum_{k=1}^n \int_{[|x-\check{\mu}_{nk}|\ge\epsilon\,\check{\sigma}_n]\cap[|x|\le c]} [x-\check{\mu}_{nk}]^2\, dF_{nk}(x)$

(d) $\le \left[\frac{c+\epsilon\, a/8}{a/2}\right]^2 \sum_{k=1}^n \int_{[|x-\check{\mu}_{nk}|\ge\epsilon\,\check{\sigma}_n]\cap[|x|\le c]} dF_{nk}(x)$ for $n \ge n_\epsilon$, by (b)

(e) $\le \left[\frac{c+\epsilon\, a/8}{a/2}\right]^2 \sum_{k=1}^n P(\,|X_{nk}| \ge \epsilon\, a/4)$ by (b) again

(f) $\le \left[\frac{c+\epsilon\, a/8}{a/2}\right]^2 2\,P(M_n \ge \epsilon\, a/4) \to 0,$ since $M_n \to_p 0$ by (4)

(noting (8.1.8) as well). (Only very minor modifications of the above are called for in order to verify this for the $\tilde{\text{W}}$inorized case.)

Suppose that (3) holds. (We now assume knowledge of theorem C.1.1.Then

(7) $\sum_1^n X_{nk}^s \to_d N(0, 2a^2)$ for the symmetrized uan rvs $X_{nk}^s \equiv X_{nk} - X'_{nk}$,

defined as in section 8.3. Applying the continuity theorem for chf's then yields

(g) $\phi_n^s(t) \equiv \prod_1^n \phi_{nk}^s(t) \to \exp(-a^2\, t^2),$ uniformly on any $|t| \le M$,

as $n \to \infty$. Since the X_{nk}^s's are symmetric, their chf's are real valued. Moreover, all $(1-\phi_{nk}^s(t)) = \mathrm{E}(1-\cos t\, X_{nk}^s) \ge 0$. Thus lemma 8.1.4 (more powerful than the first product lemma 9.6.3, for positive numbers) shows that (g) holds if and only if

(8) $\sum_1^n \mathrm{E}(1-\cos t\, X_{nk}^s) = \sum_1^n (1-\phi_{nk}^s(t)) \to a^2\, t^2$, uniformly on any $|t| \le M$.

Thus for all $n \ge$ (some $n_{\epsilon,t}$) we have (recalling that $a = b \oplus c$ means that $|a-b| \le c$)

(h) $\frac{1}{3}\, a^2 t^3 = \int_0^t a^2\, u^2\, du = \int_0^t \sum_1^n \mathrm{E}\,(1-\cos u\, X_{nk}^s)\, du \oplus \epsilon$ from (8)

$= \int_0^t \{\sum_1^n \int_{-\infty}^\infty (1-\cos ux)\, dF_{nk}^s(x)\}\, du \oplus \epsilon$

(i) $= \int_{-\infty}^\infty \{\sum_1^n \int_0^t (1-\cos ux)\, du\}\, dF_{nk}^s(x) \oplus \epsilon$ by Tonelli

$= \sum_1^n \int_{-\infty}^\infty \{u(1-\frac{\sin ux}{ux})|_0^t\}\, dF_{nk}^s(x) \oplus \epsilon$

(9) $$= \sum_1^n \int_{-\infty}^\infty t\left(1-\frac{\sin tx}{tx}\right) dF_{nk}^s(x) \oplus \epsilon \qquad \text{uniformly on any } |t| \le M$$

whenever (8) holds. (Recall the inequality 9.5.1 for "comparison.") Let $t_1 = 1$ and $t_2 = 2$. Since $\frac{1}{2}\,(8\,t_1^3 - t_2^3) = 0$, subtracting the t_2-value of expression (9) from 8 times the t_1-value of expression (9) gives

(j) $0 = \frac{1}{2} \sum_1^n \int_{-\infty}^\infty [8(1-\frac{\sin x}{x}) - 2(1-\frac{\sin 2x}{2x})]\, dF_{nk}^s(x) \oplus \epsilon$

(10) $$= \sum_1^n \int_{-\infty}^\infty \left[3 - \frac{4\,\sin x}{x} + \frac{\sin 2x}{2x}\right] dF_{nk}^s(x) \oplus \epsilon \equiv \sum_1^n \int_{-\infty}^\infty h(x)\, dF_{nk}^s(x) \oplus \epsilon.$$

Consider the function $h(x) \equiv (3 - \frac{4\sin x}{x} + \frac{\sin 2x}{2x})$. Let $g(x) \equiv \inf\{h(y) : y \ge x\}$, so that g is the $\nearrow$ function closest to $f(\cdot)$ from below. (On $[0,4]$ the function h

increases from 0 to about 4, it then oscillates ever smaller about 3, and eventually converges to 3 at ∞.) The proof of the basic inequality (3.4.18) trivially gives

(k) $\quad P(\max_{1\le k\le n}|X^s_{nk}|\ge\epsilon)\le\sum_1^n P(\,|X^s_{nk}|\ge\epsilon)\le(\,\sum_1^n \mathrm{E}\,(h(X^s_{nk}))/\,g(\epsilon)\to 0.$

Thus, (8.3.14) and (k) yield

(11) $\quad M^s_n\equiv[\,\max_{1\le k\le n}|X^s_{nk}|\,]\to_p 0, \qquad \text{while} \quad M_n\equiv[\,\max_{1\le k\le n}|X_{nk}|\,]\to_p 0$

holds true since $[\,\max_{1\le k\le n}|X_{nk}-\ddot{m}_{nk}|\,]\to_p 0$ follows from $M^s_n\to_p 0$ and the symmetrization inequality (8.3.9), and the uan condition then guarantees that $[\,\max_{1\le k\le n}|\ddot{m}_{nk}|\,]\to 0$. (See Kallenberg (1997) regarding (11) via (10).) (Note that the "if" implication in Corollary 1 is now established as well.) Since (11) yields

(l) $\quad |\,P(\,\sum_1^n \check{X}^s_{nk}\le x)-P(\,\sum_1^n X^s_{nk}\le x)\,|\le\sum_1^n P(\,|X^s_{nk}|\ge c)\to 0$

(where $\check{X}^s_{nk}$ denotes X^s_{nk} truncated at $\pm c$), we have also established from (11) that

(12) $\quad \sum_1^n \check{X}^s_{nk}\to_d N(0,2a^2), \qquad \text{for} \qquad \check{X}^s_{nk}\equiv X^s_{nk}\cdot 1_{[\,|X^s_{nk}|\le c\,]}.$

Application of the continuity theorem to (12) yields (for every value of $c>0$)

(m) $\quad a^2\,t^2\leftarrow\sum_1^n \mathrm{E}\,(1-\cos t\check{X}^s_{nk})=\sum_1^n \frac{1}{2}\,t^2\,\mathrm{E}\,(\check{X}^s_{nk})^2\oplus\sum_1^n t^4\,\mathrm{E}\,(\check{X}^s_{nk})^4$

(n) $\quad\quad = \frac{1}{2}\,t^2\sum_1^n \mathrm{E}\,(\check{X}^s_{nk})^2\oplus c^2\,t^4\sum_1^n \mathrm{E}\,(\check{X}^s_{nk})^2=\frac{1}{2}\,t^2\sum_1^n \mathrm{E}\,(\check{X}^s_{nk})^2(1\oplus 2\,c^2\,t^2).$

Let n become very large and specify $c>0$ tiny in (m)–(n) to obtain

(13) $\quad \sum_1^n \mathrm{E}\,(\check{X}^s_{nk})^2\to 2\,a^2, \qquad \text{while} \qquad (M^s_n)^2=[\,\max_{1\le k\le n}(X^s_{nk})^2\,]\to_p 0$

follows from (11). The hypothesis of (C.1.9)are thus satisfied for the rvs $n\,(X^s_{nk})^2$, with all $\nu_n\equiv 1$ (or, $2a^2$). (Recall, (13) holds for all $c>0$.) Thus (C.1.3)

(14) $\quad \sum_1^n (\check{X}^s_{nk})^2=_a\,\sum_1^n (X^s_{nk})^2\to_p 2\,a^2.$

(Recall that appendix C could have been read immediately after chapter 8.) Further, the means of the $(\sum_1^n \check{X}^s_{nk})$ form a bounded sequence, since (using (13))

(15) $\quad \{\mathrm{E}\,|\sum_1^n \check{X}^s_{nk}|\,\}^2\le\mathrm{E}\,\{(\,\sum_1^n \check{X}^s_{nk})^2\}=\sum_1^n \mathrm{E}\,(\check{X}^s_{nk})^2\to 2\,a^2.$

A further consequence of (11) is that

(16) $\quad \sum_1^n \check{X}^s_{nk}=_a\sum_1^n X^s_{nk}=\sum_1^n (X_{nk}-X'_{nk})=_a\sum_1^n (\check{X}_{nk}-\check{X}'_{nk})=\sum_1^n \check{X}^s_{nk}.$

From (16) we can further claim that

(17) $$\begin{aligned}&\textstyle\sum_1^n \mathrm{E}\,(\check{X}^s_{nk})^2\to 2\,a^2 \qquad\qquad \text{and}\\ &2\,\check{\sigma}^2_n\equiv 2\textstyle\sum_1^n \check{\sigma}^2_{nk}=2\sum_1^n \mathrm{Var}[\check{X}_{nk}]\sum_1^n \mathrm{E}\,(\check{X}^s_{nk})^2\to 2\,a^2\end{aligned}$$

using (13) for the first $\to$ and mimicking (m)–(13) for the rvs $\check{X}^s_{nk}$ for the second. It is known from hypothesis (and $M_n\to_p 0$ in (11)) that

(o) $\quad (\,\sum_1^n \check{X}_{nk}-\sum_1^n \check{\mu}_{nk})+(\,\sum_1^n \check{\mu}_{nk})\,=_a\,\sum_1^n X_{nk}\to_d N(b,a^2).$

Combining this with (17) gives

(p) $\quad \check{\mu}_n\equiv\sum_1^n \check{\mu}_{nk}\to b, \qquad \text{and} \qquad \check{\sigma}^2_n=\sum_1^n \check{\sigma}^2_{nk}\to a^2$

has already been shown. Thus (4) is established, and this gives theorem 5.1.

Consider Corollary 2. If (3) holds, then (7) holds with a^2 and (14) gives (6). If (6) holds for symmetric uan rvs, then (C.1.3) and it implies (C.1.9)and hence (4). In light of the remark below (11), we now turn to the "only if" part of Corollary 1. Since $Y_n \to_d Y$, the symmetrized versions satisfy $Y_n^s \to_d Y^s \equiv Y - Y'$. That $M_n \to_p 0$ (and hence $M_n^s \to_p 0$) means that the truncated versions also satisfy $\check{Y}_n \equiv \sum_1^n \check{X}_{nk} \to_d Y$ and $\check{Y}_n^s \to_d Y^s$. Assume that $\check{\sigma}_{n'} \to \infty$ on some subsequence n'. Then $\check{Y}_{n'}^s/\check{\sigma}_{n'} \to_p 0$, but then a trivial application of Liapunov yields the contradictory result that $\check{Y}_{n'}^s/\check{\sigma}_{n'} \to_d N(0,2)$. Thus $\limsup_n \hat{\sigma}_n < \infty$. Pick a further subsequence n'' on which $\check{\sigma}_{n''} \to$ (some a) and $\check{\mu}_{n''} \to$ (some b). (Recall (15) and (8.3.18) regarding "b.") Then (4) implies (3) gives $Y \cong N(b, a^2)$.□

Remark 5.1 In the proof of theorem 5.1, it was shown that (15) implies both

(18)
$$\begin{aligned}
&\textstyle\sum_1^n (X_{nk} - \check{\mu}_{nk})/\{\sum_1^n \check{\sigma}_{nk}^2\}^{1/2} \to N(0,1),\\
&\textstyle\sum_1^n \check{\mu}_{nk} \to b, \quad \sum_1^n \check{\sigma}_{nk}^2 \to a^2, \quad \text{and} \quad M_n \equiv \{\max_{1\le k\le n} |X_{nk}|\} \to_p 0, \quad \text{and}\\
&\check{M}_n \equiv \{\max_{1\le k\le n} |X_{nk} - \check{\mu}_{nk}|\}/\{\sum_1^n \check{\sigma}_{nk}^2\}^{1/2} \to_p 0 \qquad \text{and}\\
&\tilde{M}_n \equiv \{\max_{1\le k\le n} |X_{nk} - \tilde{\mu}_{nk}|\}/\{\sum_1^n \tilde{\sigma}_{nk}^2\}^{1/2} \to_p 0.
\end{aligned}$$

Moreover, this holds for any choice of the truncation point "c" of the given rvs X_{nk}. (We can replace $\check{M}_n \to_p 0$ by $\tilde{M}_n \to_p 0$ in (18) since the added contributions to the $\tilde{\text{W}}$insorized quantities satisfy the bounds (for every $c > 0$)

(19)
$$\begin{aligned}
&|\textstyle\sum_1^n \tilde{\mu}_{nk} - \sum_1^n \check{\mu}_{nk}| \le c \sum_1^n P(|X_{nk}| \ge c) \to 0 \qquad \text{and}\\
&|\textstyle\sum_1^n \tilde{\sigma}_{nk}^2 - \sum_1^n \check{\sigma}_{nk}^2| \le c^2 \sum_1^n P(|X_{nk}| \ge c) + 3\,c^2 \sum_1^n P(|X_{nk}| \ge c) \to 0,
\end{aligned}$$

since $M_n \to_p 0$ is equivalent to $\sum_1^n P(|X_{nk}| \ge c) \to 0$. □

Notation 5.1 Let $Y_{n1}, \ldots, Y_{nn}$ denote the "original" row independent "rvs of interest," with dfs $F_{n1}, \ldots, F_{nn}$. Let

(20)
$$\textstyle\sum_1^n X_{nk} \equiv \sum_1^n (Y_{nk} - b_{nk})/\sqrt{n}\, a_n = \sqrt{n}\,(\bar{Y}_n - \bar{b}_n)/a_n$$

for some constants $\bar{b}_n \equiv \frac{1}{n}\sum_1^n b_{nk}$ and some $a_n > 0$. Further, let

(21)
$$\textstyle\sum_1^n X_{nk}^s \equiv \sum_1^n Y_{nk}^s/\sqrt{n}\, a_n = \sqrt{n}\ \overline{Y_n^s}/a_n$$

for the symmetrized $Y_{nk}^s \equiv Y_{nk} - Y_{nk}'$; see (8.3.2). Let $\nu_n \equiv a_n^2$. $\check{\text{T}}$runcate, by letting $\check{Y}_{nk}^s \equiv Y_{nk}^s \cdot 1_{[\,|Y_{nk}^s| \le \sqrt{n}\, a_n]}$. Then

(22)
$$\sqrt{n}\ \overline{\check{Y}_n^s}/a_n = \textstyle\sum_1^n \check{Y}_{nk}^s/\sqrt{n}\, a_n \quad \text{has mean } 0 \text{ and variance } U_n^s(\sqrt{n}\, a_n)/a_n^2$$

where

(23)
$$U_n^s(x) \equiv \tfrac{1}{n}\textstyle\sum_1^n E\,(\,(Y_{nk}^s)^2 \cdot 1_{[\,|Y_{nk}^s| \le x\,]}). \qquad \square$$

Theorem 5.2 Consider the original rvs of interest Y_{nk} given in notation 5.1 above, *always* assuming the resulting X_{nk}'s to be uan. The following claims are equivalent.

(24) $\quad \sqrt{n}\,(\bar{Y}_n - \bar{b}_n)/a_n \to_d N(0,1) \qquad$ for some $\bar{b}_n$ and $a_n > 0$.

(25) $\quad \sqrt{n}\,(\bar{Y}^s_n - 0)/a_n \to_d N(0,2) \qquad$ for some $a_n > 0$.

(26) $\quad (a) \quad M^s_n/\nu_n \equiv [\,\max_k \frac{1}{n}(Y^s_{nk})^2\,]/a_n^2 = [\,\max_k |X^s_{nk}|\,]^2 \to_p 0 \qquad$ and

$\qquad (b) \quad \sum_1^n \mathrm{E}(\check{Y}^s_{nk})^2/n\,\nu_n = \mathrm{E}\,\overline{(\check{Y}^s_n)^2}/a_n^2 \to 2. \qquad$ (So $\nu_n \equiv a_n^2 \sim \frac{1}{2}\,\mathrm{E}\,\overline{(\check{Y}^s_n)^2}$.)

(27) $\quad \frac{1}{n}\sum_1^n (Y^s_{nk})^2/\nu_n = \overline{(Y^s_n)^2}/a_n^2 \to_p 2, \qquad$ with $\nu_n = a_n^2$ as in (26)(b).

(28) $\quad$ Claims (C.1.3)–(C.1.10) hold for the $(Y^s_{nk})^2$, with $\nu_n = a_n^2$ as in (26)(b).

When (26) holds, the constants in (24) (now, named $\grave{a}_n$ and $\grave{b}_n$) must satisfy both:

$$\text{(29)} \qquad \begin{aligned} &\grave{a}_n^2 \sim \tfrac{1}{n}\textstyle\sum_1^n \mathrm{Var}[\check{\ddot{Y}}_{nk}] \sim \tfrac{1}{2}\,\tfrac{1}{n}\textstyle\sum_1^n \mathrm{Var}[\check{Y}^s_{nk}] = \tfrac{1}{2}\,\mathrm{E}\,\overline{(\check{Y}^s_n)^2} = \tfrac{1}{2}\,U^s_n(\sqrt{n}\,a_n) \sim a_n^2\,. \\ &\sqrt{n}\,\{\,\grave{b}_n - \tfrac{1}{n}\textstyle\sum_1^n (\ddot{m}_{nk} + \mathrm{E}\,\check{\ddot{Y}}_{nk})\}/\,\grave{a}_n \to 0 \end{aligned}$$

for the median $\ddot{m}_{nk}$ of Y_{nk} and for the truncated $\check{\ddot{Y}}_{nk}$ as defined as in (31) below. (Specify an "inspired" truncation point, and hope that something like (30) below saves us from the circularity of $U^s_n(\sqrt{n}\,a_n) \sim 2\,a_n^2$ in (29). See remark 10.6.2.)

Proof. That (24) implies (25) is trivial. That (25), (26), and (27) are equivalent is just theorem 5.1 and its corollary 2. That (27) and (28) are equivalent is just a careful comparison of (27) with (C.1.3)(which brings (C.1.4)–(C.1.10) with it).

Before going on, it is instructive to include an (unnecessary) proof that (26) is (C.1.9)in disguise. The first statement in (C.1.9)is exactly (26)(a). Also, for the specific $\nu_n = a_n^2$ defined in (26)(b), the second claim in (C.1.9)is

$$\text{(30)} \qquad \begin{aligned} \check{\mu}^s_n/\nu_n &= U^s_n(\sqrt{n}\,a_n)/a_n^2 = \textstyle\sum_1^n \mathrm{E}\,\{(Y^s_{nk})^2 \cdot 1_{[\,|Y^s_{nk}| \le \sqrt{n}\,a_n\,]}\}/na_n^2 \\ &= \textstyle\sum_1^n \mathrm{E}\,\{(X^s_{nk})^2 \cdot 1_{[(X^s_{nk})^2 \le 1]}\} \to 2, \qquad \text{which is exactly (26)(b).} \end{aligned}$$

We now show that (26) implies (24). Let $\ddot{m}_{nk}$ denote any median of Y_{nk} and set $\ddot{Y}_{nk} \equiv Y_{nk} - \ddot{m}_{nk}$, and for the a_n of (26)(b) set

(31) $\quad \check{\ddot{Y}}_{nk} \equiv \ddot{Y}_{nk} \cdot 1_{[\,|\ddot{Y}_{nk}| \le \sqrt{n}\,a_n]}\,; \qquad$ label its mean $\check{\ddot{\mu}}_{nk}$ and variance $\check{\ddot{\sigma}}^2_{nk}$.

Let $\ddot{X}_{nk} \equiv \ddot{Y}_{nk}/\sqrt{n}\,a_n = (Y_{nk} - \ddot{m}_{nk})/\sqrt{n}\,a_n$ and set

(32) $\quad \check{\ddot{X}}_{nk} \equiv \ddot{X}_{nk} \cdot 1_{[\,|\ddot{X}_{nk}| \le 1\,]} \equiv \check{\ddot{Y}}_{nk}/\sqrt{n}\,a_n.$

Combining (8.3.9) and (26)(a) yields

(33) $\quad \ddot{M}_n \equiv [\,\max_k |\ddot{X}_{nk}|\,] = [\,\max_k |Y_{nk} - \ddot{m}_{nk}|\,]/\sqrt{n}\,a_n \to_p 0.$

Since the DCT (domination with "1") and (33) imply $[\,\max_k |\check{\ddot{\mu}}_{nk}|\,]/\sqrt{n}a_n \to 0$ (just as in line (a) of the proof of theorem 5.1), we get

(34) $\quad \check{\ddot{M}}_n \equiv [\,\max_k |\check{\ddot{X}}_{nk} - \mathrm{E}(\check{\ddot{X}}_{nk})|\,] =_a [\,\max_k |Y_{nk} - \ddot{m}_{nk} - \check{\ddot{\mu}}_{nk}|\,]/\sqrt{n}\,a_n \to_p 0.$

Note that the $\check{\ddot{X}}_{nk} - \check{\ddot{\mu}}_{nk}$ are 0 mean rvs that satisfy (34). Now X^s_{nk}, $\ddot{X}^s_{nk}$, n and $\check{\ddot{X}}{}^{s}_{nk}$ denote the symmetrized versions of X_{nk}, $\ddot{X}_{nk}$, and $\check{\ddot{X}}_{nk}$, and let $\check{X^s}_{nk}$ and $\check{\ddot{X}^s}_{nk}$ denote the truncated (at $\sqrt{n}\, a_n$) versions of X^s_{nk} and $\ddot{X}^s_{nk}$. All of the rvs $\ddot{X}^s_{nk}$ are exactly equal to the rvs X^s_{nk}; so (with $c = 1$) each $\check{\ddot{X}^s}_{nk}$ equals $\check{X^s}_{nk}$. Thus

$$(35) \qquad \textstyle\sum_1^n \check{\ddot{X}}{}^{s}_{nk} =_a \sum_1^n \check{\ddot{X}^s}_{nk} = \sum_1^n \check{X^s}_{nk} =_a \sum_1^n X^s_{nk} \to_d N(0,2),$$

using (33) for the first $=_a$ in (35) and (26)(a) for the second. Thus (using (26)(b) and then $\check{\ddot{X}^s}_{nk} = \check{X^s}_{nk}$ in line (a))

$$\text{(a)} \qquad 1 \sim \mathrm{E}\,\overline{(\check{Y^s_n})^2}/2\,a_n^2 = \tfrac{1}{2}\textstyle\sum_1^n \mathrm{E}\,(\check{X^s}_{nk})^2 = \tfrac{1}{2}\sum_1^n \mathrm{E}\,(\check{\ddot{X}^s}_{nk})^2$$

$$\text{(b)} \qquad \sim \tfrac{1}{2}\textstyle\sum_1^n \mathrm{E}\,(\check{\ddot{X}}{}^{s}_{nk})^2 \qquad \text{by (34), (35), and (26)(b) applied to the } \check{\ddot{X}}{}^{s}_{nk}$$

$$(36) \qquad = \textstyle\sum_1^n \mathrm{Var}[\check{\ddot{X}}_{nk}] = \sum_1^n \mathrm{Var}[\check{\ddot{Y}}_{nk}]/n\,a_n^2.$$

Thus the first claim in (29) holds. The second claim follows from the convergence of types theorem. (The implication of (34)–(36) is that first symmetrizing and then truncating the $\ddot{X}_{nk}$'s is asymptotically equivalent to first truncating and then symmetrizing the $\ddot{X}_{nk}$'s.) (More complete results for iid rvs are presented in the next section.) □

From a Quantile Point of View

Notation 5.2 (Weak negligibility in the CLT context) Let $X_{n1}, \ldots, X_{nn}$ be independent rvs with dfs $F_{n1}, \ldots, F_{nn}$. Fix $\theta > 0$. Define $x_{\theta n}$ by requiring $[-x_{\theta n}, x_{\theta n}]$ to be the shortest closed, symmetric interval to which $\bar{F}_n \equiv \frac{1}{n}\sum_1^n F_{nk}$ assigns probability at least $1 - \theta/n$. Let $\bar{P}_n(x) \equiv \frac{1}{n}\sum_1^n P(\,|X_{nk}| > x)$ denote the average tail probability, and then let $\overline{K}_n$ denote the qf of the df $1 - \bar{P}_n(\cdot)$. Note the quantile relationship $x_{\theta n} = \overline{K}_n(1 - \theta/n)$.

Let $\tilde{X}_{nk}$ denote X_{nk} $\tilde{\text{W}}$insorized outside $[-x_{\theta_o n}, x_{\theta_o n}]$ (commonly, use $\theta_o \equiv 1$). Let $\tilde{\mu}_{nk}$ and $\tilde{\sigma}^2_{nk}$ denote the resulting means and variances; set both $\tilde{\mu}_n \equiv \frac{1}{n}\sum_1^n \tilde{\mu}_{nk}$ and $\tilde{\sigma}^2_n \equiv \frac{1}{n}\sum_1^n \tilde{\sigma}^2_{nk}$. Applying discussion 8.3.1 to the rvs $|X_{nk}|/\sqrt{n}\,\tilde{\sigma}_n$ (whose average df has $(1-\theta/n)$th quantile $x_{\theta n}/\sqrt{n}\,\tilde{\sigma}_n$) shows that the following conditions for the weak negligibility of the rvs $X_{nk}/\sqrt{n}\,\tilde{\sigma}_n$ are equivalent:

$$(37) \qquad M_n \equiv [\,\max_{1\le k\le n} |X_{nk}|\,]/\sqrt{n}\,\tilde{\sigma}_n \to_p 0\,.$$

$$(38) \qquad x_{\theta n}/\sqrt{n}\,\tilde{\sigma}_n \to 0 \qquad\qquad \text{for all } 0 < \theta \le \theta_{\theta_0}\,.$$

$$(39) \qquad \textstyle\sum_1^n P(\,|X_{nk}| > \epsilon\sqrt{n}\,\tilde{\sigma}_n) \to 0 \qquad \text{for all } 0 < \epsilon \le 1\,.$$

□

Theorem 5.3 (CLT using $\tilde{\text{W}}$insorization) Let $X_{n1}, \ldots, X_{nn}$ be independent rvs having dfs $F_{n1}, \ldots, F_{nn}$. If any/all of (37), (38), and (39) holds, then both

$$(40) \qquad \bar{Z}_n \equiv \sqrt{n}\,[\bar{X}_n - \tilde{\mu}_n]/\tilde{\sigma}_n \to_d N(0,1) \qquad \text{and}$$

$$(41) \qquad \tilde{M}_n \equiv [\,\max_{1\le k\le n} |X_{nk} - \tilde{\mu}_{nk}|\,]/\sqrt{n}\,\tilde{\sigma}_n \to_p 0\,.$$

Proof. For the proof, we will need notation similar to that above, for any θ. For $0 < \theta \le \theta_0$ let $\tilde{X}^\theta_{nk}$ denote X_{nk} Winsorized outside $[-x_{\theta n}, x_{\theta n}]$; let $\tilde{\mu}^\theta_{nk}$ and $\tilde{\sigma}^\theta_{nk}$ denote the obvious, and define both $\tilde{\mu}_{\theta n} \equiv \frac{1}{n}\sum_1^n \tilde{\mu}^\theta_{nk}$ and $\tilde{\sigma}^2_{\theta n} \equiv \frac{1}{n}\sum_1^n (\tilde{\sigma}^\theta_{nk})^2$. Let $\tilde{X}_{\theta n} \equiv \frac{1}{n}\sum_1^n \tilde{X}^\theta_{nk}$, and let $\tilde{Z}_{\theta n} \equiv \sqrt{n}\,[\tilde{X}_{\theta n} - \tilde{\mu}_{\theta n}]/\tilde{\sigma}_{\theta n}$. Define third central moments $\tilde{\gamma}^\theta_{nk} \equiv \mathrm{E}\,|\tilde{X}^\theta_{nk} - \tilde{\mu}^\theta_{nk}\,|^3$, and let $\tilde{\gamma}_{\theta n} \equiv \sum_1^n \tilde{\gamma}^\theta_{nk}$.

Call the rvs $\tilde{Y}^\theta_{nk} \equiv (\tilde{X}^\theta_{nk} - \tilde{\mu}^\theta_{nk})/\sqrt{n}\,\tilde{\sigma}_{\theta n}$ the *associated summands*, and let $\tilde{M}_{\theta n} \equiv [\max_k |\tilde{Y}^\theta_{nk}|\,]$ denote their maximal summand. Now, all $\tilde{\mu}_{nk} \in [-x_{\theta_0 n}, x_{\theta_0 n}]$. Thus $M_n \to_p 0$ implies that $\tilde{M}_n \to_p 0$ by using (37) and (38) in the triangle inequality. Fix $0 < \theta \le \theta_0$. Observe that with $\delta_n \equiv P(\sum_1^n X_{nk} \ne \sum_1^n \tilde{X}^\theta_{nk}))$, we have

$$(42)\qquad P(\bar{Z}_n \le z) = P(\tilde{Z}_{\theta n} \times (\tilde{\sigma}_{\theta n}/\tilde{\sigma}_n) + \sqrt{n}(\tilde{\mu}_{\theta n} - \tilde{\mu}_n)/\tilde{\sigma}_n \le z) \oplus \delta_n\,.$$

Since $P(\sum_1^n X_{nk} \ne \sum_1^n \tilde{X}^\theta_{nk}) \le \theta$ is trivial, and since $\|F_{\tilde{Z}_{\theta n}} - \Phi\| \le 13\,\tilde{\gamma}_{\theta n}/\sqrt{n}\,\tilde{\sigma}^3_{\theta n}$ by the Berry–Esseen theorem, (40) will follow immediately from showing that (43) below holds. [Recall that $a = b \oplus c$ means $|a-b| \le c$.] □

Inequality 5.1 For each $0 < \theta \le \theta_0$, results (44), (47), and (45) below lead to

$$(43)\qquad \text{(i)}\ \frac{\tilde{\gamma}_{\theta n}}{\sqrt{n}\,\tilde{\sigma}^3_{\theta n}} \to 0\,,\qquad \text{(j)}\ \frac{\tilde{\sigma}^2_{\theta n}}{\tilde{\sigma}^2_n} \to 1\,,\qquad \text{(k)}\ \frac{\sqrt{n}\,|\,\tilde{\mu}_{\theta n} - \tilde{\mu}_n\,|}{\tilde{\sigma}_n} \to 0\,.$$

Proof. For convenience, set $\theta_0 = 1$. We then note that $\tilde{\sigma}_n = \tilde{\sigma}_{1n}$. Bounding one power of $|\tilde{X}^\theta_{nk} - \tilde{\mu}^\theta_{nk}|^3$ by $2\,x_{\theta n}$ in the integrand of each $\tilde{\gamma}^\theta_{nk}$ gives

$$(44)\qquad \frac{\tilde{\gamma}_{\theta n}}{\sqrt{n}\,\tilde{\sigma}^3_{\theta n}} \le 2\,\frac{x_{\theta n}}{\sqrt{n}\,\tilde{\sigma}_{\theta n}} \le 2\,\frac{x_{\theta n}}{\sqrt{n}\,\tilde{\sigma}_n} \to 0,$$

using (38). Since the probability outside $[-x_{1n}, x_{1n}]$ is at most $1/n$,

$$(45)\qquad \sqrt{n}\,|\,\tilde{\mu}_{\theta n} - \tilde{\mu}_n\,|/\tilde{\sigma}_n \le x_{\theta n}/\sqrt{n}\,\tilde{\sigma}_n \to 0$$

by (38). We need some notation before turning to (43)(j). Let $\tilde{V}_{\theta n}$ denote the average of the second moments of the $\tilde{X}^\theta_{nk}$, and set $\tilde{V}_n \equiv \tilde{V}_{1n}$. Now, $\bar{F}_n$ assigns at most $1/n$ probability to the complement of the interval $[-x_{1n}, x_{1n}]$, and on $[-x_{\theta n}, x_{\theta n}]$ the integrand of $\tilde{V}_{\theta n}$ never exceeds $x^2_{\theta n}$. And so

$$(46)\qquad 0 \le [\,\tilde{\sigma}^2_{\theta n} - \tilde{\sigma}^2_n]/\,\tilde{\sigma}^2_{\theta n} = \{\tilde{V}_{\theta n} - \tilde{V}_n\}/\tilde{\sigma}^2_{\theta n} - \tfrac{1}{n}\textstyle\sum_1^n\ \{(\tilde{\mu}^\theta_{nk})^2 - \tilde{\mu}^2_{nk}\}/\tilde{\sigma}^2_{\theta n}$$

$$\text{(a)}\qquad = \{\tilde{V}_n \oplus \tfrac{1}{n}\,x^2_{\theta n} - \tilde{V}_n\}/\tilde{\sigma}^2_{\theta n} - \tfrac{1}{n}\textstyle\sum_1^n \{[\,\tilde{\mu}_{nk} \oplus x_{\theta n}\,P(\,|X_{nk}| > x_{1n})]^2 - \tilde{\mu}^2_{nk}\}/\tilde{\sigma}^2_{\theta n}$$

$$\le x^2_{\theta n}/[\,n\,\tilde{\sigma}^2_{\theta n}] + 2\,\tfrac{1}{n}\textstyle\sum_1^n \{\,[\,|\tilde{\mu}_{nk}|\;x_{\theta n}\,P(\,|X_{nk}| > x_{1n}]/\tilde{\sigma}^2_{\theta n}\,\}$$

$$+\ x^2_{\theta n}\ \tfrac{1}{n}\textstyle\sum_1^n P^2(\,|X_{nk}| > x_{1n})/\tilde{\sigma}^2_{\theta n} \qquad (\text{where all } P^2(\cdot) \le P(\cdot))$$

$$(47)\qquad \le (1+2+1)\,\{x^2_{\theta n}/n\,\tilde{\sigma}^2_{\theta n}\} = 4\,\{x^2_{\theta n}/n\,\tilde{\sigma}^2_n\} \to 0$$

by (38). Thus (43) holds, and it gives the normality of $\bar{Z}_n$ in (40). (Note the equivalence of (C.1.5) and (C.1.6), with $r = 2$.) The proof of (43) also showed

$$(48)\qquad \{\tilde{V}_{\theta n} - \tilde{V}_n\}/\,\tilde{\sigma}^2_{\theta n} \to 0 \qquad\text{and}\qquad \tfrac{1}{n}\textstyle\sum_1^n\ \{(\tilde{\mu}^\theta_{nk})^2 - \tilde{\mu}^2_{nk}\}/\tilde{\sigma}^2_{\theta n} \to 0.$$

(Contrast these calculations with those in remark 5.1.) □

Remark 5.2 (Summary) The work above is in the context of theorem C.1.1, not the context of theorem 5.1. For illustration, we have often done extra work with the cumbersome $\tilde{\text{W}}$insorized rvs, as well as basic work with the simpler $\check{\text{T}}$runcated rvs. Here is why! Use the $\check{\text{T}}$runcated rvs to prove, then claims made for the more "natural" $\tilde{\text{W}}$insorized rvs are often the ones of true interest that "look familar."

(Deceptively simple uan requirement) In most cases centering at medians is not possible, and it looks from (26)(b) or from $U_n^s(\sqrt{n}\,a_n) \sim 2\,a_n^2$ in (29) like we have an infinite loop in deciding where to truncate. Not necessarily! The fact that $\check{\mu}_n/\nu_n \to 1$ in (30)= (C.1.9)gives us some hope. Also, we would much rather claim asymptotic results for a sum that has been "prettied up" so that it looks familiar. That is, use the simplest equivalence to prove your result. Then use all the other equivalences to claim all the familiar looking results you can think of that are in the same vein. (In the next section on iid rvs we can be much more specific about this. See remark 10.6.2.) Compare the final sentence in the statement of theorem 5.1 with the circularity of the requirement that $U_n^s(\sqrt{n}\,a_n) \sim 2\,a_n^2$ in (29). □

6 Identically Distributed RVs

Notation 6.1 ($\tilde{\text{W}}$insorization notation) Let K be a fixed qf. Let $a > 0$ be tiny. We agree that $\text{dom}(a, a)$ denotes $[0, 1-a)$, $(a, 1]$, or $(a, 1-a)$ according as $X \geq 0$, $X \leq 0$, or general X, and that $\tilde{K}_{a,a}(\cdot)$ denotes K $\tilde{\text{W}}$insorized outside $\text{dom}(a, a)$. [For example, when X takes on both positive and negative values, then $\tilde{K}_{a,a}(\cdot)$ equals $K_+(a)$, $K(t)$, $K(1-a)$ according as $t \leq a$, $a < t < 1-a$, $1-a \leq t$.] Define $q(t) \equiv K^+(1-t) + K_+^-(t)$. Then define $v(t) \equiv [K^+(1-t)]^2 + [K_+^-(t)]^2$. [Note that $v(t) \leq q^2(t) \leq 2\, v(t)$.] Let ξ be a Uniform$(0,1)$ rv. Let $X \equiv K(\xi)$, and $\tilde{X}(a) \equiv \tilde{K}_{a,a}(\xi)$. We also define notation for various moments by agreeing that

$$(1) \qquad \tilde{X}(a) \equiv \tilde{K}_{a,a}(\xi) \cong (\tilde{\mu}(a), \tilde{\sigma}^2(a)) \qquad \text{with} \qquad \tilde{\gamma}(a) \equiv \text{E}|\tilde{X}(a) - \tilde{\mu}(a)|^3\,.$$

Now define the rvs $(X_{n1}, \ldots, X_{nn}) \equiv (K(\xi_{n1}), \ldots, K(\xi_{nn}))$ for row-independent Uniform$(0,1)$ rvs $\xi_{n1}, \ldots, \xi_{nn}$; thus they are row-independent with df F and qf K. Our interest is in

$$(2) \qquad \begin{aligned} &\bar{Z}_n \equiv \sqrt{n}\,[\bar{X}_n - \tilde{\mu}_n]/\tilde{\sigma}_n, && \text{where } \tilde{\mu}_n \equiv \tilde{\mu}(1/n) \text{ and } \tilde{\sigma}_n \equiv \tilde{\sigma}(1/n)\,, \\ &\tilde{Z}(a_n) \equiv \sqrt{n}\,[\tilde{X}_n - \tilde{\mu}(a_n)]/\tilde{\sigma}(a_n), && \text{where } \tilde{X}_n \equiv \tfrac{1}{n}\textstyle\sum_1^n \tilde{X}_{nk}(a_n), \end{aligned}$$

This is sufficient for the following theorem, but we now prepare for its follow up.

Let $\tilde{\sigma}_n \equiv \tilde{\sigma}_{1n} \equiv \tilde{\sigma}(1/n)$. Let $v_{1n} \equiv V(1/n)$, where

$$(3) \qquad \begin{aligned} &V(t) \equiv \textstyle\int_{\text{dom}(t,t)} K^2(s)\, ds && \text{on } [0, 1/2] \qquad \text{and} \\ &m(t) \equiv \textstyle\int_{\text{dom}(t,t)} |K(s)|\, ds && \text{on } [0, 1/2]\,, \end{aligned}$$

which for all t sufficiently small are equal to $\int_{\text{dom}(t,t)} v(s)\, ds$ and $\int_{\text{dom}(t,t)} q(s)\, ds$. The *Winsorized second moment* (when Winsorized outside $\text{dom}(1/n, 1/n)$) is given by $\tilde{v}_{1n} \equiv \text{E}\,\tilde{K}^2_{1/n,1/n}(\xi)$.

Let $|X|$ have df $F_{|X|}(\cdot)$ and qf $K_{|X|}(\cdot)$. Let $\bar{v}(t) \equiv K^2_{|X|}(T)$. Define the *partial second moments* and the *truncated second moments* in terms of both dfs and qfs via

$$(4) \qquad \begin{aligned} &V_{|X|}(t) \equiv \textstyle\int_{[0,1-t)} K^2_{|X|}(s)\, ds && \text{on } [0, 1]\,, \\ &U(x) \equiv \textstyle\int_{[|y| \leq x]} y^2\, dF(y) && \text{on } [0, \infty)\,. \end{aligned}$$

Let $\bar{v}_{1n} \equiv V_{|X|}(1/n)$ and $u_{1n} \equiv U(x_{1n})$, and we define $x_{\theta n} \equiv K_{|X|}(1-\theta/n)$ to be the $(1-\theta/n)$-quantile of the rv $|X|$. Thus $[-x_{\theta n}, x_{\theta n}]$ is the shortest interval symmetric about 0 that contains at least the proportion $(1-\theta/n)$ of the X probability. The *Winsorized second moment* in the present context is $\tilde{u}_{1n}$, where

$$(5) \qquad \tilde{u}_{1n} \equiv u_{1n} + x^2_{1n}\, P(\,|X| > 1/n) \quad \text{equals} \quad \tilde{\bar{v}}_{1n} \equiv \bar{v}_{1n} + K^2_{|X|}(1-1/n)/n,$$

even though $u_{1n} \geq m_{1n}$ always holds (by (C.1.53)). □

Theorem 6.1 (CLT for iid rvs) Let $X_{n1}, \ldots, X_{nn}$ be row independent iid rvs. (See (2)–(5) for the possible ν_n choices in (9).) The following are equivalent.

(6) $\tilde{r}^2(t) \equiv t\, q^2(t)/\tilde{\sigma}^2(t) \to 0$ as $t \to 0$. (Note (28) below.)

(7) $\bar{Z}_n \to_d N(0,1)$.

(8) $\tilde{Z}(1/n) \to_d N(0,1)$.

(9) $[\max_{1\le k\le n} \frac{1}{n} X_{nk}^2]/\nu_n \to_p 0$ for ν_n any of $v_{1n}, \tilde{v}_{1n}, u_{1n}, \bar{v}_{1n}, \tilde{u}_{1n}$, or $\tilde{\sigma}_{1n}$.

(10) $x_{\theta n}/(\sqrt{n}\,\tilde{\sigma}_n) \to 0$ for all $0 < \theta \le 1$ (or, use any ν_n as in (27)).

(11) $S_n^2/\tilde{\sigma}_n^2 \equiv (\overline{X_n^2} - \bar{X}_n^2)/\tilde{\sigma}_n^2 \to_p 1$.

Moreover, $\breve{\mu}_n \equiv \int_{(1/n,1-1/n)} K(t)\,dt$ can replace $\tilde{\mu}_n$ in the definition of $\bar{Z}_n$ in (7).

Proof. Consider (7). We are Winsorizing the qf on $[t, 1-t]$, not the df on $[-x_{\theta n}, x_{\theta n}]$. Even so, we essentially repeat the argument used for (10.5.34). Let

(a) $\tilde{\mu}_{\theta n} \equiv \tilde{\mu}(\theta/n)$, $\tilde{\sigma}_{\theta n} \equiv \tilde{\sigma}(\theta/n)$, $\tilde{Z}_{\theta n} \equiv \sqrt{n}\,\{\frac{1}{n}[\sum_1^n \tilde{X}_{nk}(\theta/n)] - \tilde{\mu}_{\theta n}\}/\tilde{\sigma}_{\theta n}$.

(Now, $a = b \oplus c$ means that $|a-b| \le c$.) First observe, with the probability that the two rvs differ bounded by $\delta_n \equiv P(\sum_{k=1}^n X_k \ne \sum_{k=1}^n \tilde{X}_{nk}(\theta/n))$, that

(b) $P(\bar{Z}_n \le z) = P(\tilde{Z}_{\theta n} \times (\tilde{\sigma}_{\theta n}/\tilde{\sigma}_n) + \sqrt{n}(\tilde{\mu}_{\theta n} - \tilde{\mu}_n)/\tilde{\sigma}_n \le z) \oplus \delta_n$.

Since $\delta_n = P(\sum_1^n X_k \ne \sum_1^n \tilde{X}_{nk}(\theta/n)) \le 2\theta$ is trivial, and since the Berry–Esseen theorem gives $\|F_{\tilde{Z}_{\theta n}} - \Phi\| \le 9\,\tilde{\gamma}_{\theta n}/\sqrt{n}\,\tilde{\sigma}_{\theta n}^3$, the current theorem will follow from showing that for each $\theta > 0$ we have

(12) (i) $\dfrac{\tilde{\gamma}_{\theta n}}{\sqrt{n}\,\tilde{\sigma}_{\theta n}^3} \to 0$, (j) $\dfrac{\tilde{\sigma}(\theta/n)}{\tilde{\sigma}(1/n)} \to 1$, (k) $\dfrac{\sqrt{n}\,|\,\tilde{\mu}(\theta/n) - \tilde{\mu}(1/n)\,|}{\tilde{\sigma}(1/n)} \to 0$.

Recall from above that $q(a) \equiv K^+(1-a) + K_+^-(a)$. Then when $r(t) \to 0$, we have

(c) $\dfrac{\tilde{\gamma}_{\theta n}}{\sqrt{n}\,\tilde{\sigma}_{\theta n}^3} \le \dfrac{1}{\sqrt{\theta}} \times \dfrac{\sqrt{\theta/n}\, q(\theta/n)}{\tilde{\sigma}(\theta/n)} = \tilde{r}(\theta/n)/\sqrt{\theta} \to 0$ using $\tilde{r}(t) \to 0$,

(d) $\tilde{\sigma}^2(\theta/n)/\tilde{\sigma}^2(1/n) \to 1$ in analogy with (10.5.38), using $\tilde{r}(t) \to 0$,

(e) $\sqrt{n}\,|\,\tilde{\mu}(\theta/n) - \tilde{\mu}(1/n)\,|/\tilde{\sigma}(1/n) \le \{q(\theta/n) - q(1/n)\}/\sqrt{n}\,\tilde{\sigma}(1/n)$

$\le q(\theta/n)/\sqrt{n}\,\tilde{\sigma}(1/n)$

(f) $\le \{\tilde{r}(\theta/n)/\sqrt{\theta}\}\,\{\tilde{\sigma}(\theta/n)/\tilde{\sigma}(1/n)\} \to 0$ by (c) and (d).

Thus (12), and so $\bar{Z}_n \to_d N(0,1)$ in (7). Just (12)(i)($\theta = 1$) gives normality in (8).

Before completing this proof, we will present a definition and a proposition that will add a myriad of other equivalences to the list. □

Definition 6.1 (Slowly varying functions) (a) Call $L(\cdot) > 0$ *slowly varying at* 0 (written $L \in \mathcal{R}_0$ or $L \in \mathcal{L}$) provided $L(ct)/L(t) \to 1$ as $t \to 0$, for each $c > 0$.
(b) The function $l(\cdot) > 0$ on $(0, \infty)$ is called *slowly varying at* ∞ (written as $l \in \mathcal{U}_0$) if it satisfies $l(cx)/l(x) \to 1$ as $x \to \infty$, for each positive number $c > 0$.
(∗) The functions $L(t) = \log(1/t)$ and $l(x) = \log(x)$ are the prototypes.

Proposition 6.1 (Equivalent conditions that give a CLT for iid rvs) The following contains just the "best" items from a long list of equivalents. When any one (hence all) hold, we write either $F \in \mathcal{D}(\text{Normal})$ or $K \in \mathcal{D}(\text{Normal})$ and say that F (or, K) is in the *domain of attraction* of the normal distribution. [We require one specific $a_n \searrow 0$ having $\overline{\lim}\,(a_n/a_{n+1}) < \infty$) in (29); $a_n = \theta/n$ is one such.] (See notation 5.1 for $v_{1n}, \tilde{v}_{1n}, u_{1n}, \bar{v}_{1n}, \tilde{u}_{1n} = \tilde{\bar{v}}_{1n}$ or $\tilde{\sigma}_{1n}$.) The following are equivalent:

(13) $R(x) \equiv x^2\, P(|X| > x)/\,U(x) \to 0$ (or, $p\; x^2\, P(|X| \ge x)/\,U(x-) \to 0)$.

(14) $x\,[M(cx) - M(x)]/U(x) \to 0$, all $c > 0$; where $M(x) \equiv \int_{[|y| \le x]} |y|\, dF(y)$.

(15) $U(\cdot)$ is slowly varying at infinity, where $U(x) \equiv \int_{[|y| \le x]} y^2\, dF(y)$.

(16) $V(\cdot)$ is slowly varying at zero, where $V(t) \equiv \int_{\text{dom}(t,t)} K^2(s)\, ds$

(17) $V_{|X|}(\cdot)$ is slowly varying at zero, where $V_{|X|}(t) \equiv \int_{[0,1-t)} K^2_{|X|}(s)\, ds$

(18) $\tilde{\sigma}^2(\cdot)$ is slowly varying at zero, where $\tilde{\sigma}^2(t) \equiv \text{Var}\,[\tilde{K}^2_{t,t}(\xi)]$.

(19) $\overline{X^2_n}\,/\nu_n \to_p 1$ for any one specific $\nu_n > 0$.

(20) $\overline{X^2_n}\,/\,\nu_n \to_p 1$, for ν_n one of $v_{1n}, \tilde{v}_{1n}, u_{1n}, \bar{v}_{1n}$ or $\tilde{u}_{1n} = \tilde{\bar{v}}_{1n}$.

(21) $S^2_n\,/\,\tilde{\sigma}^2_n \to_p 1$, where $\tilde{\sigma}^2_{1n} \equiv \text{Var}\,[\tilde{K}^2_{1/n,1/n}(\xi)]$.

(22) $\mathcal{D}^2_n \equiv [\frac{1}{n} \max_k (X_{nk} - \bar{X}_n)^2\,]/S^2_n \to_p 0$.

(23) $[\max_k X^2_{nk}\,]\,/\,n\,\overline{X^2_n} \to_p 0$.

(24) $[\max_k \frac{1}{n} X^2_{nk}\,]\,/\,\nu_n \to_p 0$ and $U(n\,\nu_n)/\,\nu_n \to 0$.

(25) $[\max_k \frac{1}{n} X^2_{nk}\,]\,/\,\nu_n \to_p 0$ for ν_n one of $v_{1n}, \tilde{v}_{1n}, u_{1n}, \bar{v}_{1n}, \tilde{u}_{1n}$, or $\tilde{\sigma}_{1n}$.

(26) $n\,P(X^2 > c\,n\,\bar{v}_{1n}) \to 0$ for all $c > 0$ (or, use any ν_n as in (25)).

(27) $x_{\theta n}/\,[\sqrt{n}\,\tilde{\sigma}_n] \to 0$ for all $0 < \theta \le 1$ (or, use any ν_n as in (25)).

(28) $\tilde{r}^2(t) \equiv t\,q^2(t)/\tilde{\sigma}^2(t) \to 0$ as $t \to 0$. (See near (1) for $q(\cdot)$).

(29) $\tilde{r}(a_n) = \sqrt{a_n}\; q(a_n)/\tilde{\sigma}(a_n) \to 0$ for one specific $a_n \searrow 0$, as above.

(30) $t\,K^2_{|X|}(1-t)/V_{|X|}(t) \to 0$ where $V_{|X|}(t) \equiv \int_{[0,1-t)} K^2_{|X|}(s)\, ds$.

(31) $t\,[K^2_+(t) \vee K^2(1-t)]/V(t) \to 0$ where $V(t) \equiv \int_{\text{dom}(t,t)} K^2(s)\, ds$.

(32) $t\,[K^2_+(ct) \vee K^2(1-ct)]/\tilde{\sigma}^2(t) \to 0$ for all $0 < c \le 1$. (Or, use $K^2_{|X|}(ct)$.)

(33) $q(\theta/n)\,/\,[\sqrt{n}\,\tilde{\sigma}(1/n)] \to 0$ for all $\theta > 0$. (See near (1) for $q(\cdot)$).

(34) $[q(\theta/n) - q(1/n)]\,/\,[\sqrt{n}\,\tilde{\sigma}(1/n)] \to 0$ for all $\theta > 0$.

(35) $\sqrt{n}\,[m(\theta/n) - m(1/n)]\,/\,\tilde{\sigma}(1/n) \to 0$ for all $\theta > 0$. (See (3) for $m(\cdot)$).

(36) $[v(\theta/n) - v(1/n)]\,/\,[n\,\tilde{\sigma}^2_n(1/n)] \to 0$ for all $\theta > 0$. (See near (1) for $v(\cdot)$.)

(Recall the Gnedenko–Kolmogorov theorem 6.6.1, (6.6.3)–(6.6.9), and (C.2.20).) Equivalents for the $X^s \equiv X - X'$'s follow from (25), (10.5.26), (10.5.36), and (8.3.9).

Corollary 1 (Asymptotic normality of the Student-T statistic) If any one (hence, all) of (16)–(36) holds, then

(37) $T_n \equiv \sqrt{n}\,[\,\bar{X}_n - \tilde{\mu}_n\,]\,/S_n \to_d N(0,1).$

Thus, we have a confidence interval available for μ_{1n} that is asymptotically valid for any df $F \in \mathcal{D}$(Normal).

Proof. (Continued) The equivalence of conditions (13)–(36) was specifically developed in section C.1–section C.3 where much longer lists of equivalences are given. Theorem C.1.1 needed only to tie the CLT into the combined list. As earlier pointed out, many of these equivalences are found in the literature. (Be aware that having a different denominator in a condition, or being required to verify it only on some sequence of values, can be very valuable.)

That (6) implies (7) and (8) has already been established. The equivalence of (6), (9) and (10) (as well as all the equivalences in proposition 6.1 was shown in appendix C. Also, (9) (using $\nu_n = \tilde{\sigma}_{1n}$) is exactly the same as (10.5.37), which was shown in theorem 10.5.3 to imply (7). The following easy exercise will complete the loop regarding (8). □

Exercise 6.1 Show that (8) implies (7).

Remark 6.1 (Natural parameters) The conclusion $\sqrt{n}\,[\bar{X}_n - b_n]/a_n \to_d N(0,1)$ for *some* b_n and *some* $a_n > 0$ has been shown to imply that necessarily

(38) $\sqrt{n}\,[\bar{X}_n - \tilde{\mu}_n]/\tilde{\sigma}_n \to_d N(0,1).$

Statisticians like to use means and standard deviations in their conclusions, even if they chose to verify (13) or (15) as the easiest way to establish such a result. That

(39) conclusions $(7), (11), (9), (13), (15), (18), (20),$ and (22) are all equivalent

is a beautiful result. Verify the easiest of (6)–(37) to establish the rest. (Those troubling, but enabling, a_n^2's have disappeared from the verification process for many of these conditions, and they have disappeared entirely from (38).) □

Remark 6.2 (Determining $\nu_n \equiv a_n^2$) Consider the possibility

(40) $a_n^2 \equiv U(\sqrt{n}\,)$ where $U(x) \equiv \int_{[\,|y|\le x]} y^2\, dF(y).$

It is in the spirit of (10.5.29) that

(a) $\frac{U(\sqrt{n}\,a_n)}{a_n^2} = \frac{U(\sqrt{n\,U(\sqrt{n}\,)})}{U(\sqrt{n}\,)}$ "should converge" to 1,

whenever $F \in \mathcal{D}$(Normal) (that is, whenever $U(\cdot)$ is slowly varying). The prototype slowly varying function is $\log(\cdot)$. If $U(n) \sim \log n$ were true, then indeed

(b) $\frac{U(\sqrt{n}\,a_n)}{a_n^2} \sim \frac{\log(\sqrt{n\,\log(\sqrt{n}\,)})}{\log(\sqrt{n}\,)} \to 1.$

Thus the strategy (40) is worth a look. (It is even worth a try in the non-iid case of (10.5.29). Just replace $U(x)$ above by $U_n(x) \equiv \int_{[\,|y|\le x]} y^2\, d\bar{F}_n(y)$.) (Just for fun, seek the order of $U(\sqrt{n}\,a_n)/\,a_n^2$ when $a_n^2 = U(\sqrt{n}\,c_n)$ with $c_n^2 \equiv U(\sqrt{n}\,)$?) □

7 A Converse of the Classical CLT

Theorem 7.1 (Domain of normal attraction of the normal df) Consider iid rvs $X_1, X_2, \ldots$, and set $Z_n = \sum_1^n X_k/\sqrt{n}$. Then:

(1) $\quad Z_n = \mathrm{O}_p(1) \quad$ implies $\quad \mathrm{E}X_1 = 0 \quad$ and $\quad \mathrm{E}(X_1^2) < \infty$.

(2) $\quad \mathrm{E}X_1 = 0 \quad$ and $\quad \mathrm{Var}[X_1] < \infty \quad$ imply $\quad Z_n \to_d N(0, \mathrm{Var}[X_1]\,)$.

Proof. (Giné and Zinn) Now, (2) was established previously. Consider (1). Fix $t > 0$. Let $Z_n^\epsilon \equiv \sum_1^n \epsilon_k\, X_k/\sqrt{n}$ for iid Rademacher rvs $\epsilon_1, \epsilon_2, \ldots$ that are independent of the X_k's. By Giné–Zinn symmetrization of (8.3.10), we have

(a) $\quad P(Z_n^\epsilon > 2\lambda) \le 2 \sup_{n\ge1} P(Z_n > \lambda)\,;$

and thus $P(Z_n^\epsilon > \lambda) = \mathrm{O}_p(1)$ by our hypotheses. Also, Khinchin's inequality in exercise 8.3.3 (regarding the X_k's as fixed constants, and with $r = 1$) gives

(b) $\quad \mathrm{E}_\epsilon\, |Z_n^\epsilon| \ge \frac{1}{3}\, \overline{X_n^2}^{1/2} = c\, S_n, \qquad$ now with $S_n^2 \equiv \overline{X_n^2}$ and $c = \frac{1}{3}$.

Applying Paley–Zygmund's inequality 3.4.9 to Z_n^ϵ (conditioned on fixed values of the X_k's) for the first inequality and (b) for the second yields

(c) $\quad P_\epsilon(|Z_n^\epsilon| > t) \ge \left(\dfrac{(\mathrm{E}_\epsilon|Z_n^\epsilon| - t)^+}{(\mathrm{E}_\epsilon\{(Z_n^\epsilon)^2\})^{1/2}} \right)^2 \ge \left(\dfrac{(cS_n - t)^+}{S_n} \right)^2$

(d) $\quad = c^2\,(1 - t/cS_n)^2\, 1_{[S_n > t/c]} \ge (c^2/4)\, 1_{[S_n > 2t/c]}\,.$

Taking expectations across the extremes of this inequality with respect to the X_k's gives the bound

(e) $\quad P(|Z_n^\epsilon| > t) \ge (c^2/4)\, P(S_n > 2t/c)\,.$

Thus $S_n = \mathrm{O}_p(1)$, by combining (e), (a), and the hypothesis.

Fix $M > 0$. The SLLN gives

(f) $\quad \frac{1}{n}\sum_1^n X_k^2\, 1_{[X_k^2 \le M]} \to_{a.s.} \mathrm{E}\{X_1^2\, 1_{[X_1^2 \le M]}\,\}\,.$

But $\to_{a.s.}$ implies $\to_d$. Thus, applying (9.1.12) to the open set (t, ∞) gives

(g) $\quad 1_{(0, \mathrm{E}(X_1^2\, 1_{[X_1^2 \le M]}))}(t) \le \underline{\lim}_n\, P(\frac{1}{n}\sum_1^n X_k^2\, 1_{[X_k^2 \le M]} > t)$

(h) $\quad \le \sup_n P(\frac{1}{n}\sum_1^n X_k^2\, 1_{[X_k^2 \le M]} > t), \qquad$ for each $t > 0$.

It follows that for each $t > 0$ we have

(i) $\quad \sup_{M>0} 1_{(0, \mathrm{E}(X_1^2\, 1_{[X_1^2 \le M]}))}(t) \le \sup_{M>0} \sup_n P(\frac{1}{n}\sum_1^n X_k^2\, 1_{[X_k^2 \le M]} > t)\,.$

(j) $\quad \le \sup_n P(\frac{1}{n}\sum_1^n X_k^2 > t) \le \sup_n P(S_n^2 > t)\,.$

Since $S_n = O_p(1)$, we have $S_n^2 = O_p(1)$; and this implies that we can specify a t value of t_0 in (j) so large that the right-hand side of (j) at t_0 is less than 1/2. But this implies that for this t_0 the indicator function in (i) must equal zero uniformly in M. This means that

(k) $$\sup_{M>0} \mathrm{E}(X_1^2\, 1_{[X_1^2 \le M]}) \le t_0\,.$$

But this last supremum equals $\mathrm{E}(X_1^2)$, and hence we must have $\mathrm{E}(X_1^2) \le t_0 < \infty$.

To complete the proof, we must now show that $\mathrm{E}(X_1) = 0$. Since $\mathrm{E}X_1^2 < \infty$, the WLLN gives $\bar{X}_n \to_p \mathrm{E}X_1$. But the hypothesis that $Z_n = O_p(1)$ implies that $\bar{X}_n = Z_n/\sqrt{n} \to_p 0$. Combining these gives $\mathrm{E}X_1 = 0$. □

8 Bootstrapping

Suppose $X_1, X_2, \ldots$ are an iid sample from F. Denote the empirical df of the sample $\mathsf{X}_n \equiv (X_1, \ldots, X_n)'$ by $\mathbb{F}_n(\cdot)$. This empirical df $\mathbb{F}_n$ has mean $\bar{X}_n$ and variance S_n^2. Let $\mathsf{X}_n^* \equiv (X_{n1}^*, \ldots, X_{nn}^*)$ denote an iid sample from $\mathbb{F}_n$, called the *bootstrap sample*. Let $\bar{X}_n^*$ and S_n^* denote the mean and the standard deviation of the bootstrap sample. Since the moments of $\mathbb{F}_n$ exist, we will work with normed summands. Note that the normed summands of a bootstrap sample *always* constitute a uan array, since

(1) $$\max_k P^*(|X_{nk}^* - \bar{X}_n|/\sqrt{n}S_n \geq \epsilon) \leq \mathrm{Var}[X_{1n}^*]/(\epsilon^2 n S_n^2) = 1/n\epsilon^2 \to 0,$$

all $\epsilon > 0$. The maximum normed summand (when forming the bootstrap mean) is

(2) $$M_n^* \equiv [\max_k |X_{nk}^* - \bar{X}_n|\,]/\sqrt{n}\, S_n].$$

Now (random sampling $X_i - \bar{X}_n$ values), we can view M_n^* as the value of both the $M_n = \tilde{M}_n$ of (10.5.37) and of (10.5.41), and note that (for each $0 < \theta \leq \theta_0 = 1/2$) the quantity $x_{\theta n}/\sqrt{n}\,\tilde{\sigma}_n$ in (10.6.10) now has a value of $\mathcal{D}_n$, where

(3) $$\mathcal{D}_n \equiv \mathcal{D}(\mathsf{X}_n) \equiv [\,\max_k |X_k - \bar{X}_n|\,]/\sqrt{n}\, S_n\,.$$

Now, $\mathcal{D}_n$ is formed from the original sample, while M_n^* is formed from the bootstrap sample. The following theorem is in the spirit of result (10.5.40). (Note, moreover, that $0 \leq M_n^* \leq \mathcal{D}_n$, while $P(M_n^* = \mathcal{D}_n) \geq 1 - (1 - 1/n)^n \to 1 - 1/e > 0$ also holds.) The "standardized" rv of (10.5.40) is now equal to

(4) $$\bar{Z}_n^* \equiv \sqrt{n}\,[\bar{X}_n^* - \bar{X}_n]/S_n\,, \qquad \text{and also define} \qquad T_n^* \equiv \sqrt{n}\,[\bar{X}_n^* - \bar{X}_n]/S_n^*\,.$$

Agree that the *weak bootstrap* holds if

(5) $$\bar{Z}_n^* \to_d N(0,1) \quad \text{for the joint probability on } \Omega \times \Omega_n^*.$$

Agree that the *strong bootstrap* holds if

(6) $$P(\bar{Z}_n^* \leq z |\, X_1, \ldots, X_n) \to P(N(0,1) \leq z) \quad \text{for a.e. given sequence of values of } X_1, X_2, \ldots.$$

Theorem 8.1 (Bootstrapping) Consider $\bar{Z}_n^*$ in the iid case.
(i) The weak bootstrap for $\bar{Z}_n^*$ is equivalent to both

(7) $$\mathcal{D}_n \to_p 0 \qquad \text{and/or} \qquad \tilde{\sigma}^2(\cdot) \text{ is slowly varying at zero}$$

and/or any one (hence, all) of (10.5.6)–(10.5.37).
(ii) The strong bootstrap for $\bar{Z}_n^*$ is equivalent to both

(8) $$\mathcal{D}_n \to_{\text{a.s.}} 0 \qquad \text{and/or} \qquad \mathrm{Var}[X_1] = \sigma_F^2 \in (0,\infty)\,.$$

Corollary 1 (i) The weak bootstrap holds for T_n^* whenever $\mathcal{D}_n \to_p 0$.
(ii) The strong bootstrap holds for T_n^* whenever $\mathcal{D}_n \to_{a.s.} 0$.

Proof. Now (10.6.22) and (10.6.18) show that $\mathcal{D}_n \to_p 0$ is equivalent to $\tilde{\sigma}^2(\cdot)$ being slowly varying. (Thus, (7) is a true statement.) Additionally, it is known from the SLLN of theorem 8.4.1 and proposition 6.6.1 (or, from exercise 8.4.20(ii)) that $\mathcal{D}_n \to_{a.s.} 0$ is equivalent to $\mathrm{Var}[X_1] < \infty$. (Thus, (8) is a true statement.)

We next verify normality. Consider (ii). The Liapunov bound of (10.2.5) is

(a) $$\|F_{\bar{Z}_n^*} - \Phi\| \le 8\,\gamma_n/\sqrt{n}\,S_n^3 \le 8\,[\,\max_{1\le k\le n}|X_k - \bar{X}_n|\,]/S_n = 8\,\mathcal{D}_n \to 0$$

for a.e. sample value of $X_1, X_2, \ldots$. Consider (i). Well, $P(\mathcal{D}_n > \epsilon/8) < \epsilon$ for all $n \ge$ (some n_ϵ) means that $P(\,\|F_{\bar{Z}_n^*} - \Phi\| > \epsilon) < \epsilon$ for all $n \ge n_\epsilon$. That is, (5) holds.

Consider the converse of the normality statements. Suppose $\bar{Z}_n^* \to_d N(0,1)$ for a fixed array $(x_1, \ldots, x_n, \ldots)$. The summands are necessarily uan by (1). Thus (10.6.10) (for any $\theta < 1$) is equivalent to $\mathcal{D}_n \to 0$ for this same fixed array (as already noted just above (3)). Thus $\mathcal{D}_n \to_{a.s.} 0$ is implied by the strong bootstrap, and $\mathcal{D}_n \to_p 0$ is implied (by going to subsequences) by the weak bootstrap.

Consider the corollary next. Use (10.6.20) and (6.6.4) to conclude that both

(9) $$S_n^*/S_n \to_{p\times p^*} 1 \qquad \text{if} \quad \mathcal{D}_n \to_p 0 \qquad \text{and}$$

(10) $$S_n^*/S_n \to_{p^*} 1 \quad \text{for a.e. value of } (X_1, X_2, \ldots) \quad \text{if} \quad \mathcal{D}_n \to_{a.s.} 0.$$

These last two results are useful in their own right. □

Exercise 8.1 Establish all the minor details of (9), via theorem C.1.1 (if needed).

9 Bootstrapping with Slowly $\nearrow$ $\tilde{\text{W}}$insorization

Let k_n and k'_n denote *$\tilde{W}$insorization numbers*, with $k_n \wedge k'_n \to \infty$. But suppose the *$\tilde{W}$insorization fractions* $(a_n \vee a'_n) \equiv (k_n \vee k'_n)/n \to 0$. Bootstrapping such a $\tilde{\text{W}}$insorized mean from an iid sample *always* works!

Notation 9.1 Let $K_n \equiv F_n^{-1}$ denote the qf associated with some fixed nondegenerate df F_n. We can always specify that $\text{dom}(a_n, a'_n) \equiv (a_n, 1-a'_n)$ for any $n(a_n \wedge a'_n) \to \infty$. However, if $F_{n-}(0) = 0$, then specifying that $\text{dom}(0, a'_n) \equiv [0, 1-a'_n)$ is preferable; and if $F_n(0) = 1$, then specifying that $\text{dom}(a_n, 0) \equiv (a_n, 1]$ is preferable. So we agree to let $a_n \wedge a'_n$ denote $a_n \wedge a'_n$, or a'_n, or a_n according to the scheme used. Let $\tilde{K}_n(\cdot)$ denote $K_n(\cdot)$ $\tilde{\text{W}}$insorized outside of $\text{dom}(a_n, a'_n)$, and let $\tilde{\mu}_n$ and $\tilde{\sigma}_n^2$ denote the mean and variance of $\tilde{K}_n(\xi)$. Then the rvs $\tilde{X}_{nk} \equiv \tilde{K}_n(\xi_{nk})$ are (*unobservable*) row-independent with qf $\tilde{K}_n$ (when ξ_{nk} are row-independent Uniform$(0,1)$ rvs). Let $\tilde{X}_n \equiv (\tilde{X}_{n1} + \cdots + \tilde{X}_{nn})/n$ and $\tilde{S}_n^2 \equiv \sum_1^n (\tilde{X}_{nk} - \tilde{X}_n)^2/n$. The quantities of primary interest here are

(1) $$\tilde{Z}_n \equiv \sqrt{n}\,[\tilde{X}_n - \tilde{\mu}_n]/\tilde{\sigma}_n \qquad \text{and} \qquad \hat{Z}_n \equiv \sqrt{n}\,[\tilde{X}_n - \tilde{\mu}_n]/\tilde{S}_n\,.$$

Let $\tilde{\gamma}_n \equiv \text{E}|\tilde{X}_{n1} - \tilde{\mu}_n|^3$. Let $\mathcal{F}_n$ denote the collection of all dfs F_n having $\tilde{\sigma}_n > 0$. □

Theorem 9.1 (Universal studentized CLT) Suppose the $\tilde{\text{W}}$insorization fractions satisfy $n(a_n \wedge a'_n) = (k_n \wedge k'_n) \to \infty$. Then uniformly in $\mathcal{F}_n$:

(2) $$\|F_{\tilde{Z}_n} - \Phi\| \le 9\,\tilde{\gamma}_n/\sqrt{n}\,\tilde{\sigma}_n^3 \le 9/\sqrt{n(a_n \wedge a'_n)} = 9/\sqrt{k_n \wedge k'_n} \to 0\,.$$

(3) $$P(|\tilde{S}_n/\tilde{\sigma}_n - 1| \ge \epsilon) \to 0.$$

(4) $$\|F_{\hat{Z}_n} - \Phi\| \to 0.$$

Requiring $(a_n \vee a'_n) \to 0$ guarantees that every nondegenerate df F is eventually in all further $\mathcal{F}_n$. (Recall again that the rvs $\tilde{X}_{nk}$ are *unobservable* rvs.)

Example 9.1 Let all F_n be Bernoulli(10^{-10}). Then n must be huge before $\tilde{\sigma}_n(a_n) > 0$. □

Proof. That $\|F_{\tilde{Z}_n} - \Phi\| \le 9\,\tilde{\gamma}_n/\sqrt{n}\,\tilde{\sigma}_n^3$ is immediate from the Berry–Esseen theorem. Maximizing one power $|\tilde{K}_n(t) - \tilde{\mu}_n|^1$ in the integrand of $\tilde{\gamma}_n$ (but leaving $|\tilde{K}_n(t) - \tilde{\mu}_n|^2$ to integrate) gives

(5) $$\begin{aligned} &\tilde{\gamma}_n/\sqrt{n}\,\tilde{\sigma}_n^3 \le [\,|K^+(1-a'_n) - \tilde{\mu}_n| \vee |\,K_+^-(a_n) - \tilde{\mu}_n|\,]/\sqrt{n}\,\tilde{\sigma}_n \\ &\quad \le 1/\sqrt{n\,(a_n \wedge a'{}_n)} = 1/\sqrt{k_n \wedge k'_n} \to 0\,, \end{aligned}$$

as claimed. Thus (2) holds. Let $q_n \equiv K(1-a'_n) - K_+(a_n)$, as usual. Moreover,

(a) $$\tilde{S}_n^2/\tilde{\sigma}_n^2 = \{\,[\textstyle\sum_1^n (\tilde{X}_{nk} - \tilde{\mu}_n)^2/n]/\tilde{\sigma}_n^2\,\} - \{(\tilde{X}_n - \tilde{\mu}_n)/\tilde{\sigma}_n\}^2 \equiv \{I_{2n}\} - \{I_{1n}\}^2\,,$$

where Chebyshev's inequality gives both $P(|I_{1n}| \ge \epsilon) \le 1/(\epsilon^2\,n) \to 0$ and

(b) $\quad P(\,|I_{2n}-1|\ge\epsilon)\le\frac{1}{n^2\,\epsilon^2}\sum_1^n \mathrm{Var}[\left(\frac{\tilde X_{nk}-\tilde\mu_n}{\tilde\sigma_n}\right)^2-1]\le\frac{\sum_1^n \mathrm{E}[(\tilde X_{nk}-\tilde\mu_n)^4]}{\epsilon^2 n^2\,\tilde\sigma_n^4}$

(c) $\quad \le\frac{1}{\epsilon^2\,n(a_n\wedge a_n')}\cdot\frac{(a_n\wedge a_n')\,q_n^2}{\tilde\sigma_n^2}\le\frac{2}{\epsilon^2\,n(a_n\wedge a_n')}\cdot 1=\frac{2}{\epsilon^2(k_n\wedge k_n')}\to 0$

uniformly in all dfs $F\in\mathcal{F}_n$. Thus (3) holds. Writing

(d) $\quad \hat Z_n=\sqrt{n}\,(\tilde X_n-\tilde\mu_n)/\tilde S_n=\tilde Z_n+\tilde Z_n\,(\tilde\sigma_n/\tilde S_n-1)\,,$

we obtain (4); note that $P(A_{n\epsilon})\equiv P([\,|\tilde Z_n|\le M_\epsilon])$ can be made uniformly small via Chebyshev, even though this set depends on F, while $P(|\tilde\sigma_n/\tilde S_n-1|\ge\epsilon)\to 0$ uniformly in $F\in\mathcal{F}_n$ also (as exhibited in (c)). □

Notation 9.2 Let $\mathsf{X}_n\equiv(X_{n1},\ldots,X_{nn})'$ denote an iid sample from the qf K_n, and let $\bar X_n$, S_n^2, G_n, and $\mathbb{K}_n(\cdot)$ denote its sample mean, sample variance, sample third absolute central moment, and sample qf. Let $\tilde{\mathsf{X}}_n\equiv(\tilde X_{n1},\ldots,\tilde X_{nn})'$ denote the (k_n,k_n')-$\tilde{\text{W}}$insorized sample, for integers k_n and k_n' having $k_n\wedge k_n'$ going to ∞ (here $k_n\wedge k_n'$ will denote either $k_n\wedge k_n'$, or k_n', or k_n as in the scheme of notation 9.1). Let $a_n\equiv k_n/n$ and $a_n'\equiv k_n'/n$. Let $\tilde X_n$, $\tilde S_n$, $\tilde G_n$, and $\tilde{\mathbb{K}}_n$ denote the sample mean, sample variance, sample third central moment, and sample qf of the population $\tilde{\mathsf{X}}_n$. Let $\vec{X}_n^*\equiv(X_{n1}^*,\ldots,X_{nn}^*)'$ denote the iid bootstrap sample from $\tilde{\mathbb{K}}_n(\cdot)$, and let X_n^* and $\widetilde{S_n^{*2}}$ be the sample mean and sample variance of the bootstrap sample X_n^*. Let $\mathbb{P}_n^*$ denote the bootstrap probability distribution. Our rvs of interest are

(6) $\quad \widetilde{Z_n^*}\equiv\sqrt{n}\,[\,\widetilde{X_n^*}-\tilde X_n]/\tilde S_n \qquad\text{and}\qquad \widehat{Z_n^*}\equiv\sqrt{n}\,[\,\widetilde{X_n^*}-\tilde X_n]/\widetilde{S_n^*}\,.$

[We saw in the previous section that the sample mean and sample variance $\overline{X_n^*}$ and S_n^{*2} of an iid bootstrap sample from $\mathbb{K}_n$ are such that $\bar Z_n\equiv\sqrt{n}\,[\,\overline{X_n^*}-\bar X_n]/S_n$ satisfies the strong (or the weak) bootstrap if and only if $\mathrm{Var}[X]\in(0,\infty)$ (or $F\in\mathcal{D}(\text{Normal})$). But next we see the glories of Winsorizing! Winsorizing *does do* what Winsorizing was supposed to do. The bootstrap *always works*, provided that we Winsorize just a little bit.] □

Theorem 9.2 (Universal bootstrap CLT) Suppose the $\tilde{\text{W}}$insorization fractions are such that $(k_n\wedge k_n')=n(a_n\wedge a_n')\to\infty$ in the context of notation 9.2. Then uniformly in all X_n for which $\tilde S_n>0$ we have that for a.e. X_n, conditional on X_n,

(7) $\quad \|F_{\widetilde{Z_n^*}}-\Phi\|\le 9\,\tilde G_n/\sqrt{n}\,\tilde S_n^3\le 9/\sqrt{n(a_n\wedge a_n')}=9/\sqrt{k_n\wedge k_n'}\to 0\,,$

(8) $\quad \mathbb{P}_n^*(|\widetilde{S_n^*}-\tilde S_n|/\tilde S_n\ge\epsilon\,\Big|\,\mathsf{X}_n\,)\to 0\,,$

(9) $\quad \|F_{\widehat{Z_n^*}}-\Phi\|\to 0,$

(10) $\quad \overline{\lim}\,\tilde S_n>0$ if we also specify that $(a_n\vee a_n')\to 0\,,$ with F nondegenerate.

Proof. This is immediate from the previous theorem. □

Remark 9.1 If we knew how to Winsorize correctly in theorem 9.1, it would be a useful theorem. The point is, we do always know how to Winsorize correctly in the bootstrap of theorem 9.2.

But should we instead do bootstrap sampling from the empirical qf $\mathbb{K}_n$ itself, rather than $\tilde{\mathbb{K}}_n$, and then Winsorize this sample? No! Sampling from $\tilde{\mathbb{K}}_n$ gives us the analog of theorem 9.1, while sampling from $\mathbb{K}_n$ (it can be shown) does not. (Sampling from $\mathbb{K}_n$ could, however, be shown to work for any qf K in a very large class of distributions.) □

Exercise 9.1 Let $\xi_1, \xi_2, \ldots$ be iid Uniform$(0,1)$ rvs. Let $X_{nk} \equiv \check{F}_n^{-1}(\xi_k)$ with

$$\check{F}_n^{-1}(t) \equiv -\sqrt{t \vee (1/n)}^{\,-1}\, 1_{(0,1/2)}(t) + \sqrt{(1-t) \vee (1/n)}^{\,-1}\, 1_{[1/2,1)}(t).$$

Let $V_n \equiv \mathrm{Var}[X_{nk}]$, and let $\bar{X}_n \equiv \sum_{k=1}^n X_{nk}/n$. Compute V_n, as well as the higher moments $\mathrm{E}|X_{nk}|^3$ and $\mathrm{E}X_{nk}^4$.

(a) Show that $Z_n \equiv \sqrt{n}\,\bar{X}_n/\sqrt{V_n} \to_d N(0,1)$ *by verifying* the Lindeberg condition.

(b) What conclusion does the Berry–Esseen theorem imply for Z_n?

(c) Show that $\overline{X_n^2} \equiv \sum_{k=1}^n X_{nk}^2/n$ satisfies $\overline{X_n^2}/V_n \to_p 1$.

(d) Of course, this immediately implies that $T_n \equiv \sqrt{n}\,\bar{X}_n/(\,\overline{X_n^2}\,)^{1/2} \to_d N(0,1)$.

(e) Show that $(\mathrm{E}|X_{nk}|)^2/V_n \to 0$.

Exercise 9.2 Formulate and solve another example in the spirit of exercise 9.1.

Exercise 9.3 Verify that $S_n^2 \equiv \sum_{k=1}^n (X_{nk} - \bar{X}_n)^2/(n-1)$ satisfies $S_n^2/\tilde{\sigma}_n^2 \to_p 1$, and do it *by verifying* a Lindeberg type condition in the context of theorem 10.5.1.

Chapter 11

Infinitely Divisible and Stable Distributions

1 Infinitely Divisible Distributions

Definition 1.1 (Triangular arrays, and the uan condition) A *triangular array* is just a collection of rvs $X_{n1}, \ldots, X_{nn}$, $n \geq 1$, such that the rvs in the nth row are independent. Call it a *uan array* if the *uniform asymptotic negligibility* condition holds, that is

$$(1) \qquad \max_{1 \leq k \leq n} P(\,|X_{nk}| \geq \epsilon\,) \to 0 \qquad \text{as } n \to \infty \quad \text{for all } \epsilon > 0.$$

The uan condition is a natural one for preventing one term from dominating the whole sum.

The Problem: Let $S_n \equiv X_{n1} + \cdots + X_{nn}$ denote the nth row sum of a uan array.

(i) Find the family of all possible limit laws of S_n.

(ii) Find conditions for convergence to a specified law of this form.
Find specialized results for further restrictions on the uan array.

- (a) Suppose variances exist.
- (b) Suppose the limit law is normal or Poisson.
- (c) Consider $S_n = [(X_1 + \cdots + X_n) - B_n]/A_n$ for a singly subscripted sequence of iid rvs $X_1, \ldots, X_n, \ldots$.

Some of the results in this chapter are stated with only indications of the proofs. The goal in this chapter is simply to develop some rough understanding of the subject. We will see in this section that the set of all possible limit laws of row sums $S_n \equiv X_{n1} + \cdots + X_{nn}$ of a uan array of X_{nk}'s is exactly the class of infinitely divisible laws, which we now define. □

Definition 1.2 (Infinitely divisible) Call both the rv Y and its distribution *infinitely divisible* (*id*) if for every value of n it is possible to decompose Y into n iid components as

(2) $$Y \cong Y_{n1} + \cdots + Y_{nn} \qquad \text{for some iid rvs } Y_{n1}, \cdots, Y_{nn}\,.$$

We denote the class of all id distributions by $\mathcal{I}$; the subclass with finite variance is denoted by $\mathcal{I}_2$. (We remark that the Y_{ni}'s of this definition form a uan array, but this needs to be shown; note exercise 1.2.)

Exercise 1.1 (Chf expansions for the uan array $X_{n1} + \cdots + X_{nn}$) Consider a uan array of rvs X_{nk}, as in (1).
(a) Let F_{nk} and ϕ_{nk} denote the df and the chf of X_{nk}. Show that

(3) $$[\max_{1\le k\le n} |\phi_{nk}(t) - 1|] \to 0 \qquad \text{uniformly on every finite interval.}$$

[Hint. Integrate over $|x| < \epsilon$ and $|x| \ge \epsilon$ separately to obtain

$$|\phi_{nk}(t) - 1| \le \delta_\epsilon + 2P(|X_{nk}| \ge \epsilon),$$

from which point the result is minor.] We then define (as will be useful regarding an expansion of Log $(1 + (\phi_{nk}(\cdot) - 1))$ below)

(4) $$\epsilon_n(t) \equiv \sum_{k=1}^n |\phi_{nk}(t) - 1|^2.$$

(b) Verify the elementary fact that if $X_{nk} \cong (0, \sigma_{nk}^2)$, and $\sigma_n^2 \equiv \sum_1^n \sigma_{nk}^2 \le M < \infty$ with $[\max_{1\le k\le n} \sigma_{nk}^2] \to 0$, then $\epsilon_n(t) \to 0$ uniformly on each finite interval.

Exercise 1.2 (Chf expansions for the uan array $Y \cong Y_{n1} + \cdots + Y_{nn}$)
(a) If ϕ is id, then $\phi(t) \ne 0$ for any t.
(b) Let ϕ and ϕ_n denote the chf of Y and of the Y_{nk}'s, respectively, in (2). Show that these $Y_{n1}, \ldots, Y_{nn}$ form a uan array.
(c)* If $Y_m \to_d Y$ for id rvs Y_m, then Y is id. (Or, give an "approximate proof.")

Motivation 1.1 (Limits of uan arrays) Suppose that $S_n \equiv X_{n1} + \cdots + X_{nn}$ for some row independent uan array. Let F_{nk} and ϕ_{nk} denote the df and chf of the rv X_{nk}. Then for the function $\epsilon_n(\cdot)$ of (4) we necessarily have for n sufficiently large (recall that $a = b \oplus c$ means that $|a - b| \le c$) that

$$\text{Log } \phi_{S_n}(t) = \sum_1^n \text{Log } \phi_{nk}(t) = \sum_1^n \text{Log } (1 + (\phi_{nk}(t) - 1))$$

$$= \sum_1^n [\phi_{nk}(t) - 1] \oplus \epsilon_n(t) \qquad \text{with } \epsilon_n(\cdot) \text{ as in (4)}$$

(5) $$= \sum_{k=1}^n \int_{-\infty}^{\infty} (e^{itx} - 1)\, dF_{nk}(x) \oplus \epsilon_n(t).$$

If we further assume that all the X_{nk}'s have 0 means and finite variances, then we can rewrite (5) to obtain

$$\text{Log } \phi_{S_n}(t) = \sum_{k=1}^n \int_{-\infty}^{\infty} \left[\frac{e^{itx} - 1 - itx}{x^2}\right] \left[x^2\, dF_{nk}(x)\right] \oplus \epsilon_n(t)\,.$$

Thus

(6)
$$\operatorname{Log}\ \phi_{S_n}(t) = \int_{-\infty}^{\infty} \{ \tfrac{e^{itx}-1-itx}{x^2} \} \, [x^2 \textstyle\sum_1^n \, dF_{nk}(x)] \ \oplus \ \epsilon_n(t)$$
$$\equiv \int_{-\infty}^{\infty} \{ \phi(x,t) \} \, [dK_n(x)] \ \oplus \ \epsilon_n(t) \, .$$

Observe additionally that

(7) $\phi(x,t)$ is continuous on the (x,t)-plane with
each $\phi(\cdot,t)$ bounded and continuous, and equal to $-t^2/2$ at $x=0$.

Moreover, take the point of view that

(8) $K_n(x) = \int_{-\infty}^{x} y^2 \, d\sum_{k=1}^n F_{nk}(y) =$ (the contribution to variance up to x).

It is natural to hope that S_n will converge in distribution to a rv Y whose Log chf is of the form

(9) $\operatorname{Log}\ \phi_Y(t) \equiv \int_{-\infty}^{\infty} \phi(x,t) \, dK(t) \, ,$

provided that $K_n \to_d K$.

When means and variances need not exist, we define

(10) $\alpha(x) \equiv x^2 \wedge 1$ and $\beta(x) \equiv (|x| \wedge 1)\ \operatorname{sign}(x)$.

We then note that (5) can also be manipulated to give

$$\operatorname{Log}\ \phi_{S_n}(t) = \int_{-\infty}^{\infty} \left[\frac{(e^{itx} - 1 - it\beta(x))}{\alpha(x)} \right] \left[\alpha(x) \, d \sum_{k=1}^{n} F_{nk}(x) \right]$$
$$+ \ it \ [\textstyle\int_{-\infty}^{\infty} \beta(x) d \sum_{k=1}^n F_{nk}(x)] \ \oplus \epsilon_n(x)$$

(11)
$$\equiv \int_{-\infty}^{\infty} \phi(x,t) \, dH_n(x) + it \, \beta_n \ \oplus \ \epsilon_n(x),$$

where this new $\phi(x,t)$ still satisfies (7), and where we now define

(12) $H_n(x) = \int_{-\infty}^{x} \alpha(u) \, d\sum_{k=1}^n F_{nk}(u) \doteq$ (contribution to variance up to x),

at least for x near 0, which is where all the action is in any uan array. It is natural to hope that S_n will now converge in distribution to a rv Y whose chf is of the form

(13) $\operatorname{Log}\ \phi_Y(t) \equiv it\beta + \int_{-\infty}^{\infty} \phi(x,t) \, dH(x),$

provided that $H_n \to_d H$ and $\beta_n \to \beta$.

We are thus particularly interested in the behavior of K_n and/or H_n (and $\epsilon_n(\cdot)$), both for the general uan X_{nk}'s of (1) and for the special uan Y_{nk}'s of (2). □

The next example enables us to show that any chf of the form (9) or (13) is the chf of some id distribution. The details are left to the easy exercise 1.3.

Example 1.1 (Generalized Poisson and compound Poisson distributions) We suppose that the rvs $X_{n1}, \ldots, X_{nn}$ are iid with

$$X_{nk} = \begin{cases} a_j & \text{with probability } p_{nj}, \quad \text{for } 1 \le j \le J, \\ 0 & \text{otherwise,} \end{cases}$$

where $np_{nj} \to \lambda_j \in (0, \infty)$ as $n \to \infty$. Then

$$\text{(14)} \qquad \phi_{S_n}(t) \to \phi_Y(t) \equiv \exp(\textstyle\sum_{j=1}^J \lambda_j (e^{ita_j} - 1)) = \prod_{j=1}^J \exp(\lambda_j (e^{ita_j} - 1)).$$

Thus the limiting distribution is that of $Y \equiv \sum_{j=1}^J a_j Y_j$ for independent Poisson(λ_j) rvs Y_j. This is called the *generalized Poisson distribution*.

Note also that the chf on the right-hand side of (14) satisfies

$$\text{(15)} \qquad \phi_Y(t) = \exp(\lambda \textstyle\sum_{j=1}^J p_j (e^{ita_j} - 1)) = \exp(\lambda(\phi_W(t) - 1)) = \mathrm{E} e^{it(W_1 + \cdots + W_N)},$$

where $\lambda \equiv \sum_{j=1}^J \lambda_j$, $p_j \equiv P(W = a_j) = \lambda_j / \lambda$ for $1 \le j \le J$, ϕ_W is the chf of W, and $W_1, W_2, \ldots$ are iid as W and $N \cong$ Poisson(λ). The distribution of the rv Y is called the *compound Poisson distribution*, and is distributed as a Poisson sum of independent Multinomial($k; \lambda_1/\lambda, \ldots, \lambda_k/\lambda$) rvs.

The compound Poisson distribution of (15) is obviously id, as is clearly seen by using λ/n in place of λ in (15) for the iid Y_{nk}'s of (2). And thus the generalized Poisson distribution in (14) is also id. It is in the compound Poisson format that we recognize that this distribution is id, but it will be in its generalized Poisson format that we will put it to work for us. □

Exercise 1.3 Our heuristics have suggested that if Y is id, then its chf ϕ is of the form (13) (the reader should also note (29), where the normal component of the limit is removed). They have not yet suggested the converse. However,

$$i\beta t + \int_{-\infty}^{\infty} \phi(x,t)\, d\nu(x) \leftarrow \textstyle\sum_{j=1}^{m_n} \{[e^{itx_j} - 1 - it\beta(x_j)]/\alpha(x_j)\}\nu(I_j) + i\beta t$$

$$\text{(16)} \qquad = \textstyle\sum_{j=1}^{m_n} (e^{itx_j} - 1)\, \lambda_j + it\beta_n$$

with $\lambda_j \equiv [\nu(I_j)/\alpha(x_j)]$ and $\beta_n \equiv \beta - \sum_{j=1}^{m_n} \nu(I_j)\beta(x_j)/\alpha(x_j)$

and appropriate intervals I_j (with x_j, I_j and λ_j all depending on n)

$\to$ (a limit of generalized Poisson rvs)

(17) $\quad$ = (a limit of id rvs) = id,

since the limit under $\to_d$ of id rvs is also an id rv (as was stated in exercise 1.2(c)). [The present exercise is to make all this rigorous.]

Theorem 1.1 (Kolmogorov's representation theorem) Let the rv Y have chf ϕ and finite variance. We will use the symbol K to denote a generalized df with $0 = K(-\infty) < K(+\infty) < \infty$. Now, Y is id if and only if

$$\text{(18)} \qquad \operatorname{Log} \phi(t) = i\beta t + \int_{-\infty}^{\infty} \phi(x,t)\, dK(x) \quad \text{for some such gdf } K \text{ and some real } \beta,$$

with $\phi(x,t)$ as in (6) (and (7)). The representation is unique. Moreover, it holds that $\beta = \mathrm{E}Y$ and $K(+\infty) = \mathrm{Var}[Y]$.

Theorem 1.2 (Bounded variances limit theorem for $\mathcal{I}_2$) (a) We start with a triangular array of row-independent rvs $X_{nk} \cong (\mu_{nk}, \sigma_{nk}^2)$ having

(19) $$\max_{1 \le k \le n} \sigma_{nk}^2 \to 0 \qquad \text{and} \qquad \sigma_n^2 \equiv \textstyle\sum_{k=1}^n \sigma_{nk}^2 \le M < \infty$$

(such a triangular array is necessarily uan). Then

(20) $$S_n \equiv \sum_{k=1}^n X_{nk} \to_d Y,$$

where the limiting rv necessarily satisfies

(21) $$\text{Log } \phi_Y(t) = i\beta t + \int_{-\infty}^{\infty} \left[\frac{e^{itx} - 1 - itx}{x^2}\right] dK(x) = it\beta + \int_{-\infty}^{\infty} \phi(x,t)\, dK(t)$$

with $K(+\infty) \le \limsup \sigma_n^2$, if and only if

(22) $$\mu_n \equiv \sum_{k=1}^n \mu_{nk} \to \beta \qquad \text{and} \qquad K_n(\cdot) \equiv \sum_{k=1}^n \int_{-\infty}^{\cdot} y^2\, dF_{nk}(y + \mu_{nk}) \to_{sd} K(\cdot).$$

If $K_n(\infty) = \sigma_n^2 \le M < \infty$ is replaced by $K_n(+\infty) = \sigma_n^2 \to \text{Var}[Y] < \infty$, then we can claim that $\text{Var}[Y] = K(+\infty) = \lim \sigma_n^2$, and we will write $K_n \to_d K$.
(b) The family of all possible limit laws of such S_n is the family $\mathcal{I}_2$ of all possible infinitely divisible laws that have finite variance.

Proof. (If one grants exercises 1.1 and 1.3, then the proof we give will be complete. These exercises are straightforward.) Any chf of the form (18) or (21) is id, by exercise 1.3. Differentiating twice in (21) shows that this chf has mean β and variance $\int_{-\infty}^{\infty} 1\, dK = K(+\infty)$. For the uniqueness of the representation, just differentiate (21) twice, and thus obtain $-(\text{Log } \phi(t))'' = \int e^{itx}\, dK(x)$ subject to $0 = K(-\infty) < K(+\infty) = \text{Var}[Y]$; then applying the ordinary inversion formula to $\int e^{itx}\, dK(x)$ gives K in terms of $-(\text{Log } \phi)''$.

It remains to show that any id Y with mean 0 and finite variance has a chf of the form (21) with $\beta = 0$. Reconsider (6). For the special uan Y_{nk}'s of (2) we have $\text{Var}[Y_{nk}] = \text{Var}[Y]/n$, so that exercise 1.1(b) implies that $\epsilon_n(t) \to 0$ uniformly on all finite intervals. Moreover, the family of functions $\phi(\cdot, t)$ in (7) are bounded and continuous functions that converge to 0 as $|x| \to \infty$. Applying the Helly–Bray exercise 12.1.1 to each $\phi(\cdot, t)$ in (6) shows (for the first equality we need only recall that $Y \cong T_n \equiv Y_{n1} + \cdots + Y_{nn}$) that

(a) $$\text{Log } \phi_Y(t) = \text{Log } \phi_{T_n}(t) \to \textstyle\int \phi(x,t)\, dK(x), \qquad \text{provided that } K_n \to_{sd} K.$$

So we must show that $K_n \to_{sd} K$. Now, every subsequence n' has a further n'' on which $K_{n''} \to_{sd}$ (some K). But every resulting such $\int \phi(x,t)\, dK(x)$ with a limiting K inserted must equal Log ϕ_Y, and thus the first paragraph of this proof implies that K has $0 = K(-\infty) < K(+\infty) = \text{Var}[Y]$. In fact, all possible subsequential limits K must be equal, by the uniqueness of the representation in equation (21). Thus $K_n \to_{sd} K$ on the whole sequence n. Thus Log $\phi_Y = \int \phi(x,t)\, dK(t)$. This completes the proof of theorem 1.1.

We now turn to the proof of theorem 1.2. Under the basic hypothesis (19), we have in (6) that $\epsilon_n(t) \to 0$ uniformly on all finite intervals (by exercise 1.1(b)). Thus whenever $K_n \to_{sd}$ (some K) and $\mu_n \to \mu$, we have by applying Helly–Bray to each $\phi(\cdot, t)$ in (7) that Log $\phi_{S_n}(t) \to$ Log $\phi(t)$ for each t, for the ϕ of (21) with $\beta = \mu$. Thus $S_n \to_d Y$ for the id Y with chf given by (21).

Suppose $S_n \to_d Y$. We argue (as in the proof of theorem 1.1) that each subsequence n' has a further n'' on which the K_n of (6) satisfies $K_{n''} \to_{sd}$ (the same K) and $\mu_{n''} \to$ (the same μ). Thus $K_n \to_{sd} K$ and $\mu_n \to \mu$, using theorem 1.1 for this uniqueness. That is, Log $\phi_{S_n}(t) \to$ Log $\phi(t)$ for the ϕ of (21) having this K and μ. But Log $\phi_{S_n}(t) \to$ Log $\phi_Y(t)$ also. Thus Log $\phi_Y =$ Log ϕ. Moreover, $K(+\infty) = \mathrm{Var}[Y] \le \liminf \mathrm{Var}[S_n] = \liminf \sum_1^n \sigma_{nk}^2 \le M$ (using Fatou and a Skorokhod representation for which $S_n^* \to_{a.s.} Y^*$). If

(b) $\qquad K_n(+\infty) = \mathrm{Var}[S_n] = \sigma_n^2 \to \mathrm{Var}[Y] = K(+\infty)\,,$

then $K_n \to_{sd} K$ reasonably becomes $K_n \to_d K$. □

Example 1.2 (Normal convergence)
(i) (Representation) The $N(0,1)$ chf ϕ has

$$\mathrm{Log}\ \phi(t) = -t^2/2.$$

Thus $\mu = 0$ and $K = 1_{[0,\infty)}$.

(ii) (Lindeberg–Feller theorem) Suppose the triangular array with $X_{nk} \cong (\mu_{nk}, \sigma_{nk}^2)$ satisfies $\mu_{nk} = 0$ and $\sigma_n^2 \equiv \sum \sigma_{nk}^2 = 1$ for all n. Then

(23) $\qquad S_n \equiv \sum_{k=1}^n X_{nk} \to_d N(0,1) \qquad$ and $\qquad [\max_{1\le k\le n} \sigma_{nk}^2] \to 0$

if and only if

(24) $\qquad \sum_{k=1}^n \int_{[|x|\ge\epsilon]} x^2\, dF_{nk}(x) \to 0 \qquad$ for all $\epsilon > 0$. □

Exercise 1.4 Verify example 1.2.

Example 1.3 (Poisson convergence)
(i) (Representation) The Poisson(λ) chf ϕ has

$$\mathrm{Log}\ \phi(t) = \lambda(e^{it} - 1) = it\lambda + \lambda(e^{it} - 1 - it).$$

Thus $\mu = \lambda$ and $K = \lambda 1_{[1,\infty)}$.

(ii) (Convergence) Suppose the triangular array with $X_{nk} \cong (\mu_{nk}, \sigma_{nk}^2)$ satisfies $[\max_{1\le k\le n} \sigma_{nk}^2] \to 0$, and $\sum_1^n \sigma_{nk}^2 \to \lambda \in (0,\infty)$. Then

(25) $\qquad S_n \equiv \sum_1^n X_{nk} \to_d$ Poisson(λ)

if and only if

(26) $\qquad \sum_{k=1}^n \mu_{nk} \to \lambda \qquad$ and $\qquad \sum_{k=1}^n \int_{[|x-1|\ge\epsilon]} x^2\, dF_{nk}(x + \mu_{nk}) \to 0.$ □

Exercise 1.5 Verify example 1.3.

Exercise 1.6 (Decomposition of normal and Poisson distributions) Suppose that $X \cong X_1 + X_2$, where X_1 and X_2 are independent $\mathcal{I}_2$ rvs. Then:

(27) X normal implies that X_1 and X_2 are both normal.

(28) X Poisson implies that X_1 and X_2 are both Poisson.

[This is also true if $\mathcal{I}$ replaces $\mathcal{I}_2$.]

Exercise 1.7 If ϕ is a chf, then $\exp(c(\phi - 1))$ is an id chf for all $c > 0$. Thus any chf such as the one in (15) represents the log of an id chf. (We do not make explicit use of this anywhere.)

From here to the end of this section we mainly just state results, mostly by analogy.

Theorem 1.3 (Lévy–Khinchin representation theorem) Let Y have chf ϕ. Then Y is infinitely divisible (id) if and only if

(29) $$\text{Log } \phi(t) = i\beta t + \int_{-\infty}^{\infty} \phi(x,t)\, dH(x) = i\beta t - \sigma^2 t^2/2 + \int_{-\infty}^{\infty} \phi(x,t)\, d\nu(x),$$

$$\text{where } \sigma^2 = \Delta H(0) \text{ and } \nu \equiv H - \sigma^2 1_{[0,\infty)},$$

for $\phi(x,t)$ as in (11) (and (7)). The representation is unique. (We write $Y =_r (\beta, H)$ to denote this representation. We will think of $i\beta t - \sigma^2 t^2/2$ as the normal component of the limit law.)

Theorem 1.4 (General limit theorem for $\mathcal{I}$) Let the rv's X_{nk} form a uan triangular array. Then

(30) $$S_n \equiv \sum_{k=1}^{n} X_{nk} = \sum_{k=1}^{n} (X_{nk} - b_{nk}) + \sum_{k=1}^{n} b_{nk} \to_d Y,$$

where necessarily

(31) $$\text{Log } \phi_Y(t) = i\beta t + \int_{-\infty}^{\infty} [(e^{itx} - 1 - it\beta(x))/\alpha(x)]\, dH(x),$$

if and only if for some finite-measure generalized df H and for some real β we have

(32) $$\beta_n \to \beta \qquad \text{and} \qquad H_n \to_d H$$

where

(33) $$\beta_n \equiv \sum_{k=1}^{n} \left[b_{nk} + \int_{-\infty}^{\infty} \beta(x)\, dF_{nk}(x + b_{nk})\right] \qquad \text{and}$$
$$H_n(\cdot) = \sum_{k=1}^{n} \int_{-\infty}^{\cdot} \alpha(y)\, dF_{nk}(y + b_{nk}),$$

and where $b_{nk} \equiv \mathrm{E} X_{nk} 1_{(-\delta,\delta)}(X_{nk})$, with $\delta > 0$ arbitrary but fixed. The family of possible limit laws of such S_n is the family $\mathcal{I}$ of all possible infinitely divisible distributions.

Theorem 1.5 (Normal limits of uan arrays)
(a) Let $S_n \equiv X_{n1} + \cdots + X_{nn}$ for iid X_{nk}'s, and suppose that $S_n \to_d S$. Then

(34) $\quad M_n \to_p 0 \qquad$ if and only if $\qquad S$ is normal.

(b) Let $S_n \equiv X_{n1} + \cdots + X_{nn}$ for independent X_{nk}'s, and suppose $S_n \to_d S$. Then

(35) $\quad M_n \to_p 0 \qquad$ if and only if $\qquad S$ is normal and the X_{nk}'s are uan.

(Compare this to (10.5.5).)

Exercise 1.8 (a) Show that all Gamma(r, ν) rvs are infinitely divisible.
(b) Show that an infinitely divisible rv can not be concentrated on a finite interval.

Exercise 1.9 Use theorem 1.4 to prove the asymptotic normality condition of theorem 10.5.3.

Exercise 1.10 Use theorem 1.4 to prove theorem 1.5.

Exercise 1.11 Use chfs to determine necessary and sufficient conditions under which the WLLN will hold.

2 Stable Distributions

Definition 2.1 (Domain of attraction) Let $X, X_1, X_2, \ldots$ be iid F. Suppose an $A_n > 0$ and a B_n exist such that $S_n \equiv X_1 + \cdots + X_n$ satisfies

(1) $$(S_n - B_n)/A_n \to_d Y \cong G.$$

Then F is said to belong to the *domain of attraction* of G, and we write $F \in \mathcal{D}(G)$. We also say that G possesses a domain of attraction. [If $\mathrm{Var}[X] < \infty$ above, then necessarily $(S_n - n\,\mathrm{E}X)/\sqrt{n} \to_d N(0, \mathrm{Var}[X])$ by the ordinary CLT; thus the only new and interesting cases have $\mathrm{E}X^2 = \infty$.]

Definition 2.2 (Stable law) Call a df G *stable* if for all n there exist constants $a_n > 0$ and b_n with

(2) $$S_n \cong a_n X + b_n, \qquad \text{where } X, X_1, \ldots, X_n \text{ are iid as } G.$$

We call G *strictly stable* if we may take all $b_n = 0$ in (2).

Theorem 2.1 (Only stable dfs have domains of attraction) A df G will possess a domain of attraction if and only if G has a stable distribution. Moreover, the a_n of (2) must satisfy

(3) $$a_n = n^{1/\alpha} \qquad \text{for some } 0 < \alpha \le 2.$$

We call α the *characteristic exponent* of G. [To compare with section **??.??** results for $\mathcal{R}_{-\beta}$ with $\beta > 0$, we define

(4) $$\beta \equiv (2/\alpha) - 1 \qquad \text{or} \qquad \alpha = 2/(\beta + 1),$$

where $0 < \alpha \le 2$ and $0 \le \beta < \infty$.]

Definition 2.3 (Basic domain of attraction) The df F is said to belong to the *basic domain of attraction* of G (or the *domain of normal attraction* of G), which is denoted by writing $F \in \mathcal{D}_N(G)$, provided $A_n = (\text{constant}) \times n^{1/\alpha}$ works in (1).

Remark 2.1 (a) Suppose that G_1 and G_2 are both of the same type (in the sense of definition 9.1.4). Then:

(5) $$\begin{aligned} &\mathcal{D}(G_1) = \mathcal{D}(G_2) \qquad \text{and} \qquad \mathcal{D}_N(G_1) = \mathcal{D}_N(G_2), \\ &G \in \mathcal{D}(G) \quad \text{if } G \text{ is stable.} \end{aligned}$$

(Thus there is a domain of attraction for the normal type, but this is not so for a particular normal df.) □

Proof. Suppose G is stable. Then $S_n \cong a_n X + b_n$, or $(S_n - b_n)/a_n \cong X \cong G$ for all n. Thus $G \in \mathcal{D}(G)$.

Suppose G possesses a domain of attraction. Thus there exists $X_1, X_2, \ldots$ iid F where $F \in \mathcal{D}(G)$. Hence for some $A_n > 0$ and some B_n we have

(a) $$T_n \equiv (X_1 + \cdots + X_n - B_n)/A_n \to_d Y \cong G.$$

Replace n by nk in (a) and obtain

$$T_{nk} = \frac{X_1 + \cdots + X_{nk} - B_{nk}}{A_{nk}}$$

(b) $$= \left[\frac{(X_1 + \cdots + X_n) - B_n}{A_n} + \cdots + \frac{(X_{n(k-1)+1} + \cdots + X_{nk}) - B_n}{A_n}\right] \frac{A_n}{A_{nk}}$$

$$-(B_{nk} - k\,B_n)/\,A_{nk}\,,$$

which can be rewritten in the more useful format

$$\frac{A_{nk}}{A_n}\,T_{nk} + \left[\frac{B_{nk} - k\,B_n}{A_n}\right]$$

$$= \left[\frac{X_1 + \cdots + X_n - B_n}{A_n}\right] + \cdots + \left[\frac{X_{n(k-1)+1} + \cdots + X_{nk} - B_n}{A_n}\right]$$

(c) $$\to Y_1 + \cdots + Y_k \qquad \text{for } Y_1, \ldots, Y_k \text{ iid as } G.$$

Let k be fixed. Recall that the convergence of types theorem states that if

(d) $$T_{nk} \to_d Y \quad \text{(true, from (a))} \qquad \text{and also}$$
$$a_{nk}\,T_{nk} + b_{nk} \to_d Y_1 + \cdots + Y_k \qquad \text{(true, from (c))},$$

with $a_{nk} \equiv A_{nk}/A_n$ and $b_{nk} \equiv (B_{nk} - k\,B_n)/A_n$, then it must be that

(e) $$a_{nk} \to (\text{some } a_k) \in (0,\infty), \qquad b_{nk} \to (\text{some } b_k) \in (-\infty,\infty), \qquad \text{where}$$

(f) $$Y_1 + \cdots + Y_k \cong a_k\,Y + b_k.$$

From (f), we see that G is stable. This completes the proof of the first statement.

Now we further exploit equation (e). From it we determine that

(g) $$a_{mk} = \lim_{n\to\infty} \frac{A_{nmk}}{A_n} = \lim_{n\to\infty} \frac{A_{nmk}}{A_{nm}}\,\frac{A_{nm}}{A_n} = a_k a_m \qquad \text{for all } m, k \geq 1.$$

We now let $Z \equiv Y - Y'$ and $Z_n \equiv Y_n - Y_n'$, where $Y, Y', Y_1, Y_1', \ldots$ are iid as G. Then the rvs Z and Z_n are symmetric, and (1) shows that

(h) $$Z_1 + \cdots + Z_n \cong a_n Z \qquad \text{for the same } a_n\text{'s as in (2).}$$

Thus for some $x > 0$ fixed we have

$$P(Z > (a_m/a_{m+n})\,x) = P(a_{m+n}Z > a_m x)$$

$$= P(Z_1 + \cdots + Z_{m+n} > a_m x) \qquad \text{by (h)}$$

$$= P((Z_1 + \cdots + Z_m) + (Z_{m+1} + \cdots + Z_{m+n}) > a_m x)$$

$$= P(a_m Z_1 + a_n Z_2 > a_m x) \qquad \text{by (h)}$$

$$\geq P(a_n Z_2 \geq 0 \text{ and } a_m Z_1 > a_m x) = P(Z_2 \geq 0)P(Z_1 > x)$$

(i) $$\geq (\text{some } \delta) > 0 \quad \text{for all } m \text{ and } n \qquad (\text{by choice of } x).$$

Thus a_m/a_{m+n} stays bounded away from ∞ for all m and n. Letting $m = k^I$ and $m+n = (k+1)^I$, we have from (h) that

$$(a_k/a_{k+1})^I = a_{k^I}/a_{(k+1)^I} = a_m/a_{m+n}$$

(j) $$\leq (\text{ some } M) < \infty \qquad \text{for all } k \text{ and } I.$$

Thus $(a_k/a_{k+1}) \leq 1$ for all k; that is,

(k) $$a_n \text{ is } \nearrow, \qquad \text{where} \quad a_{mk} = a_m\, a_k \quad \text{for all } m, k \geq 1,$$

was shown in (a). Thus exercise 2.1 below shows that $a_n = n^{1/\alpha}$ for some $\alpha > 0$. Suppose that $\alpha > 2$. Then $\text{Var}[Y] < \infty$ by exercise 2.2 below. We can thus claim that $\sqrt{n}(\bar{Y} - \mu) \to_d N(0,1)$ by the ordinary CLT. Thus $A_n = \sqrt{n}$ and $a_k = \sqrt{k}$ work above. By the convergence of types theorem, there are no other choices. That is, when $\alpha > 2$, then only $\alpha = 2$ works. (See Breiman (1968, p. 202).) □

Exercise 2.1 Suppose that $a_n \nearrow$ with $a_1 = 1$, and suppose that $a_{mk} = a_m a_k$ for all $k, m \geq 1$. Then show that necessarily $a_n = n^{1/\alpha}$ for some $\alpha \geq 0$.

Exercise 2.2 (Moments) Suppose that $Y \cong G$ is stable with characteristic exponent α. Then

(6) $$\text{E}|Y|^r < \infty \text{ for all } 0 < r < \alpha.$$

[Hint. Use the inequalities of section 8.3 to show that $n\,P(|X| > a_n\, x)$ is bounded in n, where $a_n \equiv n^{1/\alpha}$. Then bound the appropriate integral.]

Exercise 2.3 (Strictly stable dfs) Suppose that G is stable with characteristic exponent $\alpha \neq 1$. Then there is a number b such that $G(\cdot + b)$ is strictly stable. [Hint. Show that b must satisfy $b'_n \equiv b_n + (a_n - n)b = 0$ for all n, and specify b such that $b'_2 = 0$. Or else $b'_2 = b = 0$ immediately.]

Example 2.1 (Hitting time as a stable law) Watch Brownian motion $\mathbb{S}$ until it first attains height a. The time this takes is denoted by T_a. According to the strong Markov property,

(7) $$T_{na} \cong T_a^{(1)} + \cdots + T_a^{(n)} \qquad \text{with} \qquad T_a^{(1)}, \ldots, T_a^{(n)} \quad \text{iid as } T_a\,.$$

Checking the covariance functions, we see that $\mathbb{S}(a^2\cdot)/a \cong \mathbb{S}$ on $[0,\infty)$, and thus $T_a \cong a^2 T_1$. Putting these last two equations together gives

(8) $$T_a^{(1)} + \cdots + T_a^{(n)} \cong T_{na} \cong n^2 a^2 T_1 \cong n^2 T_a\,;$$

thus T_a is strictly stable with $\alpha = \frac{1}{2}$. From (12.7.3), $P(T_a < t) = 2[1 - \Phi(a/\sqrt{t}\,)]$ for the $N(0,1)$ df Φ. Thus the df F_a and the density f_a of T_a are given by

(9) $$F_a(x) = 2[1 - \Phi(a/\sqrt{x}\,)] \quad \text{and} \quad f_a(x) = \frac{a}{\sqrt{2\pi} x^{3/2}} e^{-a^2/(2x)} \quad \text{for } x > 0. \qquad \square$$

3 Characterizing Stable Laws

Theorem 3.1 (General stable chfs) Suppose that $Y \cong G$ is stable. Then either Y is a normal rv or there is a number $0 < \alpha < 2$ and constants $m_1, m_2 \geq 0$ and β for which

$$(1) \quad \begin{aligned} \text{Log } \phi(t) = i\beta t &+ m_1 \textstyle\int_0^\infty [\, e^{itx} - 1 - it\alpha(x)\,] \frac{1}{x^{1+\alpha}}\, dx \\ &+ m_2 \textstyle\int_{-\infty}^0 [\, e^{itx} - 1 - it\alpha(x)\,] \frac{1}{|x|^{1+\alpha}}\, dx. \end{aligned}$$

For $0 < \alpha < 1$ or $1 < \alpha < 2$ this can be put in the form

$$(2) \qquad \text{Log } \phi(t) = idt - c\,|t|^\alpha \times (1 + i\theta\, C_\alpha \text{ sign}(t)),$$

where $d \in (-\infty, \infty)$, $c > 0$, $|\theta| \leq 1$, and $C_\alpha \equiv \tan(\pi\,\alpha/2)$. For $\alpha = 1$ the form is

$$(3) \qquad \text{Log } \phi(t) = idt - c\,|t| \times (1 + i\theta \text{ sign}(t) \log|t|)$$

with c, d, θ as above. In fact, $\theta = (m_1 - m_2)/(m_1 + m_2)$ measures skewness, while the constants d and $(1/c)^{(1/\alpha)}$ are just location and scale parameters. Then let $p \equiv m_1/(m_1 + m_2) = (1 + \theta)/2$.

Corollary 1 (Symmetric stable chfs) $\phi(\cdot)$ is the chf of a nondegenerate and symmetric stable distribution with characteristic exponent α if and only if

$$(4) \qquad \phi(t) = \exp(-c|t|^\alpha) \qquad \text{for some } 0 < \alpha \leq 2 \quad \text{and some } c > 0.$$

Proof. We give only a direct proof of the corollary. This keeps things simple. Let $Y \cong G$ be strictly stable with chf ϕ and let $\psi \equiv \text{Log } \phi$. Since $S_n \cong a_n Y$, we have $\phi^n(t) = \phi(a_n t)$. Thus (modulo $2\pi i$)

$$n\psi(t) = \psi(a_n t) = \psi(a_m(a_n/a_m)t) = m\psi((a_n/a_m)t)$$

$$(a) \qquad = m\psi((n/m)^{1/\alpha} t) \qquad \text{by (11.2.3).}$$

Thus for all rationals $r > 0$ we have shown that

$$(5) \qquad r\psi(t) = \psi(r^{1/\alpha} t) \qquad \text{modulo } 2\pi i,$$

and by continuity, (5) also holds for all real $r > 0$. Set $t = 1$ and $r = \tau^\alpha$ in (5) for

$$\psi(\tau) = c\,\tau^\alpha \qquad \text{for all } \tau > 0, \quad \text{with } c \equiv \psi(1).$$

$$(b) \qquad \equiv (-c_1 + ic_2)\,\tau^\alpha.$$

It must be true that $c_1 > 0$; if $c_1 < 0$ were true, then we would have the impossible situation that $\phi(\tau) \to \infty$ as $\tau \to \infty$, while $c_1 = 0$ would imply that $|\phi(\tau)| = 1$ for all τ, and that Y is degenerate by proposition 9.8.2. Thus for some $c_1 > 0$, for $t > 0$ we must have

$$(c) \qquad \phi(t) = \exp[(-c_1 + ic_2)\, t^\alpha], \qquad \text{with} \quad \phi(-t) = \overline{\phi(t)}.$$

We can summarize the two equations in (c) as

(d) $\phi(t) = \exp(-c_1|t|^\alpha \times [1 - i(c_2/c_1)\,\text{sign}(t)]\ )\,.$

Since G is symmetric, ϕ is real, and so $c_2 = 0$. Thus (5) holds. All that remains is to be sure that ϕ is a valid chf. This follows from the next two exercises.

If $\alpha \neq 1$, then exercise 11.2.3 shows that $\phi(t)\exp(idt)$ is a strictly stable chf for some real d. Thus it satisfies (d), which is almost (2). □

Exercise 3.1 Suppose that $X \cong F$ with chf ϕ that satisfies

(6) $\phi(t) = 1 - c\,|t|^\alpha + O(t^2) \qquad \text{as } t \to 0$

for some fixed $\alpha \in (0, 2)$. Suppose now that $X_1, \ldots, X_n$ are iid as F. Show that the properly normed sum satisfies $S_n/n^{1/\alpha} \to_d Z$ where $\phi_Z(t) = \exp(-c|t|^\alpha)$.

Exercise 3.2 Suppose X has density $f(x) = (\alpha/2)\,|x|^{-(\alpha+1)} \times 1_{(1,\infty)}(|x|)$ with the constant $\alpha \in (0, 2)$. Show that ϕ satisfies (6).

Exercise 3.3 (Holtzmark–Chandrasekar) Let $X_{n1}, \ldots, X_{nn}$ be iid Uniform$(-n, n)$. We now let $0 < 1/p < 2$ and $M > 0$ be fixed. Let

$$Z_n \equiv \textstyle\sum_{k=1}^n Z_{nk} \equiv \sum_{k=1}^n (M\,\text{sign}(X_{nk})/|X_{nk}|^p.$$

Regard Z_n as the sum of forces exerted on a unit mass at the origin by n stars of mass M that are uniformly distributed on $(-n, n)$ in a universe where an inverse pth power of attraction is operating. Show that $Z_n \to_d Z$, where the chf of Z is given by $\phi_Z(t) = \exp(-c|t|^\alpha)$ for appropriate c and α.

Exercise 3.4 Show that (1) can be put into the form (2) or (3).

4 The Domain of Attraction of a Stable Law

We now merely state some results that assert when convergence in distribution to a general stable law takes place.

Theorem 4.1 (Stable domain of attraction with $0 < \alpha < 2$)
(a) Now, $F \in \mathcal{D}(G)$ for some stable G with characteristic exponent $\alpha \in (0, 2)$ if and only if (as $x \to \infty$) both

(1) $\quad U(x) \equiv \int_{[|y| \le x]} y^2 \, dF(y) \in \mathcal{U}_{2-\alpha} \qquad$ [for $\alpha = 2/(\beta + 1)$]

(or equivalently)

(2) $\quad V(t) \equiv \int_t^{1-t} K^2(s) \, ds \in \mathcal{R}_{-\beta} \qquad$ [for $\beta \equiv (2 - \alpha)/\alpha$])

and also

(3) $\quad P(X > x)/P(|X| > x) \to (\text{some } p) \in [0, 1].$

Moreover, α and p determine G up to type, as follows from the theorem of types.
(b) The constants A_n of $(S_n - B_n)/A_n \to_d Y \cong G$ necessarily satisfy (according to the theorem of types)

(4) $\quad nU(A_n)/A_n^2 \sim nA_n^{-\alpha} L(A_n) \to 1 \qquad \text{as } n \to \infty.$

(c) The following are equivalent (for some constant $0 < \alpha < 2$, and then for some $c > 0$ and some $0 \le p \le 1$):

(5) $\quad F \in \mathcal{D}_N(G) \qquad$ for some stable G with characteristic exponent α.

(6) $\quad x^\alpha P(X > x) \to c\, p \quad$ and $\quad x^\alpha P(X < -x) \to c\,(1 - p) \quad$ as $x \to \infty$.

(7) $\quad t|K(1 - t)|^\alpha \to cp \quad$ and $\quad t|K(t)|^\alpha \to c(1 - p) \quad$ as $t \to 0$.

Theorem 4.2 (Domain of attraction of the normal) $F \in \mathcal{D}(\text{Normal})$ if and only if either (hence, both) $U(\cdot)$ is slowly varying at ∞ or $V(\cdot)$ is slowly varying at 0. (Theorem 10.6.1 and proposition 10.6.1 give a myriad of other equivalences.)

Exercise 4.1 Use section C.4 to show that for $0 < \alpha < 2$ the following are equivalent conditions:

(8) $\quad U \in \mathcal{U}_{2-\alpha}$.

(9) $\quad P(|X| > x) \in \mathcal{U}_\alpha$.

(10) $\quad [x^2 P(|X| > x)/\int_{[|y| \le x]} y^2 \, dF(y)] \to \frac{2-\alpha}{\alpha} \qquad$ as $x \to \infty$.

Other characterizations in terms of K can be found in or derived from theorem C.4.2. The theorems and this remark can also be proved via the Lévy–Khinchin theorem and results about regularly varying functions.

Exercise 4.2 (a) State necessary and sufficient conditions on F for $F \in \mathcal{D}(\text{Cauchy})$.
(b) Do the same for $F \in \mathcal{D}_N(\text{Cauchy})$.
(c) Show by example that $\mathcal{D}_N(\text{Cauchy})$ is a proper subset of $\mathcal{D}(\text{Cauchy})$.
(d) Observe that a symmetric df $F(\cdot)$ is in $\mathcal{D}(\text{Cauchy})$ if and only if the tail function defined by $xP(X > x) = x(1 - F(x))$ is slowly varying. (Recall the tail function $\tau(\cdot)$ of Feller used in the WLLN in (8.4.2).)

Exercise 4.3 (a) Show by example that the domain of normal attraction of the normal law $\mathcal{D}_N(N(0,1))$ is a proper subset of the domain of attraction of the normal law $\mathcal{D}(N(0,1))$. To this end, let $X_1, X_2, \ldots$ be iid with density

$$f(x) = |x|^{-3} \times (2\log|x|) \times 1_{[1,\infty)}(|x|)$$

and consider $S_n/(\sqrt{n}\,\log n)$.
(b) Give a second example that works.
(c) For both examples, determine an A_n that works.

Exercise 4.4 (a) Consider the context of theorem 11.4.2. The constants A_n used for $(S_n - B_n)/A_n \to_d Y \cong N(0,1)$ must satisfy $n\,\tilde{\sigma}_n^2/A_n^2 \to 1$ (equivalently, it must hold that $n\,U(A_n)/A_n^2 \to 1$ as $n \to \infty$) as follows from the theorem of types.
(b) It is also true that

$$(11) \qquad F \in \mathcal{D}_N(\text{Normal}) \qquad \text{if and only if} \qquad \sigma^2 < \infty.$$

(When $\sigma^2 \in (0,\infty)$, we know already that $\sqrt{n}(\bar{X}_n - \mu)/\sigma \to N(0,1)$.)

5 Gamma Approximation

If the underlying summands X_{nk} in a CLT approximation are symmetric, then a normal approximation may seem particularly appropriate. But what if the underlying distribution is positively skewed? (If X is negatively skewed, we just consider $-X$ instead.) Consider the rv $G_r \equiv [\text{Gamma}(r) - r]/\sqrt{r} \cong (0,1)$, where r is chosen to make the the third cumulants match. Might this not give a better approximation for small n? And since $G_r \to_d N(0,1)$ as $r \to \infty$, there is no contradiction at the limit. We will show that this recipe works. The fact that the sum of independent gammas is again gamma is crucial to the technical details of the proof.

Can we go one step further and match the first four cumulants? Yes, because sums of independent rvs that are distributed as "gamma-gamma" again belong to the same family. [This will also work within our ability to match up the cumulants.]

The *skewness* $\gamma_1 \equiv \mu_3/\sigma^3 = \mathrm{E}(X-\mu)^3/\sigma^3$ is defined to be the third cumulant of the standardized rv, and the *tail heaviness* (or *kurtosis*) is the fourth such cumulant defined by $\gamma_2 \equiv \mu_4/\sigma^4 - 3 = \mathrm{E}(X-\mu)^4/\sigma^4 - 3$. We will use these formulas with $\mu = 0$ and $\sigma^2 = 1$, so that the skewness becomes μ_3 and the tail heaviness becomes $\mu_4 - 3$. For the standardized gamma, the first four cumulants are given by

(1) $$G_r \equiv \frac{\text{Gamma}(r) - r}{\sqrt{r}} \cong \left(0, 1; \frac{2}{\sqrt{r}}, \frac{6}{r}\right).$$

For the difference of two independent gammas we let $p + q = 1$ and $c, d > 0$ and set $u \equiv p/c$ and $v \equiv q/d$, and then further define $r = c^2 n$ and $s = d^2 n$ and set

(2) $$\begin{aligned} G_{r,s} &\equiv -\sqrt{p}\, G_{c^2 n} + \sqrt{q}\, G_{d^2 n} \cong \left(0, 1; \tfrac{2}{\sqrt{n}}\,(-\tfrac{p^{3/2}}{c} + \tfrac{q^{3/2}}{d}), \tfrac{6}{n}\,(\tfrac{p^2}{c^2} + \tfrac{q^2}{d^2})\right) \\ &= \left(0, 1; \tfrac{2}{\sqrt{n}}\,(-\sqrt{p}\, u + \sqrt{q}\, v), \tfrac{6}{n}\,(u^2 + v^2)\right). \end{aligned}$$

[This parameterization can match all (μ_3, μ_4) pairs for which $\mu_3^2 \le \frac{2}{3}(\mu_4 - 3)$, all of which have heavier tails than normal distributions.]

Theorem 5.1 (Gamma approximation; the GLT) Let $X_{n1}, \ldots, X_{nn}$ be iid as an X having df F with cumulants $(0, 1;\ \mu_3,\ \mu_4 - 3)$, where $\mu_3 \in (0, \infty)$, so that

(3) $$Z_n \equiv \tfrac{1}{\sqrt{n}} \textstyle\sum_{k=1}^n X_{nk} \cong (0, 1;\ \mu_3/\sqrt{n},\ (\mu_4 - 3)/n).$$

Assume either (4)(a) for the df results below or (4)(b) for the density results, where:

(4) $$\text{(a)} \quad |\phi_X(t)| \to 0 \text{ as } |t| \to \infty \qquad \text{or} \qquad \text{(b)} \quad \textstyle\int_{-\infty}^{\infty} |\phi_X(t)|\, dt < \infty.$$

(i) Let $r \equiv 4n/\mu_3^2$, so that $G_r \equiv [\text{Gamma}(r) - r]/\sqrt{r} \cong (0, 1;\ \mu_3/\sqrt{n},\ 3\,\mu_3^2/2n)$. Then for some constants C_F or $C_{F,n} \to 0$ (that may depend on the df of $(X - \mu)/\sigma$),

(5) $$\|F_{Z_n} - F_{G_r}\| \le C_F/n \quad \text{and } \|f_{Z_n} - f_{G_r}\| \le C_F/n \quad \text{when } \mu_4 < \infty,$$

(6) $$\|F_{Z_n} - F_{G_r}\| \le C_{F,n}/\sqrt{n} \text{ and } \|f_{Z_n} - f_{G_r}\| \le C_{F,n}/\sqrt{n} \text{ when } \mu_3 \in (0, \infty).$$

(ii) Suppose r and s can be specified so $G_{r,s} \cong (0, 1;\ \mu_3/\sqrt{n},\ (\mu_4 - 3)/n)$. Then $n^{3/2}$ can replace n in (5) when $\mu_5 < \infty$. And n can replace $\sqrt{n}$ in (6) when $\mu_4 < \infty$.
(iii) The density $g_r(\cdot)$ of $G_r(\cdot)$ may replace the $N(0,1)$ density in the local limit theorems of section 10.3.

Proof. We initially approximate the distribution of Z_n by that of

(a) $\bar{Z}_n \equiv \frac{1}{\sqrt{2}}\,N(0,1) + \frac{1}{\sqrt{2}}\,[\text{Gamma}(\bar{r}) - \bar{r}]/\sqrt{\bar{r}} \qquad \text{with} \quad \bar{r} \equiv n/(2\,\mu_3^2)$

$\cong \frac{1}{\sqrt{2}}\,(0,1;0,0) + \frac{1}{\sqrt{2}}\,(0,1;2/\sqrt{\bar{r}},6/\bar{r}) = (0,1;1/\sqrt{2\bar{r}},3/2\bar{r})$

(b) $\cong (0,1;\,\mu_3/\sqrt{n},\,3\,\mu_3^2/n) \qquad$ matching (3) to three cumulants

(c) $\equiv \frac{1}{\sqrt{n}}\sum_1^n [\frac{1}{\sqrt{2}}\,N_k + \frac{1}{\sqrt{2}}\,W_k] \equiv \frac{1}{\sqrt{n}}\sum_1^n Y_k\,,$

where the $N_k \cong N(0,1)$ and the $W_k \cong [\text{Gamma}(a) - a]/\sqrt{a}$ with $a = 1/(2\mu_3^2)$ are independent. Let $\phi_Y(t) \equiv \text{E}e^{itY}$ and $\psi_Y \equiv \log\phi_Y$, with ϕ_X and ψ_X defined analogously. Then

$|\phi_{Z_n}(t) - \phi_{\bar{Z}_n}(t)| = |\phi_{\bar{Z}_n}(t)| \times |e^{n[\psi_X(t/\sqrt{n}\,) - \psi_Y(t/\sqrt{n}\,)]} - 1|$

$\equiv |\phi_{\bar{Z}_n}(t)| \times |e^z - 1|$

(d) $\le |\phi_{N(0,1)}(t/\sqrt{2}\,)| \times |\phi_{\sum W_k/\sqrt{n}}\,(t/\sqrt{2}\,)| \times |z|e^{|z|}$

(e) $\le e^{-t^2/4} \times 1 \times |z|e^{|z|}\,.$

Here (provided that $|t|/\sqrt{n}$ is sufficiently small for the expansion of (9.6.22) to be valid) the inequality (9.6.22) then gives (since the first three cumulants of X and Y match)

(f) $|z| = n\,|\psi_X(t/\sqrt{n}\,) - \psi_Y(t/\sqrt{n}\,)| \le t^4\,\bar{c}_4\,[\text{E}X^4 + \text{E}Y^4]/n$

$\le t^4\,\bar{c}_4\,[(3 + (\mu_4 - 3)) + (3 + 3\,\mu_3^2)\,]/n \le t^4\,7\,\bar{c}_4\,\mu_4^{3/2}/n$

(g) $\equiv t^4\,c^2/n$

(h) $\le t^2/9 \qquad \text{for} \quad |t| \le \sqrt{n}/3c\,.$

Plugging both (g) and (h) into (e) gives

(i) $|\phi_{Z_n}(t) - \phi_{\bar{Z}_n}(t)| \le (c^2\,t^4/n)\exp(-(5/36)\,t^2) \qquad \text{for} \quad |t| \le (\text{some } d\,)\sqrt{n}\,,$

where c and d depend on the β_2 value of $(X-\mu)/\sigma$. [The specification in (a) that $\bar{Z}_n$ has a normal component is not natural or practically useful, but it delivers the technically lovely exponential bound component $\exp(-t^2/4)$ in both (d) and (e). Since (a) is not useful for a practical approximation, we will overcome this objection by doing the approximation (a) again—to difference it out. (I believe this whole approach may be new.)]

Let $G_{n1},\ldots,G_{nn}$ be iid as $[\text{Gamma}(b) - b]/\sqrt{b}$ for $b \equiv 4/\mu_3^2$. Then

(j) $G_r \equiv \frac{1}{\sqrt{n}}\sum_1^n G_{nk} \cong [\text{Gamma}(r) - r]/\sqrt{r} \qquad \text{with} \quad r \equiv 4n/\mu_3^2$

(k) $\cong (0,1;\,\mu_3/\sqrt{n},\,(3/2)\,\mu_3^2/n), \quad$ matching (3) to three cumulants.

We now approximate G_r by the $\bar{Z}_n$ of (a) [just as earlier we approximated Z_n by the $\bar{Z}_n$ of (a)]. This gives (with generic constants c and d)

(l) $|\phi_{G_r}(t) - \phi_{\bar{Z}_n}(t)| \le (c^2\,t^4/n)\exp(-(5/36)\,t^2) \qquad \text{for} \quad |t| \le d\sqrt{n}\,.$

Combining (i) and (l) gives

(7) $\qquad |\phi_{Z_n}(t) - \phi_{G_r}(t)| \le (c\,t^4/n)\exp(-(5/36)\,t^2) \qquad \text{for} \quad |t| \le d\sqrt{n}\,,$

where the generic c and d may depend on the df of $(X-\mu)/\sigma$.

Consider (5)(b). The inversion formula (using (9.6.22) to expand) gives

$$|f_{Z_n}(x) - f_{G_r}(x)| = \tfrac{1}{2\pi}\,|\int_{-\infty}^{\infty} e^{-itx}\,[\phi_{Z_n}(t) - \phi_{G_r}(t)]\,dt\,| \le I_{1n} + I_{2n} + I_{3n}$$

(m) $\qquad \equiv \int_{[\,|t|\le d\sqrt{n}\,]} |\phi_{Z_n}(t) - \phi_{G_r}(t)\,|\,dt$

$$+\int_{[|t|>d\sqrt{n}\,]} |\phi_{Z_n}(t)|\,dt + \int_{[|t|>d\sqrt{n}\,]} |\phi_{G_r}(t)|\,dt\,.$$

Now (7) bounds the integrand of the lead term to give

(n) $\qquad I_{1n} \le \int (c\,t^4/n)\exp(-(5/36)t^2)\,dt \le c/n\,.$

Since $\int |\phi_X(t)|\,dt < \infty$ by (4)(b), the density inversion formula (9.4.9) shows that X has a density $f_X(\cdot)$. Since X is thus not distributed on a grid (and likewise Y), proposition 9.8.2 gives

$$\theta \equiv \|\phi_X\|_d^\infty \vee \|\phi_W\|_d^\infty < 1\,.$$

Thus the second term in (m) satisfies

$$I_{2n} = \int_{[|t|>d\sqrt{n}\,]} |\phi_X(t/\sqrt{n}\,)|^n\,dt \le \sqrt{n}\int_{[|s|>d\,]} |\phi_X(s)|^n\,ds$$

(o) $\qquad \le \theta^{n-1}\sqrt{n}\int_{-\infty}^{\infty} |\phi_X(s)|\,ds = o(n^{-r}), \qquad \text{for any } r > 0,$

since the θ^n term goes to 0 geometrically. Likewise, $I_{3n} = o(n^{-r})$, for any $r > 0$, since $|\phi_G|^k$ satisfies (4)(b), for some k. Combine (n) and (o) into (m) to get (5)(b).

Consider (5)(a). We will apply Esseen's lemma. Thus

$$\|F_{Z_n}(x) - F_{G_r}(x)\| \equiv I'_{1n} + I'_{2n} + I'_{3n}$$

(p) $\qquad \le \{\int_{[|t|\le d\sqrt{n}\,]} + \int_{[d\sqrt{n}<|t|\le d\,n\,]}\}\,\frac{1}{|t|}\,|\phi_{Z_n}(t) - \phi_{G_r}(t)|\,dt + 24\,\|g_r\|/\,\pi\,d\,n$

(q) $\qquad \le \int (c|t|^3/n)\exp(-(5/36)t^2)\,dt + 2(dn/d\sqrt{n}\,)\theta^n + 24\|g_r\|/(\pi dn)$

$$= O(1/n)\,,$$

where (4)(a) is now used to obtain θ^n.

Consider (6), when μ_4 is not assumed finite. Use of (9.6.22) at line (f) must be replaced by use of (9.6.21). The $|t|^3\,\mathrm{E}|X|^3$ in (9.6.22) would give a bound of only $C_F/\sqrt{n}$ at line (m) of the current proof; but the added $\delta_3(t/\sqrt{n}\,)$ term in (9.6.21) that is valid on $|t| \le d\sqrt{n}$ (now with a tiny d) allows a C_F to be replaced by a $C_{F,n} \to 0$. Dominated convergence is used for this, with dominating function guaranteed by $\mathrm{E}|X|^3 < \infty$.

If we knew any appropriate two-parameter family closed under convolution, we could choose those two parameters to match both third and fourth cumulants. Then $C_F/n^{3/2}$ under $\mu_5 < \infty$ and $C_{F,n}/n$ under $\mu_4 < \infty$ would be possible. The proof is essentially unchanged, and needs no further comment. The difference of two gammas can be specified in several different ways. All work. The only question is which has the greatest coverage of the (skewness, kurtosis)-plane. Using gammas, we seem stuck with positive kurtosis, which leaves out some of the least important situations. [Edgeworth expansions allow us to cover the whole (skewness, kurtosis)-plane, but they have some other deficiencies. For instance, the Edgeworth approximation to a df or density is not necessarily a df or density itself.] □

Poisson Approximation

Most discrete distributions we care about live on the integers, and the write-up here will reflect that fact and make this case fit our notation with the least effort. Rather than approximating sums of such rvs X by an appropriate gamma with a continuity correction, we will use a nice discrete analogue of the gamma.

For the standardized Poisson, the first four cumulants are

$$(8) \qquad G_r \equiv \frac{\text{Poisson}(r) - r}{\sqrt{r}} \cong \left(0, 1; \frac{1}{\sqrt{r}}, \frac{1}{r}\right).$$

For the difference of two Poissons we let $p+q=1$ and $c, d > 0$ and set $u \equiv p/c$ and $v \equiv q/d$, and then define $r = c^2 n$ and $s = d^2 n$ and set

$$(9) \qquad \begin{aligned} G_{r,s} &\equiv -\sqrt{p}\, G_{c^2 n} + \sqrt{q}\, G_{d^2 n} \cong (0, 1; \tfrac{1}{\sqrt{n}}\,(-\tfrac{p^{3/2}}{c} + \tfrac{q^{3/2}}{d}), \tfrac{1}{n}\,(\tfrac{p^2}{c^2} + \tfrac{q^2}{d^2})) \\ &= (0, 1; \tfrac{1}{\sqrt{n}}\,(-\sqrt{p}\, u + \sqrt{q}\, v), \tfrac{1}{n}\,(u^2 + v^2)) . \end{aligned}$$

[This approach can (multiply) match all (μ_3, μ_4) pairs for which $\mu_3^2 \le (\mu_4 - 3)$.]

Theorem 5.2 (Poisson approximation) Consider a rv X on the integers and let Z_n be as in (3). Let $r \equiv n/\mu_3^2$, so that

$$G_r \equiv [\text{Poisson}(r) - r]/\sqrt{r} \cong (0, 1;\, \mu_3/\sqrt{n},\, \mu_3^2/n).$$

(i) Then for some constants C_F and $C_{F,n} \to 0$ (that may depend on the df of the standardized rv $(X-\mu)/\sigma$):

$$(10) \qquad \|p_{Z_n} - p_{G_r}\| \le C_F/n^{3/2} \qquad \text{when } \mu_4 < \infty .$$

$$(11) \qquad \|p_{Z_n} - p_{G_r}\| \le C_{F,n}/n \qquad \text{when only } \mu_3 \in (0, \infty) .$$

[Most probabilities that one computes involve summing over the appropriate $M\sqrt{n}$ number of terms that are each of the type $p_{Z_n}(\cdot)$.]
(ii) Suppose r and s can be specified so that $G_{r,s} \cong (0, 1;\, \mu_3/\sqrt{n},\, (\mu_4 - 3)/n)$. Then n^2 can replace $n^{3/2}$ in (10), provided that $\mu_5 < \infty$. And $n^{3/2}$ can replace n in (11), provided that $\mu_4 < \infty$.

Proof. The appropriate inversion formula now (for a distribution on the grid $am + b$) is given by

$$(12) \qquad p_m \equiv P(X = m) = \tfrac{a}{2\pi} \int_{[\,|t| \le \pi/a\,]} \exp\left(-it(am+b)\right) \phi_X(t)\, dt .$$

By the previous proof (including the previous step (a) normal component, but now appearing in step (u)) yields

$$\text{(u)} \qquad |\phi_{Z_n}(t) - \phi_{G_r}(t)| \le (c^2\, t^4/n) \exp(-(5/36)\, t^2) \qquad \text{for} \quad |t| \le d\sqrt{n}\,,$$

for c and d that may depend on the df of $(X-\mu)/\sigma$. Applying the inversion formula in (12) now gives

$$\text{(v)} \qquad \sqrt{n}\,|P(Z_n = m/\sqrt{n}\,) - P(G_r = m/\sqrt{n}\,)\,|$$

$$= \tfrac{\sqrt{n}}{2\,\pi\,\sqrt{n}} |\int_{[\,|t| \le \pi\sqrt{n}\,]} e^{-itm/\sqrt{n}}\, [\phi_{Z_n}(t/\sqrt{n}\,) - \phi_{G_r}(t/\sqrt{n}\,)]\, dt|$$

$$\text{(w)} \qquad \le \int_{[\,|t| \le d\sqrt{n}\,]} |\phi_{Z_n}(t/\sqrt{n}\,) - \phi_{G_r}(t/\sqrt{n}\,)|\, dt$$

$$+ \int_{[d\sqrt{n} < |t| \le \pi\sqrt{n}\,]} |\phi_{Z_n}(t/\sqrt{n}\,)|\, dt + \int_{[\,d\sqrt{n} < |t| \le \pi\sqrt{n}\,]} |\phi_{G_r}(t/\sqrt{n}\,)|\, dt$$

$$\text{(x)} \qquad \le c/n + o(n^{-r}) + o(n^{-r})\,,$$

as is now easily shown with the same arguments as before (because $\theta < 1$, since $\pi\sqrt{n}$ never reaches a full period of ϕ_X; recall proposition 9.8.2). □

Exercise 5.1 Verify part (iii) of theorem 3.1.

Exercise 5.2 Show that the chf ϕ of the G_r of (1) satisfies $\int |\phi(t)|^k\, dt < \infty$ for some $k > 0$.

Exercise 5.3 We can replace the Poisson by the NegBiT(r, p) distribution with the moment structure

$$\text{(13)} \qquad G_r \equiv \frac{\text{NegBiT}(r,p) - r/p}{\sqrt{rq/p^2}} \cong \left(0, 1;\ \frac{1}{\sqrt{r}}\, \frac{1+q}{\sqrt{q}},\ \frac{1}{r}\, \frac{1+4q+q^2}{q}\right).$$

This is probably more useful than the previous theorem. (a) Verify the claim.
(b) Provide some numerical work to compare Poisson and NegBiT approximations to a situation of interest.

Remark 5.1 (Gamma approximation or Edgeworth approximation?) In the next section we will derive the classical Edgeworth approximations. The first-order Gamma (or Poisson, or NegBiT) approximations of the current section are of the same order as the first-order Edgeworth approximations. Moreover, approximation by the G_r-distribution is an approximation by a probability distribution; but this is not true of the Edgeworth approximation. Happily, gamma approximations are easily and accurately implemented in S-plus, say.

The situation is similar regarding the two second-order approximations, provided that the first two cumulants of the underlying rv can be matched within the family of G_{rs}-distributions. However, the G_{rs}-distributions are not available within any set of computer-generated routines I know, so that this would be hard to implement at present. However, this would seem to make a *nice project* for a computer-oriented statistician. □

Examples

Example 5.1 (Sampling distribution of $\bar{X}_n$ and S_n^2) Suppose $X_1, \ldots, X_n$ is a random sample from a population whose first four cumulants are $(\mu, \kappa_2; \kappa_3, \kappa_4)$. [Let K&S denote Kendall and Stuart (1977, Vol. I).] How do we apply a gamma approximation?
(a) Consider first an infinite population, in which $\kappa_2 = \sigma^2$, $\kappa_3 = \mu_3$, and $\kappa_4 = \mu_4 - 3$. Then (9.6.20) gives the first four cumulants of $\sqrt{n}\,(\bar{X}_n - \mu)$ as

$$(14) \qquad \sqrt{n}\,(\bar{X}_n - \mu) \cong (0, \sigma^2; \tfrac{1}{\sqrt{n}}\,\kappa_3,\ \tfrac{1}{n}\,\kappa_4\,).$$

Now, unbiased estimators $\hat{\kappa}_j$ of these κ_j are given (see K&S (p. 297, 300)) by

$$(15) \qquad \begin{aligned} &\hat{\kappa}_2 \equiv \tfrac{n}{n-1}\,\hat{m}_2, \quad \hat{\kappa}_3 \equiv \tfrac{n^2}{(n-1)(n-2)}\,\hat{m}_3, \\ &\hat{\kappa}_4 \equiv \tfrac{n^2}{(n-1)(n-2)(n-3)}\{(n+1)\hat{m}_4 - 3\,(n-1)\hat{m}_2^2\}, \end{aligned}$$

where $\hat{m}_j \equiv \sum_1^n (X_j - \bar{X}_n)^j / n$. We will combine these facts with the theorems of the previous sections to approximate the distribution of $\sqrt{n}(\bar{X}_n - \mu)$. Additionally (by K&S (p. 306-307),

$$(16) \qquad \sqrt{n}\,(S_n^2 - \sigma^2) \cong (0, \kappa_4 + \tfrac{2\,n}{n-1}\,\kappa_2^2; \tfrac{1}{n}\,\kappa_6 + \tfrac{12}{n-1}\,\kappa_4\,\kappa_2 + \tfrac{4(n-2)}{(n-1)^2}\,\kappa_3^2 + \tfrac{8n}{(n-1)^2}\,\kappa_2^3),$$

where correcting for skewness in (16) should probably be ignored. An unbiased estimator of the variance in (16) (unbiasedness verifiable from K&S (p. 296)) is

$$(17) \qquad \tfrac{n-1}{n+1}\,[\hat{\kappa}_4 + 2\,\tfrac{n}{n-1}\,\hat{\kappa}_2^2]\,.$$

(b) Finally (by K&S (p. 327)),

$$(18) \qquad \mathrm{Corr}[\hat{\kappa}_1, \hat{\kappa}_2] = \gamma_1 / \{\gamma_2 + \tfrac{2n}{n-1}\}^{1/2}\,.$$

(c) In approximating the bootstrap distribution of $\overline{X_n^*}$, it is exactly true that

$$(19) \qquad \sqrt{n}\,(\,\overline{X_n^*} - \bar{X}_n) \cong (0, \hat{\kappa}_2; \tfrac{1}{\sqrt{n}}\,\hat{\kappa}_3,\ \tfrac{1}{n}\,\hat{\kappa}_4\,)\,.$$

(d) Now consider a finite population $X_1, \ldots, X_N$ whose second, third, and fourth true cumulants K_j are given by (15), with N replacing n. Unbiased estimators $\hat{K}_j$ are also given by (15), now with n again (see K&S (p. 320)). It is also true (by K&S (p. 321-322)) that

$$(20) \qquad \sqrt{n}\,(\bar{X}_n - \bar{X}_N) \cong \Big(0, \tfrac{N-n}{N}\,K_2\,; \tfrac{N^2 - 3nN + 2n^2}{n^2\,N^2}\,K_3\,, \\ \{\alpha_3 - 4\,\alpha_2/N + 6\,\alpha_1/N^2 - \tfrac{3(N-1)}{N(N+1)}\,\alpha_1^2\,\}\,K_4\, + 3\,\tfrac{N-1}{N+1}\,\alpha_1^2\,K_2^2\Big),$$

where $\alpha_r \equiv (n^{-r} - N^{-r})$. Finally (by K&S (p. 323)),

$$(21) \qquad \sqrt{n}\,(S_n^2 - S_N^2) \cong (0, K_4 + 2\tfrac{n}{n-1}\,K_2^2\,)\,.$$

Then (by K&S (p. 323)) an unbiased estimator of this last variance is given by (17) (with $\hat{K}_j$ replacing $\hat{\kappa}_j$). (Though straightforward, the results cited from K&S are somewhat cumbersome.) □

Example 5.2 (Hall) A noncentral chisquare rv $\chi_n^2(\delta)$ satisfies

$$(22)\qquad T \equiv \frac{[\chi_n^2(\delta) - (n+\delta)]}{\sqrt{2(n+2\delta)}} \cong \frac{1}{\sqrt{n}} \sum_1^n X_k \cong \left(0, 1;\, \mu_3 \equiv \frac{2^{3/2}\,(1+3\,\delta/n)}{\sqrt{n}\,(1+2\,\delta/n)^{3/2}}\right).$$

So we approximate this distribution by G_r with $r \equiv (n+2\,\delta)^3/(2(n+3\,\delta)^2)$. Then

$$(23)\qquad P(\chi_n^2(\delta) \le x) \doteq P(\text{Gamma}(r) \le r + [x - (n+\delta)](n+2\,\delta)/2(n+3\,\delta))$$

$$(24)\qquad = (1-\alpha) \qquad \text{if} \quad x \equiv (n+\delta) + (\gamma^\alpha - r)2(n+3\,\delta)/(n+2\,\delta)\,,$$

where γ^α denotes the upper $1-\alpha$ percentage point of Gamma(r). This is easy to implement in Splus, for example. (Hall found that the accuracy seemed quite good, especially in relation to previous proposals.) □

Exercise 5.4 (Poisson approximation of the generalized binomial) We suppose $X_{n1}, \ldots, X_{nn}$ are independent rvs with $X_{nk} \cong$ Bernoulli(p_{nk}). Suppose further that $Y_{nk} \cong$ Poisson(p_{nk}) are independent for $1 \le k \le n$. Let P_n and Q_n denote the distributions of $X_n \equiv \sum_1^n X_{nk}$ and $Y_n \equiv \sum_1^n Y_{nk}$. Show that the total variation distance between P_n and Q_n satisfies

$$(25)\qquad d_{TV}(P_n, Q_n) \equiv \sup\{|P_n(A) - Q_n(A)| : A \in \mathcal{B}\} \le \textstyle\sum_{k=1}^n p_{nk}^2\,.$$

If $p_{nk} = \lambda_k/n$ for $1 \le k \le n$, then the bound becomes $\overline{\lambda^2}/n$.
[Hint. The first step is to replace the original Bernoulli(p_{nk}) rvs by different Bernoulli(p_{nk}) rvs, to be denoted by X_{nk} also. To this end we now define the new $Z_{nk} \cong$ Bernoulli($1 - (1-p_{nk})e^{p_{nk}}$) rvs that are independent for $1 \le k \le n$ (and they are also independent of the Y_{nk}'s). Now define

$$(26)\qquad X_{nk} \equiv 1_{[Y_{nk} \ge 1]} + 1_{[Y_{nk}=0]}\, 1_{[Z_{nk}=1]}\,,$$

and verify that it is indeed a Bernoulli(p_{nk}) rv. (This choice of the jointly distributed pair (X_{nk}, Y_{nk}) maximizes the mass on the diagonal $x = y$ of an (x, y)-coordinate system.) Now verify that

$$(27)\qquad d_{TV}(P_n, Q_n) \le P(X_n \ne Y_n) \le \textstyle\sum_1^n P(X_{nk} \ne Y_{nk}) \le \sum_1^n p_{nk}^2\,.$$

This type of proof is called a *coupling* proof, in that the (X_{nk}, Y_{nk}) pairs are coupled together as closely as possible.]

Exercise 5.5 (a)–(d) Derive the claims made in example 5.1(a)–(d).

6 Edgeworth Expansions

The Setup

Let F_0, f_0, and ϕ_0 denote the $N(0,1)$ df, density, and chf. Thus

(1) $$f_0(x) \equiv e^{-x^2/2}/\sqrt{2\pi} \qquad \text{and} \qquad \phi_0(t) \equiv e^{-t^2/2}$$

on the real line. These are related via the inversion formula for chfs as

(2) $$f_0(x) = \int_{-\infty}^{\infty} e^{-itx}\phi_0(t)\,dt/(2\pi).$$

Differentiating f_0 gives

(3) $$\begin{aligned} &f_0'(x) = -x\,f_0(x), && f_0''(x) = (x^2-1)\,f_0(x),\\ &f_0'''(x) = -(x^3-3x)\,f_0(x), && f_0^{(iv)}(x) = (x^4-6x^2+3)\,f_0(x),\\ &f_0^{(v)}(x) = -(x^5-10x^3+15x)\,f_0(x), &&\\ &f_0^{(vi)}(x) = (x^6-15x^4+45x^2-15)\,f_0(x); && \end{aligned}$$

and in general,

(4) $$f_0^{(k)}(x) = (-1)^k H_k(x) f_0(x)$$

defines what we will call the kth *Hermite orthogonal polynomial* H_k (see exercise 6.1). Equating the derivatives in (3) to derivatives of the right-hand side of (2) gives

(5) $$(-1)^k H_k(x) f_0(x) = \int_{-\infty}^{\infty} e^{-itx}(-it)^k\phi_0(t)\,dt/(2\pi),$$

which expresses $H_k f_0$ as the inverse Fourier transform of $(it)^k\phi_0(t)$. This gives the *key result* that

(6) $$(it)^k\phi_0(t) \qquad \text{is the Fourier transform of} \qquad H_k(\cdot)\,f_0(\cdot).$$

Now suppose that $X_1,\ldots,X_n$ are iid where

(7) $$X \cong (0,\sigma^2) \qquad \text{has chf } \phi(\cdot).$$

We let $S_n \equiv \sum_{k=1}^n X_k$, and agree that

(8) $$F_n(\cdot) \qquad \text{denotes the df of} \qquad S_n/(\sigma\sqrt{n}\,).$$

The idea is to expand F_n in terms of the orthogonal polynomials H_k. However, we choose instead to obtain a first-order or second-order approximation, together with an error analysis. Also,

(9) $$f_n(\cdot) \qquad \text{denotes the density of} \qquad S_n/(\sigma\sqrt{n}\,), \qquad \text{if it exists.}$$

In this latter case we will also seek to expand f_n. The expansions we will derive for f_n and F_n are known as *Edgeworth expansions*.

Edgeworth Expansions for Densities

Instead of just assuming that f_n exists, we assume instead that the chf ϕ of the rv $X \cong (0, \sigma^2)$ satisfies

(10) $\int_{-\infty}^{\infty} |\phi(t)|^m \, dt < \infty,$ for some positive integer m.

This guarantees both that f_n exists for all $n \ge m$, and that it can be found from the Fourier inversion formula (9.4.9).

Theorem 6.1 Suppose condition (10) holds. Let $\gamma_1 \equiv \mathrm{E}(X/\sigma)^3$ denote the *skewness*, and let $\gamma_2 \equiv \mathrm{E}(X/\sigma)^4 - 3$ denote the *tail heaviness* of $X \cong F(0, \sigma^2)$.
(a) Then

$$\left\| f_n(\cdot) - f_0(\cdot)\,\{1 + \frac{\gamma_1}{3!\sqrt{n}} H_3(\cdot)\} \right\|$$

(11) $= o(1/\sqrt{n}\,)$ (as a function of $\mathrm{E}|X/\sigma|^3$ and $F_{X/\sigma}$) [if $\mathrm{E}|X|^3 < \infty$]

or

(12) $= O(1/n)$ (as a function of $\mathrm{E}|X/\sigma|^4$ and $F_{X/\sigma}$) [if $\mathrm{E}X^4 < \infty$].

(b) Moreover,

$$\left\| f_n(\cdot) - f_0(\cdot) \left\{ 1 + \left[\frac{1}{\sqrt{n}} \frac{\gamma_1}{3!} H_3(\cdot) \right] + \frac{1}{n} \left[\frac{\gamma_2}{4!} H_4(\cdot) + \frac{\gamma_1^2}{2(3!)^2} H_6(\cdot) \right] \right\} \right\|$$

(13) $= o(1/n)$ (as a function of $\mathrm{E}|X/\sigma|^4$ and $F_{X/\sigma}$) [if $\mathrm{E}X^4 < \infty$]

or

(14) $= O(1/n^{3/2})$ (as a function of $\mathrm{E}|X/\sigma|^5$ and $F_{X/\sigma}$) [if $\mathrm{E}|X|^5 < \infty$].

We specifically write out that $H_0(x) \equiv 1$ and

(15)
$$\begin{array}{ll} H_1(x) \equiv x, & H_2(x) \equiv x^2 - 1, \\ H_3(x) \equiv x^3 - 3x, & H_4(x) \equiv x^4 - 6x^2 + 3, \\ H_5(x) \equiv x^5 - 10x^3 + 15x, & H_6(x) \equiv x^6 - 15x^4 + 45x^2 - 15, \end{array}$$

for use in the current set of theorems. The previous theorem was for densities. The next is for dfs. Condition (10) is used to control the extreme tails in the Fourier inversion formulas for densities. In proving analogues of (11) and (12) for dfs, we will be able to use Esseen's lemma to control these tails instead. However, the analogues of (13) and (14) run into other problems, and these are again overcome via a (now weaker) restriction on ϕ. All proofs are at the end of this section.

Exercise 6.1 Find $H_7(\cdot)$. Show that $\int_{-\infty}^{\infty} H_m H_n f_0 \, d\lambda = n!$ if $m = n$, and 0 else.

Edgeworth Expansions for Distribution Functions

Consider the potential hypothesis

(16) $\limsup_{|t|\to\infty} |\phi(t)| < 1.$

(This is weaker than (10). The Riemann–Lebesgue lemma shows that (16) holds if F has an absolutely continuous component, à la theorem 6.1.1.)

Theorem 6.2 Suppose that X is not distributed on a grid. Let $\gamma_1 \equiv \mathrm{E}(X/\sigma)^3$ denote the *skewness*, and let $\gamma_2 \equiv \mathrm{E}(X/\sigma)^4 - 3$ denote the *tail heaviness* of the rv $X \cong F(0, \sigma^2)$. (a) Then

$$\left\| F_n(\cdot) - F_0(\cdot) + f_0(\cdot) \left\{ \frac{\gamma_1}{3!\sqrt{n}} H_2(\cdot) \right\} \right\|$$

(17) $= o(1/\sqrt{n})$ (as a function of $\mathrm{E}|X/\sigma|^3$ and $F_{X/\sigma}$) [if $\mathrm{E}|X|^3 < \infty$]

or (additionally requiring (16) for (18))

(18) $= O(1/n)$ (as a function of $\mathrm{E}|X/\sigma|^4$ and $F_{X/\sigma}$) [if $\mathrm{E}X^4 < \infty$].

(b) Moreover, when (16) holds,

$$\left\| F_n(\cdot) - F_0(\cdot) + f_0(\cdot) \left\{ \frac{1}{\sqrt{n}} \frac{\gamma_1}{3!} H_2(\cdot) + \frac{1}{n} \left[\frac{\gamma_2}{4!} H_3(\cdot) + \frac{\gamma_1^2}{2 \cdot (3!)^2} H_5(\cdot) \right] \right\} \right\|$$

(19) $= o(1/n)$ (as a function of $\mathrm{E}|X/\sigma|^4$ and $F_{X/\sigma}$) [if $\mathrm{E}X^4 < \infty$]

or

(20) $= O(1/n^{3/2})$ (as a function of $\mathrm{E}|X/\sigma|^5$ and $F_{X/\sigma}$) [if $\mathrm{E}|X|^5 < \infty$].

Exercise 6.2 Let $Z_n \equiv S_n/\sigma\sqrt{n}$ as above, and let f_n denote its density under condition (10). Let $Z_n^* \equiv [\mathrm{Gamma}(r) - r]/\sqrt{r}$, with $r \equiv 4n/\gamma_1^2$, and let $g_n(\cdot)$ denote its density. Show that $\|f_n - g_n\| = O(1/n)$ as a function of $\mathrm{E}|X/\sigma|^4$ and $F_{X/\sigma}$ when $E|X|^4 < \infty$.

The Proofs

We defined and expanded the cumulant generating function $\psi(\cdot) \equiv \mathrm{Log}\ \phi(\cdot)$ in exercise 9.6.6. The first few cumulants of the standardized rv X/σ were seen to be 0, 1, γ_1, and γ_2.

Proof. Consider theorem 6.1(a). Without loss of generality we suppose that $\sigma = 1$. We now agree that

(a) $d_n \equiv f_n - \left[1 + \frac{1}{\sqrt{n}} \frac{\gamma_1}{3!} H_3 \right] f_0$

denotes the difference between the true f_n and our first approximation to it. Note from (6) that d_n has Fourier transform

$$(21)\qquad \phi_n(t) \equiv [\phi(t/\sqrt{n}\,)]^n - \phi_0(t)\left[1 + \frac{1}{\sqrt{n}}\frac{\gamma_1}{3!}(it)^3\right]$$

$$(b)\qquad = e^{-t^2/2}\,\{e^{t^2/2+n\psi(t/\sqrt{n}\,)} - [1 + (\gamma_1/3!)(it)^3/\sqrt{n}\,]\}$$

$$(c)\qquad \equiv e^{-t^2/2}\,[e^{z+\epsilon} - (1+z)].$$

Thus the Fourier inversion formula (9.4.9) gives

$$(d)\qquad d_n(x) = \textstyle\int_{-\infty}^{\infty} e^{itx}\phi_n(t)\,dt/(2\pi),$$

since (10) implies that $\int_{-\infty}^{\infty}|\phi_n(t)|\,dt < \infty$. Thus for any fixed $\theta > 0$ and all $x > 0$ we have

$$|d_n(x)| \le \textstyle\int_{-\infty}^{\infty}|\phi_n(t)|\,dt = \int_{-\infty}^{\infty} e^{-t^2/2}\,|e^{z+\epsilon} - (1+z)|\,dt$$

$$(e)\qquad = \textstyle\int_{[|t|\le\theta\sqrt{n}/\mathrm{E}|X|^3]} e^{-t^2/2}\,|e^{z+\epsilon} - (1+z)|\,dt + o(n^{-r})$$

for any $r > 0$, as in (10.3.6). Now,

$$(f)\qquad |\phi_n(t)| = e^{-t^2/2}\,|e^{n\psi(t/\sqrt{n}\,)-(-t^2/2)} - [1 + (\gamma_1/3!)(it)^3/\sqrt{n}\,]|$$

$$\equiv e^{-t^2/2}\,|e^{z+\epsilon} - (1+z)| = e^{-t^2/2}\,|e^z\,(e^\epsilon - 1) + (e^z - (1+z))|$$

$$(g)\qquad \le e^{-t^2/2}\,[|\epsilon|\,e^{|\epsilon|}\,e^{|z|} + z^2\,e^{|z|}/2]$$

using (9.6.3). Note that for all $|t| \le \theta\sqrt{n}/\mathrm{E}|X|^3$ we have

$$(h)\qquad |z| \le (|t|^3/6)\,\mathrm{E}|X|^3/\sqrt{n} \le \theta t^2/6 \le t^2/8 \qquad \text{if } \theta \le \tfrac{3}{4}\,,$$

$$|\epsilon| = |n\psi(t/\sqrt{n}\,) - [-t^2/2 + (\gamma_1/6)\,(it)^3/\sqrt{n}\,]|$$

$$(i)\qquad \le c_3\,|t|^3\,\mathrm{E}|X|^3\,\delta(\theta)/\sqrt{n} \le c_3\,t^2\,\theta\,\delta(\theta) \le t^2/8 \qquad \text{if } \theta \text{ is small enough},$$

where $\delta(\cdot)$ denotes the function $\delta_3(\cdot)$ function of (9.6.21) associated with the rv X/σ. Using (h) and (i), the bound in (e) becomes (for some θ small enough)

$$|d_n(x)| \le \textstyle\int_{[|t|\le\theta\sqrt{n}/\mathrm{E}|X|^3]} e^{-t^2/2}\,\{|\epsilon|\,e^{|\epsilon|}\,e^{|z|} + z^2\,e^{|z|}/2\}\,dt + o(n^{-r})$$

$$(j)\qquad \le c_3\,\delta(\theta)\,[\mathrm{E}|X|^3/\sqrt{n}\,]\textstyle\int_{-\infty}^{\infty}|t|^3\,e^{-t^2/4}\,dt$$

$$+[(\mathrm{E}|X|^3)^2/(72n)]\textstyle\int_{-\infty}^{\infty}|t|^6\,e^{-3t^2/8}\,dt + o(n^{-r})$$

$$(k)\qquad = o(n^{-1/2}) \qquad \text{uniformly in } x,$$

since a tiny $\delta(\theta)$ results from a sufficiently tiny θ. Thus (11) holds. For (12), we replace the bound in line (i) above by

(l) $$|\epsilon| \le \bar{c}_4\, t^4\, \mathrm{E}X^4/n \le t^2/8, \qquad \text{which is valid for } |t| \le \sqrt{n}/\sqrt{8\,\bar{c}_4\,\mathrm{E}|X|^4}\,,$$

as (9.6.22) guarantees. We then use (l) instead of (i) during (j) (now integrated over the interval $[|t| \le \sqrt{n}/\sqrt{8/\bar{c}_4\,\mathrm{E}X^4}\,]$).

We now turn to (13), and then (14). We first redefine

(m) $$d_n \equiv f_n - f_0\, \left\{1 + \frac{1}{\sqrt{n}}\frac{\gamma_1}{3!}H_3 + \frac{1}{n}\left[\frac{\gamma_2}{4!}H_4 + \frac{\gamma_1^2}{2\cdot(3!)^2}H_6\right]\right\}.$$

Taking the inverse of its Fourier transform $\phi_n(\cdot)$ gives (as in (e)) that for any fixed value of $\theta > 0$ and all x,

$$|d_n(x)| \le \int_{-\infty}^{\infty} |\phi_n(t)|\, dt$$

(n) $$= \int_{[|t|\le\theta\sqrt{n}/\mathrm{E}X^4]} e^{-t^2/2}\, |e^{t^2/2+n\psi(t/\sqrt{n})} - \{1 + z + z^2/2\}|\, dt + o(n^{-r})$$

(o) $$= \int_{[|t|\le\theta\sqrt{n}/\mathrm{E}X^4]} e^{-t^2/2}\, |e^{z+\epsilon} - (1 + z + z^2/2)|\, dt + o(n^{-r})$$

for each fixed $r > 0$, with

(p) $$z \equiv \frac{1}{\sqrt{n}}\frac{\gamma_1}{3!}(it)^3 + \frac{1}{n}\frac{\gamma_2}{4!}(it)^4.$$

The final details are nearly the same as before. □

Exercise 6.3 Finish the details of the previous proof of theorem 6.1(b).

Proof. Consider theorem 6.2(a). We note that

(q) $$D_n \equiv F_n - F_0 + f_0\left[\frac{1}{\sqrt{n}}\frac{\gamma_1}{3!}H_2\right] \quad \text{has} \quad D_n' = d_n \equiv f_n - f_0\left[1 + \frac{1}{\sqrt{n}}\frac{\gamma_1}{3!}H_3\right],$$

where d_n is as in line (a) of the previous proof (just use $x\,H_2(x) - 2x = H_3(x)$ to verify this). Esseen's lemma then gives

(r) $$\|D_n\| \le \frac{1}{\pi}\int_{[|t|\le a\sqrt{n}/\mathrm{E}|X|^3]} \frac{|\phi_n(t)|}{|t|}\, dt + \frac{24\|\, f_0\,[1 + (\gamma_1/3!)H_3/\sqrt{n}\,]\,\|}{\pi\, a\sqrt{n}/\mathrm{E}|X|^3},$$

where ϕ_n is the same ϕ_n appearing in (21). Since the norm in the second term on the right of (r) is bounded, the second term in (r) is less that $\epsilon/\sqrt{n}$ whenever $a \equiv a(\epsilon, F_{X/\sigma})$ chosen large enough. Fix this a in the limits of integration of (r), and then break this integral into two pieces: the integral over $[|t| \le \theta\sqrt{n}/\mathrm{E}|X|^3]$ with θ as in (i), and the integral over $[|t| > \theta\sqrt{n}/\mathrm{E}|X|^3]$. The integral over the set $[|t| > \theta\sqrt{n}/\mathrm{E}|X|^3]$ is $o(n^{-r})$, for any $r > 0$ (à la (11.5.6), as before at line (e)). Finally, the value of the integral over the set $[|t| \le \theta\sqrt{n}/\mathrm{E}|X|^3]$ is bounded by a term like the right-hand side of (j) (in which $|t|^3$ and t^6 are replaced in those integrals by t^2 and $|t|^5$, to account for division by $|t|$ in the integrand of (r)). This completes the proof of (17) when X is not distributed on a grid. For (18), the initial region of integration in (r) must be $[|t| \le an/\mathrm{E}X^4]$, and then $an/\mathrm{E}X^4$ will also appear below the norm term. Moreover, we will now use θ for a, since only $O(1/n)$ is required.

Consider theorem 6.2(b). We note that

(s) $$D_n(\cdot) \equiv F_n(\cdot) - F_0(\cdot) + f_0(\cdot)\left\{\frac{1}{\sqrt{n}}\frac{\gamma_1}{3!}H_2(\cdot) + \frac{1}{n}\left[\frac{\gamma_2}{4!}H_3(\cdot) + \frac{\gamma_1^2}{2\cdot(3!)^2}H_5(\cdot)\right]\right\}$$

has derivative

$$D_n'(x) = d_n$$

(t) $$= f_n(x) - f_0(x) + f_0(x)\ \left[\frac{1}{\sqrt{n}}\frac{\gamma_1}{3!}[2x - xH_2(x)]\right]$$

$$+ f_0(x)\,\frac{1}{n}\left[\frac{\gamma_2}{4!}[3(x^2-1) - xH_3(x)] + \frac{\gamma_1^2}{2(3!)^2}[5(x^4 - 6x^2 - xH_5(x)]\right]$$

(u) $$= f_n(x) - f_0(x) + f_0(x)\left\{\left[\frac{1}{\sqrt{n}}\frac{\gamma_1}{3!}H_3(x)\right] + \frac{1}{n}\left[\frac{\gamma_2}{4!}H_4(x) + \frac{\gamma_1^2}{2(3!)^2}H_6(x)\right]\right\},$$

and this is the same d_n as in (m) of the previous proof. Thus the final details are nearly the same as before. □

Exercise 6.4 Complete the details in the previous proof of theorem 6.2(b).

Exercise 6.5 Consider a non-iid case in which all dfs F_{nk} have third and/or fourth moments that are of the same order. Then all of the previous results still obtain.

Exercise 6.6 (Large deviations) Suppose the *moment generating function* (or *mgf*) $M_X(t) \equiv \mathrm{E}e^{tX}$ of the rv X is finite for $0 \le |t| < \epsilon$. Let $X_1, X_2, \ldots$ be iid $(0, \sigma^2)$. Let $F_n(\cdot)$ denote the df of $\sqrt{n}(\bar{X}_n - \mu)$ and let $F_0(\cdot)$ denote the $N(0,1)$ df. Show that

(22) $$[1 - F_n(x_n)]/[1 - F_0(x_n)] \to 1, \qquad \text{provided that} \qquad x_n = o(n^{1/6}).$$

Chapter 12

Brownian Motion and Empirical Processes

1 Special Spaces

General Metric Spaces

Let (M, d) denote an arbitrary metric space and let $\mathcal{M}_d$ denote its *Borel σ-field* (that is, the σ-field generated by the collection of all d-open subsets of M). Let $\mathcal{M}_d^B$ denote the σ-field generated by the collection of all open balls, where a *ball* is a subset of M of the form $\{y : d(y, x) < r\}$ for some $x \in M$ and some $r > 0$; call this the *Baire σ-field*. [The important concept of weak convergence is best described in the context of metric spaces.]

Exercise 1.1 Now, $\mathcal{M}_d^B \subset \mathcal{M}_d$, while

(1) $\qquad \mathcal{M}_d^B = \mathcal{M}_d \qquad$ if (M, d) is a separable metric space.

The Special Spaces $(C, \mathcal{C})$ and $(D, \mathcal{D})$

For functions x, y on $[0, 1]$, define the *uniform metric* (or *supremum metric*) by

(2) $$\|x - y\| \equiv \sup_{0 \le t \le 1} |x(t) - y(t)| .$$

Let C denote the set of all continuous functions on $[0, 1]$. Then

(3) $\qquad (C, \| \ \|)$ is a complete and separable metric space.

Here $\mathcal{C}_{\| \ \|}$ will denote the σ-field of Borel subsets of C; then $\mathcal{C}_{\| \ \|}^B$ will denote the σ-field of subsets of C generated by the open balls, and $\mathcal{C}$ will denote the σ-field generated by the finite-dimensional subsets of C (that is, all $\pi_{\mathbf{t}}^{-1}(B_k)$ for which $0 \le t_1 \le \cdots \le t_k \le 1$ and $B_k \in \mathcal{B}_k$). It can be shown that

(4) $$\mathcal{C}_{\| \ \|} = \mathcal{C}_{\| \ \|}^B = \mathcal{C} .$$

Let D denote the set of all functions on $[0,1]$ that are right continuous and possess left-hand limits at each point. (In some applications below it will be noted that D is also used to denote the set of all left-continuous functions on $[0,1]$ that have right-hand limits at each point. This point will receive no further mention. In some cases we will admit to D, and/or to C, only functions X having $X(0)=0$, etc. This, too, will receive little, if any, further mention.) In any case

(5) $\quad (D, \| \ \|)$ is a complete metric space that is not separable.

Here $\mathcal{D}_{\| \ \|}$ will denote the Borel σ-field of subsets of D, then $\mathcal{D}^B_{\| \ \|}$ will denote the σ-field of subsets of D generated by the open balls, and $\mathcal{D}$ will denote the σ-field generated by the finite-dimensional subsets of D. It can be shown that

(6) $\quad \mathcal{D} = \mathcal{D}^B_{\| \ \|}\,, \quad$ and both are proper subsets of $\quad \mathcal{D}_{\| \ \|}\,,$

and moreover,

(7) $\quad C \in \mathcal{D} \qquad \text{and} \qquad \mathcal{C} = C \cap \mathcal{D}\,.$

We now digress briefly. The proper set inclusion of (6) caused difficulties in the historical development of the theory of empirical processes (note that the uniform empirical process $\mathbb{U}_n = \sqrt{n}(\mathbb{G}_n - I)$ takes values in D). To circumvent these difficulties, various authors showed that it is possible to define a metric d on D that has nice properties (see exercise 1.4 below); thus there is a $d(\cdot,\cdot)$ for which

(8) $\quad (D, d)$ is a complete and separable metric space

whose Borel σ-field $\mathcal{D}_d$ satisfies

(9) $\quad \mathcal{D}_d = \mathcal{D}.$

Moreover, for all x, x_n in D the metric d satisfies

(10) $\quad \|x_n - x\| \to 0 \qquad \text{implies} \qquad d(x_n, x) \to 0,$

while

(11) $\quad d(x_n, x) \to 0 \quad \text{with } x \in C \qquad \text{implies} \qquad \|x_n - x\| \to 0.$

The metric d will not be important to us. We are able to replace d by $\| \ \|$ in our theorems; however, we include some information on d as an aid to the reader who wishes to consult the original literature.

Exercise 1.2 Verify (3) and (4).

Exercise 1.3 (i) Verify (5). [Hint. For each $0 \le t \le 1$ define a function x_t in D by letting $x_t(s)$ equal 0 or 1 according as $0 \le s \le t$ or $t \le s \le 1$.] (ii) Verify (6). [Hint. Consider $\cup\{O_t : 0 \le t \le 1\}$ where O_t is the open ball of radius $\frac{1}{3}$ centered at x_t.] (iii) Verify (7).

Exercise 1.4* Consult Billingsley (1968, pp. 112–115), and verify (8)–(11) for

(12) $\quad d(x,y) \equiv \inf\{\, \|x - y\circ\lambda\| \vee (\, \sup_{s\neq t} |\log \frac{\lambda(t)-\lambda(s)}{t-s}|\,) \,:\, \lambda \in \Lambda\}$,

where Λ denotes all $\uparrow$ continuous maps of $[0,1]$ onto itself. [Roughly, this metric measures how closely x and a slightly perturbed (via λ) y line up, where too much perturbation is penalized. The "log" bounds all λ-slopes away from both 0 and ∞.]

Exercise 1.5 Verify that

(13) $\quad C$ is both $\|\ \|$-separable and d-separable, viewed as a subset of D.

[We will require the $\|\ \|$-separability below.]

Let $q \geq 0$ be positive on $(0,1)$. For functions x, y on $[0,1]$ we agree that

(14) $\quad \|(x-y)/q\|$ is the $\|\cdot/q\|$-distance between x and y,

when this is well-defined (that is, when $\|x/q\|$ and $\|y/q\|$ are finite).

Exercise 1.6 It is useful to be able to view $C_\infty \equiv C_{[0,\infty)}$ as a metric space; of course, this denotes the class of all continuous functions on $[0,\infty)$. (We may sometimes require a subclass, such as the one consisting of functions that equal zero at zero; and we will make no further mention of this.) Let $\mathcal{C}_\infty \equiv \mathcal{C}_{[0,\infty)}$ denote the finite-dimensional σ-field. Consider $(C_\infty, \mathcal{C}_\infty) = (C_{[0,\infty)}, \mathcal{C}_{[0,\infty)})$.
(a) For functions x and y on $[0,\infty)$, define

(15) $$\rho_\infty(x,y) \equiv \sum_{k=1}^{\infty} 2^{-k} \frac{\rho_k(x,y)}{1+\rho_k(x,y)}\,,$$

where $\rho_k(x,y) \equiv \sup_{0\leq t\leq k} |x(t)-y(t)|$. Show that $(C_{[0,\infty)}, \rho_\infty)$ is a metric space.
(b) Show that $\rho_\infty(x,y) \to 0$ if and only if $\rho_k(x,y) \to 0$ for each $0<k<\infty$.
(c) Show that $(C_{[0,\infty)}, \rho_\infty)$ is a complete and separable metric space. Moreover, the σ-field $\mathcal{C}_{\rho_\infty}$ of Borel subsets is the same as the σ-field $\mathcal{C}_{[0,\infty)}$ of finite-dimensional subsets, as is $\mathcal{C}^B_\infty$.
(d) Verify that $(D_{[0,\infty)}, \rho_\infty)$ is a complete metric space, and that the Borel σ-field $\mathcal{D}_{\rho_\infty}$ satisfies $C_{[0,\infty)} \in \mathcal{D}_{\rho_\infty}$ and $\mathcal{C}_{\rho_\infty} = \mathcal{D}_{\rho_\infty} \cap C_{[0,\infty)}$. Also, $\mathcal{D}_{[0,\infty)} = \mathcal{D}^B_{\rho_\infty}$ is a proper subset of $\mathcal{D}_{\rho_\infty}$.
(e) Other spaces of continuous and right-continuous functions are analogously treated. They will receive no specific mention.

Independent Increments and Stationarity

If T is an interval in $(-\infty,\infty)$, then we will write

(16) $\quad \mathbb{X}(s,t] \equiv \mathbb{X}(t) - \mathbb{X}(s) \qquad$ for any $s,t \in T$,

and we will refer to this as an *increment* of $\mathbb{X}$. If $\mathbb{X}(t_0), \mathbb{X}(t_0,t_1], \ldots, \mathbb{X}(t_{k-1},t_k]$ are independent rvs for all $k \geq 1$ and all $t_0 \leq \cdots \leq t_k$ in T, then we say that $\mathbb{X}$ has *independent increments*. If $\mathbb{X}(s,t] \cong \mathbb{X}(s+h,t+h]$ for all $s,t,s+h,t+h$ in T with $h \geq 0$, then $\mathbb{X}$ is said to have *stationary increments*. If $(\mathbb{X}(t_1+h),\ldots,\mathbb{X}(t_k+h)) \cong (\mathbb{X}(t_1),\ldots,\mathbb{X}(t_k))$ for all $k \geq 1$, $h \geq 0$, and all time points in T, then $\mathbb{X}$ is said to be a *stationary process*.

2 Existence of Processes on $(C, \mathcal{C})$ and $(D, \mathcal{D})$

When dealing with processes, we would like to work with the smoothest version possible. This is the version that best models physical reality. It is important at this point to recall theorem 5.4.2 on the existence of smoother versions of processes. Roughly, if all of the sample paths of a process are shown to lie in a (useful) subset of the current image space, then we can restrict ourselves to that subset.

Theorem 2.1 (Existence of processes on $(C, \mathcal{C})$) Begin with a process

$$\mathbb{X} : (\Omega, \mathcal{A}, P) \to (R_{[0,1]}, \mathcal{B}_{[0,1]}, P_X)\,.$$

Suppose that for some $a, b > 0$ the increments of $\mathbb{X}$ satisfy

(1) $\quad \mathrm{E}|\mathbb{X}(s,t]|^b \le K \cdot F(s,t]^{1+a} \qquad$ for all $0 \le s, t \le 1$,

where F is a continuous df concentrated on $[0,1]$ and $F(s,t] \equiv F(t) - F(s)$. Then there exists an equivalent version $\mathbb{Z} : (\Omega, \mathcal{A}, P) \to (R_{[0,1]}, \mathcal{B}_{[0,1]}, P_Z)$ for which

(2) $\quad \mathbb{Z} : (\Omega, \mathcal{A}, P) \to (C, \mathcal{C}, P_Z), \quad$ with $\mathbb{Z}(t) = \mathbb{X}(t)$ a.s., for each t in $[0,1]$.

Corollary 1 (Sample path properties) For any $0 < \delta < a/b$ and any $\epsilon > 0$, there exists a constant $K_\epsilon \equiv K_{\epsilon,\delta,a,b}$ for which the process $\mathbb{Z}$ of (2) satisfies

(3) $\quad P(\,|\mathbb{Z}(s,t]\,| \le K_\epsilon \cdot F(s,t]^\delta$ for all $0 \le s \le t \le 1\,) \ge 1 - \epsilon$.

Proof. Case 1. Suppose that the df F of (1) is $F(t) = t$ on $[0,1]$. Let $0 < \delta < a/b$ be fixed. Let $\lambda \equiv (a/b - \delta)/2$. Define $t_{ni} \equiv i/2^n$ for $0 \le i \le 2^n$ and $n \ge 1$. For $n \ge 0$ define processes $\mathbb{Z}_n : (\Omega, \mathcal{A}, P) \to (C, \mathcal{C})$ by letting

(a) $\quad \mathbb{Z}_n(t) \equiv X(t_{ni}) + 2^n(t - t_{ni})\,[\mathbb{X}(t_{n,i+1}) - \mathbb{X}(t_{ni})] \qquad$ for $t_{ni} \le t \le t_{n,i+1}$,

for each $0 \le i \le 2^n - 1$; thus $\mathbb{Z}_n(\cdot)$ equals $\mathbb{X}(\cdot)$ at each t_{ni} and $\mathbb{Z}_n(\cdot)$ is linear on the intervals between these points. Define

(b) $\quad U_{ni} \equiv |\,\mathbb{X}(t_{n,i-1}, t_{ni}]\,| \qquad$ for $1 \le i \le 2^n$.

If we define

(c) $\quad \Delta_n(t) \equiv \mathbb{Z}_n(t) - \mathbb{Z}_{n-1}(t) \qquad$ for $0 \le t \le 1$,

then for $t_{n-1,i} \le t \le t_{n-1,i+1}$ we have

$$|\Delta_n(t)| \le |\,[\mathbb{X}(t_{n,2i}) + \mathbb{X}(t_{n,2i+2})]/2 - \mathbb{X}(t_{n,2i+1})\,|$$

(d) $\quad = |\mathbb{X}(t_{n,2i}, t_{n,2i+1}] - \mathbb{X}(t_{n,2i+1}, t_{n,2i+2}]|/2 \le [U_{n,2i+1} + U_{n,2i+2}]/2$

(e) $\quad \le [U_{n,2i+1} \vee U_{n,2i+2}]$

for all $n \ge 1$. Thus for all $n \ge 1$ we have

(f) $\quad \|\Delta_n\| \le V_n \equiv [\max_{1 \le i \le 2^n} U_{ni}]\,.$

Let $\theta > 0$ be arbitrary but fixed, and define

$$\text{(g)}\qquad \begin{aligned} p_n &\equiv P(\|\Delta_n\|_\delta > 2\,\theta\,2^{-n\lambda}) \\ &\equiv P(|\Delta_n(s,t]|/(t-s)^\delta > 2\,\theta\,2^{-n\lambda}\ \text{ for some } 0 \le s \le t \le 1). \end{aligned}$$

Recalling (f) shows that

$$\text{(h)}\qquad |\Delta_n(s,t]| \le 2\,V_n \qquad \text{for all } 0 \le s \le t \le 1.$$

Thus

$$\text{(i)}\qquad |\Delta_n(s,t]|/(t-s)^\delta \le 2\,V_n\,2^{n\delta} \qquad \text{for} \quad 2^{-n} \le t-s,$$

while

$$\begin{aligned} &|\Delta_n(s,t]|/(t-s)^\delta \le [|\Delta_n(s,t]|/(t-s)]\,(t-s)^{1-\delta} \le [V_n\,2^n]\,2^{-n(1-\delta)} \\ \text{(j)}\qquad &= V_n\,2^{n\delta} \qquad \text{for} \quad 0 \le t-s \le 2^{-n} \end{aligned}$$

(to see this, consider $|\Delta_n(s,t]/(t-s)|$ when s and t are both points in some $[t_{n,i-1}, t_{ni}]$). Thus for all $n \ge 1$, we have

$$\begin{aligned} p_n &\le P(2\,V_n\,2^{n\delta} > 2\,\theta\,2^{-n\lambda}) \le P(V_n > \theta\,2^{-n(\delta+\lambda)}) \qquad \text{by (g)} \\ &\le \textstyle\sum_{i=1}^{2^n} P(U_{ni} > \theta\,2^{-n(\delta+\lambda)}) \qquad \text{by (f)} \\ &\le \textstyle\sum_{i=1}^{2^n} \mathrm{E}\,U_{ni}^b/[\theta\,2^{-n(\delta+\lambda)}]^b \qquad \text{by Markov's inequality} \\ &\le 2^n[K2^{-n(1+a)}]/[\theta\,2^{-n(\delta+\lambda)}]^b \qquad \text{by (1)} \\ &= K\,\theta^{-b}\,2^{-n(a-b(\delta+\lambda))} = K\,\theta^{-b}\,2^{-nb(a/b-\delta-\lambda)} \\ \text{(k)}\qquad &= K\,\theta^{-b}\,2^{-n\lambda b} \qquad \text{since } a/b-\delta = 2\lambda > 0. \end{aligned}$$

Since $0 \le t-s \le 1$, we also have

$$\begin{aligned} p_0 &\equiv P(\,|\mathbb{Z}_0(s,t]| > 2\,\theta\,(t-s)^\delta \ \text{ for some } 0 \le s \le t \le 1) \\ &\le P(\,|\mathbb{X}(0,1]| > 2\,\theta) \le \mathrm{E}|\mathbb{X}(0,1]|^b/(2\theta)^b \\ \text{(l)}\qquad &\le K\,\theta^{-b} = K\,\theta^{-b}2^{-0\cdot\lambda\,b}. \end{aligned}$$

Now, $\lambda b = (a-\delta b)/2 > 0$, and so $2^{-\lambda\,b} < 1$; hence $\sum_{n=0}^\infty p_n < \infty$. Thus for arbitrarily small θ, we have for m sufficiently large (recall (g) for $\|\cdot\|_\delta$) that

$$\begin{aligned} &P(\max_{m\le k<\infty} \|\mathbb{Z}_k - \mathbb{Z}_m\|_\delta > \theta) = \lim_{n\to\infty} P(\max_{m\le k\le n} \|\mathbb{Z}_k - \mathbb{Z}_m\|_\delta > \theta) \\ &\le \lim_{n\to\infty} P(\max_{m\le k\le n} \|\mathbb{Z}_k - \mathbb{Z}_m\|_\delta > 2\,\theta \textstyle\sum_{m+1}^n 2^{-k\lambda}) \\ &\qquad \text{for } \textstyle\sum_{m+1}^\infty 2^{-k\lambda} < 1/2 \\ &\le \lim_{n\to\infty} P(\max_{m\le k\le n} \|\textstyle\sum_{i=m+1}^k \Delta_i\|_\delta > 2\,\theta \sum_{i=m+1}^n 2^{-i\lambda}) \\ &\le \lim_{n\to\infty} \textstyle\sum_{i=m+1}^n P(\|\Delta_i\|_\delta > 2\,\theta\,2^{-i\lambda}) \\ \text{(m)}\qquad &\le \textstyle\sum_{i=m+1}^\infty p_i \le \sum_{i=m+1}^\infty K\theta^{-b}2^{-i\lambda\,b} \to 0 \qquad \text{as } m \to \infty, \end{aligned}$$

so that

(n) $\quad \mathbb{Z}_n(t) \equiv \mathbb{Z}_0(t) + \sum_{k=1}^{n} \Delta_k(t)$

converges uniformly on $[0,1]$ for a.e. ω; call the limit function $\mathbb{Z}(t)$. Since the uniform limit of continuous functions is continuous,

(o) $\quad \mathbb{Z} = \sum_{n=0}^{\infty} \Delta_n = \lim \mathbb{Z}_n \qquad$ is a continuous function on $[0,1]$ for a.e. ω.

Now, $\mathbb{Z} = \lim \mathbb{Z}_n$, and since $\mathbb{Z}_n$ equals $\mathbb{X}$ at each t_{ni}, we have

(4) $\quad \mathbb{Z}(t_{ni}) = \mathbb{X}(t_{ni}) \qquad$ at each $t_{ni} = i/2^n$ with $0 \le i \le 2^n$ and $n \ge 0$.

Thus all finite-dimensional distributions with diadic rational coordinates are equal. For other t, we pick diadic rationals $t_1, t_2, \ldots$ such that $t_m \to t$. Then $\mathbb{X}(t_m) \to_p \mathbb{X}(t)$ as $m \to \infty$ by (1) and Markov, while $\mathbb{Z}(t_m) \to_{a.s.} \mathbb{Z}(t)$ as $m \to \infty$, since $\mathbb{Z}$ has continuous sample paths. Thus $\mathbb{Z}(t) = \mathbb{X}(t)$ a.s. by proposition 2.3.4. By redefining $\mathbb{Z} \equiv 0$ on the null set of (n), we may assume

(p) $\quad \mathbb{Z} : (\Omega, \mathcal{A}) \to (C, \mathcal{C})$

by theorem 5.4.2. So finite-dimensional distributions agree: In particular, we have $P_Z([x \in C : x_{\mathsf{t}} \in B]) = P_X([x \in R_{[0,1]} : x_{\mathsf{t}} \in B])$ for all sets $B \in \mathcal{B}_k$ and for all $\mathsf{t} \in [0,1]^k$ for any $k \ge 1$.

Case 2. General F. Define

(q) $\quad \mathbb{Y}(t) \equiv \mathbb{X}(F^{-1}(t)) \qquad$ for $0 \le t \le 1$,

where $F^{-1}(t) \equiv \inf\{x \in [0,1] : F(x) \ge t\}$. Then for $0 \le s \le t \le 1$,

$$\mathrm{E}|\mathbb{Y}(s,t]|^b = \mathrm{E}|\mathbb{X}(F^{-1}(s), F^{-1}(t)]\,|^b \le K\,[\,F \circ F^{-1}(t) - F \circ F^{-1}(s)]^{1+a}$$

(r) $$\qquad = K\,(t-s)^{1+a},$$

since $F \circ F^{-1} = I$ for continuous F by exercise 6.3.2. Now use case 1 to replace Y by an equivalent process $\bar{\mathbb{Y}} : (\Omega, \mathcal{A}) \to (C, \mathcal{C})$. Then define

(s) $\quad \mathbb{Z} = \bar{\mathbb{Y}}(F) \cong \mathbb{X}(F^{-1} \circ F) \qquad$ by (q).

Now, $F^{-1} \circ F(t) = t$, unless $F(t-\epsilon) = F(t)$ for some $\epsilon > 0$; see exercise 6.3.2. But in this case equation (1) shows that $\Delta\mathbb{X}$ is 0 across that same interval. Thus $\mathbb{X}(F^{-1} \circ F) \cong \mathbb{X}$.

For the corollary, in case 1 we have (using (o) in line 2, (g) and (k) in line 3)

$$P(|\,\mathbb{Z}(s,t]\,|/(t-s)^{\delta} > 2\,\theta/(1-2^{-\lambda}) \;\text{ for some } 0 \le s \le t \le 1)$$

$$= P(|\textstyle\sum_{k=0}^{\infty} \Delta_k(s,t]\,|/(t-s)^{\delta} > 2\,\theta \sum_{n=0}^{\infty} 2^{-n\lambda} \;\text{ for some } 0 \le s \le t \le 1)$$

$$\le \textstyle\sum_{n=0}^{\infty} p_n \le \sum_{n=0}^{\infty} K\,\theta^{-b}/2^{n\,\lambda\,b} = K\,\theta^{-b}/(1-2^{-\lambda\,b})$$

(t) $$\qquad \to 0 \qquad \text{as } \theta \to \infty.$$

Take K_ϵ to be an appropriately large value of θ. Use the transformation F^{-1} again in case 2. □

Exercise 2.1 Prove (2), by simplifying the proof of theorem 2.1 as much as possible with this simpler goal in mind.

We merely state an analogous result for the existence of processes on $(D,\mathcal{D})$.

Theorem 2.2 (Existence of processes on $(D,\mathcal{D})$; Chentsov) Let

$$\mathbb{X} : (\Omega, \mathcal{A}, P) \to (R_{[0,1]}, \mathcal{B}_{[0,1]}, P_X)$$

be a general process. Suppose that for some $K > 0$, $b > 0$, and $a > \frac{1}{2}$ we have

$$\text{(5)} \qquad \mathrm{E}|\mathbb{X}(r,s]\,\mathbb{X}(s,t]\,|^b \le K \cdot F(r,s]^a F(s,t]^a \qquad \text{for all } 0 \le r \le s \le t \le 1,$$

where F is any df concentrated on $[0,1]$. Then there exists an equivalent version $\mathbb{Z} : (\Omega, \mathcal{A}, P) \to (R_{[0,1]}, \mathcal{B}_{[0,1]}, P_Z)$, which in fact satisfies

$$\text{(6)} \qquad \mathbb{Z} : (\Omega, \mathcal{A}, P) \to (D, \mathcal{D}, P_Z), \quad \text{with } \mathbb{Z}(t) = \mathbb{X}(t) \text{ a.s., for each } t \in [0,1]\,.$$

[See Billingsley (1968, pp. 130, 134), for example.]

Exercise 2.2 Verify the existence of the Poisson process on $(D,\mathcal{D})$.

Exercise 2.3 Let $\mathbb{X} : (\Omega, \mathcal{A}, P) \to (R_T, \mathcal{B}_T)$ on some subinterval T of the line. Let T_o denote a countable dense subset of T, and suppose $P(\omega : \mathbb{X}(\cdot,\omega)$ is uniformly continuous on $T_o \cap I) = 1$ for every finite interval subinterval I of T. Then there exists a version $\mathbb{Z}$ of $\mathbb{X}$ such that every sample path $\mathbb{Z}(\cdot,\omega)$ of $\mathbb{Z}$ is continuous.

3 Brownian Motion and Brownian Bridge

Brownian Motion $\mathbb{S}$ on $[0,1]$ We define $\{\mathbb{S}(t) : 0 \le t \le 1\}$ to be a *Brownian motion* on $[0,1]$ if $\mathbb{S}$ is a normal process having the moment functions

(1) $\quad \mathrm{E}\mathbb{S}(t) = 0 \quad$ and $\quad \mathrm{Cov}[\mathbb{S}(s), \mathbb{S}(t)] = s \wedge t \quad$ for all $0 \le s, t \le 1$.

This covariance function is nonnegative definite (in the sense of (A.4.12)), and these distributions are consistent; thus Kolmogorov's consistency theorem shows that the process $\mathbb{S}$ exists as a random element on $(R_{[0,1]}, \mathcal{B}_{[0,1]})$. Modifying this $\mathbb{S}$ on a set of measure zero (as in theorem 12.2.1), we may may create a version of $\mathbb{S}$ that satisfies

(2) $\quad$ all sample paths of $\mathbb{S}$ are continuous functions on $[0,1]$ that equal 0 at 0.

Thus (as with the smoother realizations of theorem 5.4.2) there is a nice realization of $\mathbb{S}$ having smoother paths; that is,

(3) $\quad \mathbb{S}$ exists as a process on $(C, \mathcal{C})$.

So, Brownian motion exists as the coordinate map $S_t(\omega) \equiv \omega_t$ for some distribution P on $(\Omega, \mathcal{A}) = (C, \mathcal{C})$. This is a more convenient realization of $\mathbb{S}$ (than is the one guaranteed by Kolmogorov's consistency theorem). For either realization

(4) $\quad \mathbb{S}$ has stationary and independent increments.

In fact, its sample paths satisfy

(5) $\quad P(|\mathbb{S}(s,t]| \le K_\epsilon (t-s)^\delta \;$ for all $0 \le s \le t \le 1) \ge 1 - \epsilon$

for some $K_{\epsilon,\delta}$, for any fixed $\epsilon > 0$ and for any fixed $0 < \delta < \frac{1}{2}$. This follows from theorem 12.2.1 and its corollary, since for any $k \ge 1$,

(6) $\quad \mathrm{E}\,\mathbb{S}(s,t]^{2k} = [1 \cdot 3 \cdots (2k-1)]\,(t-s)^k \quad$ for all $0 \le s \le t \le 1$,

and since $a/b = (k-1)/(2k) \nearrow \frac{1}{2}$ as $k \to \infty$. (Note that (5) would allow a further application of the smoother realizations theorem using just this smaller subset of such functions in $\mathcal{C}$.) [No appeal has been made to section 1.]

Brownian Bridge $\mathbb{U}$ on $[0,1]$ Let us now define

(7) $\quad \mathbb{V}(t) \equiv \mathbb{S}(t) - t\,\mathbb{S}(1) \quad$ and $\quad \mathbb{U}(t) \equiv -\mathbb{V}(t) \quad$ for $0 \le t \le 1$.

Then both $\mathbb{U}$ and $\mathbb{V}$ are obviously normal processes on $(C, \mathcal{C})$ and satisfy (5); just observe that $\mathbb{V}(t)$ is a simple linear combination of two normal rvs. Moreover, trivial calculations give

(8) $\quad \mathrm{E}\mathbb{U}(t) = 0 \quad$ and $\quad \mathrm{Cov}[\mathbb{U}(s), \mathbb{U}(t)] = s \wedge t - st \quad$ for all $0 \le s, t \le 1$.

Call $\mathbb{U}$ a *Brownian bridge*. And $\mathbb{V}$ is also a Brownian bridge.

Brownian Motion $\mathbb{S}$ on $[0,\infty)$ Similarly, we establish the existence of Brownian motion on $(C_\infty, \mathcal{C}_\infty)$. In particular, a Brownian motion on $(C_\infty, \mathcal{C}_\infty)$ is given by

(9) $\quad \mathbb{S}(t) = (1+t)\,\mathbb{U}(t/(1+t)), \quad 0 \le t < \infty\,.$

Recall the proposition 8.6.1 LIL result. In section 12.8 we will establish the companion LIL result for Brownian motion that

$$(10)\qquad \limsup_{t\to\infty} |\mathbb{S}(t)|/[\sqrt{t}\, b(t)] = 1 \text{ a.s.} \qquad \text{(the LIL for } \mathbb{S} \text{ at infinity)},$$

where $b(t) \equiv \sqrt{2\,(1 \vee \log\log t)}$. (We will use it in very minor ways in the meantime.) The next exercise similarly defines some additional normal processes. These may provide a useful revisualization device that enables calculation.

Exercise 3.1 (Transformations of Brownian motion) Let $Z \cong \mathrm{N}(0,1)$ and the Brownian bridges $\mathbb{V}$, $\mathbb{U}^{(1)}$, and $\mathbb{U}^{(2)}$ be independent. Fix $a > 0$. Show that:

$$(11)\qquad \mathbb{S}(t) = \mathbb{V}(t) + tZ, \quad 0 \le t \le 1, \qquad \text{is a Brownian motion.}$$

$$(12)\qquad \mathbb{S}(at)/\sqrt{a}, \quad 0 \le t < \infty, \qquad \text{is a Brownian motion.}$$

$$(13)\qquad \mathbb{S}(a+t) - \mathbb{S}(a), \quad t \ge 0, \qquad \text{is a Brownian motion.}$$

$$(14)\qquad \sqrt{1-a}\,\mathbb{U}^{(1)} \pm \sqrt{a}\,\mathbb{U}^{(2)} \qquad \text{is a Brownian bridge,} \qquad \text{if } 0 \le a \le 1.$$

$$(15)\qquad \mathbb{Z}(t) \equiv [\mathbb{U}^{(1)}(t) + \mathbb{U}^{(2)}(1-t)]/\sqrt{2}, \quad 0 \le t \le \frac{1}{2}, \qquad \text{is a Brownian bridge.}$$

$$(16)\qquad \mathbb{U}(t) = (1-t)\,\mathbb{S}(t/(1-t)), \quad 0 \le t \le 1, \qquad \text{is a Brownian bridge;}$$

use the LIL at infinity of (10) to show that this $\mathbb{U}(\cdot)$ converges to 0 at $t = 1$.

$$(17)\qquad t\,\mathbb{S}(1/t), \quad 0 \le t < \infty, \qquad \text{is a Brownian motion;}$$

apply the LIL of (10) to verify that these sample paths converge to 0 at $t = 0$.

Exercise 3.2 (LIL for $\mathbb{S}$ and $\mathbb{U}$ at 0) Use (10), (17), and then (7) to show that

$$(18)\qquad \limsup_{t\to 0} |\mathbb{S}(t)|/[\sqrt{t}\, b(1/t)] = 1 \text{ a.s.}; \quad \text{and} \quad \limsup_{t\to 0} |\mathbb{U}(t)|/[\sqrt{t}\, b(1/t)] = 1 \text{ a.s.}$$

Exercise 3.3 (Integrals of normal processes are normal rvs)
(a) Suppose $\mathbb{X}$ is a normal process on $(C, \mathcal{C})$. Let $\mathbb{X}$ have mean function $m(\cdot)$ continuous on $I \equiv [0,1]$ and covariance function $\mathrm{Cov}(\cdot,\cdot)$ continuous on $I \times I$. Let $g(\cdot) \ge 0$ on I and $q > 0$ on I both be continuous on I. Let $K(\cdot)$ be an $\nearrow$ and left continuous function for which $\int_0^1 q\,|g|\,dK < \infty$. Show that the integrated process

$$(19)\qquad \begin{aligned} &\int_0^1 \mathbb{X}(t)\,g(t)\,dK(t) \\ &\quad \cong N\Big(\int_0^1 m(t)\,g(t)\,dK(t),\ \int_0^1\!\!\int_0^1 \mathrm{Cov}(s,t)\,g(s)g(t)\,dK(s)\,dK(t)\Big)\,, \end{aligned}$$

provided that both $m(s)/q(s)$ and $\mathrm{Cov}[s,s]/q^2(s)$ are continuous for $s \in I$.
(b) Determine the distribution of $\int_0^1 \mathbb{U}(t)\,dt$.
(c) Develop results for $\int_0^1 \mathbb{S}\,g\,dK$, for appropriate functions g and K.
[Hint. (a) The Riemann–Stieltjes sums are normally distributed.]

Exercise 3.4 Let $Z_0, Z_1, Z_2, \ldots$ be iid $N(0,1)$. Let $f_j(t) \equiv \sqrt{2}\sin(j\pi t)$, for $j \geq 1$; these are orthogonal functions. Verify that

(20) $\mathbb{U}(t) \equiv \sum_{j=1}^{\infty} Z_j\, f_j(t)\,/\,j\pi, \quad 0 \leq t \leq 1,$ is a Brownian bridge.

Thus the process $\mathbb{S}(t) \equiv \mathbb{U}(t) + t\,Z_0$ is a Brownian motion on $[0,1]$. Moreover,

(21) $W^2 \equiv \int_0^1 \mathbb{U}^2(t)\,dt \cong \sum_{j=1}^{\infty} Z_j^2/(j^2\,\pi^2).$

This rv has a distribution that is well tabled (the asymptotic null distribution of the Cramér–von Mises statistic).

Exercise 3.5 Show that $\mathbb{Z}$ is a Brownian motion on $[0,1]$, where

(22) $\mathbb{Z}(t) \equiv \mathbb{U}(t) + \int_0^t [\mathbb{U}(s)/(1-s)]\,ds \qquad \text{for } 0 \leq t \leq 1.$

Hint. Since the Reimann sums of normal rvs that define the integral are necessarily normal, the process $\{\mathbb{Z}(t) : 0 \leq t \leq 1\}$ will be a normal process. Then, its mean and covariance function will determine which normal process.

Exercise 3.6 (White noise) (a) Suppose that h and $\tilde{h}$ on $[0,1]$ are in $\mathcal{L}_2$. View *white noise* as an operator $d\mathbb{S}$ that takes the function h into a rv $\int_{[0,1]} h(t)\,d\mathbb{S}(t)$ in the sense of $\to_{\mathcal{L}_2}$. Define this integral first for step functions, and then use exercise 4.4.5 to define it in general. Then show that $\int_{[0,1]} h(t)\,d\mathbb{S}(t)$ exists as such an $\to_{\mathcal{L}_2}$ limit for all h in $\mathcal{L}_2$.
(b) In case h has a bounded continuous derivative h' on $[0,1]$, show that

(23) $\int_{[0,1]} h(t)\,d\mathbb{S}(t) \equiv h\mathbb{S}|_{0-}^{1+} - \int_{[0,1]} \mathbb{S}(t)\,h'(t)\,dt\,.$

(c) Determine the joint distribution of $\int_{[0,1]} h(t)\,d\mathbb{S}(t)$ and $\int_{[0,1]} \tilde{h}(t)\,d\mathbb{S}(t)$.
(d) Define $\int_{[0,1]} h(t)\,d\mathbb{U}(t)$ (appeal first to (7) for the definition), and obtain the marginal and joint distributions of all three of the rvs in (c) and (d).

Exercise 3.7 (Conditional Brownian motion) Let $0 \leq r < s < t$. Determine the conditional distribution of $\mathbb{S}(s)$ given that $\mathbb{S}(r) = y$ and $\mathbb{S}(t) = z$. Draw a figure for this situation, and then put your answer in a format that allows some insight to be offered as to an interpretation.

Exercise 3.8 Find the solution $V(t)$ of the stochastic differential equation with $V'(t) = -kV(t) + \sigma\mathbb{S}'(t)$. Determine its covariance function. (Think of a tiny particle suspended in a liquid whose velocity is impeded by the viscosity of the liquid and is additionally subjected to random changes from collisions with particles in the medium.) [Hint. Rewrite the equation first as $e^{kt}\,[V'(t) + kV(t)] = \sigma\,e^{kt}\,\mathbb{S}'(t)$, then transform it to

$$V(t) = V(0)e^{-t} + \sigma \int_0^t e^{-\sigma(t-s)}\,d\mathbb{S}(s)\,,$$

and then use integration by parts to give meaning to $d\mathbb{S}(\cdot)$.]

Exercise 3.9 Verify Chentsov's condition (12.2.5) for Brownian bridge, that

(24) $\quad \mathrm{E}\{\mathbb{U}_n(r,s]^2\,\mathbb{U}_n(s,t]^2\} \le (\text{some } K)(s-r)(t-s) \qquad \text{for all } 0 \le r \le s \le t \le 1.$

Specify a suitable specific K.

Exercise 3.10 The partial sum process $\{\mathbb{S}_n(t) : 0 \le t < \infty\}$ is defined below in (12.8.1). Verify Chentsov's condition for the case of iid $(0, \sigma^2)$ rvs, that

(25) $\quad \mathrm{E}\{\mathbb{S}_n(r,s]^2\,\mathbb{S}_n(s,t]^2\} \le (\text{some } K)\,\sigma^4\,(s-r)(t-s)$

on the grid with $r = i/n, s = j/n, t = k/n$ and $0 \le i < j < k < \infty$. Specify a suitable specific K. (Then note theorem 14.1.6.)

4 Stopping Times

We first paraphrase the main result of this section. If we observe a right-continuous process at a random time that depends on the process only through its past, then the result is a rv (that is, it is measurable).

Notation 4.1 Let $(\Omega, \mathcal{A}, P)$ denote our basic probability space. We suppose that our time set is a linearly ordered set such as $[0,1]$, $[0,\infty)$, $[0,\infty]$, $\{0,1,2,\ldots\}$, $\{0,1,2,\ldots,\infty\}$. Let X denote a process with such an index set, defined on $(\Omega, \mathcal{A}, P)$. We now suppose that the $\mathcal{A}_t$'s are an $\nearrow$ collection of sub σ-fields of $\mathcal{A}$, in that $\mathcal{A}_s \subset \mathcal{A}_t$ whenever $s < t$. Call such a collection of $\mathcal{A}_t$'s a *filtration*. If it further holds that each X_t is an $\mathcal{A}_t$-measurable rv, then we say that the X-process is *adapted* to the $\mathcal{A}_t$'s. The minimal such collection of $\nearrow$ σ-fields is the *histories* $\sigma_t \equiv \sigma[X_s^{-1}(\mathcal{B}) : s \le t]$. Roughly, σ_t denotes all events for the process up to time t. We let $\mathcal{A}_{t+} \equiv \cap_{n=1}^{\infty} \mathcal{A}_{t+1/n}$; and if $\mathcal{A}_{t+} = \mathcal{A}_t$ for all $t \ge 0$, then we call the σ-fields $\mathcal{A}_t$ *right continuous*. Let $\mathcal{A}_{t-} \equiv \sigma[\mathcal{A}_s : s < t]$. Then let $\mathcal{A}_\infty \equiv \sigma[\cup_{t<\infty} \mathcal{A}_t]$. □

Definition 4.1 (Stopping times) An (extended) rv $\tau \ge 0$ will be called an (*extended*) *stopping time* with respect to the $\mathcal{A}_t$'s if $[\tau \le t] \in \mathcal{A}_t$ for all $t \ge 0$. (That is, one can determine whether $\tau \le t$ using only current knowledge $\mathcal{A}_t$.)

Roughly, whether τ "stops" or "occurs" by time t or not depends only on those events $\mathcal{A}_t$ with probabilities within our knowledge base up through time t. We define the *pre-τ σ-field*

(1) $\qquad \mathcal{A}_\tau \equiv \{A \in \mathcal{A} : A \cap [\tau \le t] \in \mathcal{A}_t \text{ for all } t \ge 0\};$

roughly, at any instant we can decide whether or not A has occurred yet. Note that if $\tau(\omega) \equiv t$ for all ω, then $\mathcal{A}_\tau = \mathcal{A}_t$; that is, the fixed time t is a stopping time whose σ-field is $\mathcal{A}_t$. We now develop some other technical properties.

Proposition 4.1 (Preservation of stopping times) Suppose the rvs $T_1, T_2, \ldots$ are stopping times. Then:

(2) $\qquad T_1 \vee T_2 \quad \text{and} \quad T_1 \wedge T_2 \qquad$ are stopping times.

(3) $\qquad$ If $T_n \nearrow$, then $T \equiv \lim T_n$ is a stopping time.

(4) $\qquad$ If $T_n \searrow$ and $\mathcal{A}_t$'s are right continuous, then $T \equiv \lim T_n$ is a stopping time.

This proposition is also true for extended stopping times.

Proof. Note that these four rvs satisfy

(a) $\qquad [T_1 \vee T_2 \le u] = [T_1 \le u] \cap [T_2 \le u] \in \mathcal{A}_u,$

(b) $\qquad [T_1 \wedge T_2 \le u] = [T_1 \le u] \cup [T_2 \le u] \in \mathcal{A}_u,$

(c) $\qquad [T \le u] = \cap_{n=1}^{\infty} [T_n \le u] = \cap_{n=1}^{\infty} (\text{events in } \mathcal{A}_u) \in \mathcal{A}_u,$

(d) $\qquad [T \le u] = \cap_{m=1}^{\infty} \cup_{n=1}^{\infty} [T_n \le u + 1/m] = \cap_{m=1}^{\infty} (\mathcal{A}_{u+1/m} \text{ events}) \in \mathcal{A}_{u+} = \mathcal{A}_u.$

No change is needed for extended stopping times. □

Proposition 4.2 (Integral stopping times) Integer-valued $T \geq 0$ is a stopping time if and only if $[T = n] \in \mathcal{A}_n$ for all $0 \leq n < \infty$. This result is also true for extended stopping times.

Proof. If T is a stopping time, then $[T = n] = [T \leq n] \cap [T \leq n-1]^c$ is in $\mathcal{A}_n$, since $[T \leq n-1]^c \in \mathcal{A}_{n-1} \subset \mathcal{A}_n$. Also, $[T = \infty] = (\cup_{n=1}^{\infty}[T \leq n])^c \in \mathcal{A}_\infty$. Going the other way, $[T \leq n] = \cup_{m \leq n}[T = m] \in \mathcal{A}_n$, since $[T = m] \in \mathcal{A}_m \subset \mathcal{A}_n$. Also, $[T \leq \infty] = \Omega \in \mathcal{A}_\infty$. □

Exercise 4.1 (Properties of stopping times) Let $T_1, T_2, \ldots$ be (extended) stopping times; no ordering is assumed. Then (using (9) and/or (10) below is ok):

(5) $T_1 + T_2$ is an (extended) stopping time if the $\mathcal{A}_t$'s are right continuous.

(6) $A \in \mathcal{A}_{T_1}$ implies $A \cap [T_1 \leq T_2] \in \mathcal{A}_{T_2}$. [Hint. $[T_1 \wedge t \leq T_2 \wedge t] \in \mathcal{A}_t$.]
$[T_1 < T_2],\ [T_1 = T_2],\ [T_1 > T_2]$ are all in both $\mathcal{A}_{T_1}$ and $\mathcal{A}_{T_2}$.

(7) $T_1 \leq T_2$ implies $\mathcal{A}_{T_1} \subset \mathcal{A}_{T_2}$. Also, $\mathcal{A}_{T_1} \cap [T_1 \leq T_2] \subset \mathcal{A}_{T_1 \wedge T_2} = \mathcal{A}_{T_1} \cap \mathcal{A}_{T_2}$.

(8) If $T_n \searrow T_0$ and the $\mathcal{A}_t$'s are right continuous, then $A_{T_0} = \cap_{n=1}^{\infty} \mathcal{A}_{T_n}$.

[Hint. $[T_1 + T_2 \leq u] = \cap_{m=1}^{\infty}\{\cup [T_1 \leq a + \frac{1}{m}] \cap [T_2 \leq b + \frac{1}{m}] : a + b \leq u,\ \text{rational } a, b\}$.]

Proposition 4.3 (Stopping time measurability) Suppose τ is a stopping time with respect to the $\mathcal{A}_t$'s. Then:

(9) $\mathcal{A}_\tau$ is a σ-field.

(10) τ is $\mathcal{A}_\tau$-measurable.

Proof. Consider (10). For example, for each real number r we have $[\tau \leq r] \in \mathcal{A}_\tau$, since $[\tau \leq r] \cap [\tau \leq t] = [\tau \leq r \wedge t] \in \mathcal{A}_{r \wedge t} \subset \mathcal{A}_t$ for each $t \geq 0$. Thus (10) holds.

Consider (9). Let $A_1, A_2, \ldots$ be in $\mathcal{A}_\tau$. Then

(a) $(\cup A_k) \cap [\tau \leq t] = \cup (A_k) \cap [\tau \leq t]) = \cup (\text{events in } \mathcal{A}_t) \in \mathcal{A}_t$.

Also, $A_1^c \cap [\tau \leq t] = [\tau \leq t] \setminus (A_1 \cap [\tau \leq t]) \in \mathcal{A}_t$. Thus, (9) holds. □

Definition 4.2 (Progressively measurable) Let $\{X(t) : t \geq 0\}$ be a process. Let $\mathcal{B}_t$ denote the Borel subsets of $[0, t]$. Call X *progressively measurable*, and denote this type of measurability by writing

$$X : ([0, t] \times \Omega, \mathcal{B}_t \times \mathcal{A}_t) \to (R, \mathcal{B}) \qquad \text{for each } t \geq 0,$$

provided that for each t in the index set we have

(11) $\{(s, \omega) : 0 \leq s \leq t \text{ and } X(s, \omega) \in B\}$ is in $\mathcal{B}_t \times \mathcal{A}_t$.

Proposition 4.4 (Measurability of stopped right-continuous processes) Let τ be a stopping time with respect to the $\mathcal{A}_t$'s. Also, let $X : (\Omega, \mathcal{A}, P) \to (D_{[0,\infty)}, \mathcal{D}_{[0,\infty)})$ be a process adapted to the $\mathcal{A}_t$'s. Then:

(12) X is progressively measurable.

(13) $X(\tau)$ is $\mathcal{A}_\tau$-measurable.

We may replace $[0, \infty)$ by any $[0, \theta)$ or $[0, \theta]$ with $0 < \theta < \infty$.
(Essentially, one can work nicely with right-continuous processes.)

Proof. That $\mathcal{A}_\tau$ is a σ-field and that τ is $\mathcal{A}_\tau$-measurable are trivial for any image space. (For example, for each real number r we have $[\tau \le r] \in \mathcal{A}_\tau$, since $[\tau \le r] \cap [\tau \le t] = [\tau \le r \wedge t] \in \mathcal{A}_{r\wedge t} \subset \mathcal{A}_t$ for each $t \ge 0$.)

Fix $t > 0$. Now, (12) holds, since $X = \lim X_n$ (by right continuity), where

$$(14)\quad X_n(s,\omega) \equiv \begin{cases} X((k+1)/2^n \wedge t, \omega) & \text{for } k/2^n \le s < (k+1)/2^n,\ k \ge 0,\ s < t, \\ X(t,\omega) & \text{for } s = t \end{cases}$$

clearly satisfies $X_n : ([0,t] \times \Omega, \mathcal{B}_t \times \mathcal{A}_t) \to (R, \mathcal{B})$ is measurable. That is, X_n is progressively measurable. Thus the process X is also, by proposition 2.2.2, since $X_n(\tilde{\omega}) \to X(\tilde{\omega})$ for each $\tilde{\omega} \equiv (s,\omega)$. That is, $X : ([0,t] \times \Omega, \mathcal{B}_t \times \mathcal{A}_t) \to (R, \mathcal{B})$ is measurable (or, X is progressively measurable).

The following type of truncation argument with stopping times is common; learn it. We must show that $[X(\tau) \in B] \cap [\tau \le t] \in \mathcal{A}_t$, for all Borel sets $B \in \mathcal{B}$. But setting $\tau^* \equiv \tau \wedge t$, we see that

$$(15)\quad [X(\tau) \in B] \cap [\tau \le t] = [X(\tau^*) \in B] \cap [\tau \le t] \qquad \text{for} \quad \tau^* \equiv \tau \wedge t,$$

and hence it suffices to show that $[X(\tau^*) \in B] \in \mathcal{A}_t$. Note that the mapping $\omega \to (\tau^*(\omega), \omega)$ is a measurable mapping from $(\Omega, \mathcal{A}_t)$ to $([0,t] \times \Omega, \mathcal{B}_t \times \mathcal{A}_t)$, since for $A \in \mathcal{A}_t$ and the identity function I we have (for sets $[0,s] \times A$ generating $\mathcal{B}_t \times \mathcal{A}_t$)

$$(a)\quad [(\tau^*, I) \in [0,s] \times A] = [\tau^* \le s] \cap A = [\tau \wedge t \le s] \cap A \in \mathcal{A}_t.$$

Combining this $(\tau^*, I) : (\Omega, \mathcal{A}_t) \to ([0,t] \times \Omega, \mathcal{B}_t \times \mathcal{A}_t)$ measurability with the progressive measurability $X : ([0,t] \times \Omega, \mathcal{B}_t \times \mathcal{A}_t) \to (R, \mathcal{B})$ shown above, we see that the composition map $\omega \to X(\tau^*(\omega), \omega)$ is $\mathcal{B}$-$\mathcal{A}_t$-measurable. We express this by writing $[X(\tau^*) \in B] \in \mathcal{A}_t$. (Both (12) and (13) have been established.) □

Exercise 4.2 Let $T \ge 0$ be a rv and let $\{\mathcal{A}_t : t \ge 0\}$ be an increasing sequence of σ-fields. Establish the following facts.
(a) If T is a stopping time, then $[T < t] \in \mathcal{A}_t$ for all $t \ge 0$.
(b) If $[T < t] \in \mathcal{A}_t$ for all $t \ge 0$ and the $\mathcal{A}_t$'s are right continuous, then T is a stopping time.

Exercise 4.3 Let T_1 be a stopping time, and suppose $T_2 \ge T_1$ where T_2 is an $\mathcal{A}_{T_1}$-measurable rv, then T_2 is a stopping time.

Exercise 4.4 Let $0 < a < b < 1$. Let $T_a \equiv \inf\{t : \mathbb{S}(t) = a\}$ for a Brownian motion $\mathbb{S}$ on $(C, \mathcal{C})$. Then T_a is a stopping time, but $\frac{1}{2}T_a$ is not. Also, $A \equiv [\,\|\mathbb{S}\| \ge a]$ is in $\mathcal{A}_{T_a}$, but $B \equiv [\,\|\mathbb{S}\| \ge b]$ is not.

Definition 4.3 (Augmented filtration) Let $(\Omega, \hat{A}, P)$ denote the completion of the probability space $(\Omega, \mathcal{A}, P)$. Let $\mathcal{N} \equiv \{N \in \mathcal{A} : P(N) = 0\}$ denote all null sets. Let $\{\mathcal{A}_t : t \ge 0\}$ be an $\nearrow$ sequence of σ-fields; that is, the $\mathcal{A}_t$'s form a filtration. If all $\mathcal{A}_t = \mathcal{A}_{t+}$, the $\mathcal{A}_t$'s are called *right-continuous*. If all $\mathcal{A}_t = \sigma[\mathcal{A}_t, \mathcal{N}]$, then they are said to be a *complete filtration*. If a filtration $\{\mathcal{A}_t : t \ge 0\}$ is both complete and right-continuous, such a collection of σ-fields is called an *augmented filtration*.

Proposition 4.5 (Kallenberg) Let $\{\mathcal{A}_t : t \ge 0\}$ denote an $\nearrow$ sequence of σ-fields on the probability space $(\Omega, \mathcal{A}, P)$; that is, the $\mathcal{A}_t$'s form a filtration. Consider the null sets $\mathcal{N} \equiv \{N \in \mathcal{A} : P(N) = 0\}$. Define $\hat{\mathcal{A}}_t \equiv \sigma[\mathcal{A}_t, \mathcal{N}]$.

(a) Then the $\nearrow$ collection of σ-fields $\hat{\mathcal{A}}_{t+}$ necessarily equals the completion $\widehat{\mathcal{A}_{t+}}$ of the right-continuous filtration $\mathcal{A}_{t+}$, and so forms an augmented filtration for $(\Omega, \hat{\mathcal{A}}, P)$. Moreover, this is the minimal augmented filtration.

(b) If the $\mathcal{A}_t \equiv \sigma_t$ denote the histories of a right-continuous process $X : (\Omega, \mathcal{A}, P) \to (D_{[0,\infty)}, \mathcal{D}_{[0,\infty)})$, then the completion of the right-continuized histories necessarily forms the minimal augmented filtration. (That is, complete each $\sigma_{t+} \equiv \cap_{n=1}^{\infty} \sigma_{t+1/n}$.)

(c) If $S \le T$ a.s. then $\mathcal{A}_S \subset \mathcal{A}_T$ relative to the augmented filtration $\widehat{\sigma_{t+}} = \hat{\sigma}_{t+}$.

Proof. It is trivial that $\widehat{\mathcal{A}_{t+}} \subset \widehat{\hat{\mathcal{A}}_{t+}} = \hat{\mathcal{A}}_{t+}$. To show the converse, consider a set $A \in \hat{\mathcal{A}}_{t+}$. Then for each $n \ge 1$ we have $A \in \hat{\mathcal{A}}_{t+1/n}$, so $P(A \triangle A_n) = 0$ for some set $A_n \in \mathcal{A}_{t+1/n}$. Note the $A^* \equiv \overline{\lim}_n A_n$ is in $\mathcal{A}_{t+}$, while $P(A \triangle A^*) = 0$ since $A \triangle A^* \subset \cup_1^{\infty}(A \triangle A_n) = \cup\{\text{null}\}_n = \{\text{null}\}$; thus $A \in \widehat{\mathcal{A}_{t+}}$. Thus the main claim in (a) is established. Let $\mathcal{F}_t$ denote any other augmented filtration for which all $\mathcal{F}_t \supset \mathcal{A}_t$. Then $\widehat{\mathcal{A}_{t+}} = \hat{\mathcal{A}}_{t+} \subset \hat{\mathcal{F}}_{t+} = \mathcal{F}_{t+} = \mathcal{F}_t$, as claimed. Part (b) follows at once. Part (c) follows from exercise 12.4.1 on properties of stopping times. □

Example 4.1 (Haeusler) Let both A and A^c be measurable subsets of some $(\Omega, \mathcal{A}, P)$ that have probability exceeding 0. Define $X_t(\omega)$ on $0 \le t \le 1$ to be identically 0 if $\omega \in A$ and to equal $(t - 1/2) \cdot 1_{[1/2,1]}(t)$ if $\omega \in A^c$. All paths of this X-process are continuous. Since X_t is always 0 for $0 \le t \le 1/2$, we have $\sigma_t = \{\emptyset, \Omega\}$ for $0 \le t \le 1/2$. However, $\sigma_t = \{\emptyset, A, A^c, \Omega\}$ for $1/2 < t \le 1$. These histories σ_t are not right continuous at $t = 1/2$. The right continuized histories σ_{t+} equal $\{\emptyset, \Omega\}$ for $0 \le t < 1/2$ and equal $\{\emptyset, A, A^c, \Omega\}$ for $1/2 \le t \le 1$. They are already complete, so $\hat{\sigma}_{t+} = \widehat{\sigma_{t+}} = \sigma_{t+}$. Now, proposition 4.5(c) could be applied. □

5 Strong Markov Property

We now extend the strong Markov property (which was proved for discrete-time processes in section 8.6) to processes with stationary and independent increments.

Theorem 5.1 (Strong Markov property) Consider the stochastic process $X : (\Omega, \mathcal{A}, P) \to (D_{[0,\infty)}, \mathcal{D}_{[0,\infty)})$ adapted to right-continuous $\mathcal{A}_t$'s. Suppose that $X(0) = 0$, X has stationary and independent increments, and suppose that the increment $X(t+s) - X(t)$ is independent of $\mathcal{A}_t$ for all $s \geq 0$. Let τ be an extended stopping time for the $\mathcal{A}_t$'s, and suppose $P(\tau < \infty) > 0$. For some $t \geq 0$ we define

$$(1) \qquad Y(t) \equiv \begin{cases} X(\tau + t) - X(\tau) & \text{on } [\tau < \infty], \\ 0 & \text{on } [\tau = \infty]. \end{cases}$$

Then $Y : \Big([\tau < \infty] \cap \Omega,\ [\tau < \infty] \cap \mathcal{A},\ P(\,\cdot\,\Big|[\tau < \infty])\Big) \to (D_{[0,\infty)}, \mathcal{D}_{[0,\infty)})$ and

$$(2) \qquad P\Big(Y \in F\Big|[\tau < \infty]\Big) = P(X \in F) \qquad \text{for all } F \in \mathcal{D}_{[0,\infty)}.$$

Moreover, for all $F \in \mathcal{D}_{[0,\infty)}$ and for all $A \in \mathcal{A}_\tau$, we have

$$(3) \qquad P\Big([Y \in F] \cap A\Big|[\tau < \infty]\Big) = P(\,[X \in F]\,) \times P\Big(A\Big|[\tau < \infty]\Big).$$

Thus if $P(\tau < \infty) = 1$, then X and Y are equivalent processes and the process Y is independent of the σ-field $\mathcal{A}_\tau$.

Proof. That $Y : (\Omega \cap [\tau < \infty], \mathcal{A} \cap [\tau < \infty], P(\cdot|[\tau < \infty]) \to (D_{[0,\infty)}, \mathcal{D}_{[0,\infty)})$ follows from proposition 12.4.3. This proposition and exercise 12.4.1 show that

$$(4) \qquad \mathcal{A}'_t \equiv \mathcal{A}_{\tau+t} \text{ are } \nearrow \text{ and right continuous, with } Y \text{ adapted to the } \mathcal{A}'_t\text{'s.}$$

Case 1. Suppose the finite part of the range of τ is a countable subset $\{s_1, s_2, \ldots\}$ of $[0, \infty)$. Let $t_1, \ldots, t_m \geq 0$, let $B_1, \ldots, B_m$ be Borel subsets of the real line, and let $A \in \mathcal{A}_\tau$. Then

$$\begin{aligned}
&P([Y(t_1) \in B_1, \ldots, Y(t_m) \in B_m] \cap A \cap [\tau < \infty]) \\
&\quad = \textstyle\sum_k P([Y(t_1) \in B_1, \ldots] \cap A \cap [\tau = s_k]) \\
&\quad = \textstyle\sum_k P([X(t_1 + s_k) - X(s_k) \in B_1, \ldots] \cap A \cap [\tau = s_k]) \\
&\quad = \textstyle\sum_k P(X(t_1 + s_k) - X(s_k) \in B_1, \ldots)\, P(A \cap [\tau = s_k]) \\
&\quad = P(X(t_1) \in B_1, \ldots) \textstyle\sum_k P(A \cap [\tau = s_k]) \\
\text{(a)} \qquad &\quad = P(X(t_1) \in B_1, \ldots, X(t_m) \in B_m)\, P(A \cap [\tau < \infty]),
\end{aligned}$$

where the third equality holds as $A \cap [\tau = s_k] = (A \cap [\tau \leq s_k]) \cap [\tau = s_k]$ is in $\mathcal{A}_{s_k}$, and is thus independent of the other event by the independent increments of X.

Putting $A = [\tau < \infty]$ in (a) yields

$$P(Y(t_1) \in B_1, \ldots, Y(t_m) \in B_m \,|[\tau < \infty])$$

$$\text{(b)} \qquad = P(X(t_1) \in B_1, \ldots, X(t_m) \in B_m);$$

substituting (b) into (a) and dividing by $P(\tau < \infty)$ yields

$$P([Y(t_1) \in B_1, \ldots, Y(t_m) \in B_m] \cap A|[\tau < \infty])$$

$$\text{(c)} \qquad = P(Y(t_1) \in B_1, \ldots, Y(t_m) \in B_m|[\tau < \infty])\, P(A|[\tau < \infty]).$$

Thus (b) and (c) hold for the class $\mathcal{G}$ of sets of the form $[Y(t_1) \in B_1, \ldots, Y(t_m) \in B_n]$ and for all sets A in $\mathcal{A}_\tau$. Since $\mathcal{G}$ generates $Y^{-1}(\mathcal{D}_{[0,\infty)})$, equation (b) implies (2). Since $\mathcal{G}$ is also closed under finite intersections (that is, it is a $\bar{\pi}$-system), (c) and proposition 7.1.1 imply (3).

Case 2. Now consider a general stopping time τ. For $n \geq 1$, define

$$\text{(d)} \qquad \tau_n \equiv \begin{cases} k/n & \text{for } (k-1)/n < \tau \leq k/n \text{ and } k \geq 1, \\ 1/n & \text{for } \tau = 0, \\ \infty & \text{for } \tau = \infty\ . \end{cases}$$

Note that $\tau_n(\omega) \searrow \tau(\omega)$ for $\omega \in [\tau < \infty]$. For $k/n \leq t < (k+1)/n$ we have

$$[\tau_n \leq t] = [\tau \leq k/n] \in \mathcal{A}_{k/n} \subset \mathcal{A}_t$$

(so that τ_n is a stopping time), and also for A in $\mathcal{A}_\tau$ that

$$A \cap [\tau_n \leq t] = A \cap [\tau \leq k/n] \in \mathcal{A}_{k/n} \subset \mathcal{A}_t$$

(so that $\mathcal{A}_\tau \subset \mathcal{A}_{\tau_n}$). Define

$$\text{(e)} \qquad Y_n(t) = X(\tau_n + t) - X(\tau_n) \qquad \text{on } [\tau_n < \infty] = [\tau < \infty],$$

and let it equal 0 elsewhere. By case 1 results (b) and (c), both

$$\text{(f)} \qquad P(Y_n \in F\,|[\tau < \infty]) = P(X \in F) \qquad \text{and}$$

$$\text{(g)} \qquad P([Y_n \in F] \cap A\,|[\tau < \infty]) = P(Y_n \in F\,|[\tau < \infty])\, P(A\,|[\tau < \infty])$$

hold for all F in $\mathcal{D}_{[0,\infty)}$ and all A in $\mathcal{A}_\tau$ (recall that $\mathcal{A}_\tau \subset \mathcal{A}_{\tau_n}$ as above, and $[\tau < \infty] = [\tau_n < \infty]$). Let $(r_1, \ldots, r_m)$ denote any continuity point of the joint df of the finite dimensional random vector $(Y(t_1), \ldots, Y(t_m))$, and define

$$\text{(h)} \qquad \begin{aligned} G_n &\equiv [Y_n(t_1) < r_1, \ldots, Y_n(t_m) < r_m, \tau < \infty], \\ G &\equiv [Y(t_1) < r_1, \ldots, Y(t_m) < r_m, \tau < \infty], \\ G^* &\equiv [Y(t_1) \leq r_1, \ldots, Y(t_m) \leq r_m, \tau < \infty], \\ H &\equiv [X(t_1) < r_1, \ldots, X(t_m) < r_m]. \end{aligned}$$

By the right continuity of the sample paths, $Y_n(t) \to Y(t)$ for every t and every ω in $[\tau < \infty]$; thus

$$\text{(i)} \qquad G \subset \underline{\lim}\, G_n \subset \overline{\lim}\, G_n \subset G^*.$$

Thus

$$P(G|\tau<\infty) \le P(\underline{\lim}\, G_n|[\tau<\infty]) \le \underline{\lim}\, P(G_n|\tau<\infty) \quad \text{by (i), then DCT}$$

$$= P(H) = \overline{\lim}\, P(G_n|\tau<\infty) \qquad \text{by using (f) twice}$$

$$\le P(\overline{\lim}\, G_n|\tau<\infty) \le P(G^*|\tau<\infty) \qquad \text{by the DCT and (i)}$$

$$\le P(G|\tau<\infty) + \textstyle\sum_{i=1}^m P(Y(t_i)=r_i|\tau<\infty)$$

(j) $$= P(G|\tau<\infty),$$

since $(r_1,\ldots,r_m)$ is a continuity point. Thus (j) implies

(k) $$P(G|\tau<\infty) = P(H),$$

and this is sufficient to imply (2). Likewise, for $A \in \mathcal{A}_\tau \subset \mathcal{A}_{\tau_n}$,

$$P(G\cap A|[\tau<\infty]) \le P(\underline{\lim}\, G_n \cap A|\tau<\infty) \qquad \text{by (i)}$$

$$\le \underline{\lim}\, P(G_n\cap A|\tau<\infty) \quad \text{by the DCT}$$

$$= \underline{\lim}\, P(G_n|\tau<\infty)\, P(A|\tau<\infty) \qquad \text{by (c), with } [\tau<\infty]=[\tau_n<\infty]$$

$$= P(G|\tau<\infty)\, P(A|\tau<\infty) \qquad \text{by (j)}$$

$$= \overline{\lim}\, P(G_n|\tau<\infty)\, P(A|\tau<\infty) \qquad \text{by (j)}$$

$$= \overline{\lim}\, P(G_n\cap A|\tau<\infty) \qquad \text{by (c), with } [\tau<\infty]=[\tau_n<\infty]$$

$$\le P(\overline{\lim}\, G_n\cap A|\tau<\infty) \le P(G^*\cap A|\tau<\infty) \qquad \text{by the DCT, then (i)}$$

$$\le P(G\cap A|\tau<\infty) + \textstyle\sum_{i=1}^m P(Y(t_i)=r_i|\tau<\infty)$$

(l) $$= P(G\cap A|\tau<\infty),$$

since $(r_1,\ldots,r_m)$ is a continuity point. Thus (l) implies

(m) $$P(G\cap A|\tau<\infty) = P(G|\tau<\infty)\, P(A|\tau<\infty);$$

and using proposition 7.1.1 again, we see that this is sufficient to imply (3).

The final statement is immediate, since when $P(\tau<\infty)=1$ we necessarily have $P(A|\tau<\infty)=P(A)$ for all $A\in\mathcal{A}$. □

6 Embedding a RV in Brownian Motion

Let $a, b > 0$. For a Brownian motion $\mathbb{S}$ on $(C_\infty, \mathcal{C}_\infty)$, we define

(1) $\qquad \tau \equiv \tau_{ab} \equiv \inf\{t : \mathbb{S}(t) \in (-a, b)^c\}$

to be the first time $\mathbb{S}$ hits either $-a$ or b. Call τ a *hitting time*. [Show that τ is a stopping time.] Note figure 6.1.

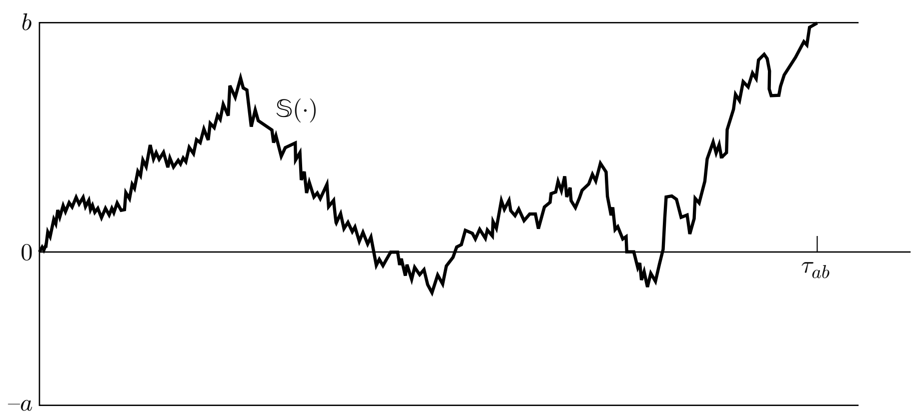


Figure 6.1 The stopping time τ_{ab}.

Theorem 6.1 (Embedding via τ_{ab}) Let $\tau \equiv \tau_{ab}$. Then:

(2) $\qquad \mathrm{E}\mathbb{S}(\tau) = 0\,.$

(3) $\qquad P(\mathbb{S}(\tau) = -a) = b/(a+b) \qquad$ and $\qquad P(\mathbb{S}(\tau) = b) = a/(a+b)\,.$

(4) $\qquad \mathrm{E}\tau = a\,b = \mathrm{E}\mathbb{S}^2(\tau) \qquad$ and $\qquad \mathrm{E}\,\tau^2 \le 4ab(a+b)\,.$

(5) $\qquad \mathrm{E}\tau^r \le r\,\Gamma(r)\,2^{2r}\,\mathrm{E}\mathbb{S}^{2r}(\tau) \le r\,\Gamma(r)\,2^{2r}\,ab\,(a+b)^{2r-2} \qquad$ for all $r \ge 1\,.$

Definition 6.1 (Martingale) A process $\{M(t) : t \ge 0\}$ is a continuous parameter *martingale* (*mg*) if $\mathrm{E}|M(t)| < \infty$ for all t, M is adapted to the $\mathcal{A}_t$'s, and

(6) $\qquad \mathrm{E}\{M(t)\,|\mathcal{A}_s\} =_{a.s.} M(s) \qquad$ for all $0 \le s \le t.$

Definition 6.2 (Stopping time) If τ is a random time (just a rv that is ≥ 0) for which the event $[\tau \le t] \in \mathcal{A}_t$ for all t, then we call τ a *stopping time*.

Future theorem Let τ be a stopping time. With appropriate regularity conditions on a mg M, we can claim that

(7) $\quad \mathrm{E}M(\tau) = \mathrm{E}M(0)\,.$

Our present applications are simple special cases of a result called the *optional sampling theorem* for mgs. The general version will be proven in chapter 18. We will use it for such simple special cases now. □

Proof. The independent increments of $\mathbb{S}$ lead to satisfaction of the mg property stated in (6). Also, $\mathbb{S}$ is suitably integrable (we will see later) for (7) to hold (note (13.6.9) and (13.6.16)). Thus, with $p \equiv P(\mathbb{S}(\tau) = b)$, we have

(a) $\quad 0 = \mathrm{E}\mathbb{S}(\tau) = bp - a(1-p), \qquad \text{or} \qquad p = a/(a+b)\,.$

Also, the process

(8) $\quad \{\mathbb{S}^2(t) - t : t \ge 0\} \qquad$ is a mg adapted to the σ-fields $\mathcal{A}_t \equiv \sigma_t\,,$

since

$$\mathrm{E}\{\mathbb{S}^2(t) - t|\mathcal{A}_s\} = \mathrm{E}\{[\mathbb{S}(t) - \mathbb{S}(s) + \mathbb{S}(s)]^2 - t|\mathcal{A}_s\}$$

$$= \mathrm{E}\{[\mathbb{S}(t) - \mathbb{S}(s)]^2 + 2\,\mathbb{S}(s)[\mathbb{S}(t) - \mathbb{S}(s)] + \mathbb{S}^2(s) - t|\mathcal{A}_s\}$$

$$= \mathrm{E}\{[\mathbb{S}(t) - \mathbb{S}(s)]^2\} + 2\,\mathbb{S}(s)\mathrm{E}\{\mathbb{S}(t) - \mathbb{S}(s)\} + \mathbb{S}^2(s) - t$$

$$= t - s + 2\,\mathbb{S}(s)\cdot 0 + \mathbb{S}^2(s) - t$$

(b) $\quad = \mathbb{S}^2(s) - s\,.$

This process is also suitably integrable, so that optional sampling can be used to imply $\mathrm{E}[\mathbb{S}(\tau)^2 - \tau] = 0$. Thus

(c) $\quad \mathrm{E}\tau = \mathrm{E}\mathbb{S}^2(\tau) = (-a)^2 \cdot b/(a+b) + b^2 \cdot a/(a+b) = ab\,.$

We leave (5) to exercise 12.7.3 below. □

Theorem 6.2 (Skorokhod embedding of a zero-mean rv) Suppose X is a rv with df F having mean 0 and variance $0 \le \sigma^2 \le \infty$. Then there is a stopping time τ such that the stopped rv $\mathbb{S}(\tau)$ is distributed as X; that is,

(9) $\quad \mathbb{S}(\tau) \cong X.$

Moreover,

(10) $\quad \mathrm{E}\tau = \mathrm{Var}[X] \qquad \text{and} \qquad \mathrm{E}\,\tau^2 \le 16\,\mathrm{E}\,X^4.$

and for any $r \ge 1$ we have

(11) $\quad \mathrm{E}\tau^r \le K_r\,\mathrm{E}|X|^{2r} \qquad \text{with} \qquad K_r \equiv r\,\Gamma(r)\,2^{4r-2}\,.$

Proof. For degenerate F, just let $\tau \equiv 0$. Thus suppose F is nondegenerate. Let (A, B) be independent of $\mathbb{S}$, with joint df H having

$$(12) \qquad dH(a,b) = (a+b)\, dF(-a)\, dF(b)/EX^+ \qquad \text{for } a \geq 0, b > 0\,.$$

The procedure is to observe $(A, B) = (a, b)$ according to H, and then to observe τ_{ab}, calling the result τ. (Clearly, $\tau_{ab} = 0$ if $a = 0$ is chosen.) Note that $[\tau \leq t]$ can be determined by (A, B) and $\{\mathbb{S}(s) : 0 \leq s \leq t\}$, and hence is an event in $\mathcal{A}_t \equiv \sigma[A, B, \mathbb{S}(s) : 0 \leq s \leq t]$. For $t \geq 0$,

$$P(\mathbb{S}(\tau) > t) = \mathrm{E}(P\{\mathbb{S}(\tau) > t | A = a, B = b\})$$

$$\text{(a)} \qquad = \int_{[0,\infty)} \int_{(0,t]} 0 \cdot dH(a,b) + \int_{[0,\infty)} \int_{(t,\infty)} (a/(a+b))\, dH(a,b) \qquad \text{by (3)}$$

$$\text{(b)} \qquad = \int_{(t,\infty)} \int_{[0,\infty)} a\, dF(-a)\, dF(b) \,/\, \mathrm{E}X^+ = \int_{(t,\infty)} dF(b)\, \mathrm{E}X^- \,/\, \mathrm{E}X^+$$

$$\text{(c)} \qquad = 1 - F(t)\,,$$

since $EX = 0$ with X nondegenerate implies $\mathrm{E}X^+ = \mathrm{E}X^-$. Likewise, for $t \geq 0$,

$$\text{(d)} \qquad P(\mathbb{S}(\tau) \leq -t) = \int_{[0,t)} \int_{(0,\infty)} 0 \cdot dH(a,b) + \int_{[t,\infty)} \int_{(0,\infty)} (b/(a+b))\, dH(a,b)$$

$$\text{(e)} \qquad = \int_{[t,\infty)} \int_{(0,\infty)} b\, dF(b)\, dF(-a) \,/\, \mathrm{E}X^+ = \int_{[t,\infty)} dF(-a)$$

$$\text{(f)} \qquad = F(-t)\,.$$

Thus $\mathbb{S}(\tau) \cong X$. Moreover,

$$\mathrm{E}\tau = \mathrm{E}(\mathrm{E}\{\tau | A = a, B = b\}) = \mathrm{E}(\mathrm{E}\{\mathbb{S}^2(\tau) | A = a, B = b\}) = \mathrm{E}\mathbb{S}^2(\tau)$$

$$\text{(g)} \qquad = \mathrm{E}X^2 = \mathrm{Var}[X]\,.$$

Note that $(a+b)^{2r-1} \leq 2^{2r-2}[a^{2r-1} + b^{2r-1}]$ by the C_r-inequality. Thus

$$\mathrm{E}\tau^r = \mathrm{E}(\mathrm{E}\{\tau^r | A = a, B = b\})$$

$$\text{(h)} \qquad \leq 2^{2r}\, r\, \Gamma(r)\, \mathrm{E}(AB(A+B)^{2r-2}) \qquad \text{by (5)}$$

$$\leq 2^{2r}\, r\, \Gamma(r)\, \mathrm{E}(AB(A+B)^{2r-1}/(A+B))$$

$$\text{(i)} \qquad \leq r\, \Gamma(r)\, 2^{4r-2}\, \mathrm{E}\left(\frac{B}{A+B} A^{2r} + \frac{A}{A+B} B^{2r} \right)$$

$$\text{(j)} \qquad = K_r\, \mathrm{E}(\mathrm{E}\{\mathbb{S}^{2r}(\tau) | A = a, B = b\}) = K_r\, \mathrm{E}(\mathbb{S}^{2r}(\tau)) = K_r\, \mathrm{E}X^{2r}\,,$$

as claimed. □

7 Barrier Crossing Probabilities

For $-a < 0 < b$ we defined the *hitting time*

(1) $\tau_{ab} \equiv \inf\{t : \mathbb{S}(t) \in (-a,b)^c\}$,

where $\mathbb{S}$ denotes Brownian motion on $(C_\infty, \mathcal{C}_\infty)$. We also considered the rv $\mathbb{S}(\tau_{ab})$, which is called Brownian motion *stopped* at τ_{ab}. We saw that it took on the two values b and $-a$ with the probabilities $p \equiv a/(a+b)$ and $q \equiv 1-p = b/(a+b)$.

For $a > 0$ we define the stopping time (the *hitting time* of a)

(2) $\tau_a \equiv \inf\{t : \mathbb{S}(t) \ge a\}$.

[Now, $[\tau_a < c] = \cap_{q<a} \cup_{r<c} [\mathbb{S}(r) > q]$ (over rational p and q) shows that τ_a is a stopping time.] The LIL of (8.6.1) shows that both τ_{ab} and τ_a are finite a.s.

Theorem 7.1 (The reflection principle; Bachelier) Both

(3) $P(\sup_{0\le t\le c} \mathbb{S}(t) > a) = P(\tau_a < c) = 2\,P(\mathbb{S}(c) > a)$

$= 2\,P(N(0,1) \ge a/\sqrt{c}\,)$ for $a > 0$ and

(4) $P(\|\mathbb{S}\|_0^1 > a) = 4\sum_{k=1}^\infty P((4k-3)\,a < N(0,1) < (4k-1)\,a)$

(5) $$= 1 - \frac{4}{\pi}\sum_{k=0}^\infty \frac{(-1)^k}{2k+1}\,\exp\left(-\frac{(2k+1)^2\pi^2}{8\,a^2}\right) \qquad \text{for } a > 0.$$

Proof. Define the stopping time $\tau'_a \equiv \tau_a \wedge c$, and note that $\tau_a = \tau'_a$ on the event $[\mathbb{S}(c) > a]$. Now, $[\tau'_a < c] \in \mathcal{A}_{\tau'_a}$ is independent of the Brownian motion $\{\mathbb{Y}(t) \equiv \mathbb{S}(\tau'_a + t) - \mathbb{S}(\tau'_a) : t \ge 0\}$, by strong Markov, with $P(\tau'_a < \infty) = 1$. In (b) below we will use that $\mathbb{S}(\tau'_a) = a$ on $[\mathbb{S}(c) > a]$. We have

$P(\tau_a < c) = P(\tau'_a < c)$

(a) $= P([\tau'_a < c] \cap [\mathbb{S}(c) > a]) + P([\tau'_a < c] \cap [\mathbb{S}(c) < a]) + 0$

(b) $= P([\tau'_a < c] \cap [\mathbb{S}(c) - \mathbb{S}(\tau'_a) > 0]) + P([\tau'_a < c] \cap [\mathbb{S}(c) - \mathbb{S}(\tau'_a) < 0])$

(c) $= 2\,P([\tau'_a < c] \cap [\mathbb{S}(c) - \mathbb{S}(\tau_a) > 0])$ using the strong Markov property

(d) $= 2\,P(\mathbb{S}(c) > a)$,

since the events in (c) and (d) are identical.

The two-sided boundary of formula (4) follows from a more complicated reflection principle. Let $A_+ \equiv [\|\mathbb{S}^+\| > a] = [\mathbb{S}$ exceeds a somewhere on $[0,1]]$ and $A_- \equiv [\,\|\mathbb{S}^-\| > a] = [\mathbb{S}$ falls below $-a$ somewhere on $[0,1]]$. Though $[\,\|\mathbb{S}\| > a] = A_+ \cup A_-$, we have $P(\|\mathbb{S}\| > a) < P(A_+) + P(A_-)$, since we included paths that go above a and then below $-a$ (or vice versa) twice. By making the first reflection in figure 7.1, we see that the probability of the former event equals that of $A_{+-} = [\,\|\mathbb{S}^+\| > 3a]$, while that of the latter equals that of $A_{-+} = [\,\|\mathbb{S}^-\| > 3a]$. But subtracting out these probabilities from $P(A^+) + P(A^-)$ subtracts out too much, since the path may then have recrossed the other boundary; we compensate for this by adding back in the probabilities of $A_{+-+} \equiv [\,\|\mathbb{S}^+\| > 5a]$ and $A_{-+-} \equiv [\,\|\mathbb{S}^-\| > 5a]$, which a second reflection shows to be equal to the appropriate probability. But we must continue this process ad infinitum. Thus

(e) $$P(\|\mathbb{S}\|_0^1 > a) = \begin{cases} P(A_+) - P(A_{+-}) + P(A_{+-+}) - \cdots + \\ P(A_-) - P(A_{-+}) + P(A_{-+-}) - \cdots \end{cases}$$

(f) $$= 2\,[P(A_+) - P(A_{+-}) + P(A_{+-+}) - \cdots] \qquad \text{by symmetry}$$

$$= 2\sum_{k=1}^{\infty} (-1)^{k+1}\, 2\, P(N(0,1) > (2k-1)\,a) \qquad \text{by (3)}$$

(g) $$= 4\sum_{k=1}^{\infty} P((4k-3)\,a < N(0,1) < (4k-1)\,a)$$

as claimed. The final expression (5) is left for the reader; it is reputed to converge more quickly. □

Exercise 7.1 Prove (5). (See Chung (1974, p. 223).)

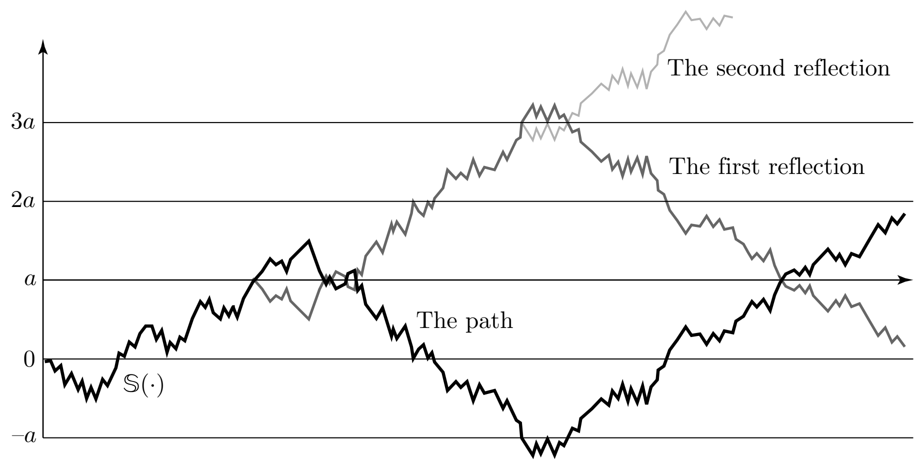


Figure 7.1 The reflection principle for Brownian motion.

Theorem 7.2 (The reflection principle for linear boundaries; Doob) Consider the line $c\,t + d$ with $c \ge 0, d > 0$. Then:

(6) $$P(\mathbb{S}(t) \ge ct + d \quad \text{for some } t \ge 0) = \exp(-2\,c\,d)\,.$$

(7) $$P(|\mathbb{S}(t)| \ge c\,t + d \quad \text{for some } t \ge 0) = 2\sum_{k=1}^{\infty} (-1)^{k+1} \exp(-2\,k^2 c\,d)\,.$$

Proof. Now, for any $\theta \ne 0$ the process

(8) $$\{V(t) \equiv \exp(\theta\,[\mathbb{S}(t) - \theta\,t/2]) : t \ge 0\} \qquad \text{is a mg} \quad (\text{with } V(0) \equiv 1).$$

This holds with $\sigma_t \equiv \sigma[\mathbb{S}(s) : s \le t]$ (using the mgf of a normal rv), since

$$\mathrm{E}\{V(t)|\sigma_s\} = \mathrm{E}\{\exp(\theta[\mathbb{S}(s) - \theta\,s/2] + \theta\,[\mathbb{S}(s,t] - \theta(t-s)/2])\,|\,\sigma_s\}$$

(a) $$= V(s)\,\mathrm{E}\{\exp(\theta\,N(0,t-s))\}\ \exp(-\theta^2\,(t-s)/2)$$

(b) $$= V(s).$$

Thus if we now redefine τ_{ab} as $\tau_{ab} \equiv \inf\{t : \mathbb{X}(t) \equiv \mathbb{S}(t) - \theta t/2 \in (-a,b)^c\}$, where we have $a > 0, b > 0$, then $V(t) = e^{\theta\,\mathbb{X}(t)}$. Hence the "future theorem" gives

(c) $\quad 1 = \mathrm{E}V(\tau_{ab}) = P(\mathbb{X}(\tau_{ab}) = -a)\, e^{-\theta a} + P(\mathbb{X}(\tau_{ab}) = b)\, e^{\theta b},$

so that

(9) $\quad P(\mathbb{X}(\tau_{ab}) = b) = (1 - e^{-\theta a})/(e^{\theta b} - e^{-\theta a})$

(d) $\quad \to e^{-\theta b} \qquad$ if $\theta > 0$ and $a \to \infty$

(e) $\quad = e^{-2cd} \qquad$ if $\theta = 2c$ and $b = d$.

But this same quantity also satisfies (by proposition 1.1.2)

(f) $\quad P(\mathbb{X}(\tau_{ab}) = b) \to P(\mathbb{X}(t) \ge b \quad \text{for some } t) \qquad \text{as } a \to \infty$

$\quad = P(\mathbb{S}(t) - \theta t/2 \ge b \quad \text{for some } t) = P(\mathbb{S}(t) \ge \theta t/2 + b \quad \text{for some } t)$

(g) $\quad = P(\mathbb{S}(t) \ge ct + d \quad \text{for some } t) \qquad$ if $c = \theta/2$ and $d = b$.

Equating (g) to (e) (via (f) and (9)) gives (6). □

Exercise 7.2 Prove (7). (See Doob (1949).)

Theorem 7.3 (Kolmogorov–Smirnov distributions) Both

(10) $\quad P(\|\mathbb{U}^{\pm}\| > b) = \exp(-2b^2) \qquad$ for all $b > 0 \qquad$ and

(11) $\quad P(\|\mathbb{U}\| > b) = 2\sum_{k=1}^{\infty} (-1)^{k+1} \exp(-2\,k^2\, b^2) \qquad$ for all $b > 0$.

Proof. Now, $\|\mathbb{U}^-\| \cong \|\mathbb{U}^+\|$ and

$\quad P(\|\mathbb{U}^+\| > b) = P(\mathbb{U}(t) > b \quad \text{for some } 0 < t < 1)$

(12) $\quad = P((1-t)\,\mathbb{S}(t/(1-t)) > b) \qquad$ for some $0 \le t \le 1, \qquad$ by (12.3.16)

$\quad = P(\mathbb{S}(r) > b + r\,b \quad \text{for some } r \ge 0) \qquad$ letting $r = t/(1-t)$

(a) $\quad = \exp(-2\,b^2) \qquad$ by theorem 7.2.

Likewise,

(b) $\quad P(\|\mathbb{U}\| > b) = P(|\mathbb{S}(r)| > b + r\,b \quad \text{for some } r \ge 0)$

(c) $\quad = 2\sum_{k=1}^{\infty} (-1)^{k+1} \exp(-2\,k^2\, b^2)$

by theorem 7.2. □

Exercise 7.3 (a) Prove (12.6.5) for $r = 2$.
(b) Prove (12.6.5) for integral r. (This is unimportant.)
[Hint. The $V_\theta \equiv \exp(\theta\,[\mathbb{S}(t) - \theta\, t^2/2]), t \geq 0$ of (8) are martingales on $[0, \infty)$. Differentiate formally under the integral sign in the martingale equality

$$(13) \qquad \int_A V_\theta(t)\, dP = \int_A V_\theta(s)\, dP \qquad \text{for all } A \in \mathcal{A}_s\,.$$

Then conclude that $[\partial^k/\partial\theta^k\, V_\theta(t)]\big|_{\theta=0}$ is a martingale for each $k \geq 1$. For $k = 4$ this leads to $\mathbb{S}^4(t) - 6t\,\mathbb{S}^2(t) + 3t^2 = t^2 H_4(\mathbb{S}(t)/\sqrt{t}\,)$ being a martingale on $[0, \infty)$; here $H_4(t) = t^4 - 6t^2 + 3$ is the "fourth Hermite polynomial." The reader needs to work only with the single specific martingale in part (a); the rest of this hint is simply an intuitive explanation of how this martingale arises.]

8 Embedding the Partial Sum Process

The Partial Sum Process

Let $X_{n1}, \ldots, X_{nn}$ be row-independent rvs having a common $F(0,1)$ distribution, and let $X_{n0} \equiv 0$. We define the *partial sum process* $\mathbb{S}_n$ on $(D, \mathcal{D})$ by

$$(1) \qquad \mathbb{S}_n(t) \equiv \frac{1}{\sqrt{n}} \sum_{i=0}^{[nt]} X_{ni} = \frac{1}{\sqrt{n}} \sum_{i=0}^{k} X_{ni} \qquad \text{for } \frac{k}{n} \le t < \frac{k+1}{n},\ 0 \le k \le n$$

(or for all $k \ge 0$, in case the nth row is $X_{n1}, X_{n2}, \ldots$). Note that

$$(2) \qquad \begin{aligned} &\operatorname{Cov}[\mathbb{S}_n(s), \mathbb{S}_n(t)] = \textstyle\sum_{i=1}^{[ns]} \sum_{j=1}^{[nt]} \operatorname{Cov}[X_{ni}, X_{nj}]/n \\ &\quad = [n\,(s \wedge t)]/n \qquad \text{for } 0 \le s, t \le 1 \end{aligned}$$

for the greatest integer function $[\cdot]$. We suspect that $\mathbb{S}_n$ "converges" to $\mathbb{S}$. We will establish this shortly.

Embedding the Partial Sum Process

Notation 8.1 Let $\{\mathbb{S}(t) : t \ge 0\}$ denote a Brownian motion on $(C_\infty, \mathcal{C}_\infty)$. Then

$$(3) \qquad \mathbb{Z}_n(t) \equiv \sqrt{n}\, \mathbb{S}(t/n) \quad \text{for } t \ge 0 \qquad \text{is also such a Brownian motion.}$$

By using the Skorokhod embedding technique of the previous section repeatedly on the Brownian motion $\mathbb{Z}_n$, we may guarantee that for appropriate stopping times $\tau_{n1}, \ldots, \tau_{nn}$ (with all $\tau_{n0} \equiv 0$) we obtain that

$$(4) \qquad X_{nk} \equiv \mathbb{Z}_n(\tau_{n,k-1}, \tau_{nk}], \quad \text{for } 1 \le k \le n, \quad \text{are iid } F(0,1) \text{ rvs.}$$

Let $\mathbb{S}_n$ denote the partial sum process of these X_{nk}'s. Then, for $t \ge 0$ we have

$$(5) \qquad \begin{aligned} &\mathbb{S}_n(t) = \tfrac{1}{\sqrt{n}} \textstyle\sum_{k=1}^{[nt]} X_{nk} = \tfrac{1}{\sqrt{n}} \mathbb{Z}_n(\tau_{n,[nt]}) = \mathbb{S}\Big(\frac{\tau_{n,[nt]}}{n}\Big) \\ &\quad = \mathbb{S}\Big(\tfrac{1}{n} \textstyle\sum_{k=1}^{[nt]} T_{nk}\Big) = \mathbb{S}(I_n(t)) \end{aligned}$$

with $T_{nk} \equiv (\tau_{nk} - \tau_{n,k-1})$ and $I_n(t) \equiv \frac{1}{n}\tau_{n,[nt]}$. Observe that:

$$(6) \qquad X_{n1}, \ldots, X_{nn} \quad \text{are iid } F(0,1), \text{ in each row.}$$

$$(7) \qquad T_{n1}, \ldots, T_{nn} \quad \text{are iid with means } = 1 = \operatorname{Var}[X], \text{ in each row.}$$

$$(8) \qquad \mathrm{E}T_{nk}^r \le K_r \cdot \mathrm{E}|X_{nk}|^{2r}, \qquad \text{with} \qquad K_r \equiv r\,\Gamma(r)\, 2^{4r-2}. \qquad \Box$$

Theorem 8.1 (Skorokhod's embedding theorem) The partial sum process $\mathbb{S}_n$ on $(D, \mathcal{D})$ of row-independent $F(0,1)$ rvs formed as above satisfies

$$(9) \qquad \|\mathbb{S}_n - \mathbb{S}\| \to_p 0 \qquad \text{as } n \to \infty.$$

Notice: The joint distributions of any $\mathbb{S}_m, \mathbb{S}_n$ in theorem 8.1 are not the same as they would be if formed from a single sequence of iid rvs. In fact, we have no idea of what these joint distributions may be. However, the partial sums of an iid sequence do not generally converge to their limit in the sense of $\to_p$, so we have gained a great deal via the embedding.

Theorem 8.2 (Embedding at a rate) Suppose that $\mathrm{E}X^4 < \infty$. Let I denote the identity function. Then for each $0 \le \nu < \frac{1}{4}$, the process $\mathbb{S}_n$ of (5) satisfies

(10) $\quad n^{\nu} \, \| \, (\mathbb{S}_n - \mathbb{S})/I^{1/2-\nu} \, \|_{1/n}^{1} = O_p(1).$

Proof. Consider theorem 8.1. Let I denote the identity function. Suppose we now show that

(a) $\quad \|I_n - I\|_0^1 = \sup_{0 \le t \le 1} |\tau_{n,[nt]}/n - t| \to_p 0.$

Then on any subsequence n' where $\to_p 0$ in (a) may be replaced by $\to_{a.s.} 0$, the continuity of the paths of $\mathbb{S}$ will yield

(b) $\quad \|\mathbb{S}_{n'}(\cdot) - \mathbb{S}(\cdot)\| = \|\mathbb{S}(I_{n'}) - \mathbb{S}\| \to_{a.s.} 0,$

and thus (9) will follow. This is a useful argument; learn it. It therefore remains to prove (a). The WLLN gives

(c) $\quad I_n(t) = \tau_{n,[nt]}/n \to_p t \qquad$ for any fixed t.

Using the diagonalization technique, we can extract from any subsequence a further subsequence n' on which

(d) $\quad I_{n'}(t) \to_{a.s.} t \qquad$ for all rational t.

But since all functions involved are monotone, and since the limit function is continuous, this implies that a.s.

(e) $\quad I_{n'}(t) \to t$ uniformly on $[0, 1]$.

Thus (a) follows from (e), since every n has a further n' with the same limit. Thus the conclusion (9) holds.

In the proof just given, the conclusion (9) can trivially be replaced by

(f) $\quad \sup_{0 \le t \le m} |\mathbb{S}_n(t) - \mathbb{S}(t)| \to_p 0.$

Appealing to exercise 12.1.6(b) for the definition of $\| \cdot \|_{\infty}$, we thus obtain

(11) $\quad \rho_{\infty}(\mathbb{S}_n, \mathbb{S}) \to_p 0 \qquad$ on $(C_{\infty}, \mathcal{C}_{\infty})$,

provided that the rvs $X_{n1}, X_{n2}, \ldots$ are appropriately iid $(0, \sigma^2)$. [We consider the proof of theorem 8.2 at the end of this section.] □

Let $g : (D, \mathcal{D}) \to (R, \mathcal{B})$ and let Δ_g denote the set of all $x \in D$ for which g is not $\| \cdot \|$-continuous at x. If there exists a set $\Delta \in \mathcal{D}$ having $\Delta_g \subset \Delta$ and $P(\mathbb{S} \in \Delta) = 0$, then we say that g is *a.s.* $\| \cdot \|$-*continuous* with respect to the process $\mathbb{S}$.

Theorem 8.3 (Donsker) Let $g : (D, \mathcal{D}) \to (R, \mathcal{B})$ denote an a.s. $\|\cdot\|$-continuous mapping that is $\mathcal{D}$-measurable. Then $g(\mathbb{S}_n) : (\Omega, \mathcal{A}, P) \to (R, \mathcal{B})$, and both

(12) $\quad g(\mathbb{S}_n) \to_p g(\mathbb{S}) \qquad$ as $n \to \infty \qquad$ for the $\mathbb{S}_n$ of (5) $\qquad$ and

(13) $\quad g(\mathbb{S}_n) \to_d g(\mathbb{S}) \qquad$ as $n \to \infty \qquad$ for any $\mathbb{S}_n$ having the same distribution.

($\mathcal{D}$-measurability is typically trivial, and hypothesizing it avoids the measurability difficulties discussed in section 12.1.) [Theorem 8.2 allows (13) for $\mathcal{D}$-measurable functionals g that are continuous in other $\|\cdot/q\|$-metrics.]

Proof. Now, $\|\mathbb{S}_n - \mathbb{S}\|$ is a $\mathcal{D}$-measurable rv, and $\|\mathbb{S}_n - \mathbb{S}\| \to_p 0$ for the $\mathbb{S}_n$ of (5). Thus any subsequence n' has a further subsequence n'' for which $\|\mathbb{S}_{n''} - \mathbb{S}\| \to 0$ for all $\omega \notin A''$, where $P(A'') = 0$. Moreover,

(a) $\quad P(A'' \cup [\mathbb{S} \in \Delta]) \le P(A'') + P(\mathbb{S} \in \Delta) = 0\,,$

and if $\omega \notin A'' \cup [\mathbb{S} \in \Delta]$, then $g(\mathbb{S}_{n''}(\omega)) \to g(\mathbb{S}(\omega))$ holds, since $\|\mathbb{S}_{n''}(\omega) - \mathbb{S}(\omega)\| \to 0$ and g is $\|\,\|$-continuous at $\mathbb{S}(\omega)$. Thus $g(\mathbb{S}_n) \to_p g(\mathbb{S})$ as $n \to \infty$ for the $\mathbb{S}_n$ of (5). Thus $g(\mathbb{S}_n) \to_d g(\mathbb{S})$ for the $\mathbb{S}_n$ of (5), and hence of (13) also. [Note that we are dealing only with functionals for which the compositions $g(\mathcal{S}_n)$ and $g(\mathcal{S})$ are $(\Omega, \mathcal{A}, P) \to (R, \mathcal{B})$ measurable.] □

Example 8.1 Since the functionals $\|\cdot\|$ and $\|\cdot^+\|$ are a.s. $\|\cdot\|$-continuous,

(14) $\quad \|\mathbb{S}_n^+\| \to_d \|\mathbb{S}^+\| \qquad$ and $\qquad \|\mathbb{S}_n\| \to_d \|\mathbb{S}\|\,.$

The limiting distributions are those given in theorem 7.1. □

Exercise 8.1 Let $X_0 \equiv 0$ and $X_1, X_2, \ldots$ be iid $(0, \sigma^2)$. Define $S_k \equiv X_1 + \cdots + X_k$ for each integer $k \ge 0$.
(a) Find the asymptotic distribution of $(S_1 + \cdots + S_n)/c_n$ for an appropriate c_n.
(b) Determine a representation for the asymptotic distribution of the "absolute area" under the partial sum process, as given by $(|S_1| + \cdots + |S_n|)/c_n$.

The LIL

Recall the (8.6.1) LIL for a single sequence of iid $F(0,1)$ rvs $X_1, X_2, \ldots$ with partial sums $S_n \equiv X_1 + \cdots + X_n$; that is

(15) $\quad \overline{\lim}_{n\to\infty} |S_n|/\sqrt{2\,n\,\log\log n} = 1$ a.s.

The two LILs for Brownian motion (recall (12.3.7) and (12.3.18)) are

(16) $\quad \overline{\lim}_{t\to\infty} |\mathbb{S}(t)|/\sqrt{2\,t\,\log\log t} = 1$ a.s.,

(17) $\quad \overline{\lim}_{t\to 0} |\mathbb{S}(t)|/\sqrt{2\,t\,\log\log(1/t)} = 1$ a.s.

Notation 8.2 Define stopping times $T_1, T_2, \ldots$ (with $T_0 = 0$) having mean 1 for which the rvs

(18) $X_k \equiv \mathbb{S}(\tau_{k-1}, \tau_k]$ are iid as F.

Let $\tau_k \equiv T_0 + T_1 + \cdots + T_k$ for $k \geq 0$, and define the partial sums

(19) $S_n \equiv \sum_{k=1}^{n} X_k = \mathbb{S}(\tau_n) = \mathbb{S}(n) + [\mathbb{S}(\tau_n) - \mathbb{S}(n)]$.

[Note that *this embedding differs* from that in notation 8.1. This one is based on a single sequence of rvs $X_1, X_2, \ldots$.] □

Exercise 8.2 (The LIL) (a) First prove (15), while assuming that (16) is true. [Hint. By proposition 8.6.1, we want to show (roughly) that

(20)
$$|\,\mathbb{S}(\tau_n) - \mathbb{S}(n)|/\sqrt{2\,n\,\log\log n} \to_{a.s.} 0 \qquad \text{or that}$$
$$|\,\mathbb{S}(\tau_{[t]}) - \mathbb{S}(t)|/\sqrt{2\,t\,\log\log t} \to_{a.s.} 0.$$

We will now make rigorous this approach to the problem. First apply the SLLN to $\tau[t]/t$ as $t \to \infty$. Then define $\Delta_k \equiv \sup\{|\mathbb{S}(t) - \mathbb{S}(t_k)| : t_k \leq t \leq t_{k+1}\}$, with $t_k \equiv (1+a)^k$ for some suitably tiny $a > 0$. Use a reflection principle and Mills' ratio to show that $P(\Delta_k \geq (\text{an appropriate } c_k)) < \infty$. Complete the proof using Borel– Cantelli.]

(b) Now that you know how to deal with the "blocks" Δ_k, model a proof of (16) on the proof of proposition 8.6.1.

Proof for Embedding at a Rate

Proof. Consider theorem 8.2. Let $d^2 \equiv \mathrm{Var}[T]$. Let $\mathrm{Log}\ k \equiv 1 \vee (\log k)$. Let $M \equiv M_\epsilon$ be specified below, and define

(a) $A_n^c \equiv [\max_{1\leq k\leq n} |\sum_{i=1}^{k} (T_{ni} - 1)|/(d\sqrt{k}\ \mathrm{Log}\ k) \geq 2M/d]$.

Then the monotone inequality gives

$$A_n^c \subset [\max_{1\leq k\leq n} |\textstyle\sum_{i=1}^{k} \{(T_{ni} - 1)/(d\sqrt{i}\ \mathrm{Log}\ i)\}| \geq M/d]$$

(b) $\qquad \equiv [\max_{1\leq k\leq n} |\sum_{i=1}^{k} Y_{ni}| \geq M/d]$,

where the Y_{ni}'s are independent with mean 0 and variance $(i\ \mathrm{Log}^2 i)^{-1}$. Thus the Kolmogorov inequality gives

$$P(A_n^c) \leq (d/M)^2\, \mathrm{Var}[\textstyle\sum_1^n Y_{ni}] = (d/M)^2 \sum_1^n (i\ \mathrm{Log}^2 i)^{-1}$$

(21) $\qquad \leq (d/M)^2 \sum_1^\infty (i\ \mathrm{Log}^2 i)^{-1} \equiv (d/M)^2\, v^2 < \epsilon^2 \qquad \text{if } M > dv/\epsilon$

(c) $\qquad < \epsilon$.

Thus

(d)
$$P(B_n) \equiv P\left(\max_{1\leq k\leq n} \frac{n^\nu\, |\,\mathbb{S}(\sum_1^k T_{ni}/n) - \mathbb{S}(k/n)|}{(k/n)^{1/2-\nu}} \geq \frac{2M}{\sqrt{dv}}\right)$$

$\le P(B_n \cap A_n) + P(A_n^c)$

(e) $\le \sum_{k=1}^{n} P\left(\left[\frac{|\mathbb{S}(\sum_1^k T_{ni}/n) - \mathbb{S}(k/n)|}{[2M\sqrt{k}\,(\text{Log } k)/n]^{1/2}} \ge \frac{2M}{\sqrt{dv}}\,\frac{k^{1/2-\nu}}{\sqrt{n}}\,\frac{1}{[2M\sqrt{k}\,(\text{Log } k)/n]^{1/2}}\right] \cap A_n\right) + \epsilon$

(f) $\le \sum_{k=1}^{n} P(\sup_{0\le|r|\le a} |\mathbb{S}(r+k/n) - \mathbb{S}(k/n)|/\sqrt{a} \ge b) + \epsilon$

with $a \equiv 2M\sqrt{k}\,(\text{Log } k)/n$ (as in A_n in (a)),

and with $\ge b$ as on the right in (e)

(g) $\le 3\sum_{k=1}^{n} P(\sup_{0\le r\le a} |\mathbb{S}(t, t+r]|/\sqrt{a} \ge b/3) + \epsilon$ using (k) below

(h) $\le 12\sum_{k=1}^{n} P(N(0,1) \ge b/3) + \epsilon$ by the reflection principle

(i) $\le 12\sum_{k=1}^{n} \exp(-(b/3)^2/2) + \epsilon$ by Mills' ratio

$$\le 12\sum_{k=1}^{n} \exp\left(-\frac{M}{9dv}\,\frac{k^{1/2-2\nu}}{\text{Log } k}\right) + \epsilon \qquad (22)$$

(j) $< 2\epsilon\,,$

if $M \equiv M_\epsilon$ is large enough and if $0 \le \nu < \frac{1}{4}$ (this final step holds, since we have $\int_0^\infty \exp(-c\,x^\delta)\,dx \to 0$ as $c \to \infty$). The inequality (g) used

$\sup_{0\le|r|\le a} |\mathbb{S}(r+k/n) - \mathbb{S}(k/n)|$

(k) $\le \sup_{0\le r\le a} |\mathbb{S}(r+k/n) - \mathbb{S}(k/n)| + 2\sup_{0\le r\le a} |\mathbb{S}(r+k/n-a) - \mathbb{S}(k/n-a)|$

[with t in (g) equal to k/n or $k/n - a$, and with a as above (see (f))].

Now, $P(B_n) \le 2\epsilon$ shows that (10) is true, provided that the sup over all of $[1/n, 1]$ is replaced by the max over the points k/n with $1 \le k \le n$. We now "fill in the gaps". Thus (even a crude argument works here)

$P(\sqrt{n}\max_{1\le k\le n-1} \sup_{0\le t\le 1/n} |\mathbb{S}(t+k/n) - \mathbb{S}(k/n)| \,/\, k^{1/2-\nu} \ge M)$

$\le \sum_{k=1}^{n-1} P(\|\,\mathbb{S}\|_0^{1/n} \ge M\,k^{1/2-\nu}/\sqrt{n}\,)$

$\le 4\sum_{k=1}^{n-1} P(N(0,1) \ge M\,k^{1/2-\nu})$ by the reflection principle

(23) $\le 4\sum_{k=1}^{n-1} \exp(-M^2\,k^{1-2\nu}/2)$ by Mills' ratio

(l) $< \epsilon\,,$

if $M \equiv M_\epsilon$ is large enough and if $0 \le \nu < \frac{1}{2}$ (even). □

Exercise 8.3 Suppose $\mathrm{E}X^4 < \infty$. Show that the process $\mathbb{S}_n$ of (5) satisfies

(24) $(n^{1/4}/\log n)\,\|\,\mathbb{S}_n - \mathbb{S}\| = O_p(1)\,.$

[Hint. Replace $n^\nu/(k/n)^{1/2-\nu}$ by $n^{1/4}/\log n$ in the definition of B_n in (d). Now determine the new form of the bounds in (20) and (21).] [While interesting and often quoted in the literature, this formulation has little value for us.]

9 Other Properties of Brownian Motion

Here we collect some selected sample path properties of Brownian motion, just to illustrate a sample of what is known. Some proofs are outlined in the exercises.

Definition 9.1 (Variation) For a sequence of *partitions*

$$\mathcal{P}_n \equiv \{(t_{n,k-1}, t_{nk}] : k = 1, \ldots, n\} \text{ of } [0,1] \text{ (with } 0 \equiv t_{n0} < \cdots < t_{nn} \equiv 1),$$

define the rth *variation* of $\mathbb{S}$ corresponding to $\mathcal{P}_n$ by

(1) $\quad V_n(r) \equiv \sum_{k=1}^n |\mathbb{S}(t_{nk}) - \mathbb{S}(t_{n,k-1})|^r.$

We call these partitions *nested* if $\mathcal{P}_n \subset \mathcal{P}_{n+1}$ for all $n \geq 1$. We further define the *mesh* of the partitions to be $\|\mathcal{P}_n\| \equiv \sup_{1\leq k\leq n} |t_{nk} - t_{n,k-1}|$.

Theorem 9.1 (Nondifferentiability)
(a) Almost every Brownian path is nowhere differentiable.
(b) In fact, $V_n(1) \to \infty$ a.s. if $\|\mathcal{P}_n\| \to 0$.
(c) (Finite squared variation) $V_n(2) \to_{\mathcal{L}_2} 1$ if $\|\mathcal{P}_n\| \to 0$.
(d) (Finite squared variation) $V_n(2) \to_{a.s.} 1$
if either (i) $\sum_{n=1}^\infty \|\mathcal{P}_n\| < \infty$ or (ii) The $\mathcal{P}_n$ are nested with mesh approaching 0.
(e) (Dudley) $V_n(2) \to_{a.s.} 1$ if and only if $(\log n)\,\|\mathcal{P}_n\| \to 0$.

Theorem 9.2 (Lévy) The Hölder condition is true:

$$\limsup_{\{0\leq s<t\leq 1 \text{ and } t-s=a\searrow 0\}} \frac{|\mathbb{S}(t) - \mathbb{S}(s)|}{\sqrt{2\,a\,\log(1/a)}} = 1 \quad \text{a.s.} \tag{2}$$

Theorem 9.3 (The zeros of $\mathbb{S}$ in $[0,1]$) Define

$$\text{Zeros}(\omega) \equiv \{t \in [0,1] : \mathbb{S}(t,\omega) = 0\}.$$

For almost all ω, the set Zeros(ω) is a closed and perfect set of Lebesgue measure zero. [A set A is called *dense in itself* if every point x of A is such that every neighborhood of x contains another point of A beyond x. A compact set that is dense in itself is called *perfect*.]

Theorem 9.4 (Strassen) Let $\mathbb{Z}_n(t) \equiv \mathbb{S}(nt)/\sqrt{2n \log\log n}$ for $0 \leq t \leq 1$. Let $\mathcal{K}$ be the set of absolutely continuous functions f on $[0,1]$ with $f(0) = 0$ and $\int_0^1 [f'(t)]^2\, dt \leq 1$; equivalently,

$$\mathcal{K} \equiv \left\{ f \in C[0,1] : f(0) = 0,\ f(t) = \int_0^t f'(s)\, ds,\ \int_0^1 [f'(s)]^2\, ds \leq 1 \right\}.$$

For almost all ω the sequence $\{\mathbb{Z}_n(\cdot,\omega) : n = 3, 4, \ldots\}$ visualized within $C[0,1]$ is relatively compact with limit set $\mathcal{K}$. That is,

(3) $\quad P(\,\overline{\lim}_n \|\mathbb{Z}_n - \mathcal{K}\| = 0) = 1 \qquad \text{and} \qquad P(\,\cap_{f\in\mathcal{K}}\,[\,\underline{\lim}_n \|\mathbb{Z}_n - f\| = 0]) = 1.$

We will write this conclusion symbolically as $\mathbb{Z} \hookrightarrow \mathcal{K}$. [This can be used to establish a LIL for various functionals g of $\mathbb{S}_n$, by determining the extreme values of $g(\mathcal{K})$.]

Exercise 9.1 Prove theorem 9.1(a).

Exercise 9.2 (α) Let $Z \cong \text{N}(0,1)$. Let $r > 0$. Show that:

(a) $C_r \equiv \text{E}|Z|^r = 2^{r/2}\,\Gamma(\frac{r+1}{2})/\sqrt{\pi}$.

(b) $|\mathbb{S}(t_{n,k-1}, t_{nk}]\,|^r \cong (\,C_r\,|t_{nk} - t_{n,k-1}|^{r/2}\,,\,(C_{2r} - C_r^2)\,(t_{nk} - t_{n,k-1})^r\,)\,.$

(β) Now show that $\text{E}\,V_n(2) = 1$ and $\text{Var}[V_n(2)] \le (C_{2r} - C_r^2)\,\|\mathcal{P}_n\|$, giving

(c) $\sum_1^\infty P(\,|V_n(2) - 1| \ge \epsilon) \le (C_{2r} - C_r^2)\,\epsilon^{-2}\sum_1^\infty \|\mathcal{P}_n\| < \infty$.

(γ) Finally, demonstrate the truth of theorem 9.1(d), case (i).

Exercise 9.3 Prove theorem 9.1(b) when all $t_{nk} = k/2^n$.
[Hint. Let $0 < \lambda < 1$. The Paley–Zygmund inequality gives

$$P(V_n(1) > \lambda \text{E}\,V_n(1)) \ge (1-\lambda)^2\,\text{E}^2 V_n(1)/\text{E}(V_n^2(1)) \to (1-\lambda)^2,$$

where $\text{E}\,V_n(1) \to \infty$.]

10 Various Empirical Process

Suppose that $\xi_{n1}, \ldots, \xi_{nn}$ are iid Uniform$(0,1)$. Their *empirical df* $\mathbb{G}_n$ is defined by

$$(1) \qquad \mathbb{G}_n(t) \equiv \frac{1}{n}\sum_{k=1}^{n} 1_{[0,t]}(\xi_{nk}) \qquad \text{for } 0 \le t \le 1$$

$$(2) \qquad = k/n \quad \text{for} \quad \xi_{n:k} \le t < \xi_{n:k+1} \quad \text{and} \quad 0 \le k \le n\,,$$

where $0 \equiv \xi_{n:0} \le \xi_{n:1} \le \cdots \le \xi_{n:n} \le \xi_{n:n+1} \equiv 1$ are the order statistics; see figure 10.1. Note that $n\mathbb{G}_n(t) \cong$ Binomial$(n,t) \cong (t, t(1-t))$. The Glivenko–Cantelli theorem shows that $\mathbb{G}_n$ converges uniformly to the true df I; that is,

$$(3) \qquad \|\mathbb{G}_n - I\| \to_{a.s.} 0 \qquad \text{(even for the present triangular array of } \xi_{nk}\text{'s).}$$

(The Cantelli proof of the SLLN based on fourth moments shows that $\mathbb{G}_n(t) \to_{a.s.} t$ for each fixed t; even for triangular arrays. The rest of the proof is identical.) The *uniform empirical process* $\mathbb{U}_n$ is defined by

$$(4) \qquad \mathbb{U}_n(t) \equiv \sqrt{n}\,[\mathbb{G}_n(t) - t] = \frac{1}{\sqrt{n}}\sum_{k=1}^{n}[1_{[\xi_{nk}\le t]} - t] \qquad \text{for } 0 \le t \le 1.$$

This process is also pictured in figure 10.1. The means and covariances of $\mathbb{U}_n$ are the same as those of Brownian bridge $\mathbb{U}$, in that

$$(5) \qquad \mathrm{E}\mathbb{U}_n(t) = 0 \qquad \text{and} \qquad \mathrm{Cov}[\mathbb{U}_n(s), \mathbb{U}_n(t)] = s\wedge t - st \qquad \text{for all } 0 \le s,t \le 1;$$

this follows easily from

$$(6) \qquad \mathrm{Cov}[1_{[0,s]}(\xi_{nk}), 1_{[0,t]}(\xi_{nk})] = s\wedge t - st \qquad \text{for } 0 \le s,t \le 1.$$

{Moreover, for any d_k's and e_k's we have immediately from this that

$$(7) \qquad \mathrm{Cov}[\,\textstyle\sum_1^n d_k\, 1_{[0,s]}(\xi_{nk}),\, \sum_1^n e_k\, 1_{[0,t]}(\xi_{nk}))] = (\,\sum_1^n d_k\, e_k) \times [s\wedge t - st]\,;$$

we would have $\sum_1^n \mathrm{E}[d_k, e_k]$ instead, if these were rvs independent of the ξ_{nk}'s.} We note that $\mathbb{G}_n^{-1}(t) \equiv \inf\{x \in [0,1] : \mathbb{G}_n(x) \ge t\}$ is left continuous, with

$$(8) \qquad \mathbb{G}_n^{-1}(t) = \xi_{n:k} \text{ for } (k-1)/n < t \le k/n$$

and $\mathbb{G}_n^{-1}(0) = 0$, as in figure 10.1. The *uniform quantile process* $\mathbb{V}_n$ is defined by

$$(9) \qquad \mathbb{V}_n(t) \equiv \sqrt{n}\,[\mathbb{G}_n^{-1}(t) - t] \qquad \text{for } 0 \le t \le 1.$$

The key identities relating $\mathbb{U}_n$ and $\mathbb{V}_n$ are (with I the identity function) the trivial

$$(10) \qquad \mathbb{U}_n = -\mathbb{V}_n(\mathbb{G}_n) + \sqrt{n}\,[\mathbb{G}_n^{-1}\circ\mathbb{G}_n - I] \qquad \text{on } [0,1]\,,$$

$$(11) \qquad \mathbb{V}_n = -\mathbb{U}_n(\mathbb{G}_n^{-1}) + \sqrt{n}\,[\mathbb{G}_n\circ\mathbb{G}_n^{-1} - I] \qquad \text{on } [0,1]\,.$$

Note that

$$(12) \qquad \|\mathbb{G}_n\circ\mathbb{G}_n^{-1} - I\| = 1/n \qquad \text{and} \qquad \|\mathbb{G}_n^{-1}\circ\mathbb{G}_n - I\| = [\max_{1\le k\le n+1} \delta_{nk}]\,;$$

here $\delta_{nk} \equiv (\xi_{n:k} - \xi_{n:k-1})$, for $1 \le k \le n+1$, denotes the kth of the $n+1$ *uniform spacings*.

It is sometimes convenient to use the smoothed versions $\ddot{\mathbb{G}}_n$ and $\ddot{\mathbb{G}}_n^{-1}$ defined by

(13) $\ddot{\mathbb{G}}_n(\xi_{n:k}) = k/(n+1)$ and $\ddot{\mathbb{G}}_n^{-1}(k/(n+1)) = \xi_{n:k}$ for $0 \le k \le n+1$,

connected linearly between points. Upon occasion the *smoothed uniform quantile process* $\ddot{\mathbb{V}}_n(t) \equiv \sqrt{n}\,[\ddot{\mathbb{G}}_n^{-1}(t) - t]$ is a useful variation on $\mathbb{V}_n$. The Glivenko–Cantelli theorem implies that

(14) $\|\mathbb{G}_n - I\| \to_{a.s.} 0$, $\|\mathbb{G}_n^{-1} - I\| \to_{a.s.} 0$, $\|\ddot{\mathbb{G}}_n - I\| \to_{a.s.} 0$, $\|\ddot{\mathbb{G}}_n^{-1} - I\| \to_{a.s.} 0$;

see figure 10.1. Coupling these with the identities (10) and (11) shows that

(15) $\|\mathbb{U}_n - \mathbb{U}\| \to_p 0$ if and only if $\|\mathbb{V}_n - \mathbb{V}\| \to_p 0$.

Let $\mathsf{c}_n \equiv (c_{n1}, \ldots, c_{nn})'$ denote a vector of known constants normalized so that

(16) $c_{n\cdot} \equiv \frac{1}{n}\sum_{k=1}^n c_{nk} = 0$ and $\sigma_{c,n}^2 \equiv \frac{1}{n}\sum_{k=1}^n (c_{nk} - c_{n\cdot})^2 = 1$, and let $\overline{c_{n\cdot}^4} \equiv \frac{1}{n}\sum_{k=1}^n [c_{nk} - c_{n\cdot}]^4$.

We suppose that these constants also satisfy the *uan condition*

(17) $\max_{1\le k\le n} |c_{nk} - \bar{c}_n|/[\sqrt{n}\,\sigma_{c,n}] = \left[\max_{1\le k\le n} |c_{nk}|/\sqrt{n}\right] \to 0$ as $n \to \infty$.

The *weighted uniform empirical process* is defined by

(18) $\mathbb{W}_n(t) \equiv \frac{1}{\sqrt{n}} \sum_{k=1}^n c_{nk}\,[1_{[\xi_{nk}\le t]} - t]$ for $0 \le t \le 1$.

The $\mathbb{W}_n$ process is pictured in figure 10.1. It is trivial from (7) that

(19) $\mathrm{Cov}[\mathbb{W}_n(s), \mathbb{W}_n(t)] = s \wedge t - st$ for $0 \le s, t \le 1$.

It is easy to show that $\mathbb{W}_n \to_{fd} \mathbb{W}$, where $\mathbb{W}$ denotes another Brownian bridge, one that is independent of $\mathbb{U}$.

Let $\mathsf{R}_n \equiv (R_{n1}, \ldots R_{nn})'$ denote the *ranks* of $\xi_{n1}, \ldots, \xi_{nn}$; and then denote the *antiranks* by $\mathsf{D}_n \equiv (D_{n1}, \ldots, D_{nn})'$. Then R_n is a random permutation of the vector $(1, \ldots, n)'$, while D_n is the inverse permutation. These satisfy

(20) $\xi_{nD_{nk}} = \xi_{n:k}$ and $\xi_{nk} = \xi_{n:R_{nk}}$.

As observed in example 7.5.3,

(21) $(\xi_{n:1}, \ldots, \xi_{n:n})$ and $(R_{n1}, \ldots, R_{nn})$ are independent rvs.

The empirical *finite sampling process* $\mathbb{R}_n$ is defined by

(22) $\mathbb{R}_n(t) \equiv \frac{1}{\sqrt{n}} \sum_{k=1}^{[(n+1)t]} c_{nD_{nk}}$ for $0 \le t \le 1$.

The $\mathbb{R}_n$ process is also pictured in figure 10.1. The key identities are

(23) $\mathbb{W}_n = \mathbb{R}_n(\mathbb{G}_n)$ or $\mathbb{R}_n = \mathbb{W}_n(\ddot{\mathbb{G}}_n^{-1})$ on $[0,1]$.

These identities give

(24) $\|\mathbb{W}_n - \mathbb{W}\| \to_p 0$ if and only if $\|\mathbb{R}_n - \mathbb{R}\| \to_p 0$,

as with (15). Because of (21), we see that

(25) $\mathbb{R}_n$ and $\mathbb{V}_n$ are independent processes.

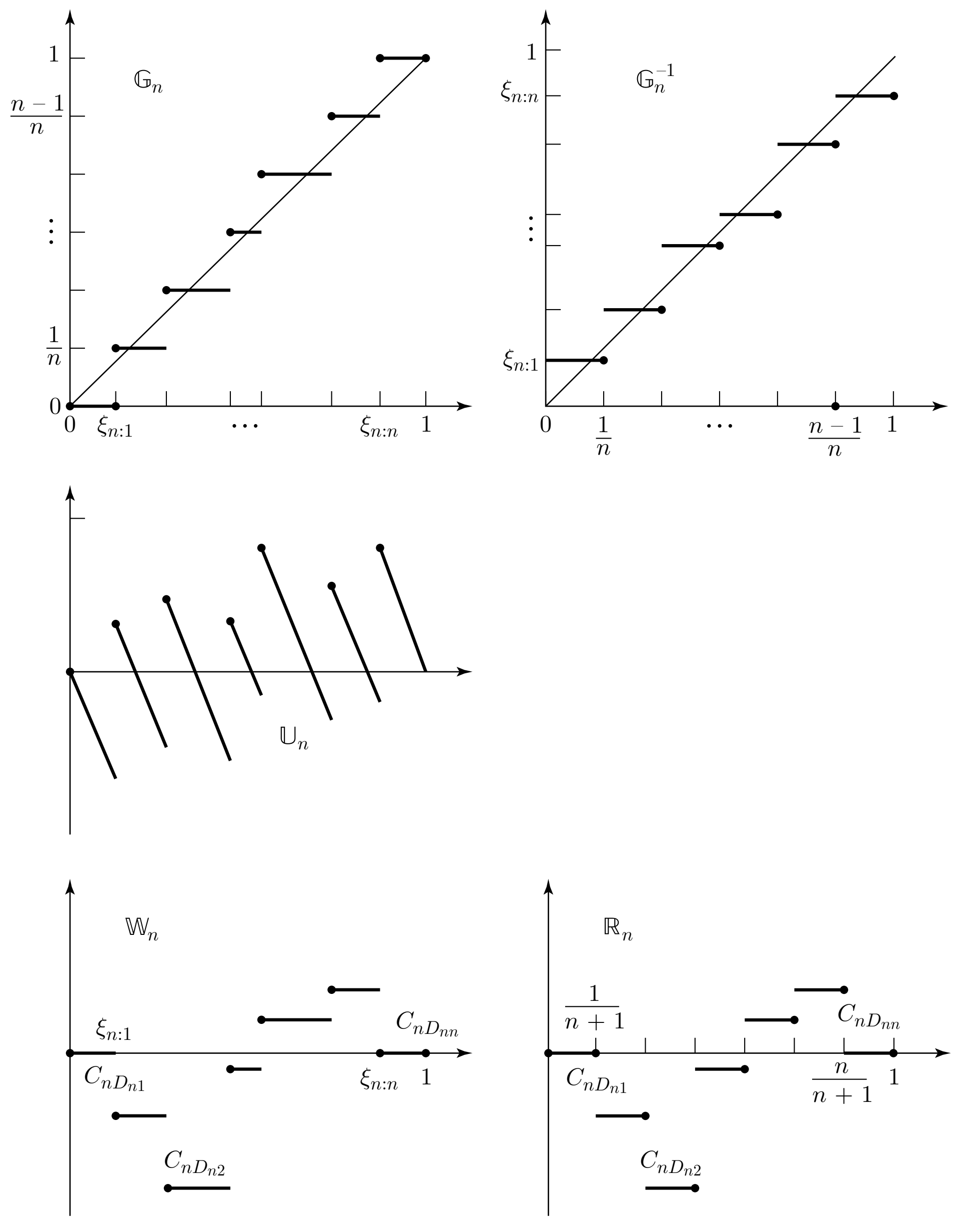


Figure 10.1 $\mathbb{G}_n$, $\mathbb{G}_n^{-1}$, $\mathbb{U}_n$, $\mathbb{W}_n$ and $\mathbb{R}_n$.

We reiterate that

(26) $\quad \mathbb{W}$ and $\mathbb{V} = -\mathbb{U}$ are independent Brownian bridges;

this is further corroborated, since (7) with $\sum_1^n d_k\, e_k = \sum_1^n c_{nk}/n = 0$ imply that the cross covariance

(27) $\quad \mathrm{Cov}[\mathbb{U}_n(s), \mathbb{W}_n(t)] = 0 \qquad$ for all $0 \le s, t \le 1$.

We will prove only part of theorems 10.1 and 10.3 (namely, that (28) holds). For the believability of the rest, we will rely on (28), our earlier proof that $\mathbb{S}_n$ can be embedded at a rate, and the proof of theorem 10.2. (Shorack(1991) contains these proofs, written in the current style and notation.) See section 12.11 for proofs of theorems 10.2 and 10.4.

Theorem 10.1 (Convergence of the uniform processes) We can define independent Brownian bridges $\mathbb{U} = -\mathbb{V}$ and $\mathbb{W}$ and row-independent Uniform$(0,1)$ rvs $\xi_{n1}, \ldots, \xi_{nn}$ on a common probability space $(\Omega, \mathcal{A}, P)$ in such a way that

(28) $\quad \|\mathbb{U}_n - \mathbb{U}\| \to_p 0 \qquad$ and $\qquad \|\mathbb{V}_n - \mathbb{V}\| \to_p 0$,

(29) $\quad \|\mathbb{W}_n - \mathbb{W}\| \to_p 0 \qquad$ and $\qquad \|\mathbb{R}_n - \mathbb{R}\| \to_p 0$,

provided that the c_{nk}'s are uan with $\bar{c}_n = 0$, $\sigma_c^2 = 1$, and $\overline{c_n^4} < \infty$.

Theorem 10.2 (Pyke–Shorack) Let $q > 0$ on $(0,1)$ be $\nearrow$ on $[0, \frac{1}{2}]$, $\searrow$ on $[\frac{1}{2}, 1]$, and have $\int_0^1 [q(t)]^{-2}\, dt < \infty$. Then:

(30) $\quad \|\,(\mathbb{U}_n - \mathbb{U})/q\,\| \to_p 0 \qquad$ and $\qquad \|\,(\mathbb{V}_n - \mathbb{V})/q\,\| \to_p 0$.

(31) $\quad \|\,(\mathbb{W}_n - \mathbb{W})/q\,\| \to_p 0 \qquad$ and $\qquad \|\,(\mathbb{R}_n - \mathbb{W})/q\,\| \to_p 0$.

Corollary 1 (Csörgő–Révész) We may replace $1/q$ in the previous theorem by K, for any qf K having $\mathrm{Var}[K(\xi)] < \infty$.

Theorem 10.3 (Weighted approximation of the uniform processes) The embeddings of the previous theorem are such that for any $0 \le \nu < \frac{1}{4}$ we have
(a) (M. Csörgő, S. Csörgő, Horváth, Mason)

(32) $$\Delta_{\nu n} \equiv \|n^{\nu}(\mathbb{U}_n - \mathbb{U})/[I \wedge (1-I)]^{1/2-\nu}\,\|_{1/n}^{1-1/n} = O_p(1)\,,$$

(33) $$\bar{\Delta}_{\nu n} \equiv \|n^{\nu}(\mathbb{V}_n - \mathbb{V})/[I \wedge (1-I)]^{1/2-\nu}\,\|_{1/2n}^{1-1/2n} = O_p(1)\,.$$

(b) (Shorack) Suppose $\overline{\lim}\ \overline{c_n^4} < \infty$. Then

(34) $$\dot{\Delta}_{\nu n} \equiv \|n^{\nu}(\mathbb{W}_n - \mathbb{W})/[I \wedge (1-I)]^{1/2-\nu}\,\|_{1/n}^{1-1/n} = O_p(1)\,,$$

(35) $$\ddot{\Delta}_{\nu n} \equiv \|n^{\nu}(\mathbb{R}_n - \mathbb{W})/[I \wedge (1-I)]^{1/2-\nu}\,\|_{1/2n}^{1-1/2n} = O_p(1)\,.$$

[The supremum limits in (32) and (34) may be changed to c/n and $1 - c/n$ for any constant $c > 0$. This relates to exercise 10.3 below.]

Theorem 10.4 (Weighted approximation of $\mathbb{G}_n$; Mason) For *any* realization of $\mathbb{G}_n$, any $n \ge 1$, any $0 < \nu < \frac{1}{2}$, and all $\lambda > 0$ we have

$$(36) \qquad \Delta^0_{\nu n} \equiv \left\| \frac{n^\nu(\mathbb{G}_n - I)}{[I \wedge (1-I)]^{1-\nu}} \right\| = O_p(1)\,.$$

We may replace $\mathbb{G}_n$ by $\ddot{\mathbb{G}}_n^{-1}$ in (36).

Example 10.1 (R-statistics) Consider the *simple linear rank statistics*

$$(37) \qquad T_n \equiv \frac{1}{\sqrt{n}} \sum_{k=1}^n c_{nk}\, K\left(\frac{R_{nk}}{n+1}\right) = \frac{1}{\sqrt{n}} \sum_{k=1}^n K\left(\frac{i}{n+1}\right) c_{nD_{nk}}$$

$$(38) \qquad = \int_0^1 K\, d\mathbb{R}_n = -\int_0^1 \mathbb{R}_n\, dK\,,$$

where the last step holds if $K = K_1 - K_2$ with each $K_i \nearrow$ and left continuous on $(0,1)$. As in (12.3.19), this suggests that

$$(39) \qquad T_n \to_p \int_0^1 \mathbb{W}\, dK \cong N(0, \mathrm{Var}[K(\xi)]),$$

provided that the uan condition holds and provided that $\mathrm{Var}[K_i(\xi)] < \infty$ for $i = 1, 2$. Indeed, this can be shown to be true. (Writing

$$T_n = -\int_0^1 \mathbb{W}\, dK \oplus \|(\mathbb{R}_n - \mathbb{W})/q\| \int_0^1 [q(t)]^{-2}\, d|K|$$

provides a simple proof in case this integral is finite for some square integrable function q and for total variation measure $d|K|$.) We will return to this in chapter 16 below. □

Proof. Consider $\mathbb{V}_n$. We will represent our uniforms rvs as a normed sum of Exponential(1) rvs. Thus we really begin with a Skorokhod embedding of iid Exponential(1) rvs.

Let $F(x) = 1 - \exp(-(x+1))$ for $x \ge -1$, so that F is a $(0,1)$ df; and if $X \cong F$, then $X + 1 \cong$ Exponential(1). According to Skorokhod's embedding theorem, there exist row-independent rvs $X_{n1}, \ldots, X_{nn}$ with df F such that the partial sum process $\mathbb{S}_n$ of the nth row satisfies $\|\mathbb{S}_n - \mathbb{S}\| \to_p 0$ for some Brownian motion $\mathbb{S}$. We now define

$$(40) \qquad \eta_{nk} \equiv k + X_{n1} + \cdots + X_{nk} \quad \text{and} \quad \xi_{n:k} \equiv \eta_{nk}/\eta_{n,n+1} \qquad \text{for } 1 \le k \le n+1.$$

It is an elementary exercise below to show that these ξ_{nk}'s are distributed as n row-independent Uniform$(0,1)$ order statistics. Let $\mathbb{G}_n$ denote their empirical df and $\mathbb{U}_n$ their uniform empirical process. The key identity relating $\mathbb{V}_n$ to $\mathbb{S}_n$ is

$$(a) \qquad \mathbb{V}_{n-1}\left(\frac{k}{n-1}\right) = \sqrt{n-1}\left[\frac{\eta_k}{\eta_n} - \frac{k}{n-1}\right]$$

$$= \frac{n}{\eta_n}\sqrt{\frac{n-1}{n}}\left[\frac{\eta_k - k}{\sqrt{n}} - \frac{k}{n}\,\frac{\eta_n - n}{\sqrt{n}}\right] - \sqrt{n-1}\left[\frac{k}{n-1} - \frac{k}{n}\right]$$

(41) $$= \frac{n}{\eta_n}\sqrt{\frac{n-1}{n}}\,[\mathbb{S}_n(k/n) - (k/n)\,\mathbb{S}_n(1)\,] - \frac{1}{\sqrt{n-1}}\,\frac{k}{n}\,,$$

so that for $0 \le t \le 1$,

(42) $$\mathbb{V}_{n-1}(t) = \frac{n}{\eta_n}\sqrt{\frac{n-1}{n}}\;[\,\mathbb{S}_n(I_n(t)) - I_n(t)\,\mathbb{S}_n(1)\,] - \frac{1}{\sqrt{n-1}}\,I_n(t),$$

where $I_n(t) \equiv k/n$ for $(k-1)/(n-1) < t \le k/(n-1)$ and $1 \le k \le n-1$ with $I_n(0) \equiv 0$ satisfies $\|I_n - I\| \to 0$. Note that $\eta_n/n \to_p 1$ by the WLLN and that $\|\mathbb{G}_n - I\| \to_p 0$. Thus

(43) $$\|\mathbb{V}_n - \mathbb{V}\| \to_p 0 \qquad \text{for} \qquad \mathbb{V} \equiv \mathbb{S} - I\,\mathbb{S}(1)$$

follows from the identity (42), $\|I_n - I\| \to 0$, and the fact that

(b) $$\|\mathbb{S}(I_n) - \mathbb{S}\| \le \|\mathbb{S}(I_n) - \mathbb{S}(I_n)\| + \|\mathbb{S}(I_n) - \mathbb{S}\| \le \|\mathbb{S}_n - \mathbb{S}\| + \|\mathbb{S}(I) - \mathbb{S}\| \to_p 0\,,$$

by continuity of all sample paths of the $\mathbb{S}$ process.

All sample paths of $\mathbb{V}$ are continuous, and the maximum jump size of $|\mathbb{V}_n - \mathbb{V}|$ is bounded above by $[\sqrt{n}\max_{1\le i\le n+1}\delta_{ni}]$; so $\|\mathbb{V}_n - \mathbb{V}\| \to_p 0$ and (12) imply

(44) $$[\sqrt{n}\max\nolimits_{1\le k\le n+1}\delta_{ni}] = \sqrt{n}\,\|\mathbb{G}_n^{-1}\circ\mathbb{G}_n - I\| \to_p 0 \qquad \text{as } n \to \infty.$$

Thus

$$\|\mathbb{U}_n - \mathbb{U}\| = \|-\mathbb{V}_n(\mathbb{G}_n) + \sqrt{n}\,[\mathbb{G}_n\circ\mathbb{G}_n^{-1} - I\,] + \mathbb{V}\|$$

(c) $$\le \|\mathbb{V}_n(\mathbb{G}_n) - \mathbb{V}(\mathbb{G}_n)\| + \|\mathbb{V}(\mathbb{G}_n) - \mathbb{V}\| + \sqrt{n}\,\|\mathbb{G}_n\circ\mathbb{G}_n^{-1} - I\|$$

(d) $$\le \|\mathbb{V}_n - \mathbb{V}\| + \|\mathbb{V}(\mathbb{G}_n) - \mathbb{V}\| + o_p(1)$$

$$= \|\mathbb{V}(\mathbb{G}_n) - \mathbb{V}\| + o_p(1)$$

(e) $$= o_p(1),$$

using $\|\mathbb{G}_n - I\| \to_p 0$ and uniform continuity of the sample paths of $\mathbb{V}$.

We will prove Mason's theorem in the next section. □

Exercise 10.1 Establish the claim made just below (40).

Example 10.2 (The supremum functionals) Suppose $g : (D, \mathcal{D}) \to (R, \mathcal{B})$ is a.s. $\|\cdot\|$-continuous. Then

(45) $$g(\mathbb{U}_n) \to_p g(\mathbb{U}) \qquad \text{and} \qquad g(\mathbb{V}_n) \to_p g(\mathbb{V})$$

for the special constructions of theorem 10.1. Moreover, convergence in distribution holds for any versions of these processes. Letting # denote $+, -,$ or $|\cdot|$, we can thus claim the convergence in distribution

(46) $$\|\mathbb{U}_n^{\#}\| \to_d \|\mathbb{U}^{\#}\| \qquad \text{and} \qquad \int_0^1 \mathbb{U}_n^2(t)\,dt \to_d \int_0^1 \mathbb{U}^2(t)\,dt$$

for any versions of these processes. These limiting distributions of $\|\mathbb{U}^{\pm}\|$ were given in theorem 12.7.2, while that of $\int_0^1 \mathbb{U}^2(t)\,dt$ will be given in (12.12.6). □

Exercise 10.2 (The two-sample uniform process) (i) Let $\mathbb{G}_m$ and $\mathbb{H}_n$ be the empirical dfs of two independent Uniform$(0,1)$ special constructions. Let

$$\mathbb{U}_m = \sqrt{m}\,(\mathbb{G}_m - I) \qquad \text{and} \qquad \mathbb{V}_n \equiv \sqrt{n}\,(\mathbb{H}_n - I)$$

denote the corresponding empirical process, and let $\lambda_{mn} \equiv n/(m+n)$. Then

$$\mathbb{W}_{mn} \equiv \sqrt{\tfrac{mn}{m+n}}\,(\mathbb{G}_m - \mathbb{H}_n) = (\sqrt{\lambda_{mn}}\,\mathbb{U}_m - \sqrt{1-\lambda_{mn}}\,\mathbb{V}_n)$$

has

$$\text{(47)} \quad \begin{aligned} &\|\mathbb{W}_{mn} - \mathbb{W}^0_{mn}\| \to_p 0 \quad \text{as } m \wedge n \to \infty, \qquad \text{where} \\ &\mathbb{W}^0_{mn} \equiv (\sqrt{\lambda_{mn}}\,\mathbb{U} - \sqrt{1-\lambda_{mn}}\,\mathbb{V}\,) \text{ is a Brownian bridge.} \end{aligned}$$

We thus have $\|\mathbb{W}^{\#}_{mn}\| \to_d \|\mathbb{W}^{\#}\|$, for Brownian bridge $\mathbb{W}$. Write out the details.

(ii) Now use a discrete reflection principle to compute the *exact* distribution of $P(\|\mathbb{W}^{+}_{nn}\| \ge a)$, and pass to the limit in the resulting expression to obtain (12.7.9). (This provides an alternative to the earlier method.) [Hint. Go through the order statistics of the combined sample from smallest to largest. If it is from sample 1, step up one unit as you go to the right one unit. If it is from sample 2, step down one unit as you go to the right one unit. In this way, perform a random walk from $(0,0)$ to $(2n,0)$. What is the chance you ever cross a barrier of height a?]

Example 10.3 (The Kolmogorov–Smirnov and Cramér–von Mises statistics) Let $\xi_{n1},\ldots,\xi_{nn}$ be the iid Uniform$(0,1)$ rvs of the special construction, and let F denote an arbitrary df. Then $X_{nk} \equiv F^{-1}(\xi_{nk}), 1 \le k \le n$, are iid F. Let $\mathbb{F}_n$ denote the empirical df of $X_{n1},\ldots,X_{nn}$ and let $\mathbb{E}_n$ denote the empirical process defined by $\mathbb{E}_n(x) \equiv \sqrt{n}\,[\mathbb{F}_n(x) - F(x)]$. Now, $\mathbb{E}_n = \mathbb{U}_n(F)$. Thus (28) implies $\|\,\mathbb{E}_n - \mathbb{U}(F)\| \le \|\mathbb{U}_n - \mathbb{U}\| \to_p 0$, where equality holds if F is continuous. Thus

$$\text{(48)} \quad \sqrt{n}\,D^{\#}_n \equiv \sqrt{n}\,\|(\mathbb{F}_n - F)^{\#}\| = \|\mathbb{U}^{\#}_n\| \to_d \|\mathbb{U}^{\#}\| \qquad \text{if } F \text{ is continuous.}$$

Likewise, a change of variable allows elimination of F, and gives

$$\text{(49)} \quad W^2_n \equiv \int n(\mathbb{F}_n - F)^2\,dF = \int_0^1 \mathbb{U}^2_n(t)\,dt \to_d \int_0^1 \mathbb{U}^2(t)\,dt \qquad \text{if } F \text{ is continuous.}$$

These statistics are used to test whether F is really the true df, and $\sqrt{n}D^{\#}_n$ and W^2_n all measure how far the estimate $\mathbb{F}_n$ of the true df differs from the hypothesized df F. [The percentage points of the asymptotic distributions of $\sqrt{n}D^{\#}_n$ and W^2_n, under the null hypothesis when F is really the true df, are available.]

Consider now the two-sample problem in which the rvs $X^{(i)}_{nj} \equiv F^{-1}(\xi^{(i)}_{nj})$, for $i = 1,2$ and $1 \le j \le n_i$, of independent special constructions have empirical dfs $\mathbb{F}^{(1)}_{n_1}$ and $\mathbb{F}^{(2)}_{n_2}$. Note that for independent uniform empirical processes

$$\text{(50)} \quad \sqrt{\tfrac{n_1 n_2}{n_1+n_2}}\,[\mathbb{F}^{(1)}_{n_1} - \mathbb{F}^{(2)}_{n_2}] = \sqrt{\tfrac{n_2}{n_1+n_2}}\,\mathbb{U}^{(1)}_{n_1}(F) - \sqrt{\tfrac{n_1}{n_1+n_2}}\,\mathbb{U}^{(2)}_{n_2}(F) \equiv \mathbb{W}_{n_1,n_2}(F)$$

$$\text{(51)} \quad =_a \mathbb{W}_{n_1,n_2}(F) \qquad \text{if } F \text{ is continuous,}$$

where

$$\text{(52)} \quad \mathbb{W}_{n_1,n_2} \equiv \sqrt{\tfrac{n_1}{n_1+n_2}}\,\mathbb{U}^{(1)} - \sqrt{\tfrac{n_2}{n_1+n_2}}\,\mathbb{U}^{(2)} \cong \mathbb{U} \qquad \text{for all } n_1 \text{ and } n_2.$$

This gives the asymptotic null distribution for the various supremum and integral functionals with which we have dealt, no matter which version of these processes is considered. □

Exercise 10.3 Show that $n\,\xi_{n:n} \to_d$ Exponential(1).

Exercise 10.4 ($\int_0^1 g\,d\mathbb{U}_n$) Suppose Var[$g(\xi)$] and Var[$h(\xi)$] are finite.
(a) Show that there exist rvs (to be labeled $\int_0^1 g\,d\mathbb{U}$ and $\int_0^1 h\,d\mathbb{W}$) for which

(53) $$\int_0^1 g\,d\mathbb{U}_n \to_p \int_0^1 g\,d\mathbb{U} \qquad \text{and} \qquad \int_0^1 h\,d\mathbb{W}_n \to_p \int_0^1 h\,d\mathbb{W}\,.$$

(b) Show also that

(54) $$\int_0^1 g\,d\mathbb{V}_n \to_p -\int_0^1 g\,d\mathbb{U} \qquad \text{and} \qquad \int_0^1 h\,d\mathbb{R}_n \to_p \int_0^1 h\,d\mathbb{W}\,.$$

Exercise 10.5 (Mason) Consider the $\Delta_{n\nu}$ of (32). For some $a > 0$,

(55) $$\sup_{n\ge 2} \mathrm{E}\exp(a\,\Delta_{n\nu}) < \infty.$$

[Hint. This is too hard to be an "exercise," but it is a very nice bound.]

11 Inequalities for Various Empirical Processes

We wish to apply the Birnbaum–Marshall and Hájek–Rényi inequalities to various martingales (mgs) associated with the processes of the previous section.

Proposition 11.1 (Various martingales)

(1) $\{\mathbb{U}_n(t)/(1-t) : 0 \le t < 1\}$ is a mg.

(2) $\{\mathbb{W}_n(t)/(1-t) : 0 \le t < 1\}$ is a mg.

(3) $\{\mathbb{U}(t)/(1-t) : 0 \le t < 1\}$ is a mg.

(4) $\{\mathbb{V}_n(k/(n+1)) : (1-k/(n+1)) : 0 \le k \le n\}$ is a mg.

(5) $\{\mathbb{R}_n(k/(n+1))/(1-k/n) : 0 \le k \le n-1\}$ is a mg.

Proof. Let $\mathcal{A}_t \equiv \sigma[1_{[\xi \le s]} : 0 \le s \le t]$. Then

(a) $\mathrm{E}\{1_{[\xi \le t]} - t|\mathcal{A}_s\} = 1_{[\xi \le s]} + \frac{t-s}{1-s} 1_{[\xi > s]} - t$

$= 1_{[\xi \le s]} + \frac{t-s}{1-s}\{1 - 1_{[\xi \le s]}\} - t$

(b) $= \frac{1-t}{1-s}\{1_{[\xi \le s]} - s\}$,

so that

(6) $[1_{[\xi \le t]} - t]/(1-t), \quad 0 \le t \le 1,$ is a mg.

Noting (12.10.7), summing (6) shows that (1) and (2) hold.

Let $\mathcal{A}_t \equiv \sigma[\mathbb{U}(s) : 0 \le s \le t]$. Letting $\Sigma_{st} \equiv \mathrm{Cov}[\mathbb{U}(s), \mathbb{U}(t)] = s \wedge t - st$,

(c) $\mathrm{E}(\mathbb{U}(t)|\mathcal{A}_s) = \{[s(1-t)]/[s(1-s)]\}\mathbb{U}(s)$,

since $\mathbb{U}(t)|\mathbb{U}(s)$ is normal with mean $\mu_t + \Sigma_{ts}\Sigma_{ss}^{-1}[\mathbb{U}(s) - \mu_s]$. Thus (3) holds.

Consider (5). Let $Z_{nk} \equiv \mathbb{R}_n(k/(n+1))/(1-k/n)$, and set $\Delta Z_{nk} \equiv Z_{nk} - Z_{n,k-1}$ for integers $1 \le k \le n-1$. Then

$$\Delta Z_{nk} = \frac{n}{n-k}\mathbb{R}\left(\frac{k}{n+1}\right) - \frac{n}{n-k}\,\frac{n-k}{n-k+1}\,\mathbb{R}\left(\frac{k-1}{n+1}\right)$$

(d) $$= \frac{1}{\sqrt{n}}\,\frac{n}{n-k}\left[c_{nD_{nk}} + \frac{1}{n-k+1}\sum_{j=1}^{k-1} c_{nD_{nj}}\right].$$

Let $\mathcal{A}_k \equiv \sigma[D_{n1}, \ldots, D_{nk}]$. Then

$$\mathrm{E}(\Delta Z_{nk}|\mathcal{A}_{k-1}) = \frac{1}{\sqrt{n}}\,\frac{n}{n-k}\,[\mathrm{E}(c_{nD_{nk}}|\mathcal{A}_{k-1}) + \frac{1}{n-k+1}\sum_{j=1}^{k-1} c_{nD_{nj}}]$$

(e) $$= \frac{1}{\sqrt{n}}\,\frac{n}{n-k}\left[\frac{1}{n-k+1}\sum_{j=k}^{n} c_{nD_{nj}} + \frac{1}{n-k+1}\sum_{j=1}^{k-1} c_{nD_{nj}}\right]$$

$= 0, \quad$ since $\bar{c}_n = 0$.

Apply the finite sampling results (A.1.8) and (A.1.9) to (d) to conclude that

$$\text{(f)} \qquad \mathrm{Var}[\Delta Z_{nk}] = \tfrac{\sigma^2_{c,n}}{n}\left(\tfrac{n}{n-k}\right)^2\left\{1 - \tfrac{2}{n-k+1}\,\tfrac{k-1}{n-1} + \tfrac{k-1}{(n-k+1)^2}\,\left[1 - \tfrac{(k-1)-1}{n-1}\right]\right\}$$

$$\text{(7)} \qquad = \frac{\sigma^2_{c,n}}{n-1}\,\frac{n^2}{(n-k)(n-k+1)}\,.$$

Thus (5) holds. Consider (4).

Let $\mathcal{A}_k \equiv \sigma[\xi_{n:1}, \ldots \xi_{n:k}]$. Then

$$\text{(g)} \qquad \mathrm{E}(\xi_{n:k}|\mathcal{A}_i) - \tfrac{k}{n+1} = \xi_{n:i} + \tfrac{k-i}{n-i+1}\,[1 - \xi_{n:i}] - \tfrac{k}{n+1} = \tfrac{n-k+1}{n-i+1}\,\left[\xi_{n:i} - \tfrac{i}{n+1}\right],$$

since the conditional distribution of $\xi_{n:k}$ given $\xi_{n:i}$ is that of the $(k-i)$th order statistic in a sample of size $n-i$ from Uniform$(\xi_{n:i}, 1)$, and (A.1.32) can be applied. Thus (4) holds. □

Inequality 11.1 (Pyke–Shorack) Let $\mathbb{X}$ denote one of the processes $\mathbb{U}_n$, $\ddot{\mathbb{V}}_n$, $\mathbb{W}_n$, $\mathbb{R}_n$, or $\mathbb{U}$. Let $q > 0$ on $[0, \theta]$ be $\nearrow$ and right continuous. Then for all $\lambda > 0$ we have the probability bound

$$\text{(8)} \qquad P(\|\mathbb{X}/q\|_0^\theta \geq \lambda) \leq (16/\lambda^2)\int_0^\theta [q(t)]^{-2}\,dt\,.$$

Proof. Let $\mathbb{X}$ denote any one of $\mathbb{U}_n$, $\mathbb{W}_n$, or $\mathbb{U}$. Then $\mathbb{X}(t)/(1-t)$ is a mg with mean 0 and variance $\nu(t) \equiv t/(1-t)$. Thus the Birnbaum–Marshall inequality gives

$$P(\|\,\mathbb{X}(t)/q(t)\,\|_0^\theta \geq \lambda) = P(\|\,[\mathbb{X}(t)/(1-t)]/[q(t)/(1-t)]\,\|_0^\theta \geq \lambda)$$

$$\text{(a)} \qquad \leq (4/\lambda)^2 \int_0^\theta [(1-t)/q(t)]^2\, d\,[t/(1-t)] = (4/\lambda)^2 \int_0^\theta [q(t)]^2\,dt\,.$$

Let $\mathbb{X}$ denote $\mathbb{R}_n$. Then, with $b_k \equiv q(k/(n+1))$ and $m \equiv [(n+1)\theta]$,

$$P(\|\,\mathbb{R}_n(t)/q(t)\,\|_0^\theta \geq \lambda) = P(\max_{1\leq k\leq m} |\,\mathbb{R}_n(k/(n+1))\,|/b_k \geq \lambda)$$

$$\leq P\left(\max_{1\leq k\leq m}\, \tfrac{|\mathbb{R}_n(k/(n+1))/(1-k/n)|}{b_k/(1-k/n)} \geq \lambda\right)$$

$$\text{(b)} \qquad \leq \tfrac{4}{\lambda^2}\sum_{k=1}^m \tfrac{\mathrm{Var}[\Delta Z_{nk}]}{[b_k/(1-k/n)]^2} \qquad \text{by (b) and Hájek–Rényi}$$

$$\leq \tfrac{4}{\lambda^2}\sum_{k=1}^m \tfrac{1}{n-1}\,\tfrac{n^2}{(n-k)(n-k+1)}\,\tfrac{(n-k)^2}{n^2\,b_k^2} \qquad \text{by (7)}$$

$$\text{(c)} \qquad \leq \tfrac{4}{\lambda^2}\,\tfrac{1}{n}\sum_{k=1}^m b_k^{-2} \leq \tfrac{16}{\lambda^2}\int_0^\theta [q(t)]^{-2}\,dt\,.$$

(We can improve (a) and (c) by a factor of 4, as stated in the Hájek–Rényi inequality, but there is no real point to this.) □

Exercise 11.1 Verify (8) for $\ddot{\mathbb{V}}_n$.

Inequality 11.2 (In probability linear bounds on $\mathbb{G}_n$ and $\mathbb{G}_n^{-1}$) For all $\epsilon > 0$ there exists $\lambda \equiv \lambda_\epsilon$ so small that the event $A_{n\epsilon}$ on which

(9) $\mathbb{G}_n(t) \le t/\lambda \quad \text{on } [0,1], \qquad \mathbb{G}_n(t) \ge \lambda t \quad \text{on } [\xi_{n:1}, 1],$

(10) $\mathbb{G}_n(1-t) \le 1-\lambda(1-t) \quad \text{on } [0, \xi_{n:n}), \qquad \mathbb{G}_n(1-t) \ge 1-(1-t)/\lambda \quad \text{on } [0,1]\,,$

(11) $|\mathbb{G}_n(t) - t| \le 1/(\lambda\sqrt{n}\,) \quad \text{on } [0,1]$

has $P(A_{n\epsilon}) \ge 1 - \epsilon$ for all $n \ge 1$. Let $1_{n\epsilon}$ denote the indicator function of $A_{n\epsilon}$. (These conclusions hold for *any* realization of $\mathbb{G}_n$ and $\mathbb{G}_n^{-1}$.) (Note that linear bounds on $\mathbb{G}_n^{-1}$ are also established by this result.)

Proof. Now, $(\xi_{n:1}, \ldots \xi_{n:n})$ has joint density $n!$ on its domain. Thus

(a) $P(\mathbb{G}_n(t) \le t/\lambda \text{ for } 0 \le t \le 1) = P(\xi_{n:k} \ge \lambda k/n \text{ for } 1 \le k \le n)$

$= \int_\lambda^1 \int_{\lambda(n-1)/n}^{t_n} \cdots \int_{\lambda 2/n}^{t_3} \int_{\lambda/n}^{t_2} n!\, dt_1 \cdots dt_n = \cdots$

(b) $= n! \left[\frac{t^n}{n!} - \frac{\lambda t^{n-1}}{n!}\right]\Big|_\lambda^1 = 1 - \lambda$

(c) $\ge 1 - \epsilon/3 \qquad \text{for all } \lambda \le \lambda_\epsilon \equiv \epsilon/3,$

and for all n. This gives the upper bound of (9). And (8.3.20) gives

Daniels' equality

$P(\|\mathbb{G}_n/I\| \le \lambda) = P(\mathbb{G}_n(t) \le t/\lambda \text{ for } 0 \le t \le 1)$

(12) $= P(\xi_{n:k} \ge \lambda k/n \text{ for } 1 \le k \le n) = P(\mathbb{G}_n^{-1}(t) \ge \lambda t \text{ for all } 0 \le t \le 1)$

$= 1 - \lambda \quad \text{for all } 0 \le \lambda \le 1\,.$

We now turn to the lower bound of (9). Now,

(13) $S_k \equiv n\,\xi_{n:k+1}/k, \qquad 1 \le k \le n-1,$ is a reversed mg,

as a rearrangement of $\mathrm{E}(\xi_{n:k}|\xi_{n:k+1}) = [k/(k+1)]\,\xi_{n:k+1}$ shows. So, immediately,

(d) $\tilde{S}_k \equiv S_{(n-1)-k+1} = S_{n-k} = n\,\xi_{n:n-k+1}/(n-k)$ is a mg for $1 \le k \le n-1$.

Now calculate

$1 - P(\mathbb{G}_n(t) \ge \lambda t \text{ everywhere on } [\xi_{n:1}, 1]) \qquad [\text{or } = P(\|I/\mathbb{G}_n\|_{\xi_{n:1}}^1 > 1/\lambda)]$

(e) $= P(\xi_{n:k+1} > (k/n)/\lambda \ \text{ for some } 1 \le k \le n-1)$

$= P(\max_{1\le k\le n-1} S_k > 1/\lambda) = P(\max_{1\le k\le n-1} \tilde{S}_k > 1/\lambda)$

$= P(\max_{1\le k\le n-1} \exp(r\tilde{S}_k) > \exp(r/\lambda))$

(f) $\le \inf_{r>0}\, [e^{-r/\lambda}\, \mathrm{E}e^{r\tilde{S}_{n-1}}] = \inf_{r>0}\, [e^{-r/\lambda}\, \mathrm{E}e^{rn\,\xi_{n:2}}]$ by Doob's (8.9.3)

(g) $= \inf_{r>0} \int_0^1 e^{-r/\lambda}\, e^{rnt}\, n(n-1)\, t(1-t)^{n-2}\, dt$

$= \inf_{r>0} e^{-r/\lambda} \int_0^n e^{rs}\, n(n-1)\,(s/n)\,(1-s/n)^{n-2}\, ds/n$

$\le \inf_{r>0} e^{-r/\lambda} \int_0^n s e^{rs}\,(1-s/n)^{n-2}\, ds$

$\le \inf_{r>0} e^{-r/\lambda} \int_0^n s\, e^{rs}\, e^{-s}\, e^2\, ds$ since $1 - s/n \le e^{-s/n}$

$\le e^2 \inf_{r>0} e^{-r/\lambda} \int_0^\infty s\, \exp(-(1-r)s)\, ds$

(h) $= e^2 \inf_{r>0} e^{-r/\lambda}/(1-r)^2$ from the mean of an exponential density

(i) $= (e^2/4\lambda^2)\exp(-1/\lambda)$ since differentiation gives $r = 1 - 2\lambda$

(j) $< \epsilon/3$

for $\lambda \equiv \lambda_\epsilon$ small enough. Thus the lower bound in (9) holds. Then (10) follows from (9) by symmetry. Finally, (11) holds since $\|\mathbb{U}_n\| = O_p(1)$. In fact, we have

Chang's inequality

(14)
$$P(\|I/\mathbb{G}_n^{-1}\|_{\xi_{n:1}}^1 \le x) = P(\mathbb{G}_n(t) \ge t/x \text{ on all of } [\xi_{n:1}, 1])$$
$$\ge 1 - 2\, x^2\, e^{-x} \quad \text{for all } x \ge 1. \qquad \Box$$

Proof. Consider Mason's theorem 12.10.4. Apply the Pyke–Shorack inequality with divisor $q(t) \equiv (a \vee t)^{1-\nu}$ to obtain

$$P(n^\nu \|\,(\mathbb{G}_n(t) - t)/t^{1-\nu}\,\|_a^b \ge \lambda) = P(\|\,\mathbb{U}_n/t^{1-\nu}\,\|_a^b \ge \lambda\, n^{(1/2)-\nu})$$

(a) $\le P(\|\mathbb{U}_n/q\|_0^b \ge \lambda n^{1/2-\nu}) \le 4\int_0^b (a \vee t)^{-(2-2\nu)}\, dt/(\lambda^2\, n^{1-2\nu})$

$= \frac{4}{\lambda^2\,(an)^{1-2\nu}} + \frac{4}{\lambda^2\, n^{1-2\nu}} \int_a^b t^{-(2-2\nu)}\, dt$

$= \frac{4}{\lambda^2\,(an)^{1-2\nu}} - \frac{4}{\lambda^2\, n^{1-2\nu}} \cdot \frac{1}{(1-2\nu)t^{1-2\nu}}\Big|_a^b$

(b) $\le 8(1-2\nu)^{-1}/[\lambda^2\,(an)^{1-2\nu}]\,.$

Using $a = 1/n$, $b = \frac{1}{2}$ and the symmetry about $\frac{1}{2}$ gives

(15) $$P\left(\left\|\frac{n^\nu[\mathbb{G}_n(t) - t]}{[t \wedge (1-t)]^{1-\nu}}\right\|_{1/n}^{1-1/n} \ge \lambda\right) \le \left[\frac{16}{(1-2\nu)}\right]\frac{1}{\lambda^2}\,.$$

But $[0, 1/n]$ is easy (and $[1 - 1/n, 1]$ is symmetric), since on $[0, 1/n]$ we have

(c) $n^\nu|\mathbb{G}_n(t) - t|/t^{1-\nu} \le (nt)^\nu[1 + \mathbb{G}_n(t)/t] \le 1 + \mathbb{G}_n(t)/t\,;$

and thus (9) gives

(16) $$P\left(\left\|\frac{n^\nu[\mathbb{G}_n(t) - t]}{[t \wedge (1-t)]^{1-\nu}}\right\|_0^{1/n} \ge \lambda\right) \le \frac{1}{\lambda - 1} \qquad \text{for } \lambda > 1.$$

We can repeat this same proof up to (13) with $\ddot{\mathbb{G}}_n^{-1}$ and $\ddot{\mathbb{V}}_n$ replacing $\mathbb{G}_n$ and $\mathbb{U}_n$, because of the Pyke–Shorack inequality. Then $(0, 1/n]$ is trivial for $\ddot{\mathbb{G}}_n^{-1}$, as the values on this whole interval are deterministically related to the value at $1/n$. □

Exercise 11.2 Prove the Pyke–Shorack theorem 12.10.2. [Hint. Model your proof on (a) of the previous proof, with $a = 0$ and b sufficiently small, and with theorem 12.10.1 sufficient on $[b, 1 - b]$.]

12 Applications

Theorem 12.1 (Donsker) Let $g : (D, \mathcal{D}) \to (R, \mathcal{B})$ denote an a.s. $\|\cdot\|$-continuous mapping that is $\mathcal{D}$-measurable. Then $g(\mathbb{U}_n) : (\Omega, \mathcal{A}, P) \to (R, \mathcal{B})$, and

(1) $\quad g(\mathbb{U}_n) \to_p g(\mathbb{U}) \quad$ as $n \to \infty \quad$ for the $\mathbb{U}_n$ of (12.10.28),

(2) $\quad g(\mathbb{U}_n) \to_d g(\mathbb{U}) \quad$ as $n \to \infty \quad$ for an arbitrary $\mathbb{U}_n$.

[These conclusions hold for $\mathcal{D}$-measurable functionals g that are continuous in other $\|\cdot/q\|$-metrics as well.]

Exercise 12.1 Write out the easy details to prove this Donsker theorem.

Example 12.1 (Tests of fit) (i) Call F *stochastically larger* than F_0 whenever $P_F(X > x) \ge P_{F_0}(X > x)$ for all x (with strict inequality for at least one x), and write $F \ge_s F_0$. To test the null hypothesis H_0 that $F = F_0$ is true against the alternative hypothesis H_a that $F \ge_s F_0$ it is reasonable to reject the H_0 claim that for large values of *Birnbaum's statistic* $Z_n \equiv \int_{-\infty}^{\infty} \sqrt{n}\,[\,\mathbb{F}_n(x) - F_0(x)\,]\,dF_0(x)$. Now suppose that H_0 is true, with a continuous df F_0. Then

(3) $$Z_n \cong \int_{-\infty}^{\infty} \mathbb{U}_n(F_0)\,dF_0 = \int_0^1 \mathbb{U}_n(t)\,dt \to_d Z \equiv \int_0^1 \mathbb{U}(t)\,dt \cong N(0, \tfrac{1}{12})\,.$$

(ii) Alternatively, one could form the *Cramér–von Mises statistic*

(4) $$W_n \equiv \int_{-\infty}^{\infty} \{\sqrt{n}\,[\,\mathbb{F}_n(x) - F_0(x)\,]\}^2\,dF_0(x)$$

$$\cong \int_{-\infty}^{\infty} \mathbb{U}_n^2(F_0)\,dF_0 \qquad \text{by (6.5.22)}$$

$$= \int_0^1 \mathbb{U}_n^2(t)\,dt \qquad \text{when } F_0 \text{ is continuous, by (6.3.10)}$$

(5) $$\to_d \int_0^1 \mathbb{U}^2(t)\,dt$$

$$= \int_0^1 \{\textstyle\sum_{k=1}^{\infty} \phi_k(t) \frac{1}{\pi k}\, Z_k\}\{\sum_{j=1}^{\infty} \phi_j(t) \frac{1}{\pi j}\, Z_j\}\,dt \qquad \text{(see below)}$$

for the orthonormal functions $\phi_k(t) \equiv \sqrt{2}\,\sin(\pi k t)$ on $[0, 1]$

and iid N$(0, 1)$ rvs Z_k

$$= \textstyle\sum_{k=1}^{\infty} \sum_{j=1}^{\infty} \frac{1}{\pi^2\, jk}\, Z_j Z_k \int_0^1 \phi_k(t)\phi_j(t)\,dt$$

(6) $$= \textstyle\sum_{k=1}^{\infty} \frac{1}{\pi^2\, k^2}\, Z_k^2\,.$$

This shows that W_n is asymptotically distributed as an infinite weighted sum of independent χ_1^2 rvs. This representation of the limiting distribution has been used to provide tables. If $W_n \equiv W_n(F_0)$ is computed but a different df F is true, then

(7) $$\tfrac{1}{n}\, W_n = \int_{-\infty}^{\infty} [\,\mathbb{F}_n - F_0\,]^2\,dF_0 \to_{a.s.} \int_{-\infty}^{\infty} [\,F - F_0\,]^2\,dF_0 > 0\,.$$

[In statistical parlance, this shows that the W_n-test is consistent against any df alternative $F \neq F_0$.]
(iii) A third possibility is the *Anderson–Darling statistic*

$$(8) \qquad A_n \equiv \int_{-\infty}^{\infty} \frac{\{\sqrt{n}\,[\,\mathbb{F}_n - F_0\,]\}^2}{F_0(1-F_0)}\, dF_0 = \int_0^1 \frac{\mathbb{U}_n^2(t)}{t(1-t)}\, dt$$

for F_0 continuous

$$(9) \qquad \to_d \int_0^1 \frac{\mathbb{U}^2(t)}{t(1-t)}\, dt \cong \sum_{k=1}^{\infty} \tfrac{1}{k(k+1)}\, Z_k^2 \,. \qquad \square$$

Proof. (i) Consider Birnbaum's Z_n.

Method 1: By (6.5.22) and then the change of variable theorem of (6.3.10) (with identity function H) one obtains the first two steps of (3). Apply Donsker for the third step. Appeal to (12.3.19) for the $\to_d$ to a normal rv Z. Finally, appeal to Fubini's theorem for both

$$(a) \qquad \mathrm{E}Z = \mathrm{E}\int_0^1 \mathbb{U}(t)\, dt = \int_0^1 \mathrm{E}(\mathbb{U}(t))\, dt = \int_0^1 0\, dt = 0 \qquad \text{and}$$

$$\mathrm{E}(Z^2) = \mathrm{E}\{\int_0^1\int_0^1 \mathbb{U}(s)\,\mathbb{U}(t)\, ds\, dt\}$$

$$= \int_0^1\int_0^1 \mathrm{E}\{\mathbb{U}(s)\,\mathbb{U}(t)\}\, ds\, dt = \int_0^1\int_0^1 [s\wedge t - s\,t]\, ds\, dt$$

$$(b) \qquad = \int_0^1\int_0^t s(1-t)\, ds\, dt = 1/12\,.$$

Method 2: Apply (12.10.28) for

$$(c) \qquad \int_0^1 |\mathbb{U}_n(t) - \mathbb{U}(t)|\, dt \le \int_0^1 1\, dt \times \|\mathbb{U}_n - \mathbb{U}\| \to_p 0$$

to replace step three in the above. Thus $Z_n \to_d Z$.

The rest of the justification of example 12.1 is outlined in exercise 12.2 and exercise 12.3. $\square$

Exercise 12.2 Consider the Cramér–von Mises statistic W_n.
(I) Verify step (5). Use (12.10.28).
(II) We now seek to justify the step representing $\mathbb{U}$ as an infinite series. To this end formally write

$$(p) \qquad \mathbb{U}(t) = \sum_1^{\infty} \phi_k(t)\tfrac{1}{\pi\, k}\, Z_k$$

for iid N(0, 1) rvs Z_k and the orthonormal functions $\phi_k(\cdot)$. First recall the group of trigonometric identities

$$(q) \qquad \begin{aligned} &\sin(A+B) = \sin A\cos B + \cos A\sin B\,,\\ &2\sin A\cos B = \sin(A+B) + \sin(A-B)\,,\\ &\cos(A+B) = \cos A\cos B - \sin A\sin B\,,\\ &2\cos A\sin B = \sin(A+B) - \sin(A+B)\;. \end{aligned}$$

Use these to verify that $\int_0^1 \phi_j(t)\phi_k(t)\,dt$ equals 0 or 1 according as $j \neq k$ or $j = k$. [Think of this formal $\mathbb{U}(t)$ as an odd function on $[-1, 1]$, and thus only these $\phi_k(\cdot)$ are needed.] Then note that the Fourier coefficients and the Fourier series are

(r) $\langle \mathbb{U}, \phi_k \rangle = \int_0^1 \mathbb{U}(t)\,\phi_k(t)\,dt = \frac{1}{\pi\,k}\,Z_k\,,$

(s) $\mathbb{U} = \sum_{k=1}^\infty \langle \mathbb{U}, \phi_k \rangle\,\phi_k = \sum_{k=1}^\infty \phi_k\,\frac{1}{\pi\,k}\,Z_k\,.$

So, verify that the series in (p) converges a.s. and then everything so far for the formal $\mathbb{U}$ is rigorous. Then Parseval's identity (note theorem **??.??**) gives

(t) $\int_0^1 \mathbb{U}^2(t)\,dt = \|\,\mathbb{U}\,\|^2 = \sum_1^\infty |\langle \mathbb{U}, \phi_k \rangle|^2 = \sum_1^\infty \frac{1}{\pi^2\,k^2}\,Z_k^2\,.$

Finally, one needs to verify the step (u) in the identity

$$\mathrm{E}\{\,\sum_{j=1}^\infty \phi_j(s)\,\frac{1}{\pi\,j}\,Z_j\,\times\,\sum_{k=1}^\infty \phi_k(t)\,\frac{1}{\pi\,k}\,Z_k\,\} = \sum_1^\infty \frac{1}{\pi^2\,k^2}\,\phi_k(s)\,\phi_k(t)$$

$$= \frac{2}{\pi^2}\sum_1^\infty \frac{1}{k^2}\,\sin(\pi\,k\,s)\,\sin(\pi\,k\,t)$$

(u) $$= s \wedge t - s\,t\,,$$

and thus the (originally formal) process $\mathbb{U}$ is in fact a Brownian bridge. Where did this idea come from? Verifying that

(v) $\int_0^1 \mathrm{Cov}[s,t]\,\phi_k(s)\,ds = \frac{1}{\pi\,k}\,\phi_k(t)$ on $[0,1]$

shows that $\mathrm{Cov}[s,t] \equiv \mathrm{Cov}[\mathbb{U}(s), \mathbb{U}(t)] = s \wedge t - st$ has eigenvalues $\frac{1}{\pi\,k}$ with associated eigenfunctions $\phi_k(\cdot)$ for $k = 1, 2, \ldots$. [Recall the spectral decomposition of matrices in (A.3.2)–(A.3.4).]

Exercise 12.3 Verify the results claimed for the Anderson–Darling statistic A_n. [Verifying $\to_d$ will be a little trickier this time, since (12.10.30) will now be needed in place of (12.10.28).] The rest is roughly similar in spirit, but the details are now a geat deal more complicated. Fundamentally, one must now represent the covariance function

$$\mathrm{Cov}[s,t] = (s \wedge t - s\,t)/\sqrt{s(1-s)\;t(1-t)}$$

as a convergent infinite series of orthonormal functions. (Hopefully, at least the approach is now clear. Providing the details is hard work.)

Chapter 13

Martingales

1 Basic Technicalities for Martingales

Notation 1.1 We will work with processes on the following time sets I: $\{0,\ldots,n\}$, $\{0,1,\ldots\}$, $\{0,1,\ldots,\infty\}$, $\{\ldots,-1,0\}$, $\{-\infty,\ldots,-1,0\}$ in the discrete case and $[0,t]$, $[0,\infty)$, $[0,\infty]$, $(-\infty,0]$, $[-\infty,0]$ in the continuous case. In the continuous cases we will consider only processes of the type $X:(\Omega,\mathcal{A},P)\to(D_I,\mathcal{D}_I)$ that are adapted to an $\nearrow$ sequence of sub σ-fields $\mathcal{A}_t$ of $\mathcal{A}$. We will use the notation $\{a_n\}_{n=0}^{\infty}$, $\{a_n\}_{n=0}^{\overline{\infty}}$, $\{a_n\}_{n=-\infty}^{0}$, $\{a_n\}_{n=-\underline{\infty}}^{0}$ to denote sequences over $\{0,1,\ldots\}$, $\{0,1,\ldots,\infty\}$, $\{\ldots,-1,0\}$, $\{-\infty,\ldots,-1,0\}$, respectively. □

Definition 1.1 (Martingale and submartingale) Suppose $\mathrm{E}|X_t|<\infty$ for all t. Call $\{X_t,\mathcal{A}_t\}_{t\in I}$ a *martingale* (abbreviated *mg*) if

(1) $\quad E(X_t|\mathcal{A}_s)=X_s \quad$ a.s. $\qquad$ for each pair $s\le t$ in I.

Call $\{X_t,\mathcal{A}_t\}_{t\in I}$ a *submartingale* (abbreviated *submg*) if

(2) $\quad \mathrm{E}(X_t|\mathcal{A}_s)\ge X_s \quad$ a.s. $\qquad$ for each pair $s\le t$ in I.

(If the inequality in (2) is reversed, the process $\{X_t,\mathcal{A}_t\}_{t\in T}$ is then called a *supermartingale*.) When the index set I is a subset of the negative numbers $[-\infty,0]$, we refer to such a process as a *reversed martingale* or *reversed submartingale*.
[Most results in the first seven sections of this chapter are due to Doob.]

Basic Technicalities

Proposition 1.1 (Equivalence) Now, $\{X_t,\mathcal{A}_t\}_{t\in I}$ is a submg if and only if the moments $\mathrm{E}|X_t|<\infty$ for all $t\in I$ and for every pair $s\le t$ we have

(3) $\quad \int_A (X_t-X_s)\,dP\ge 0 \qquad$ for all $A\in\mathcal{A}_s$.

Similarly, $\{X_t,\mathcal{A}_t\}_{t\in I}$ is a mg if and only if equality holds in (3).

Notation As in section 8.9, we combine these two statements by writing

(4) $\quad \{X_t,\mathcal{A}_t\}_{t\in I}$ is a s-mg $\quad$ iff $\quad \int_A (X_t-X_s)\,dP \stackrel{\geq}{=} 0 \quad$ for all $A\in\mathcal{A}_s$.

Proof. As in section 8.9, for each pair $s \le t$,

(a) $\int_A (X_t - X_s)\, dP = \int_A \mathrm{E}\,(X_t - X_s|\mathcal{A}_s)\, dP \geqq 0$ for all $A \in \mathcal{A}_s$

if and only if $\mathrm{E}(X_t|\mathcal{A}_s) - X_s = \mathrm{E}(X_t - X_s|\mathcal{A}_s) \geqq 0$ a.s. $\mathcal{A}_s$. □

Definition 1.2 (a) Call $\{X_t, \mathcal{A}_t\}_{t\in I}$ *integrable* if $\sup\{\mathrm{E}|X_t| : t \in I\} < \infty$.
(b) If $\{X_t^2, \mathcal{A}_t\}_{t\in I}$ is integrable, then $\{X_t, \mathcal{A}_t\}_{t\in I}$ is called *square-integrable.*

Proposition 1.2 Let $\phi : (R, \mathcal{B}) \to (R, \mathcal{B})$ have $\mathrm{E}|\phi(X_t)| < \infty$ for all $t \in I$.

(a) If ϕ is convex and $\{X_t, \mathcal{A}_t\}_{t\in I}$ is a mg, then $\{\phi(X_t), \mathcal{A}_t\}_{t\in I}$ is a submg.

(b) If ϕ is convex and $\nearrow$ and $\{X_t, \mathcal{A}_t\}_{t\in I}$ is a submg, then $\{\phi(X_t), \mathcal{A}_t\}_{t\in I}$ is a submg.

Proof. Clearly, $\phi(X_t)$ is adapted to $\mathcal{A}_t$. Let $s \le t$. For the mg case,

(a) $\mathrm{E}[\phi(X_t)|\mathcal{A}_s] \ge \phi(\mathrm{E}(X_t|\mathcal{A}_s))$ by the conditional Jensen inequality

(b) $= \phi(X_s)$ a.s. (since $\{X_t, \mathcal{A}_t\}_{t\in I}$ is a mg)

For the submg case,

(c) $\mathrm{E}[\phi(X_t)|\mathcal{A}_s] \ge \phi(\mathrm{E}(X_t|\mathcal{A}_s))$ by the conditional Jensen inequality

(d) $\ge \phi(X_s)$ a.s.,

since ϕ is $\nearrow$ and $\mathrm{E}(X_t|\mathcal{A}_s) \ge X_s$ a.s. □

Example 1.1 Let $\{X_t, \mathcal{A}_t\}_{t\in I}$ be a martingale. Then:

(5) $\{|X_t|^r, \mathcal{A}_t\}_{t\in I}$ is a submg, for any $r \ge 1$ having $\mathrm{E}|X_t|^r < \infty$ for all $t \in I$,

(6) $\{X_t^-, \mathcal{A}_t\}_{t\in I}$ is a submg,

(7) $\{X_t^+, \mathcal{A}_t\}_{t\in I}$ is a submg (even if $\{X_t, \mathcal{A}_t\}_{t\in I}$ is only a submg).

[Note that $\phi(x) = |x|^r$ and $\phi(x) = x^-$ are convex, while $\phi(x) = x^+$ is also $\nearrow$.] □

Proposition 1.3 If $\{X_t, \mathcal{A}_t\}_{t\in I}$ and $\{Y_t, \mathcal{A}_t\}_{t\in I}$ are s-mgs, then (trivially)

(8) $\{X_t + Y_t, \mathcal{A}_t\}_{t\in I}$ is a s-mg.

Exercise 1.1 If $\{X_t^c, \mathcal{A}_t\}_{t\in I}$ is a submg for all c in some index set C, then the maximum $\{X_t^{c_1} \vee X_t^{c_2}, \mathcal{A}_t\}_{t\in I}$ is necessarily a submg for any $c_1, c_2 \in C$. Likewise, $\{\sup_{c\in C} X_t^c, \mathcal{A}_t\}_{t\in I}$ is a submg, provided that $\mathrm{E}|\sup_{c\in C} X_t^c| < \infty$ for all $t \in I$.

Definition 1.3 (Augmented filtration) Let $(\Omega, \mathcal{A}, P)$ be a complete probability space. Let $\mathcal{N} \equiv \{N \in \mathcal{A} : P(N) = 0\}$. Let $\{\mathcal{A}_t : t \geq 0\}$ be such that the $\mathcal{A}_t$'s are an $\nearrow$ sequence of σ-fields with $\mathcal{A}_t = \hat{\mathcal{A}}_t = \mathcal{A}_{t+}$ for all $t \geq 0$ (here $\hat{\mathcal{A}}_t \equiv \sigma[\mathcal{A}_t, \mathcal{N}]$ and $\mathcal{A}_{t+} \equiv \cap\{\mathcal{A}_r : r > t\}$). [That is, they are complete and right continuous.] Such a collection of σ-fields is called an *augmented filtration*.

Notation 1.2 (Completeness assumption) In this chapter we will assume that the σ-fields $\mathcal{A}_t$ form an augmented filtration in that completion has already been performed on the σ-fields labeled $\mathcal{A}_t$. Thus, from proposition 12.4.4(c), we see that $S \leq T$ a.s. implies $\mathcal{A}_S \subset \mathcal{A}_T$. For right-continuous processes on $(D_{[0,\infty)}, \mathcal{D}_{[0,\infty)})$ this effectively comes for free; see proposition 12.4.4. If $\{X_t, \mathcal{A}_t\}_{t\in I}$ is a s-mg, then $\{X_t, \hat{\mathcal{A}}_t\}_{t\in I}$ is also a s-mg; note exercise 1.2 below. □

Exercise 1.2 Verify the claim made in the previous assumption.

Exercise 1.3 If X is a process on $(D, \mathcal{D})$ or $(D_{[0,\infty)}, \mathcal{D}_{[0,\infty)})$, then the histories $\sigma_t \equiv \sigma[X_s : s \leq t]$ are right continuous, as are the $\hat{\sigma}_t \equiv \sigma[\sigma_t \cup \mathcal{N}]$. (Recall (12.4.13) of proposition 12.4.3, proposition 12.4.4, and exercise 1.2.1.)

Remark 1.1 All definitions and results in this section make sense for processes on the measurable space $(R_I, \mathcal{B}_I)$. □

Some Examples

Example 1.2 (Sums of iids) Let $X_1, X_2, \ldots$ be iid with $\mathrm{E}(X_i) = 0$, and let $S_n \equiv X_1 + \cdots + X_n$ and $\mathcal{A}_n \equiv \sigma[S_1, \ldots, S_n]$. Then $\mathrm{E}|S_n| \leq \sum_{i=1}^n \mathrm{E}|X_i| < \infty$ and so $\{S_n, \mathcal{A}_n\}_{n=1}^\infty$ is a mg. □

Example 1.3 As in example 1.2, but now assume that $\mathrm{E}(X_k^2) \equiv \sigma^2 < \infty$. Let $Y_n \equiv S_n^2 - n\sigma^2$. Then $\{Y_n, \mathcal{A}_n\}_{n=1}^\infty$ is a mg. Note also that S_n^2 is a submg by proposition 1.2, and that we have written

$$S_n^2 = (S_n^2 - n\sigma^2) + n\sigma^2 = (\text{martingale}) + (\text{increasing process}).$$

This is an example of the Doob decomposition of a submartingale, which we will establish in section 5. □

Example 1.4 Suppose $\mu \equiv \mathrm{E}(X_i) > 0$ in example 1.2. Then the partial sums $\{S_n, \mathcal{A}_n\}_{n=1}^\infty$ form a submg. □

Example 1.5 (Wald's mg) Consider again example 1.2, but now suppose that the X_k's have a mgf $\phi(t) = \mathrm{E}\exp(tX)$. Let $Y_n \equiv \exp(cS_n)/\phi(c)^n$. Then $\{Y_n, \mathcal{A}_n\}_{n=1}^\infty$ is a mg. Note that the mg of example 1.2 is recovered by differentiating once with respect to c and setting $c = 0$; the mg of example 1.4 is recovered by differentiating twice with respect to c and setting $c = 0$. □

Example 1.6 (Brownian motion) Let $\{\mathbb{S}(t) : t \geq 0\}$ be standardized Brownian motion, and let $\mathcal{A}_t \equiv \sigma[\mathbb{S}(s) : s \leq t]$. Then $\{\mathbb{S}(t), \mathcal{A}_t : t \geq 0\}$ is a mg. □

Example 1.7 Let $Y(t) \equiv \mathbb{S}(t)^2 - t$ in example 1.6. Then the Brownian motion transform $\{Y(t), \mathcal{A}_t : t \ge 0\}$ is a mg. □

Example 1.8 (The exponential mg for Brownian motion) As in example 1.6, let $\mathbb{S}$ denote standard Brownian motion, and much as in example 1.5, set

$$Y(t) = \exp(c\,\mathbb{S}(t))/\exp(c^2 t/2) = \exp(c\,\mathbb{S}(t) - c^2 t/2)\,.$$

Differentiating once with respect to c and setting $c = 0$ yields the mg of example 1.6; differentiating twice with respect to c and setting $c = 0$ yields the mg that appears in example 1.7; higher-order derivatives yield mgs based on the *Hermite polynomials*. (Recall (12.7.13) and (10.5.15).) □

Example 1.9 Let $X \in \mathcal{L}_1$ and $\mathcal{A}_n$ be $\nearrow$ σ-fields. Then $Y_n \equiv \mathrm{E}(X|\mathcal{A}_n)$ is a mg. □

Example 1.10 (Cumulative hazard $\Lambda(\cdot)$, and a simple counting process) (a) Let X have df F on the reals R. Then

(9) $\qquad \Lambda(t) \equiv \int_{(-\infty,t]} [1 - F_-(r)]^{-1}\, dF(r) \qquad$ for all $t \in R$

is called the *cumulative hazard function*. Note that

(10) $\qquad 0 \le \Lambda(t) < \infty$ for all $t < \tau_o \equiv F^{-1}(1)$, while $\Delta\Lambda(\tau_o) = \frac{\Delta F(\tau_o)}{1 - F_-(\tau_o)} < \infty$.

Moreover, Λ is a generalized df on R that assign measure $\Lambda(a, b]$ to $(a, b]$ whenever $-\infty \le a < b < \tau_o$, that assign measure $\Delta\Lambda(\tau_o)$ to $\{\tau_o\}$, and that assigns measure 0 to any $(a, b]$ for which $\tau_o \le a < b \le \infty$. It is common that $\Lambda(t) \nearrow \infty$ as $t \nearrow \tau_o$. This holds when $X \cong$ Uniform$(0, 1)$, and for all dfs for which $F_-(F^{-1}(t)) = t$ in some neighborhood with right endpoint τ_o. Still, note that when $X \cong$ Bernoulli(p) with $0 < p < 1$, then $\Lambda(t) = (1 - p)\, 1_{[0,\infty)}(t) + 1_{[1,\infty)}(t)$. Roughly, $1 - F_-(t)$ is the probability that Y still "lives" just prior to time t. Given this, $dF(t)$ is the "instantaneous probability" of a failure at time t. Thus $d\Lambda(t) = dF(t)/[1 - F_-(t)]$ represents the *instantaneous hazard* at time t.
(b) Define the counting process

(11) $\qquad \mathbb{N}_t \equiv \mathbb{N}(t) \equiv 1_{[X \le t]} \qquad$ and let $\qquad \mathcal{A}_t \equiv \sigma[\mathbb{N}(r) : r \le t], \qquad$ for all $t \in R$.

Note that $\mathbb{N}$ is an $\nearrow$ and right-continuous process on R that is adapted to the history σ-fields $\mathcal{A}_t$, and hence is a submg.
(c) The class $\mathcal{C}_s \equiv \{[X > r] : -\infty \le r \le s\}$ is a $\bar{\pi}$-system that generates $\mathcal{A}_s$. So any two finite measures that agree on $\mathcal{C}_s$ also agree on $\mathcal{A}_s$ by the Dynkin π-λ theorem.
(d) We start with a bit of practice. The reader is to show in exercise 1.4 below that

(12) $\qquad \mathrm{E}\{\mathbb{N}_t \,|\, \mathcal{A}_s\} =_{a.s.} 1_{[X \le s]} + 1_{[X > s]}\, \frac{F(s,t]}{1 - F(s)} \qquad$ for all $-\infty < s < t < \infty$

(where $\frac{0}{0}$ is interpreted as 0 and $F(s, t] \equiv F(t) - F(s)$) by verifying that for every set A in the $\bar{\pi}$-system $\mathcal{C}_s$ the relationship $\int_A (\mathbb{N}_t - \mathrm{E}\{\mathbb{N}_t \,|\, \mathcal{A}_s\})\, dP = 0$ holds for the candidate for $\mathrm{E}\{\mathbb{N}_t \,|\, \mathcal{A}_s\}$ that is specified in (12). In like fashion verify that

(13) $\qquad \mathrm{E}\{1_{[X \ge u]} \,|\, \mathcal{A}_s\} =_{a.s.} 1_{[X > s]}\, \frac{1 - F_-(u)}{1 - F(s)} \qquad$ for all $-\infty < s < u \le t < \infty$.

(e) Next, the process

(14) $\mathbb{M}_t \equiv \mathbb{M}(t) \equiv \mathbb{N}(t) - \int_{(-\infty,t]} 1_{[X\ge r]}\, d\Lambda(r) \equiv \mathbb{N}(t) - \mathbb{A}(t)$ is a mg on R

adapted to the $\mathcal{A}_t$'s. Note first that for all $t \in R$,

(15) $\mathrm{E}|\mathbb{M}_t| \le \mathrm{E}|\mathbb{N}_t| + \int_{-\infty}^{t} \mathrm{E}\, 1_{[X\ge r]}\, d\Lambda(r) \le F(t) + \int_{-\infty}^{t} \frac{1-F_-}{1-F_-}\, dF \le 2F(t) \le 2.$

Then verify that $\int_A (\mathbb{N}_t - \mathbb{N}_s)\, dP$ is equal to $\int_A (\mathbb{A}_t - \mathbb{A}_s)\, dP$ for all sets $A \in \mathcal{C}_s$. Statement (12) gives the so called "Doob–Meyer decomposition" of the submg $\mathbb{N}$ into the mg $\mathbb{M}$ and the $\nearrow$ process $\mathbb{A}$. The motivation for the definition of $\mathbb{A}$ is found in (13.5.3) and (13.8.3). □

Example 1.11 (Another counting process) Suppose that the rvs $\xi_1, \xi_2, \ldots$ are iid Uniform(0, 1), and let $\mathbb{N}_n(t) \equiv n\mathbb{G}_n(t) \equiv$ (the number of ξ_i's $\le t$). Then $\mathbb{N}_n$ is a *counting process*, since it is $\nearrow$, and it increases only by jumps upward of size +1. Hence it is a submartingale. The reader will be asked to show (giving another Doob–Meyer decomposition) that the uniform empirical process $\mathbb{U}_n$ satisfies

(16)
$$\begin{aligned} \mathbb{M}_n(t) &\equiv \mathbb{N}_n(t) - \int_0^t \{n[1 - \mathbb{G}_{n-}(r)]/(1-r)\}\, dr \\ &= \sqrt{n}\,\{\mathbb{U}_n(t) + \int_0^t [\mathbb{U}_{n-}(r)/(1-r)]\, dr\} \qquad \text{is a martingale.} \end{aligned}$$

The covariance function of this process is $s \wedge t - st$ for all $0 \le s, t \le 1$. □

Example 1.12 (Poisson process) Suppose $\mathbb{N}(t)$ is a Poisson process with rate $\lambda > 0$. It is a counting process and hence a submartingale. Moreover, the process $\mathbb{M}(t) \equiv \mathbb{N}(t) - \lambda t$ is a martingale, and the process $\mathbb{M}^2(t) - \lambda t$ is also a martingale. □

Example 1.13 (Likelihood ratios) Let $(\Omega, \mathcal{A}, P)$ and $(\Omega, \mathcal{A}, Q)$ be probability spaces for Q and P. Suppose that $\mathcal{A}_n$ is an $\nearrow$ sequence of sub σ-fields of $\mathcal{A}$. Suppose Q_n and P_n denote the measures Q and P, respectively, restricted to $\mathcal{A}_n$, and suppose that $Q_n \ll P_n$. Let $X_n \equiv dQ_n/dP_n$. Then for $A \in \mathcal{A}_m$ and $n > m$ we have $\int_A X_n\, dP = Q_n(A) = Q_m(A) = \int_A X_m\, dP$, so that $\int_A (X_n - X_m)\, dP = 0$. This shows that

(17) $\{X_n, \mathcal{A}_n\}_{n=1}^{\infty}$ is a mg of likelihood ratios. □

Example 1.14 (Kakutani's mg) Let $X_1, X_2, \ldots$ be independent rvs with each $X_k \ge 0$ and $\mathrm{E}\, X_k = 1$. Let $M_n \equiv \prod_1^n X_k$, for $1 \le k \le n$, with $M_0 \equiv 1$. Then M_n is a mg with all $\mathrm{E}\, M_n = 1$. □

Exercise 1.4 Verify the claims made in example 1.10.

Exercise 1.5 Verify the claims made in example 1.11.

Exercise 1.6 Find the exponential martingale that corresponds to the mg $\mathbb{M}(t)$ in example 1.12. Then differentiate this twice with respect to c, set $c = 0$ each time, and obtain the two mgs given in the example.

2 Simple Optional Sampling Theorem

The following proposition gives a particularly simple special case of the optional sampling theorem. (What are the implications for gambling?)

Proposition 2.1 (Optional sampling) If $\{X_n, \mathcal{A}_n\}_{n=0}^{\infty}$ is a s-mg and stopping times S and T (relative to these $\mathcal{A}_n$'s) satisfy $0 \le S \le T \le N$ a.s. for some fixed integer N, then

(1) $\quad \mathrm{E}(X_T|\mathcal{A}_S) \geqq X_S \quad$ a.s. $\qquad$ (so $\mathrm{E}(X_T) \ge \mathrm{E}(X_T)$).

Thus $(X_0, \mathcal{A}_0), (X_S, \mathcal{A}_S), (X_T, \mathcal{A}_T), (X_N, \mathcal{A}_N)$ is a s-mg.

Proof. Case 1: $0 \le T - S \le 1$. Let $A \in \mathcal{A}_S$. Then

$$\int_A \{\mathrm{E}(X_T \mid \mathcal{A}_s) - X_S\}\, dP = \int_A (X_T - X_S)\, dP$$

$$= \int_{A \cap [T=S+1]} (X_T - X_S)\, dP = \sum_{i=0}^N \int_{A\cap[T=S+1]\cap[S=i]} (X_T - X_S)\, dP$$

(a) $$= \sum_{i=0}^N \int_{A\cap[S=i]\cap[T=i+1]} (X_{i+1} - X_i)\, dP$$

$$= \sum_{i=0}^N \int_{A_i} (X_{i+1} - X_i)\, dP \qquad \text{where } A_i \equiv A \cap [S=i] \cap [T = i+1]$$

(b) $$\geqq 0$$

by (13.1.4), provided that we show that $A_i \in \mathcal{A}_i$. We have

(c) $$A \cap [S=i] \cap [T=i+1] = (A \cap [S \le i]) \cap ([S=i] \cap [T \le i]^c)$$

(d) $$= (\text{an event in } \mathcal{A}_i) \cap (\text{an event in } \mathcal{A}_i) \in \mathcal{A}_i\,.$$

Thus (b) holds, and case 1 is completed.

Case 2: $0 \le S \le T \le N$ a.s. Define the stopping times

(e) $$R_k \equiv T \wedge (S+k) \qquad \text{for } k = 0, 1, \ldots, N;$$

recall that sums and minima of such stopping times are stopping times, as in eexercise 8.7.1. Note that

(f) $$0 \le S = R_0 \le \cdots \le R_i \le R_{i+1} \le \cdots \le R_N = T \le N \qquad \text{holds a.s.,}$$

where $0 \le R_{i+1} - R_i \le 1$ for all i. Thus,

(g) $$\mathrm{E}(X_T|\mathcal{A}_S) = \mathrm{E}(X_{R_N}|\mathcal{A}_{R_0}) = \mathrm{E}(\mathrm{E}(X_{R_N}|\mathcal{A}_{R_{N-1}})|\mathcal{A}_{R_0})$$

$$\geqq \mathrm{E}(X_{R_{N-1}}|\mathcal{A}_{R_0}) \qquad \text{by case 1, and stepwise smoothing}$$

$$\geqq \cdots \geqq \mathrm{E}(X_{R_0}|\mathcal{A}_{R_0}) = X_{R_0}$$

(h) $$= X_S\,.$$ □

3 The Submartingale Convergence Theorem

Theorem 3.1 (S-mg convergence theorem) Let $\{X_n, \mathcal{A}_n\}_{n=0}^{\infty}$ be a s-mg.
(A) Suppose $\mathrm{E}X_n^+ \nearrow M < \infty$ (i.e., the X_n^+–submg is integrable). Then

(1) $\qquad X_n \to X_\infty \quad$ a.s. $\qquad$ for some $X_\infty \in \mathcal{L}_1$.

(B) For uniformly integrable X_n's, this X_∞ *closes the s-mg* in that

(2) $\qquad \{X_n, \mathcal{A}_n\}_{n=0}^{\overline{\infty}}$ is a s-mg with $\mathrm{E}X_n \nearrow \mathrm{E}X_\infty$, and with $\mathcal{A}_\infty \equiv \sigma[\cup_{n=1}^{\infty} \mathcal{A}_n]$.

In fact, supposing $\{X_n, \mathcal{A}_n\}_{n=0}^{\infty}$ is a submg, the conclusions (aa), (bb), and (ee) of part (C) below are equivalent. If all $X_n \geq 0$, then (cc) and (dd) are also equivalent. (Closing the s-mg means that $\int_A X_n \, dP \overset{\leq}{=} \int_A X_\infty \, dP$ for all $A \in \mathcal{A}_n$ and all n, with the terminal rv $X_\infty \in \mathcal{L}_1$.)

(C) If the $\{X_n, \mathcal{A}_n\}_{n=0}^{\infty}$ of (A) is actually a mg, then the following are equivalent.

(aa) X_n's are uniformly integrable. $\qquad$ (bb) $X_n \to_{\mathcal{L}_1} X_\infty$.

(3) $\qquad$ (cc) Some rv Y closes the mg. $\qquad$ (dd) X_∞ closes the mg.

(ee) X_n's are integrable and $\overline{\lim}_n \mathrm{E}|X_n| \leq \mathrm{E}|X_\infty| < \infty$.

(D) In all the above, if $\{X_t, \mathcal{A}_t\}_{t\in[0,\infty)}$ is a process on $(D_{[0,\infty)}, \mathcal{D}_{[0,\infty)})$, then

(4) $\qquad n, X_n, \mathcal{A}_n, \{0, 1, \ldots\} \qquad$ may be replaced by $\qquad t, X_t, \mathcal{A}_t, [0, \infty)$.

[Closing an X-martingale on $[0,\infty)$ by X_∞ is the same as closing an X-martingale on $[0,\theta)$ by the limiting rv X_θ.]

Notation 3.1 If a sequence does not converge to an extended real value, then it must necessarily be oscillating. If it does so oscillate, then some interval must be "upcrossed" infinitely often. We seek to take advantage of this. Let $X_1, X_2, \ldots$ be a sequence of rvs. Let $a < b$. Then:

$$U_{[a,b]}^{(n)}(\omega) \equiv (\text{the number of upcrossings of } [a,b] \text{ in the first } n \text{ steps})$$

(5) $$\equiv \begin{bmatrix} \text{number of integer pairs } (i,j) \text{ with } 0 \leq i < j \leq n \text{ having} \\ X_i(\omega) \leq a, \;\; a < X_k(\omega) < b \text{ for } i < k < j, \;\; \text{and } X_j(\omega) \geq b \end{bmatrix},$$

(6) $$U_{[a,b]}^{(\infty)}(\omega) \equiv \lim_{n\to\infty} U_{[a,b]}^{(n)}(\omega)\,. \qquad \square$$

Inequality 3.1 (Upcrossing inequality; Doob) If $\{X_k, \mathcal{A}_k\}_{k=0}^{n}$ is a submg, then

(7) $$\mathrm{E}U_{[a,b]}^{(n)} \leq \frac{1}{b-a}\{\mathrm{E}(X_n - a)^+ - \mathrm{E}(X_0 - a)^+\} \leq \frac{1}{b-a}\{\mathrm{E}X_n^+ + |a|\}.$$

If $\{X_n, \mathcal{A}_n\}_{n=0}^{\overline{\infty}}$ is a submg, then

(8) $$\mathrm{E}U_{[a,b]}^{(\infty)} \leq \frac{1}{b-a}\{\mathrm{E}(X_\infty - a)^+ - E(X_0 - a)^+\} \leq \frac{1}{b-a}\{\mathrm{E}X_\infty^+ + |a|\}.$$

Proof. The number of upcrossings of $[a, b]$ made by X_k and the number of upcrossings of $[0, b-a]$ made by $(X_k - a)^+$ are identical; since $(X_k - a)^+$ is also a submg, we may assume that $X_k \ge 0$ and $a = 0$ in this proof. Define

$$\text{(a)}\qquad \begin{array}{l} T_0 \equiv 0, \quad T_1 \equiv \min\{n \ge 0 : X_n = 0\}, \quad T_2 \equiv \min\{n > T_1 : X_n \ge b\}, \\ T_3 \equiv \min\{n > T_2 : X_n = 0\}, \ldots, T_{n+2} \equiv n\,; \end{array}$$

here we use the convention that $\min \emptyset \equiv n$. Clearly, these are stopping times that do satisfy $0 = T_0 \le T_1 \le \cdots \le T_{n+2} = n$. Thus proposition 13.2.1 shows that the process $\{X_{T_i}, \mathcal{A}_{T_i}\}_{i=0}^{n+2}$ is a submg. Now,

$$X_n - X_0 = \textstyle\sum_{i=1}^{n+2} (X_{T_i} - X_{T_{i-1}})$$

$$\text{(b)}\qquad = \textstyle\sum_{i\text{ odd}} (\text{same}) + \sum_{i\text{ even}} (\text{same}) \equiv I_1 + I_2\,,$$

and since X_{T_i} is a submg, we have

$$\text{(c)}\qquad \mathrm{E}I_1 \ge 0 \qquad \text{and} \qquad \mathrm{E}I_2 \ge 0.$$

Now suppose that $U_{[0,b]}^{(n)}(\omega) = k$; then

$$I_2(\omega) = [X_{T_2(\omega)}(\omega) - X_{T_1(\omega)}(\omega)] + \cdots + [X_{T_{2k}(\omega)}(\omega) - X_{T_{2k-1}(\omega)}(\omega)] + \cdots$$

$$\text{(d)}\qquad \ge [b] + \cdots + [b] + 0 = b\, U_{[0,b]}^{(n)}(\omega).$$

Thus (recall (2.1.3))

$$\text{(e)}\qquad \mathrm{E}(X_n - X_0) = \mathrm{E}I_1 + \mathrm{E}I_2 \ge \mathrm{E}I_2 \ge b\,\mathrm{E}U_{[0,b]}^{(n)},$$

and this is (7) in disguise (since X_k really denotes $(X_k - a)^+$). Finally, note the positive part inequality $(X_n - a)^+ \le X_n^+ + |a|$.

Now, $0 \le U_{[a,b]}^{(n)} \nearrow U_{[a,b]}^{(\infty)}$. Thus,

$$\text{(f)}\qquad \mathrm{E}\,U_{[a,b]}^{(\infty)} = \lim \mathrm{E}\,U_{[a,b]}^{(n)} \qquad \text{by the MCT}$$

$$\text{(g)}\qquad \le \lim \tfrac{1}{b-a} \{\mathrm{E}(X_n - a)^+ - \mathrm{E}(X_0 - a)^+\} \qquad \text{by (7)}$$

$$\text{(h)}\qquad \le \tfrac{1}{b-a} \{\mathrm{E}(X_\infty - a)^+ - \mathrm{E}(X_0 - a)^+\}\,,$$

since $\mathrm{E}(X_n - a)^+ \le \mathrm{E}(X_\infty - a)^+$ follows from $\{(X_n - a)^+, \mathcal{A}_n\}_{n=0}^{\overline{\infty}}$ being a submg. □

Proof. (Proof of theorem 3.1) Now,

$$[\omega : \lim X_n(\omega) \text{ exists as a number in } [-\infty, \infty]\,]^c = [\,\underline{\lim}\, X_n < \overline{\lim}\, X_n\,]$$

$$\text{(a)}\qquad = \textstyle\bigcup_{\{r<s\text{ rational}\}} [\underline{\lim}\, X_n < r < s < \overline{\lim}\, X_n] \equiv \bigcup_{\{r<s\text{ rational}\}} A_{rs}.$$

It suffices to prove that $P(A_{rs}) = 0$ for all r, s. Now,

(b) $\quad A_{rs} \subset B_{rs} \equiv [\omega : U^{(\infty)}_{[r,s]} = \infty]$.

It thus suffices to show that $P(B_{rs}) = 0$ for all r, s. But

(c) $\quad \mathrm{E}\, U^{(\infty)}_{[r,s]} = \mathrm{E} \lim U^{(n)}_{[r,s]} = \lim \mathrm{E}\, U^{(n)}_{[r,s]} \qquad$ by the MCT

(d) $\quad \le \lim \frac{1}{s-r} \{\mathrm{E}X_n^+ + |r|\} \qquad$ by (7)

(e) $\quad \le (M + |r|\,)/(s-r) < \infty, \qquad$ since the X_n^+'s are integrable;

hence we must have $P(B_{rs}) = 0$. Thus $X_\infty \equiv \lim X_n$ exists a.s. with values in $[-\infty, \infty]$, and X_∞ is $\mathcal{A}_\infty$-measurable, since all X_n's are. Now (recall (2.1.3)),

(f) $\quad \mathrm{E}|X_n| = 2\mathrm{E}X_n^+ - \mathrm{E}X_n \le 2\mathrm{E}X_n^+ - \mathrm{E}X_0 \le M < \infty$.

Thus Fatou's lemma implies

(g) $\quad \mathrm{E}|X_\infty| = \mathrm{E}(\lim |X_n|) \le \underline{\lim}\, \mathrm{E}|X_n| \le M < \infty$;

thus $X_\infty \in \mathcal{L}_1$ with its values in $(-\infty, \infty)$ a.s. Thus (A) holds.

Consider (B). Now $X_n \to$ (some X_∞) a.s. by (A) under any of (aa), (bb), or (ee). Vitali shows that (aa) is equivalent to (ee) and the $\mathcal{L}_1$-convergence of (bb) with the rv X_∞. (If $X_n \to_{\mathcal{L}_1} Y$ in (bb), then this Y must equal X_∞ a.s. by going to subsequences.) Thus for $n \ge m$, we have from (13.1.4) and $\mathcal{L}_1$-convergence that

(h) $\quad \int_A X_m\, dP \overset{\le}{=} \int_A X_n\, dP \to \int_A X_\infty\, dP \qquad$ for all $A \in \mathcal{A}_\infty$.

Thus $\{X_n, \mathcal{A}_n\}_{n=0}^{\overline{\infty}}$ is a s-mg by condition (13.1.4). That is, X_∞ closes the s-mg (and thus (B) holds, except for the equivalence of (cc) and (dd)).

Consider (C). For (aa)–(ee), we lack only that (cc) implies (aa). For a mg with X_n (or for submg with $X_n \ge 0$), the $|X_n|$ form a submg. Thus

(i) $\quad \mathrm{E}\{\,|X_n| \times 1_{[|X_n| \ge \lambda]}\,\} \le \mathrm{E}\{\,|Y| \times 1_{[|X_n| \ge \lambda]}\,\} \qquad$ since $|Y|$ closes the submg

(j) $\quad \to 0 \qquad$ by absolute continuity of the integral,

and since the sets satisfy

(k) $\quad P(|X_n| \ge \lambda) \le \mathrm{E}|X_n|/\lambda \le \mathrm{E}|Y|/\lambda \to 0 \qquad$ uniformly in n as $\lambda \to \infty$.

This completes the entire proof in the case of discrete time.

Consider (D). (Continuous time) Our preliminaries will not assume the s-mg structure. Now, $\overline{\lim}\, X_t(\omega)$ could be $+\infty$ for some ω's, and this will cause difficulties with the present approach. Thus (following Doob) define

(l) $\quad Y_t(\omega) = (2/\pi)\, \tan^{-1}(X_t(\omega))$

to transform the range space from $[-\infty, \infty]$ to $[-1, 1]$. For each m choose rational numbers $t_{m1}, \ldots, t_{m_{k_m}}$ in $[m, \infty)$ so that (remember, $X : (\Omega, \mathcal{A}, P) \to (D, \mathcal{D})$)

(m) $\quad P(\sup_{t \in [m,\infty)} Y_t - \sup_{t_{mj} \in [m,\infty)} Y_{t_{mj}} > \frac{1}{m}) < \frac{1}{2^m}$.

This is possible, since the sup over all rationals r in $[m,\infty)$ equals the sup over all reals t in $[m,\infty)$, and the rationals are the limit of $\{r_1,\ldots,r_k\}$ for any ordering $\{r_1,r_2,\ldots\}$ of the rationals; thus,

(n)
$$0 = P(\sup_{t\in[m,\infty)} Y_t - \sup_{r_j\in[m,\infty)} Y_{r_j} > \tfrac{1}{m})$$
$$= \lim_{k\to\infty} P(\sup_{t\in[m,\infty)} Y_t - \sup_{r_j\in[m,\infty);j\le k} Y_{r_j} > \tfrac{1}{m})\,.$$

We may assume that the t_{mj}'s were chosen so as to simultaneously satisfy

(o) $\quad P(\inf_{t\in[m,\infty)}Y_t - \inf_{t_{mj}\in[m,\infty)}Y_{t_{mj}} < -\frac{1}{m}) < \frac{1}{2^m}$.

Thus, if we

(p) $\quad$ let $t_1 < t_2 < \cdots$ denote an ordering of $\cup_{m=1}^{\infty}\{\, t_{m1},\ldots,t_{mk_m}\,\}$

(which does exist, since all $t_{mj}\ge m$), then

(q)
$$P(\sup_{t\in[m,\infty)} Y_t - \sup_{t_i\in[m,\infty)} Y_{t_i} > \tfrac{1}{m}) < \tfrac{1}{2^m} \qquad \text{and}$$
$$P(\inf_{t\in[m,\infty)} Y_t - \inf_{t_i\in[m,\infty)} Y_{t_i} < -\tfrac{1}{m}) < \tfrac{1}{2^m}\,.$$

Letting A_m and B_m denote the events in (q) we see that

$$\textstyle\sum_1^\infty P(A_m\cup B_m) \le \sum_1^\infty 2/2^m < \infty\,,$$

so that $P(A_m\cup B_m \text{ i.o.}) = 0$. Thus

(r) $\quad \overline{\lim}_{t\to\infty}Y_t = \overline{\lim}_{i\to\infty}Y_{t_i}$ a.s. $\quad$ and $\quad \underline{\lim}_{t\to\infty}Y_t = \underline{\lim}_{i\to\infty}Y_{t_i}$ a.s.

Now, transforming back via $X_t(\omega) = \tan((\pi/2)\,Y_t(\omega))$, (r) implies the next lemma:

Lemma 3.1 For every $X : (\Omega,\mathcal{A},P) \to (D_{[0,\infty)},\mathcal{D}_{[0,\infty)})$ there exist rational numbers $t_1 < t_2 < \cdots$ such that

(9) $\quad \overline{\lim}_{t\to\infty} X_t = \overline{\lim}_{i\to\infty} X_{t_i}$ a.s. $\quad$ and $\quad \underline{\lim}_{t\to\infty} X_t = \underline{\lim}_{i\to\infty} X_{t_i}$ a.s.

[Note that $t_i\to\infty$ in (9) could be replaced by $t_i \nearrow \theta$ for any finite θ.]

Armed with the (9) "lemma," it is now easy to use the discrete version of this theorem to prove the continuous version. We will refer to the continuous versions of conclusions (1)–(3) as (1′)–(3′). We return to the proof of the theorem. (Now, we again assume s-mg structure in what follows.)

Let $Y_i \equiv X_{t_i}$ and $\tilde{\mathcal{A}}_i \equiv \mathcal{A}_{t_i}$ for the t_i's in (9). Then $(Y_i,\tilde{\mathcal{A}}_i)_{i=0}^\infty$ is a s-mg to which the discrete theorems can be applied. Thus,

(s) $\quad \overline{\lim}_{t\to\infty}X_t = \overline{\lim}_{i\to\infty}X_{t_i}$ a.s. $\quad$ by (9)

(t) $\quad = (\text{a.s., some } X_\infty \text{ in } \mathcal{L}_1)$ $\quad$ by (1) applied to $(Y_i,\tilde{\mathcal{A}}_i)_{i=1}^\infty$

$\quad = \underline{\lim}_{i\to\infty}X_{t_i}$ $\quad$ by (1)

(u) $\quad = \underline{\lim}_{t\to\infty}X_t$ $\quad$ by (9),

so that

(v) $\quad X_t \to X_\infty$ a.s. $\quad$ where $X_\infty\in\mathcal{L}_1$.

That is, (1′) holds. The rest is left to the reader in exercise 3.8. □

Exercise 3.1 Complete the proof of the continuous part of theorem 3.1.

Exercise 3.2 Let $Y_t \equiv \mathrm{E}(X|\mathcal{D}_t)$, for $X \in \mathcal{L}_1(\Omega, \mathcal{A}, P)$ and for an arbitrary collection of sub σ-fields $\mathcal{D}_t$. Show that these Y_t's are uniformly integrable.

Exercise 3.3 Let $\mathcal{A}_{-\infty} \subset \cdots \subset \mathcal{A}_{-1} \subset \mathcal{A}_0 \subset \mathcal{A}_1 \subset \cdots \subset \mathcal{A}_\infty$ be sub σ-fields of the basic σ-field $\mathcal{A}$. Suppose the rv $X \in \mathcal{L}_1(\Omega, \mathcal{A}, P)$. Let $Y_n \equiv \mathrm{E}(X|\mathcal{A}_n)$. Then the process $(Y_n, \mathcal{A}_n)_{n=\underline{-\infty}}^{\overline{\infty}}$ is necessarily a uniformly integrable mg.

Exercise 3.4 Let $\{X_n, \mathcal{A}_n\}_{n=0}^{\infty}$ be a submg for which $X_n \le 0$. Then (1) holds, and $Y \equiv 0$ closes the submg.

Exercise 3.5 Let $\{X_n, \mathcal{A}_n\}_{n=0}^{\infty}$ be a submg. The following are equivalent:
(a): The X_n^+'s are uniformly integrable.
(b): There exists a rv Y that closes the submg.
(c) When these hold, then X_∞ (which necessarily exists a.s., and is in $\mathcal{L}_1$) closes the submg. [Hint. Do what you can with X_n^+. Then apply it to $Y_n^{(a)} \equiv (X_n \vee a) + |a|$, and let $a \to -\infty$.]

Exercise 3.6 Let $\{X_n, \mathcal{A}_n\}_{n=0}^{\infty}$ be a submg with $X_n \ge 0$. Let $r > 1$. Then the X_n^r's are uniformly integrable if and only if the X_n^r-process is integrable.

Exercise 3.7 (Martingale $\mathcal{L}_r$-convergence theorem) (i) Let $\{X_n, \mathcal{A}_n\}_{n=0}^{\infty}$ be a mg sequence. Let $r > 1$. Then the following are equivalent:

(10) The $|X_n|^r$-process is integrable.

(11) $X_n \to_{\mathcal{L}_r} X_\infty$.

(12) The X_n's are uniformly integrable
(thus $X_n \to_{a.s.}$ (some X_∞)) and $X_\infty \in \mathcal{L}_r$.

(13) The $|X_n|^r$'s are uniformly integrable.

(14) $\{|X_n|^r, \mathcal{A}_n\}_{n=0}^{\overline{\infty}}$ is a submg and $\mathrm{E}|X_n|^r \nearrow \mathrm{E}|X_\infty|^r < \infty$.

$(*)$ $M^* \equiv \sup\{|X_n| : 0 \le n \le \infty\} \in \mathcal{L}_r$ (via Doob's $\mathcal{L}_r$-inequality).

(ii) This theorem also holds for a submg when all $X_n \ge 0$ a.s.

Exercise 3.8 (a) Show that $t, X_t, \mathcal{A}_t, [0, \infty)$ may replace $n, X_n, \mathcal{A}_n, \{0, 1, \ldots\}$ in all of exercise 3.3–exercise 3.7. [Also, $[0, \theta)$ may replace $[0, \infty)$.]
(b) Prove (D) of theorem 3.2 below.

Definition 3.1 (Reversed s-mg) LetX_n be adapted to $\mathcal{A}_n$ with $n \in \{\ldots, -1, 0\}$, and define the σ-field $\mathcal{A}_{-\infty} \equiv \cap_{n=-\infty}^{0} \mathcal{A}_n$. The process $\{X_n, \mathcal{A}_n\}_{n=-\infty}^{0}$ is a *reversed s-mg* (as defined earlier) if all $\mathrm{E}|X_n| < \infty$ and

(15) $\quad X_n \overset{\leq}{=} \mathrm{E}(X_m|\mathcal{A}_n) \quad$ a.s. $\quad$ for all $n \leq m$.

(This is like the second law of thermodynamics run backward in time, since $\mathcal{A}_n$ brings more stability as $n \searrow$.)

Theorem 3.2 (Reversed s-mg convergence theorem) Let $\{X_n, \mathcal{A}_n\}_{n=-\infty}^{0}$ be a s-mg sequence, or a reversed s-mg.
(A) It necessarily holds that

(16) $\quad X_n \to X_{-\infty}$ a.s. $\quad$ as $n \to -\infty$
for some $X_{-\infty} \in [-\infty, \infty)$ a.s. that is $\mathcal{A}_{-\infty}$-measurable.

(B)–(C) Furthermore, the following are equivalent (and all yield an $X_\infty \in \mathcal{L}_1$):

(17) $\quad \mathrm{E}X_n \searrow M > -\infty \quad$ as $n \to -\infty$. $\quad$ (This is trivial if X_n is a mg.)

(18) $\quad X_n$'s are uniformly integrable.

(19) $\quad X_n \to_{\mathcal{L}_1} X_{-\infty}$.

(20) $\quad \{X_n, \mathcal{A}_n\}_{n=\underline{-\infty}}^{0}$ is a s-mg, $\quad$ where $\mathcal{A}_{-\infty} \equiv \cap_{n=-\infty}^{0} \mathcal{A}_n$.

(D) In all the above, if $\{X_t, \mathcal{A}_t\}_{t \in (-\infty, 0]}$ is a process on $(D_{(-\infty,0]}, \mathcal{D}_{(-\infty,0]})$, then

(21) $\quad n, X_n, \mathcal{A}_n, \{\ldots, -1, 0\} \quad$ may be replaced by $\quad t, X_t, \mathcal{A}_t, (-\infty, 0]$.

Proof. Consider (16). Let $U_{[r,s]}^{(n)}$ now denote the upcrossings of $[r, s]$ by the process $X_{-n}, \ldots, X_{-1}, X_0$. Replace line (d) of the proof of theorem 3.1 by

(a) $\quad \mathrm{E}U_{[r,s]}^{(\infty)} \leq \frac{1}{s-r}\,(\mathrm{E}X_0^+ + |r|) \leq (\text{some constant}) < \infty,$

and conclude that $X_{-\infty} \equiv \lim X_n$ exists a.s. with values in $[-\infty, +\infty]$. Since the sequence $\{X_n^+, \mathcal{A}_n\}_{n=-\infty}^{0}$ is necessarily a submg, we obtain from Fatou that

(b) $\quad \mathrm{E}X_{-\infty}^+ = \mathrm{E}(\lim X_n^+) \leq \underline{\lim}\,\mathrm{E}X_n^+ \leq \underline{\lim}\,\mathrm{E}X_0^+ = \mathrm{E}X_0^+ < \infty.$

Thus $X_{-\infty}$ takes values in $[-\infty, +\infty)$ a.s., and is $\mathcal{A}_{-\infty}$-measurable. Thus (16) does hold.

Suppose that (17) holds. Then (b) and (17) give

$$\mathrm{E}|X_{-\infty}| = \mathrm{E}[\lim |X_n|\,] \le \underline{\lim}\,\mathrm{E}|X_n| = \underline{\lim}\,[2\,\mathrm{E}X_n^+ - \mathrm{E}X_n]$$

(c) $$\le 2\,\mathrm{E}X_0^+ - M < \infty;$$

thus $X_{-\infty}$ takes values in $(-\infty,\infty)$ a.s., and $X_{-\infty} \in \mathcal{L}_1$ (using only Fatou on the right hand side). Also, from (c),

(d) $$P(|X_n| \ge \lambda) \le \mathrm{E}|X_n|/\lambda \le [2\,\mathrm{E}X_0^+ - M]/\lambda \to 0$$

uniformly in $n \in \{-\infty, \ldots, -1, 0\,\}$ as $\lambda \to \infty$. Thus (an analogous X_n^- proof works only in the case of a mg),

(e) $$\int_{[X_n^+ \ge \lambda]} X_n^+ \, dP \leqq \int_{[X_n^+ \ge \lambda]} X_0^+ \, dP \qquad \text{by (13.1.3)}$$

implies that the X_n^+'s are uniformly integrable. Now, for $n < m$ we have

$$0 \ge -\int_{[X_n \le -\lambda]} X_n^- \, dP = \mathrm{E}\,(X_n - X_m) + \mathrm{E}X_m - \int_{[X_n > -\lambda]} X_n \, dP$$

$$\geqq \mathrm{E}(X_n - X_m) + \mathrm{E}X_m - \int_{[X_n > -\lambda]} X_m \, dP$$

$$= \mathrm{E}(X_n - X_m) + \int_{[X_n \le -\lambda]} X_m \, dP$$

(f) $$\ge -\epsilon \;+\; \int_{[X_n \le -\lambda]} X_m \, dP$$

for all $n \le$ (a fixed m that is large enough), since $\mathrm{E}X_n \searrow M$

$$\ge -\epsilon \;-\; \int_{[X_n \le -\lambda]} |X_m| \, dP$$

(g) $$\ge -2\,\epsilon \qquad \text{for } \lambda \text{ large enough, as in (d), with } m \text{ now fixed.}$$

Thus, the X_n^- are uniformly integrable. Thus, the X_n are uniformly integrable; that is, (18) holds.

Then (18) implies (19) by Vitali.

Suppose (19) holds. For any $n \le m$ we have from (18.1.4) that

(h) $$\int_A X_m \, dP \geqq \int_A X_n \, dP \to \int_A X_{-\infty} \, dP \qquad \text{for all } A \in \mathcal{A}_{-\infty},$$

since $\mathcal{L}_1$-convergence gives $\int |X_n - X_{-\infty}| \, dP \to 0$. Thus $\{X_n, \mathcal{A}_n\}_{n=-\underline{\infty}}^0$ is a s-mg, so (20) holds.

Note that (20) trivially implies (17).

The extension of this theorem from n to t uses (9) to extend (16), just as in the case of theorem 3.1. Then the proof that the t versions of (16)–(20) hold adds nothing new. □

Exercise 3.9 (a) Let $\{X_n, \mathcal{A}_n\}_{n=-\infty}^{0}$ be a mg. If $\mathrm{E}|X_0|^r < \infty$ for some $r \geq 1$, then necessarily

(22) $\quad X_n \to_{\mathcal{L}_r} X_{-\infty} \qquad$ as $n \to -\infty$.

(b) Let $\{X_n, \mathcal{A}_n\}_{n=-\infty}^{0}$ be a submg. If $X_n \geq 0$ and $\mathrm{E}|X_0|^r < \infty$ for some $r \geq 1$, then (22) again holds.

Theorem 3.3 Let $\{\mathcal{A}_n\}_{n=-\infty}^{+\infty}$ be $\nearrow$ sub σ-fields of $\mathcal{A}$. Suppose that the rv X is integrable, in that $X \in \mathcal{L}_1(\Omega, \mathcal{A}, P)$. Then the following hold.

(23) $\quad \mathrm{E}(X|\mathcal{A}_n) \to_{a.s. \text{ and } \mathcal{L}_1} \mathrm{E}(X|\mathcal{A}_\infty) \qquad$ as $n \to \infty$.

(24) $\quad \mathrm{E}(X|\mathcal{A}_n) \to_{a.s. \text{ and } \mathcal{L}_1} \mathrm{E}(X|\mathcal{A}_{-\infty}) \qquad$ as $n \to -\infty$.

Proof. Now,

(25) $\quad \{Y_n, \mathcal{A}_n\}_{n=-\infty}^{\infty}$ is a mg $\quad$ for $\quad Y_n \equiv \mathrm{E}(X|\mathcal{A}_n)$,

since $\mathrm{E}(Y_m|\mathcal{A}_n) = \mathrm{E}\{\mathrm{E}(X|\mathcal{A}_m)|\mathcal{A}_n\} = \mathrm{E}(X|\mathcal{A}_n) = Y_n$ for $n \leq m$. Moreover,

(26) $\quad$ these $Y_n = \mathrm{E}(X|\mathcal{A}_n)$ are uniformly integrable,

since the tails yield (since $[\pm Y_n \geq \lambda] \in \mathcal{A}_n$)

(a) $\quad 0 \leq \int_{[\pm Y_n \geq \lambda]} \pm Y_n \, dP = \int_{[\pm Y_n \geq \lambda]} \pm \mathrm{E}(X|\mathcal{A}_n) \, dP$

(b) $\quad = \int_{[\pm Y_n \geq \lambda]} \pm X \, dP \leq \int_{[\,|Y_n| \geq \lambda]} |X| \, dP$

(c) $\quad \to 0 \qquad$ as $\lambda \to \infty$, by absolute continuity of the integral,

using $P(|Y_n| \geq \lambda) \leq \mathrm{E}|Y_n|/\lambda = \mathrm{E}|\mathrm{E}(X|\mathcal{A}_n)|/\lambda \leq \mathrm{E}|X|/\lambda \to 0$ as $\lambda \to \infty$. Thus

(27) $\quad Y_{\pm\infty} \equiv \lim_{n \to \pm\infty} Y_n \quad$ exists a.s., $\qquad$ and
$\quad \{Y_n, \mathcal{A}_n\}_{n=-\underline{\infty}}^{+\overline{\infty}}$ is a mg, and $\quad Y_n \to_{\mathcal{L}_1} Y_{\pm\infty} \qquad$ as $n \to \pm\infty$.

We just applied the s-mg convergence theorem (theorem 3.1) as $n \to \infty$ and the reversed s-mg convergence theorem (theorem 3.2) as $n \to -\infty$.

We must now show that $Y_\infty = \mathrm{E}(X|\mathcal{A}_\infty)$. Now, for $A \in \cup_{n=1}^{\infty} \mathcal{A}_n$ we have

(d) $\quad \int_A Y_\infty \, dP = \int_A Y_{n_o} \, dP \qquad$ by (13.1.4), as A is in some $\mathcal{A}_{n_o}$

(e) $\quad = \int_A X \, dP \qquad$ by definition of $\mathrm{E}(\cdot|\mathcal{A}_{n_o})$.

That is, $\int_A Y_\infty \, dP = \int_A X \, dP$ for all $A \in \cup_{n=1}^{\infty} \mathcal{A}_n$, where $\cup_{n=1}^{\infty} \mathcal{A}_n$ is a field and a $\bar{\pi}$-system that generates $\mathcal{A}_\infty$; thus equality also holds for all $A \in \mathcal{A}_\infty$, by the Carathéodory extension theorem. Thus $Y_\infty = \mathrm{E}(X|\mathcal{A}_\infty)$, by only the zero function.

We must also show that $Y_{-\infty} = \mathrm{E}(X|\mathcal{A}_{-\infty})$. Now, if $A \in \mathcal{A}_{-\infty}$, then

(f) $\quad \int_A Y_{-\infty} \, dP = \int_A Y_n \, dP \qquad$ by (13.1.4), as A is in $\mathcal{A}_n$

$\quad = \int_A X \, dP \qquad$ by definition of $\mathrm{E}(\cdot|\mathcal{A}_n)$, since $A \in \mathcal{A}_n$

(g) $\quad = \int_A \mathrm{E}(X|\mathcal{A}_{-\infty}) \, dP \qquad$ by definition of $\mathrm{E}(\cdot|\mathcal{A}_{-\infty})$;

and thus $Y_{-\infty} = \mathrm{E}(X|\mathcal{A}_{-\infty})$. $\square$

4 Applications of the S-mg Convergence Theorem

The following examples give just a few selected applications to show the power of the various s-mg convergence theorems.

Example 4.1 (SLLN) Let $X_1, X_2, \ldots$ be iid μ. Then the partial sum process $S_n \equiv X_1 + \cdots + X_n$ satisfies

(1) $\bar{X}_n \equiv S_n/n \to \mu$ a.s. and $\mathcal{L}_1$ as $n \to \infty$.

□

Proof. Let

(a) $\mathcal{A}_{-n} \equiv \sigma[S_n, S_{n+1}, \ldots] = \sigma[S_n, X_{n+1}, X_{n+2}, \ldots]$.

Now, $Y_{-n} \equiv \mathrm{E}(X_1|\mathcal{A}_{-n})$ is a reversed mg on $\ldots, -2, -1$ and

(b) $\mathrm{E}(X_1|\mathcal{A}_{-n}) \to \mathrm{E}(X_1|\mathcal{A}_{-\infty})$ a.s. and $\mathcal{L}_1$

as $n \to -\infty$, by theorem 13.3.3. Now,

$$\mathrm{E}(X_1|\mathcal{A}_{-n}) = \mathrm{E}(X_1|S_n, X_{n+1}, \cdots)$$

$$= \mathrm{E}(X_1|S_n) \qquad \text{by } (7.4.23)$$

$$= \textstyle\sum_{k=1}^{n} \mathrm{E}(X_k|S_n)/n \qquad \text{by symmetry}$$

$$= \mathrm{E}(S_n|S_n)/n$$

(c) $= S_n/n.$

Combining (b) and (c) gives

(d) $S_n/n = \mathrm{E}(X_1|\mathcal{A}_{-n}) \to \mathrm{E}(X_1|\mathcal{A}_{-\infty})$ a.s. and $\mathcal{L}_1$ as $n \to \infty$.

But $\lim(S_n/n)$ is measurable with respect to the symmetric σ-field, and so it is a.s. a constant by the Hewitt–Savage 0-1 law of exercise 7.2.1; hence $\mathrm{E}(X_1|\mathcal{A}_{-\infty})$ is a.s. a constant, by (d). But $\mathrm{E}[\mathrm{E}(X|\mathcal{A}_{-\infty})] = \mu$, so that the constant must be μ; that is $\mathrm{E}(X_1|\mathcal{A}_{-\infty}) = \mu$ a.s. Thus (d) implies $S_n/n \to \mu$ a.s. and $\mathcal{L}_1$. □

Exercise 4.1 (SLLN for U-statistics) Let $Y_{-n} \equiv U_n$ be a U-statistic based on $X_1, X_2, \ldots$, with a symmetric kernel H for which $\mathrm{E}H(X_1, X_2)$ is finite. (Thus, $H(x, y) = H(y, x)$ for all x, y.) Consider the σ-field $\mathcal{A}_{-n} \equiv \sigma[\mathsf{X}_{n:n}, X_{n+1}, X_{n+2}, \ldots]$, for the vector $\vec{X}_{n:n}$ of the first n order statistics of the sequence. [Hint. As with the SLLN above, the proof will again be based of the Hewitt–Savage 0-1 law for the symmetric σ-field.].

(a) Show that $\{Y_n, \mathcal{A}_n\}_{n=-\infty}^{-2}$ is a reversed mg.

(b) Use this to show that $U_n \to_{a.s.\text{ and }\mathcal{L}_1} \mathrm{E}H(X_1, X_2)$.

(c) Extend this to higher-dimensional kernels.

Example 4.2 (Kolmogorov's 0-1 law) Suppose that $Y_1, Y_2, \ldots$ are iid rvs and let $\mathcal{A}_n \equiv \sigma[Y_1, \ldots, Y_n]$. Suppose that $A \in \mathcal{T} \equiv$ (tail σ-field) $= \cap_{n=1}^{\infty} \sigma[Y_{n+1}, Y_{n+2}, \ldots]$. Since $\mathcal{A}_n$ is independent of $\mathcal{T}$,

$$P(A) =_{a.s.} P(A|\mathcal{A}_n) = \mathrm{E}(1_A|\mathcal{A}_n) \qquad \text{for every } n.$$

But by theorem 13.3.3 we have

$$\mathrm{E}(1_A|\mathcal{A}_n) \to_{a.s.} \mathrm{E}(1_A|\mathcal{A}_\infty) =_{a.s.} 1_A\,.$$

Thus $P(A)$ must equal 0 or 1 (as $P(A) =_{a.s.} 1_A$ implies). □

Example 4.3 (Approximation of $\mathcal{L}_1$ and $\mathcal{L}_2$ functions) Fix the function $f \in \mathcal{L}_1([0,1], \mathcal{B}, \text{Lebesgue})$; thus $\int_0^1 |f(u)|\,du < \infty$. Let

$$\mathcal{A}_n \equiv \sigma[((i-1)/2^n, i/2^n] : i = 1, \ldots, 2^n\} \nearrow \mathcal{B}[0,1]\,.$$

Define $X \equiv f(\xi)$, where $\xi \cong \text{Uniform}(0,1)$. Now let

$$X_n \equiv \mathrm{E}(X|\mathcal{A}_n) = \mathrm{E}(f(\xi)|\mathcal{A}_n) = \textstyle\sum_{k=1}^{2^n} C_{nk}(f)\, 1_{[(k-1)/2^n < \xi \le k/2^n]}\,,$$

with $C_{nk}(f) \equiv 2^n \int_{(k-1)/2^n}^{k/2^n} f(u)\,du$. Since $\mathrm{E}|X| < \infty$, theorem 13.3.3 gives

$$X_n \to \mathrm{E}(X|\mathcal{A}_\infty) = \mathrm{E}(X|\mathcal{B}) = X \quad \text{a.s. and } \mathcal{L}_1\,.$$

Summary Let $f \in \mathcal{L}_1$ and define the step function $f_n^s(\cdot)$ by

(2) $$f_n^s(t) \equiv 2^n \int_{(k-1)/2^n}^{k/2^n} f(u)\,du \qquad \text{for } \tfrac{(k-1)}{2^n} < t \le \tfrac{k}{2^n} \text{ and } 1 \le k \le 2^n\,.$$

Then for every $f \in \mathcal{L}_1$,

(3) $$f_n^s(\cdot) \to f(\cdot) \quad \text{a.s. Lebesgue} \qquad \text{and} \qquad \textstyle\int_0^1 |f_n^s(t) - f(t)|\,dt \to 0\,.$$

Now suppose that $f \in \mathcal{L}_2$. Then $f \in \mathcal{L}_1$ also, and so (3) still holds; and this implies (13.3.12). Thus (the equivalent) (13.3.11) gives

(4) $$f_n^s \to_{\mathcal{L}_2} f \qquad \text{for every } f \in \mathcal{L}_2\,.$$

[So in both cases $f_n^s(\cdot)$ can be thought of as approximating the derivative of the indefinite integral $F(x) \equiv \int_0^x f(t)\,dt$.] □

Example 4.4 (Kakutani's mg) Let $X_1, X_2, \ldots$ be independent with each $X_k \ge 0$ and $\mathrm{E}\,X_k = 1$. Define

(5) $$M_n \equiv \textstyle\prod_1^n X_k, \qquad \text{for } 1 \le k \le n,$$

with $M_0 \equiv 1$. Then $\{M_n, \mathcal{A}_n\}_1^\infty$ is a mg for which all $\mathrm{E}\,M_n = 1$, where $\mathcal{A}_n$ is an appropriate $\nearrow$ sequence of σ-fields (such as the histories). Since M_n is bounded in the space $\mathcal{L}_1$, the sm-g convergence theorem of (13.3.1) shows that

(6) $$M_n \to_{a.s.} M_\infty \in \mathcal{L}_1 \qquad \text{is always true,}$$

for the appropriate rv M_∞. We now show that the following are equivalent:

(7) $c_\infty \equiv \mathrm{E}M_\infty = 1$,

(8) M_n's are uniformly integrable,

(9) $M_n \to_{\mathcal{L}_1} M_\infty$,

(10) $\prod_1^\infty a_n > 0$, where $a_n \equiv \mathrm{E}(X_n^{1/2}) \le 1$,

(11) $\sum_1^\infty (1 - a_n) < \infty$, where $a_n \equiv \mathrm{E}(X_n^{1/2}) \le 1$.

Whenever one (hence all) of these equivalent statements fails, then necessarily

(12) $P(M_\infty = 0) = 1$ and $c_\infty = 0$. □

Proof. Because of (6), equivalence of (7)–(9) follows from Vitali's theorem (or from the submartingale convergence theorem). Equivalence of (10) and (11) is called for in the easy exercise 4.2 below. We first show that (10) implies (8). Suppose (10) holds. Define the normalized product

(13)
$$N_n \equiv \prod_1^n X_k^{1/2} \Big/ \prod_1^n a_k\,, \qquad \text{with all} \qquad \mathrm{E}N_n = 1 \qquad \text{and with}$$
$$\mathrm{E}(N_n^2) = 1/(\textstyle\prod_1^n a_k)^2 \le 1/(\prod_1^\infty a_k)^2 < \infty \qquad \text{for all } n.$$

Thus $\{N_n, \mathcal{A}_n\}_1^\infty$ is a mean-1 mg that is bounded in $\mathcal{L}_2$. Since all $\prod_1^n a_k \le 1$, Doob's $\mathcal{L}_2$-inequality (inequality 8.10.5) and the MCT give

(a) $\mathrm{E}(\sup_n M_n) = \lim_n \mathrm{E}(\sup_{1\le k\le n} M_k)$ by the MCT

(b) $\le \lim_n \mathrm{E}(\sup_{1\le k\le n} N_k^2) \times 1$

(c) $\le (\frac{2}{2-1})^2\, \mathrm{E}(N_n^2) < \infty$ by Doob's $\mathcal{L}_r$-inequality.

Thus $M^* \equiv \sup_n M_n$ is a rv in $\mathcal{L}_1$ for which $0 \le M_n \le M^*$. Hence the rvs $\{M_n : 1 \le n \le \infty\}$ are uniformly integrable. That is, (8) holds.

We next show that when (10) fails (that is, when $\prod_1^\infty a_n = 0$), then (7) fails (and that, in fact, (12) holds). Now (13) notes that the N_n all have mean 1, and hence they form an integrable mg. Thus $N_n \to_{a.s.}$ (some N_∞) $\in \mathcal{L}_1$ by the submartingale convergence theorem. Hence,

(d) $M_n^{1/2} = (\prod_1^n a_k)\, N_n \to 0$ a.s.,

impling that $M_\infty = 0$ a.s. and thus that $c_\infty = 0$. This contradicts (7), and implies the truth of (12). □

Exercise 4.2 Show the equivalence of (10) and (11). (Recall lemma 8.1.4.)

Exercise 4.3 (Borel–Cantelli) Let $\mathcal{A}_n$ be an $\nearrow$ sequence of σ-fields in $\mathcal{A}$. Show that $[A_n \text{ i.o.}] = [\omega : \sum_{n=1}^\infty P(A_n|\mathcal{A}_{n-1}) = \infty]$ a.s.

A Branching Process Model

Example 4.5 (Branching Processes) Let X denote the number of offspring of a particular type of individual, and let $p_k \equiv P(X=k)$ for $k=0,1,\ldots$. We start at generation zero with a single individual $Z_0 = 1$, and it produces the individuals in a first generation of size Z_1. These in turn produce a second generation of size Z_2, and so forth. Thus,

$$(14) \qquad Z_{n+1} \equiv \sum_{j=1}^{Z_n} X_{nj} \qquad \text{for } n \geq 0, \quad \text{with } Z_0 \equiv 1,$$

where X_{nj} denotes the number of offspring of the jth individual present in the nth generation. We assume that all X_{nj}'s are iid as the X above. Also, we suppose

$$(15) \qquad m \equiv \mathrm{E}X = \textstyle\sum_{k=0}^{\infty} k\, p_k < \infty, \qquad \text{with} \quad p_0 > 0, \quad \text{and} \quad p_0 + p_1 < 1.$$

We call this a simple *branching process* model. Let

$$(16) \qquad W_n \equiv Z_n / m^n \qquad \text{and} \qquad \mathcal{A}_n \equiv \sigma[W_1, \ldots, W_n].$$

Proposition 4.1 The process

$$(17) \qquad \{W_n, \mathcal{A}_n\}_{n=0}^{\infty} \qquad \text{is a mg} \qquad \text{with mean} \qquad \mathrm{E}W_n = 1\,,$$

and

$$(18) \qquad \mathrm{Var}[W_n] = \begin{cases} n\sigma^2 & \text{if } m = 1, \\ \sigma^2 \frac{1-m^{-n}}{m(m-1)} & \text{if } m \neq 1, \end{cases}$$

provided that $\sigma^2 \equiv \mathrm{Var}[X] < \infty$.

Proof. We note that

$$(a) \qquad \mathrm{E}Z_{n+1} = \mathrm{E}[\mathrm{E}(Z_{n+1}|Z_n)] = \textstyle\sum_{k=0}^{\infty} \mathrm{E}(Z_{n+1}|Z_n = k)\, P(Z_n = k)$$

$$= \textstyle\sum_{k=0}^{\infty} \mathrm{E}(\sum_{j=1}^{k} X_{nj})\, P(Z_n = k) = \sum_{k=0}^{\infty} m\, k\, P(Z_n = k)$$

$$(b) \qquad = m\, \mathrm{E}(Z_n) = \cdots = m^{n+1},$$

while the mg property follows from

$$(c) \qquad \mathrm{E}(W_{n+1}|\mathcal{A}_n) = m^{-(n+1)}\, \mathrm{E}(Z_{n+1}|Z_n) = m^{-(n+1)}\, m Z_n = W_n.$$

We leave (18) to the following exercise. □

Exercise 4.4 Verify the variance formula (18). Verify (20) below.

Notation 4.1 We define the generating functions f and f_n of X and Z_n by

$$(19) \qquad f(s) \equiv \textstyle\sum_{k=0}^{\infty} s^k\, p_k \qquad \text{and} \qquad f_n(s) \equiv \sum_{k=0}^{\infty} s^k\, P(Z_n = k).$$

It is easy to verify that

$$(20) \qquad f_{n+1}(s) = f_n(f(s)) = f(f_n(s)) \qquad \text{for } |s| \leq 1.$$ □

Theorem 4.1 (Branching process) (i) Suppose that $m = \mathrm{E}X > 1$ and also $\sigma^2 \equiv \mathrm{Var}[X] < \infty$. Then

(21) $W_n \to_{a.s. \text{ and } \mathcal{L}_2} W_\infty \cong (1, \sigma^2/[m(m-1)]),$

where $(W_\infty, \mathcal{A}_\infty)$ closes the mg. Also,

(22) $P(W_\infty = 0) = (\text{the probability of ultimate extinction}) = \pi,$

where

(23) $\pi \in (0,1)$ is the unique solution of $f(\pi) = \pi$.

Moreover, the chf ϕ of W_∞ is characterized as the unique solution of

(24) $\phi(mt) = f(\phi(t))$ for $t \in R$ subject $\phi(0) = 1$ and $\phi'(0) = i\,m$.

(ii) If $m \le 1$, then $W_n \to_{a.s.} 0$ as $n \to \infty$.

Proof. (i) Now, $\mathrm{E}W_n^2 \le 1 + \sigma^2/[m(m-1)]$ for all n, so that the mg $\{W_n, \mathcal{A}_n\}_{n=1}^\infty$ is square-integrable. Thus the mg $\mathcal{L}_r$ convergence of exercise 13.3.7 gives (21).

We let $\pi^* \equiv P(W_\infty = 0)$. Then

(a) $\pi^* = \sum_{k=0}^\infty P(W_\infty = 0 | Z_1 = k)\, P(Z_1 = k) = \sum_{k=0}^\infty P(W_\infty = 0)^k\, p_k$

(b) $\quad = \sum_{k=0}^\infty {\pi^*}^k\, p_k = f(\pi^*).$

Now, $f(0) = p_0 > 0$, $f(1) = 1$, $f'(1-) = m > 1$, and $f'(s)$ is $\nearrow$ in s for $0 < s < 1$; draw a figure. Thus $f(\pi) = \pi$ has a unique solution in $(0,1)$. The solution $\pi = 1$ is ruled out by $\mathrm{Var}[W_\infty] > 0$, since $\pi = 1$ would imply $W_\infty \equiv 0$. (Note that (22) also follows from (20).)

We now turn to (24). Now,

(c) $\phi_{n+1}(t) \equiv \mathrm{E}\, e^{itW_{n+1}} = \sum_{j=0}^\infty \mathrm{E}(\exp(itZ_{n+1}/m^{n+1})|Z_1 = j) P(Z_1 = j)$

$\quad = \sum_{j=0}^\infty \phi_n(t/m)^j\, p_j$

(d) $\quad = f(\phi_n(t/m)).$

Since $W_n \to_{a.s.} W_\infty$ implies $W_n \to_d W_\infty$, we have $\phi_n \to \phi$ on R. Applying this to the identity (b) gives

$\phi(t) = \lim \phi_{n+1}(t) = \lim f(\phi_n(t/m))$

$\quad = f(\lim \phi_n(t/m))$ since f is continuous on $|r| \le 1$

(e) $\quad = f(\phi(t/m)).$

Suppose now that ψ is any chf that satisfies $\psi(t) = f(\psi(t/m))$. Then

$\gamma(t) \equiv [\psi(t) - \phi(t)]/t = [(\psi(t) - 1) - (\phi(t) - 1)]/t$

$\quad \to \psi'(0) - \phi'(0)$ if $\psi(0) = 1$ and $\psi'(0)$ exists

(f) $\quad = 0$ if $\psi'(0) = i\,m$

as $t \to 0$. Also,

$$|t\,m| \times |\gamma(tm)| = |\psi(tm) - \phi(tm)| = |f(\psi(t)) - f(\phi(t))|$$

$$\leq |f'(t^*)| \times |\psi(t) - \phi(t)| \qquad \text{for } t^* \in (0,1) \text{ by the mean value theorem}$$

$$\leq m\,|\psi(t) - \phi(t)|$$

(g) $$= |tm| \times |\gamma(t)|,$$

and iterating (e) gives

(h) $$|\gamma(t)| \leq |\gamma(t/m)| \leq \cdots \leq |\gamma(t/m^n)| \to 0,$$

so that $\gamma(t) = 0$ for all $t \neq 0$. Trivially, $\gamma(t) = 0$ for $t = 0$. Thus, (24) holds.

(ii) Set $s = 0$ in (20) to get $P(Z_{n+1} = 0) = f(P(Z_n = 0))$, here $P(Z_n = 0)$ is necessarily $\nearrow$. Passing to the limit gives $\pi = \lim P(Z_n = 0) = f(\pi)$. But if $m \leq 1$, then $\pi = 1$ is the only solution of $\pi = f(\pi)$. $\square$

5 Decomposition of a Submartingale Sequence

Definition 5.1 (Predictable process) A *predictable* process $\{A_n, \mathcal{A}_n\}_{n=0}^{\infty}$ is one in which each A_n is $\mathcal{A}_{n-1}$-measurable for each $n \geq 0$; here A_0 is a constant (or A_0 is $\{\emptyset, \Omega\}$-measurable). [Especially interesting are processes that are both $\nearrow$ and predictable, since any submg can be decomposed as the sum of a mg and such a predictable process.]

Theorem 5.1 (Decomposition of a submg) Let $\{X_n, \mathcal{A}_n\}_{n=0}^{\infty}$ be a submg. Then X_n can be decomposed as

(1) $\quad X_n = Y_n + A_n = [\text{a mg}] + [\text{an } \nearrow \text{ and predictable process}],$

where $\{Y_n, \mathcal{A}_n\}_{n=0}^{\infty}$ is a 0-mean mg and A_n is a predictable process satisfying

(2) $\quad A_0 \equiv \mathrm{E}X_0 \leq A_1 \leq \cdots \leq A_n \leq \cdots \qquad$ a.s.

This decomposition is a.s. unique. Conversely, if $X_n = Y_n + A_n$ as above, then $\{X_n, \mathcal{A}_n\}_{n=0}^{\infty}$ is a submg. [Call A_n the *compensator*.]

Proof. (Doob) Suppose that $\{X_n, \mathcal{A}_n\}_{n=0}^{\infty}$ is a submg. Let $A_0 \equiv \mathrm{E}X_0$, and for $n \geq 1$ define (the compensator candidate)

(3) $$A_n \equiv \sum_{i=1}^{n} [\mathrm{E}(X_k|\mathcal{A}_{k-1}) - X_{k-1}] + \mathrm{E}X_0 = \sum_{i=1}^{n} \mathrm{E}(\Delta X_k \,|\, \mathcal{A}_{k-1}) + \mathrm{E}X_0,$$

with $\Delta X_k \equiv X_k - X_{k-1}$. Clearly, A_n is an $\nearrow$ process and each of the A_n is $\mathcal{A}_{n-1}$-measurable. So, it remains only to show that $Y_n \equiv X_n - A_n$ is a mg. Now,

$$\mathrm{E}(Y_n|\mathcal{A}_{n-1}) = \mathrm{E}(X_n|\mathcal{A}_{n-1}) - \mathrm{E}(A_n|\mathcal{A}_{n-1}) = \mathrm{E}(X_n|\mathcal{A}_{n-1}) - A_n$$

(a) $\quad = \mathrm{E}(X_n|\mathcal{A}_{n-1}) - [\mathrm{E}(X_n|\mathcal{A}_{n-1}) - X_{n-1}] - A_{n-1}$

(b) $\quad = X_{n-1} - A_{n-1} = Y_{n-1},$

so $\{Y_n, \mathcal{A}_n\}_{n=0}^{\infty}$ is indeed a mg. Consider the uniqueness. Suppose $X_n = Y_n + A_n$ is one such decomposition that works. Elementary computations give

(c) $\quad \mathrm{E}(X_n|\mathcal{A}_{n-1}) = \mathrm{E}(Y_n|\mathcal{A}_{n-1}) + A_n = Y_{n-1} + A_n \qquad$ a.s.;

but the specification of the decomposition also states that

(d) $\quad X_{n-1} = Y_{n-1} + A_{n-1} \qquad$ a.s.

Subtracting (d) from (c) gives uniqueness via

(e) $\quad A_n - A_{n-1} = \mathrm{E}(X_n|\mathcal{A}_{n-1}) - X_{n-1} \qquad$ a.s.

The converse holds, since

(f) $\quad \mathrm{E}(Y_n + A_n|\mathcal{A}_{n-1}) = Y_{n-1} + A_n \geq Y_{n-1} + A_{n-1} \qquad$ a.s., for $n \geq 1$

as required. □

Exercise 5.1 If the X-process is integrable, then the A-process is uniformly integrable (in either theorem 5.1 above or theorem 5.2 below).

Theorem 5.2 (Decomposition of a reversed submg) Let $\{X_n, \mathcal{A}_n\}_{n=-\infty}^0$ be a submg such that $\mathrm{E}(X_n) \searrow M > -\infty$ as $n \searrow -\infty$ (thus, the X_n-process is uniformly integrable with $X_n \to$ (some $X_{-\infty}$) a.s. and $\mathcal{L}_1$ where $\mathrm{E}X_{-\infty} = M$). Then X_n can be decomposed as

(4) $\quad X_n = Y_n + A_n = [\text{a mg}] + [\text{an } \nearrow \text{ and predictable process that is } \geq 0],$

where $\{Y_n, \mathcal{A}_n\}_{n=-\underline{\infty}}^0$ is a mean-M mg and A_n is an $\mathcal{A}_{n-1}$-measurable function with

(5) $\quad 0 = A_{-\infty} \equiv \lim_{n \to -\underline{\infty}} A_n \leq \cdots \leq A_n \leq \cdots \leq A_0 \qquad \text{a.s.}$

This decomposition is a.s. unique. Conversely, if $X_n = Y_n + A_n$ as above, then $\{X_n, \mathcal{A}_n\}_{n=-\underline{\infty}}^0$ is a submg. [Call A_n the *compensator*.]

Proof. (Doob) We define

(6) $\quad A_n \equiv \sum_{k=-\infty}^{n} [\mathrm{E}(X_k|\mathcal{A}_{k-1}) - X_{k-1}] \qquad \text{for } n \leq 0;$

then A_n is clearly ≥ 0, $\nearrow$, and $\mathcal{A}_{n-1}$-measurable, provided that it can be shown to be well-defined (that is, provided the sum converges a.s.). Now, with $n \leq m$,

(a) $\quad \mathrm{E}(A_m - A_n) = \sum_{n+1}^m \mathrm{E}(X_k - X_{k-1}) = \mathrm{E}X_n - \mathrm{E}X_m \leq \mathrm{E}X_0 - \mathrm{E}X_m$

(b) $\quad \leq (\mathrm{E}X_0 - M) < \infty,$

by hypothesis. Also,

(c) $\quad \tilde{A}_m \equiv \lim_{n \to -\infty} (A_m - A_n) = \lim_{n \to -\infty} \sum_{n+1}^m [\mathrm{E}(X_k|\mathcal{A}_{k-1}) - X_{k-1}]$

is ≥ 0 and $\nearrow$, so that the MCT gives

(d) $\quad \mathrm{E}\tilde{A}_m = \lim_{n \to -\infty} \mathrm{E}(A_m - A_n) = \mathrm{E}X_m - \lim_{n \to -\infty} \mathrm{E}X_n = \mathrm{E}X_m - M < \infty$

with a well-defined finite limit. Since $\tilde{A}_m \geq 0$ and $\mathrm{E}\tilde{A}_m < \infty$, we know that $\tilde{A}_m$ is finite a.s.; so (6) is well-defined The $\tilde{A}_m$'s are $\nearrow$ and bounded below by 0. Thus $A_{-\infty} \equiv \lim_{m \to -\infty} \tilde{A}_m$ exists a.s., and it is ≥ 0. Moreover, the equalities in (d) show that $\mathrm{E}\tilde{A}_m \to 0$ as $m \to -\infty$; just use the MCT via

(e) $\quad \mathrm{E}A_{-\infty} = \lim_{m \to -\infty} \mathrm{E}\tilde{A}_m = \lim_m \mathrm{E}X_m - M = M - M = 0.$

Thus $A_{-\infty} = 0$ a.s., and each $\tilde{A}_m = A_m$ a.s.

Let $Y_n \equiv X_n - A_n$. Lines (a)–(f) of the previous proof complete this proof, using (13.3.20) about mgs for the existence of $Y_{-\infty}$. □

Example 5.1 (Predictable variation, or conditional variance of a mg) Let $\{X_n, \mathcal{A}_n\}_{n=0}^{\infty}$ be a mg with each $\mathrm{E}X_n^2 < \infty$. Then $\{X_n^2, \mathcal{A}_n\}_{n=0}^{\infty}$ is a submg by proposition 13.1.2. By theorem 5.1, there is a decomposition for which

(7) $\quad Z_n \equiv X_n^2 - A_n \quad$ is a 0-mean mg adapted to the $\mathcal{A}_n$'s for $n \ge 0$.

Here A_n is the predictable process (with $A_0 \equiv \mathrm{E}X_0^2 \ge 0$) defined by

$$A_n \equiv \sum_{k=1}^n \{\mathrm{E}(X_k^2|\mathcal{A}_{k-1}) - X_{k-1}^2\} + \mathrm{E}X_0^2$$

(8) $$= \sum_{k=1}^n \mathrm{E}\{X_k^2 - X_{k-1}^2|\mathcal{A}_{k-1}\} + \mathrm{E}X_0^2$$

(9) $$= \sum_{i=1}^n \mathrm{E}\{(\Delta X_k)^2|\mathcal{A}_{k-1}\} + \mathrm{E}X_0^2\,, \qquad \text{for } n \ge 1,$$

where $\Delta X_k \equiv X_k - X_{k-1}$ and $\Delta X_0 = X_0$. The compensator term A_n (of the X_n^2-process) given in (9) is called the *conditional variance* or the *predictable variation* of the X_n-process. Note that (for $\langle X\rangle_n \equiv A_n$),

(10) $$\mathrm{E}X_n^2 = \mathrm{E}\langle X\rangle_n = \mathrm{E}A_n = \sum_{k=1}^n \mathrm{Var}[\Delta X_k] + \mathrm{E}X_0^2\,,$$

since we agree to also use the notation $\langle X\rangle_n$ to denote the predictable variation process A_n that corresponds to the mg $\{X_n, \mathcal{A}_n\}_{n=0}^{\infty}$.

Summary For any mg $\{X_n, \mathcal{A}_n\}_{n=0}^{\infty}$ having all $\mathrm{E}X_n^2$ finite,

(11) $$\langle X\rangle_n \equiv A_n = \sum_{k=1}^n \mathrm{E}\{(\Delta X_k)^2|\mathcal{A}_{k-1}\} + \mathrm{E}X_0^2$$

is always the *predictable variation* (or *conditional variance*, or *compensator*), and the conditionally centered process

(12) $\quad Z_n \equiv X_n^2 - \langle X\rangle_n \quad$ is a 0-mean mg with respect to the $\mathcal{A}_n$'s, for $n \ge 0$.□

Martingale Transforms

Definition 5.2 (H-transforms) Let $\{H_n\}_{n=0}^{\infty}$ be a predictable process with respect to the filtration $\{\mathcal{A}_n\}_{n=0}^{\infty}$. [Think of H_n being the amount a gambler will wager at stage n, based only on complete knowledge of the outcomes of the game up through time $n-1$ (but not, of course, through time n).] For some other process $\{X_n\}_{n=0}^{\infty}$, define the H-*transform* of X (to be denoted by $\{(H\cdot X)_n\}_{n=0}^{\infty}$) by

(13) $$(H\cdot X)_n \equiv \sum_{k=1}^n H_k\,(X_k - X_{k-1}) + H_0\,X_0 = \sum_{k=0}^n H_k\,\Delta X_k\,.$$

(We agree that $\Delta X_0 \equiv X_0$, and that H_0 is a constant.)

Theorem 5.3 (S-mg transforms) (i) Let $\{X_n, \mathcal{A}_n\}_{n=0}^{\infty}$ be a s-mg (or supermg). If $\{H_n\}_{n=0}^{\infty}$ is predictable with each $H_n \ge 0$ and bounded, then $\{(H\cdot X)_n, \mathcal{A}_n\}_{n=0}^{\infty}$ is a s-mg (or supermg). (The supermartingale case shows that there is no system for beating the house in an unfavorable game.)

(ii) If $\{X_n, \mathcal{A}_n\}_{n=0}^{\infty}$ is a mg and $\{H_n\}_{n=0}^{\infty}$ is predictable and bounded, then the process $\{(H\cdot X)_n, \mathcal{A}_n\}_{n=0}^{\infty}$ is a mg with mean $H_0\,\mathrm{E}X_0$.

Proof. We compute

(a) $\mathrm{E}[(H\cdot X)_{n+1}|\mathcal{A}_n] = (H\cdot X)_n + \mathrm{E}[H_{n+1}\,(X_{n+1}-X_n)|\mathcal{A}_n]$

(b) $= (H\cdot X)_n + H_{n+1}\,\mathrm{E}[\Delta X_{n+1}|\mathcal{A}_n]$

(c) $\geqq (H\cdot X)_n\,,$

since $H_{n+1}\geq 0$ and $\mathrm{E}(\Delta X_{n+1}|\mathcal{A}_n)\geqq 0$. Note that $\mathrm{E}\Delta X_0 = \mathrm{E}X_0$ in the mg case. (The supermg case just reverses the inequality.) □

Corollary 1 If $\{X_n,\mathcal{A}_n\}_{n=0}^{\infty}$ is a s-mg and T is a stopping time, then:

(a) $H_n \equiv 1_{[T\geq n]}$ is ≥ 0, bounded, and predictable.

(b) $(H\cdot X)_n = X_{T\wedge n}$ = (the stopped process) is a s-mg.

Proof. Now, H_n is predictable, since $[T\geq n] = [T\leq n-1]^c \in \mathcal{A}_{n-1}$ for a stopping time T. Furthermore,

(a) $(H\cdot X)_n = \sum_{k=0}^{n} 1_{[T\geq k]}\,(X_k - X_{k-1}) = \sum_{k=0}^{n} 1_{[T\geq k]}\,\Delta X_k = X_{T\wedge n}$

(the sum ends at m if $T(\omega)=m\in[0,n]$; else, at n). Then, apply theorem 5.3. □

Notation 5.1 Suppose $\{X_n,\mathcal{A}_n\}_{n=0}^{\infty}$ is a mg with $\mathrm{E}X_n^2<\infty$, so that $\{X_n^2\}_{n=0}^{\infty}$ is a submg with predictable variation process $\{\langle X\rangle_n\}_{n=0}^{\infty}$. Let $\{H_n\}_{n=0}^{\infty}$ denote a predictable, bounded, and ≥ 0 process. Then we know that $\{(H\cdot X)_n^2\}$ is a submg. We will now give the form taken by its predictable variation process $\langle H\cdot X\rangle_n$. Also, we will summarize everything so far in one place. □

Theorem 5.4 (Martingale transforms) Suppose $\{X_n,\mathcal{A}_n\}_{n=0}^{\infty}$ is a mg with $\mathrm{E}X_n^2<\infty$ for each n, and let $\{H_n\}_{n=0}^{\infty}$ be bounded, predictable, and ≥ 0. Then the predictable variation process $\langle X\rangle_n$ is given by

$$(14)\qquad \langle X\rangle_n \equiv A_n = \sum_{k=1}^{n} \mathrm{E}\{(\Delta X_k)^2|\mathcal{A}_{k-1}\} + \mathrm{E}X_0^2\,.$$

Then the conditionally centered process

(15) $Z_n \equiv X_n^2 - \langle X\rangle_n$ is a 0-mean mg with respect to the $\mathcal{A}_n$'s,

for $n\geq 0$. The martingale transform

(16) $W_n \equiv (H\cdot X)_n \equiv \sum_{k=0}^{n} H_k\,\Delta X_k$
is a mg with mean $H_0\,\mathrm{E}X_0$ with respect to the $\mathcal{A}_n$'s,

for $n\geq 0$. Its *predictable variation process* $\langle W\rangle_n$ is

$$(17)\qquad \langle W\rangle_n \equiv \langle H\cdot X\rangle_n = \sum_{k=1}^{n} H_k^2\,\mathrm{E}\{\,(\Delta X_k)^2\,|\mathcal{A}_{k-1}\} + H_0^2\,\mathrm{E}\,X_0^2\,.$$

Moreover, for $n\geq 0$, the sequence

(18) $L_n \equiv W_n^2 - \langle W\rangle_n \equiv \{(H\cdot X)_n\}^2 - \langle (H\cdot X)\rangle_n$
is a 0-mean mg with respect to the $\mathcal{A}_n$'s (with $L_0 = H_0^2(X_0^2 - \mathrm{E}X_0^2)$).

Proof. Recall example 5.1 and theorem 5.3 for the first parts. Then by straightforward calculation from (13), we have

(a) $\langle H \cdot X \rangle_n = \sum_{k=1}^n \mathrm{E}\{[\Delta(H \cdot X)_k]^2 | \mathcal{A}_{k-1}\} + H_0^2 \, \mathrm{E}X_0^2$

since $(H \cdot X)_0 = H_0 X_0$

$= \sum_{k=1}^n \mathrm{E}\{[H_k \, (X_k - X_{k-1})]^2 | \mathcal{A}_{k-1}\} + H_0^2 \, \mathrm{E}X_0^2$

$= \sum_{k=1}^n H_k^2 \times \mathrm{E}\{(\Delta X_k)^2 | \mathcal{A}_{k-1}\} + H_0^2 \, \mathrm{E}X_0^2$

since H_k is $\mathcal{A}_{k-1}$-measurable

(b) $= \sum_{k=0}^n H_k^2 \, \Delta \langle X \rangle_k \, .$

Note that $L_0 = W_0^2 - \langle W_0 \rangle = [H_0 \, \Delta X_0]^2 - H_0^2 \, \mathrm{E}(X_0^2) = H_0^2 \, (X_0^2 - \mathrm{E}X_0^2)$ has mean 0, while L_n is a mg by example 5.1. □

Exercise 5.2 Let $(X_n, \mathcal{A}_n)_{n=0}^{\infty}$ be a submg. Let $M_n \equiv \sup\{|X_n| : n \geq 1\}$. Use the Doob-Meyer decomposition and Doob's inequality 8.10.2 to show that

(19) $P(M_n \geq \lambda) \leq \frac{3}{\lambda} \sup_n \mathrm{E}|X_n|$ for all $\lambda > 0$.

6 Optional Sampling

We now extend the simple optional sampling theorem of section 13.2. Our aim will be to relax the restrictive assumption used there, that the stopping times are bounded.

Discrete Time

Notation 6.1 Suppose that

(1) $\{X_n, \mathcal{A}_n\}_{n=0}^{\infty}$ is a s-mg

and that

(2) $0 \le_{a.s.} T_0 \le_{a.s.} T_1 \le_{a.s.} \cdots <_{a.s.} \infty$ for stopping times $T_0, T_1, \ldots$.

We define

(3) $\tilde{X}_n \equiv X_{T_n}$ and $\tilde{\mathcal{A}}_n \equiv \mathcal{A}_{T_n}$,

so that the $\tilde{X}_n$'s are adapted to the $\tilde{\mathcal{A}}_n$'s. We would like to prove that $\{\tilde{X}_n, \tilde{\mathcal{A}}_n\}_{n=0}^{\infty}$ is a s-mg, but this requires hypotheses. The weakest such hypotheses presented are

(4) $\mathrm{E}|\tilde{X}_n| < \infty$ for all n,

(5) $\liminf_{k\to\infty} \int_{[T_n>k]} X_k^+ \, dP = 0$ for each n.

(But these conditions (4) and (5) need to be replaced by *useful* conditions that are more easily verifiable.) □

Theorem 6.1 (Optional sampling theorem) Let (1)–(3) define the sequence $\{\tilde{X}_n, \tilde{\mathcal{A}}_n\}_{n=0}^{\infty}$ with respect to the s-mg $\{X_n, \mathcal{A}_n\}_{n=0}^{\infty}$. Suppose (4) and (5) hold. Then the optionally sampled process

(6) $\{\tilde{X}_n, \tilde{\mathcal{A}}_n\}_{n=0}^{\infty}$ is a s-mg

for which

(7) $\mathrm{E}X_0 \leqq \mathrm{E}\tilde{X}_0 \leqq \cdots \leqq \mathrm{E}\tilde{X}_n \leqq \cdots \leqq \sup_k \mathrm{E}X_k \le \infty$.

Corollary 1 (A) Condition (4) holds if $\{X_n, \mathcal{A}_n\}_{n=0}^{\infty}$ is integrable.
(B) Conditions (4) and (5) both hold if any of the following conditions holds:

(8) Each T_n is a.s. bounded by some fixed integer N_n.

(9) The X_n's are uniformly integrable.

(10) $X_n \le_{a.s.}$ (some M) $< \infty$ for all n.

(11) $\mathrm{E}T_j < \infty$ for all j, and there exists a constant K such that for all j,

$$\mathrm{E}(|X_n - X_{n-1}| \,\big|\, \mathcal{A}_{n-1})(\omega) \le K \text{ for all } n \le T_j(\omega) \text{ holds a.s.}$$

Notation 6.2 The theorem becomes much cleaner if our s-mg includes an entry at ∞ that closes the s-mg. Suppose

(12) $\{X_n, \mathcal{A}_n\}_{n=0}^{\infty}$ is a s-mg

and

(13) $0 \le_{a.s.} T_0 \le_{a.s.} T_1 \le_{a.s.} \cdots \le T_\infty \le_{a.s.} \infty$

for extended stopping times $T_0, T_1, \ldots, T_\infty$. We again define

(14) $\tilde{X}_n \equiv X_{T_n}$ and $\tilde{\mathcal{A}}_n \equiv \mathcal{A}_{T_n}$,

so that the $\tilde{X}_n$'s are adapted to the $\tilde{\mathcal{A}}_n$'s (for $0 \le n \le \infty$). □

Theorem 6.2 (Optional sampling theorem) Suppose (12)–(14) hold. Then

(15) $\{\tilde{X}_n, \tilde{\mathcal{A}}_n\}_{n=0}^{\overline{\infty}}$ is a s-mg

with $\mathrm{E}X_0 \leqq \mathrm{E}X_{T_0} \leqq \mathrm{E}X_{T_1} \leqq \cdots \leqq \mathrm{E}X_{T_\infty} \leqq \mathrm{E}X_\infty < \infty$.

Continuous Time

Theorem 6.3 (Optional sampling theorem) Suppose that the X-process is integrable and satisfies $X : (\Omega, \mathcal{A}, P) \to (D_{[0,\infty)}, \mathcal{D}_{[0,\infty)})$, and that it is adapted to some filtration $\{\mathcal{A}_t\}_{t\in[0,\infty)}$. Then

(16) $n, X_n, \mathcal{A}_n, \{0,1,\ldots\}, \{0,1,\ldots,\infty\}, k$ may be replaced by $t, X_t, \mathcal{A}_t, [0,\infty), [0,\infty], s$

in theorem 6.1, corollary 1 (only (11) must be omitted from the list of things that carry over with no change), and theorem 6.2.

Proofs

Proof. Consider theorem 6.1. Let $A \in \mathcal{A}_{T_{n-1}}$. It suffices to show that

(a) $\int_A X_{T_{n-1}} dP \leqq \int_A X_{T_n}\, dP.$

Basically, we wish to use proposition 13.2.1 and the DCT. To this end we define

(b) $T_n^{(k)} \equiv T_n \wedge k =$ (a bounded stopping time) $\nearrow T_n$.

Now $A \cap [T_{n-1} \le k] \in \mathcal{A}_{T_{n-1}^{(k)}}$, since for each $m \ge 0$ we have

$$A \cap [T_{n-1} \le k] \cap [T_{n-1}^{(k)} \le m] = A \cap [T_{n-1} \le k] \cap [T_{n-1} \wedge k \le m \wedge k]$$

$$= A \cap [T_{n-1} \le m \wedge k] \in \mathcal{A}_{m\wedge k} \subset \mathcal{A}_m.$$

Thus, for n fixed we have

(c) $\int_{A\cap[T_{n-1}\le k]} X_{T_{n-1}}\,dP$

$= \int_{A\cap[T_{n-1}\le k]} X_{T_{n-1}^{(k)}}\,dP$ as the integrands are equal on the set

(d) $\overset{\le}{=} \int_{A\cap[T_{n-1}\le k]} X_{T_n^{(k)}}\,dP$ by proposition 13.2.1

(e) $\overset{\le}{=} \int_{A\cap[T_{n-1}\le k]\cap[T_n\le k]} X_{T_n}\,dP + \int_{A\cap[T_{n-1}\le k]\cap[T_n>k]} X_k\,dP\,,$

since $X_{T_n^{(k)}} = X_k$ on $[T_n > k]$. Let $k \to \infty$ in (c) and (e); and since $T_n < \infty$ a.s. and since $\mathrm{E}|X_{T_{n-1}}|$ and $\mathrm{E}|X_{T_n}|$ are finite, the DCT implies (recall that $a = b \oplus c$ means that $|a - b| \le c$)

(f) $\int_A X_{T_{n-1}}\,dP$

(g) $\overset{\le}{=} \int_A X_{T_n}\,dP \oplus \underline{\lim} \int_{[T_n>k]} X_k^+\,dP$

(h) $\equiv \int_A X_{T_n}\,dP \oplus \underline{\lim}\, a_k,$ where $\underline{\lim}_{\,k\to\infty}\, a_k = 0$ by (5).

Equate the terms in (f) and (h) on a subsequence k' having $a_{k'} \to 0$ to obtain

(i) $\int_A X_{T_{n-1}}\,dP \overset{\le}{=} \int_A X_{T_n}\,dP$ for all $A \in \mathcal{A}_{T_{n-1}}$.

This is equivalent to (6) by (13.1.4).

Letting $A = \Omega$ in (i) shows that $\mathrm{E}X_{T_n}$ is $\nearrow$. Introducing the new stopping time $\tau \equiv 0 \le T_0$ and applying this result shows that $\mathrm{E}X_0 \equiv \mathrm{E}X_\tau \overset{\le}{=} \mathrm{E}X_{T_n}$.

It remains to show that $\mathrm{E}X_{T_n} \overset{\le}{=} \sup_k \mathrm{E}X_k$. Now,

(j) $\mathrm{E}\,X_{T_n} = \int_{[T_n\le k]} X_{T_n}\,dP + \int_{[T_n>k]} X_{T_n}\,dP$

$= \int_{[T_n\le k]} X_{T_n^{(k)}}\,dP + \int_{[T_n>k]} X_{T_n}\,dP$

(k) $\overset{\le}{=} \int_{[T_n\le k]} X_k\,dP + \int_{[T_n>k]} X_{T_n}\,dP \pm \int_{[T_n>k]} X_k\,dP$

by proposition 13.2.1, since $[T_n \le k] \in \mathcal{A}_{T_n^{(k)}}$

as $[T_n \le k] \cap [T_n \wedge k \le m] = [T_n \le m \wedge k] \in \mathcal{A}_{k\wedge m} \in \mathcal{A}_m,$ for $m \ge 0$

(l) $\overset{\le}{=} \mathrm{E}X_k \oplus \int_{[T_n>k]} X_k^+\,dP \oplus \int_{[T_n>k]} X_{T_n}\,dP.$

For the second term in (l) we recall (5). For the third term in (l) we note that $\mathrm{E}|X_{T_n}| < \infty$ and that $T_n < \infty$ a.s. We thus conclude from (l) that

(m) $\mathrm{E}X_{T_n} \overset{\le}{=} \overline{\lim}_{\,k\to\infty}\, EX_k + 0 + 0$

(n) $\overset{\le}{=} \sup_k \mathrm{E}X_k\,.$

This gives (7). □

Proof. Consider the first claim made in the corollary. That is, we verify that (4) holds if $\sup_k \mathrm{E}|X_k| < \infty$. Let $T_n^{(k)} \equiv T_{n\wedge k}$. Now, both $\{X_n, \mathcal{A}_n\}_{n=0}^{\infty}$ and $\{X_n^+, \mathcal{A}_n\}_{n=0}^{\infty}$ are submgs. Since $0 \le T_n^{(k)} \le k$, proposition 13.2.1 implies that both

(a) $\quad X_0, X_{T_n^{(k)}}, X_k \quad$ and $\quad X_0^+, X_{T_n^{(k)}}^+, X_k^+ \quad$ are submgs.

Thus,

(b) $\quad \mathrm{E}|X_{T_n^{(k)}}| = \mathrm{E}[2\, X_{T_n^{(k)}}^+ - X_{T_n^{(k)}}] \qquad$ since $|x| = 2\,x^+ - x$

(c) $\quad \le 2\,\mathrm{E}X_k^+ - \mathrm{E}X_0 \qquad$ using (a)

$\quad \le 2\,\mathrm{E}|X_k| - \mathrm{E}X_0$

$\quad \le 2\,\sup_k \mathrm{E}|X_k| - \mathrm{E}X_0$

(d) $\quad \le (\text{some } M) < \infty \qquad$ by hypothesis.

Thus, Fatou's lemma and then (d) gives

(e) $\quad \mathrm{E}\,|\tilde{X}_n| = \mathrm{E}\,|X_{T_n}| = \mathrm{E}\,|\,\underline{\lim}\, X_{T_n^{(k)}}| \le \underline{\lim}\, \mathrm{E}\, X_{T_{n\wedge k}}$

$\quad \le \underline{\lim}\, \mathrm{E}\,|X_k|$

(f) $\quad = \sup \mathrm{E}\,|X_k| < \infty$ □

Proof. Consider (8). Now, $\mathrm{E}\,|\tilde{X}_n| \le \sum_0^{N_n} \mathrm{E}\,|X_j|$ implies (4), and (5) is trivial, since the integral equals 0 for $k \ge N_n$. □

Proof. Consider (9). Uniformly integrable X_n's are uniformly bounded in $\mathcal{L}_1$ and are uniformly absolutely continuous, by theorem 3.5.4. Uniformly bounded in $\mathcal{L}_1$ means $\sup_{k\ge 1} \mathrm{E}\,|X_k| < \infty$; hence by part (A) of the corollary, we have (4). Since $T_n < \infty$ a.s., we have $P(T_n > k) \to 0$ as $k \to \infty$, and hence uniform absolute continuity implies (5). □

Proof. Consider (10). Now, (5) is trivial, since $X_k^+ \le M$, and (4) holds, since

(a) $\quad 0 \le \mathrm{E}(M - X_{T_n}) = \mathrm{E}\,[\lim(M - X_{T_n^{(k)}})]$

(b) $\quad \le \underline{\lim}\, \mathrm{E}(M - X_{T_n^{(k)}}) \qquad$ by Fatou, since all $M - X_n \ge 0$

(c) $\quad \le \underline{\lim}\, \mathrm{E}(M - X_0) \qquad$ by proposition 13.2.1

(d) $\quad = M - \mathrm{E}(X_0) < \infty,$

giving $\mathrm{E}X_0 \le \mathrm{E}X_{T_n} \le M$. □

Proof. Consider (11). Let $Y_0 \equiv |X_0|$ and $Y_j \equiv |X_j - X_{j-1}|$ for $j \ge 1$. Define the sums $Z_n \equiv \sum_0^n Y_j$ and $\tilde{Z}_n \equiv \sum_0^{T_n} Y_j$. Then (4) holds since

(a) $\quad \mathrm{E}|\tilde{X}_n| \le \mathrm{E}|\tilde{Z}_n| = \sum_{k=0}^{\infty} \int_{[T_n=k]} Z_k \, dP = \sum_{k=0}^{\infty} \int_{[T_n=k]} \sum_{j=0}^{k} Y_j \, dP$

(b) $\quad = \sum_{j=0}^{\infty} \int_{[T_n \ge j]} Y_j \, dP \qquad$ by Fubini

(c) $\quad = \sum_{j=0}^{\infty} \int_{[T_n \ge j]} \mathrm{E}(Y_j|\mathcal{A}_{j-1}) \, dP$

since $[T_n \ge j] = [T_n \le j-1]^c \in \mathcal{A}_{j-1}$

(d) $\quad \le K \sum_{j=0}^{\infty} P(T_n \ge j) = K\,(1 + \mathrm{E}T_n) \qquad$ using (11) and then (8.2.1)

(e) $\quad < \infty.$

Also, (5) holds, since $\mathrm{E}T_n < \infty$ implies $P(T_n > k) \to 0$ as $k \to \infty$; hence $\mathrm{E}\tilde{Z}_n < \infty$ (as follows from (e)) gives

(f) $\quad \int_{[T_n > k]} |X_k| \, dP \le \int_{[T_n > k]} \tilde{Z}_n \, dP \to 0 \qquad$ as $k \to \infty$. □

Proof. Consider theorem 6.2. Since $\{X_n, \mathcal{A}_n\}_{n=0}^{\overline{\infty}}$ is a s-mg, we see from the s-mg convergence theorem that the X_n's are uniformly integrable and $X_n \to X_\infty$ a.s. and $\mathcal{L}_1$, for some $X_\infty \in \mathcal{L}_1$. Now, for any $n \in \{0, 1, \ldots, \infty\}$ we have

$$\mathrm{E}|X_{T_n}| \le \mathrm{E}(\lim_{k\to\infty} |X_{T_n^{(k)}}| + |X_\infty|) \qquad \text{since } T_n^{(k)} \equiv T_n \wedge k \nearrow T_n$$

$$\le \underline{\lim}\, \mathrm{E}|X_{T_n^{(k)}}| + \mathrm{E}|X_\infty| \qquad \text{by Fatou}$$

$$\le \underline{\lim}\, \mathrm{E}|X_k| + \mathrm{E}|X_\infty| \qquad \text{by proposition 13.2.1}$$

(a) $\quad < \infty, \qquad$ since the process is integrable and $X_\infty \in \mathcal{L}_1$.

Let $A \in \mathcal{A}_{T_{n-1}}$; recall that T_{n-1} could now equal $+\infty$. Even so, we do have $A \cap [T_{n-1} \le k] \in \mathcal{A}_{T_{n-1}^{(k)}}$, as shown at the start of the proof of theorem 6.1 . Thus,

(b) $\quad \int_{A\cap[T_{n-1}\le k]} X_{T_{n-1}} \, dP$

$$= \int_{A\cap[T_{n-1}\le k]} X_{T_{n-1}^{(k)}} \, dP \qquad \text{since } T_{n-1}^{(k)} = T_{n-1} \text{ on } [T_{n-1} \le k]$$

$$\leqq \int_{A\cap[T_{n-1}\le k]} X_{T_n^{(k)}} \, dP \qquad \text{by proposition 13.2.1}$$

(c) $\quad = \int_{A\cap[T_{n-1}\le k]\cap[T_n>k]} X_k \, dP + \int_{A\cap[T_{n-1}\le k]\cap[T_n\le k]} X_{T_n} \, dP,$

since $X_{T_n^{(k)}} = X_k$ on $[T_n > k]$. Let $k \to \infty$ in (b) and (c); since (a) shows that $\mathrm{E}|X_{T_{n-1}}|$ and $\mathrm{E}|X_{T_n}|$ are finite, the DCT gives

(d) $\int_{A\cap[T_{n-1}<\infty]} X_{T_{n-1}}\,dP$

$\leqq \underline{\lim}_{k\to\infty} \int_{A\cap[T_{n-1}\le k]\cap[T_n>k]} X_k\,dP + \int_{A\cap[T_n<\infty]} X_{T_n}\,dP$

(e) $= \int_{A\cap[T_{n-1}<\infty]\cap[T_n=+\infty]} X_\infty\,dP + \int_{A\cap[T_n<\infty]\cap[T_{n-1}<\infty]} X_{T_n}\,dP$

using $X_k \to_{\mathcal{L}_1} X_\infty$

(f) $= \int_{A\cap[T_{n-1}<\infty]} X_{T_n}\,dP.$

We add to each of (d) and (f) the equal terms of the equation

(g) $\int_{A\cap[T_{n-1}=\infty]} X_{T_{n-1}}\,dP = \int_{A\cap[T_{n-1}=\infty]} X_{T_n}\,dP,$

and obtain

(h) $\int_A X_{T_{n-1}}\,dP \leqq \int_A X_{T_n}\,dP \qquad$ for all $A \in \mathcal{A}_{T_{n-1}}$.

Replace T_n by T_∞ in the previous paragraph to see that

(i) $\int_A X_{T_{n-1}}\,dP \leqq \int_A X_{T_\infty}\,dP \qquad$ for all $A \in \mathcal{A}_{T_{n-1}}$.

Finally, (h), (i), and (13.1.4) show that $\{X_{T_n}, \mathcal{A}_{T_n}\}_{n=0}^{\overline{\infty}}$ is a s-mg. Add in $T_a \equiv 0$ and $T_b \equiv \infty$ for the expectation claim, with $A = \Omega$ in (h) and (i). □

Proof. Consider theorem 6.3. We must consider the extension of theorem 6.2. It suffices to consider the stopping times a pair at a time; we will do so, relabeling them so that $S \le T$ a.s. Let $D_n \equiv \{k/2^n : k = 0, 1, \ldots,\}$, and note that

(a) $\{X_t, \mathcal{A}_t\}_{t\in D_n} \qquad$ is a s-mg.

Define extended stopping times $T^{(n)}$ by

(b) $$T^{(n)}(\omega) = \begin{cases} k/2^n & \text{whenever} \quad (k-1)/2^n < T(\omega) \le k/2^n \text{ and } k \ge 0, \\ \infty & \text{whenever } T(\omega) = \infty, \end{cases}$$

and make an analogous definition for $S^{(n)}$; it is trivial that these rvs are extended stopping times. Note that a.s.

(c) $S^{(n)} \le T^{(n)}, \quad S \le S^{(n)}, \quad T \le T^{(n)} \quad S^{(n)} \searrow S, \quad T^{(n)} \searrow T.$

We can apply theorem 6.2 to $S^{(n)}$ and $T^{(n)}$ to conclude (as in (h), just above) that

(d) $\int_A X_{S^{(n)}}\,dP \leqq \int_A X_{T^{(n)}}\,dP \qquad$ for all $A \in \mathcal{A}_S \subset \mathcal{A}_{S^{(n)}}$.

Now, by right continuity of the paths,

(e) $X_{S^{(n)}} \to_{a.s.} X_S \qquad$ and $\qquad X_{T^{(n)}} \to_{a.s.} X_T \qquad$ as $n \to \infty$.

Thus Vitali's theorem allows us to pass to the limit in (d) and obtain

(f) $\int_A X_S\,dP \leqq \int_A X_T\,dP \qquad$ for all $A \in \mathcal{A}_S$,

provided that we show that

(g) the $X_{T^{(n)}}$'s (and analogously the $X_{S^{(n)}}$'s) are uniformly integrable.

Since $0 \le T^{(n)} \le T^{(n-1)}$ with both taking values in D_n, theorem 6.2 gives

(h) $X_{T^{(n)}} \leqq \mathrm{E}(X_{T^{(n-1)}}|\mathcal{A}_{T^{(n)}})$ a.s.

Thus

(17) $\{Y_n, \mathcal{B}_n\}_{n=-\infty}^{0}$ is a reversed s-mg, where $Y_n \equiv X_{T^{(-n)}}$ and $\mathcal{B}_n \equiv \mathcal{A}_{T^{(-n)}}$.

From this we need only the rather minor fact (since $\mathrm{E}X_0 \leqq \mathrm{E}X_{T^{(n)}}$ for this s-mg pair) that

(i) $\lim_{n\to-\infty} \mathrm{E}Y_n = \lim_{n\to-\infty} \mathrm{E}(X_{T^{(-n)}}) \geqq \mathrm{E}X_0 > -\infty.$

Thus the reversed s-mg theorem implies (g). Let $A = \Omega$ in (f) for

(j) $-\infty < \mathrm{E}X_0 \leqq \mathrm{E}X_S \leqq \mathrm{E}X_T \leqq \mathrm{E}X_\infty < \infty$

(we also apply (f) to $T_0 \equiv 0$ and $T_\infty \equiv \infty$). Then (f) and (j) finish the proof. □

Exercise 6.1 Prove theorem 6.3 (for the case of integrable X_t in theorem 6.1).

Exercise 6.2 Prove theorem 6.3 (for the corollary to theorem 6.1 case).

Exercise 6.3 Write out all the details of step (h), in the context ot theorem 6.2.

7 Applications of Optional Sampling

Example 7.1 (Gambler's ruin) Suppose $Y_1, Y_2, \ldots$ are iid with $p \equiv P(Y_1 = 1)$ and with $q \equiv P(Y_1 = -1)$. Let $S_n \equiv Y_1 + \cdots + Y_n$. Let $-a < 0 < b$ be integers, and define the stopping time

(1) $\tau \equiv \inf\{n : S_n = -a \text{ or } b\}$.

Define $\mathcal{A}_n \equiv \sigma[X_1, \ldots, X_n]$. Let $\phi(t) = pe^t + qe^{-t}$ denote the mgf of Y. Let $c_0 \equiv \log(q/p)$, and note that $\phi(c_0) = 1$. We now apply the examples of section 13.1. When $p = q = \frac{1}{2}$:

(2) S_n is a mean-0 mg,

(3) $Z_n^{(1)} \equiv S_n^2 - n$ is a mean-0 mg.

For general p and q with $p \in (0, 1)$:

(4) $Z_n^{(2)} \equiv S_n - n(p - q)$ is a mean-0 mg,

(5) $Z_n \equiv (q/p)^{S_n} = \exp(c_0\, S_n) = \exp(c_0\, S_n)/\phi^n(c_0)$ is a mean-1 mg.

We now make the claim (see exercise 7.1 below)

(6) $\mathrm{E}S_\tau = 0, \quad \mathrm{E}Z_\tau^{(1)} = 0, \quad \mathrm{E}Z_\tau^{(2)} = 0, \quad \mathrm{E}Z_\tau = 1.$

With probability 1, the rv S_τ takes on one of the values $-a$ or b. Now, $\tau \wedge m \nearrow \tau$ a.s. and $S_{\tau\wedge m} \to S_\tau$ a.s., while $(\tau \wedge m)$ is a bounded stopping time to which proposition 13.2.1 or theorem 6.1 necessarily applies. Thus, for $p = q = \frac{1}{2}$ we can conclude that

$$0 = \lim_m 0 = \lim_m \mathrm{E}S_{\tau\wedge m} = \mathrm{E}\,S_\tau = -aP(S_\tau = -a) + b[1 - P(S_\tau = -a)]$$

by proposition 13.2.1 and the DCT with dominating function $a + b$; and

$$\mathrm{E}\tau = \lim \mathrm{E}(\tau \wedge m) = \mathrm{E}S_{\tau\wedge m}^2 \to \mathrm{E}S_\tau^2 = a^2 P(S_\tau = -a) + b^2 P(S_\tau = b)$$

by the MCT, proposition 13.2.1, and the DCT. Solving these gives

(7) $P(S_\tau = -a) = b/(a + b)$ and $\mathrm{E}\tau = ab$ when $p = q = \frac{1}{2}$.

Justifying the other two equations in (6) ($Z_n^{(2)}$ is analogous to $Z_n^{(1)}$, while Z_n uses condition (13.6.11)),

(8) $$P(S_\tau = -a) = \frac{1 - (p/q)^b}{1 - (p/q)^{a+b}} \qquad \text{if } p \neq q$$

and, with $\mu \equiv p - q$,

(9) $$\mathrm{E}\,\tau = \frac{b}{\mu} - \frac{b + a}{\mu}\, \frac{1 - (p/q)^b}{1 - (p/q)^{a+b}} \qquad \text{if } p \neq q.$$

Note that if $\mu \equiv p - q < 0$, then $[\max_{0 \le n < \infty} S_n] \cong \text{Geometric}(p/q)$. That is,

(10) $P(\max_{0 \le n < \infty} S_n \ge b) = (p/q)^b$ for all integers b, when $p < q$.

(Just let $a \to \infty$ in the formula for $P(S_\tau = -a)$ to obtain the complementary probability.) □

Example 7.2 (Gambler's ruin for Brownian motion) Suppose that $\mathbb{S}_\mu$ is Brownian motion with drift: $\mathbb{S}_\mu(t) = \mathbb{S}(t) + \mu t$ for $t \geq 0$. Define the stopping time $\tau_{ab} \equiv \tau \equiv \inf\{t \geq 0 : \mathbb{S}_\mu(t) = -a \text{ or } b\}$, where $-a < 0 < b$. An easy argument will show that $\mathrm{E}\tau < \infty$. Observe first that

$$(11) \qquad \mathbb{S}_0(t), \quad \mathbb{S}_0^2(t) - t, \quad \mathbb{S}_\mu(t) - \mu t \qquad \text{are 0-mean mgs.}$$

Then set $\theta = -2\mu$, and recall (12.7.8) to conclude that

$$(12) \qquad \exp(\theta[\mathbb{S}_\mu(t) - \mu t] - \theta^2 t/2) = \exp(-2\mu[\mathbb{S}(t) + \mu t]) \qquad \text{is a mean-1 mg.}$$

Applying the optional sampling theorem to (11) and (12), we obtain

$$(13) \qquad P(\mathbb{S}(\tau) = -a) = b/(a+b) \qquad \text{if } \mu = 0\,,$$

$$(14) \qquad \mathrm{E}\tau = ab \qquad \text{if } \mu = 0\,,$$

$$(15) \qquad P(\mathbb{S}_\mu(\tau) = -a) = \frac{1 - e^{2\mu b}}{1 - e^{2\mu(a+b)}} \qquad \text{if } \mu \neq 0\,,$$

$$(16) \qquad \mathrm{E}\,\tau = \frac{b}{\mu} - \frac{a+b}{\mu}\,\frac{1 - e^{2\mu b}}{1 - e^{2\mu(a+b)}} \qquad \text{if } \mu \neq 0\,.$$

Note that if $\mu < 0$, then $\|\mathbb{S}_\mu\|_0^\infty \cong \text{Exponential}(2|\mu|)$. Let $a \to \infty$ in (14) to obtain

$$(17) \qquad P(\|\mathbb{S}_\mu\|_0^\infty \geq b) = \exp(-2\,|\mu|\,b) = \exp(-\theta\,b) \qquad \text{for all } b > 0.$$

Note the complete analogy with example 7.1. □

Exercise 7.1 Give all details in justifying the final two equalities in (6).

Exercise 7.2 Verify completely the claims of example 7.2. (recall theorem 12.7.1 and theorem 12.7.2.)

Exercise 7.3 Derive an analogue of the previous example 7.2 that is based on the Poisson process $\{\mathbb{N}(t) : t \geq 0\}$.

8 Introduction to Counting Process Martingales

Heuristic Treatment of Counting Process Martingales

Suppose now that the process $\{M(x), \mathcal{A}_x\}_{x\in R}$ is a martingale. Then for every increment $M(x+h) - M(x)$ we have $\mathrm{E}\{M(x+h) - M(x)|\mathcal{A}_x\} = 0$. Operating *heuristically*, this *suggests* that

(1) $\qquad \mathrm{E}\{dM(x)|\mathcal{A}_{x-}\} = 0 \quad$ a.s. $\qquad$ for any martingale $\{M(x), \mathcal{A}_x\}_{x\in R}$,

where $\mathcal{A}_{x-}$ is the σ-field generated by everything up to (but not including) time x. With this background, we now turn to our problem.

Suppose now that

(2) $\qquad N(x) \qquad$ is a counting process;

a *counting process* is (informally) an $\nearrow$ process that can increase only by taking jumps of size $+1$. Incremental change is modeled via

(3) $\qquad \mathrm{E}\{dN(x)|\mathcal{A}_{x-}\} = dA(x)$ a.s.; $\qquad$ here $dA(x)$ is $\mathcal{A}_{x-}$-measurable.

It then seems that

(4) $\qquad M(x) \equiv N(x) - A(x) \quad$ for $x \in R \qquad$ is a martingale;

we call $A(\cdot) = \int_{-\infty}^{x} dA(y) \equiv \int_{(-\infty,x]} dA(y)$ the *compensator* of N. Note that $A(\cdot)$ is an $\nearrow$ and $\mathcal{A}_{x-}$-measurable process.

We compute the *predictable variation process* $\langle M\rangle$ of the martingale M (as suggested by (13.5.3) or (13.5.8), and using integration by parts) via

$$
\begin{aligned}
(5)\qquad d\langle M\rangle(x) &\equiv \mathrm{E}\{dM^2(x)|\mathcal{A}_{x-}\} = \mathrm{E}\{M_-(x)\,dM(x) + M(x)\,dM(x)|\mathcal{A}_{x-}\}\\
&= \mathrm{E}\{2\,M_-(x)\,dM(x) + [dM(x)]^2|\mathcal{A}_{x-}\}\\
&= 2\,M_-(x)\,\mathrm{E}\{dM(x)|\mathcal{A}_{x-}\} + \mathrm{E}\{[dM(x)]^2|\mathcal{A}_{x-}\}\\
&= 2\,M_-(x)\cdot 0 + \mathrm{E}\{[dM(x)]^2|\mathcal{A}_{x-}\} \qquad \text{by (1)}\\
(6)\qquad &= \mathrm{E}\{[dM(x)]^2|\mathcal{A}_{x-}\} \qquad \text{as is also suggested immediately by (13.5.11)}\\
&\qquad \text{(so, the heuristics of both (13.5.8) and (13.5.11) give the same thing)}\\
&= \mathrm{E}\{[dN(x) - dA(x)]^2|\mathcal{A}_{x-}\}\\
&= \mathrm{E}\{[dN(x)]^2 - 2[dA(x)]\,[dN(x)] + [dA(x)]^2|\mathcal{A}_{x-}\}\\
&= \mathrm{E}\{[dN(x)]^2|\mathcal{A}_{x-}\} - 2\,dA(x)\,\mathrm{E}\{dN(x)|\mathcal{A}_{x-}\} + \mathrm{E}\{[dA(x)]^2|\mathcal{A}_{x-}\}\\
&= \mathrm{E}\{dN(x)|\mathcal{A}_{x-}\} - 2\,[dA(x)]^2 + [dA(x)]^2 \text{ by (3) and } [dN(x)]^2 = dN(x)\\
(7)\qquad &= dA(x) - [dA(x)]^2 \qquad \text{by (3)}\\
(8)\qquad &= [1 - \Delta A(x)]\,dA(x) \qquad \text{where } \Delta A(x) \equiv A(x) - A_-(x) \equiv A(x) - A(x-);
\end{aligned}
$$

note that $[dN(x)]^2 = dN(x)$, since $dN(x)$ takes on only the values 0 and 1. When we combine (5) = (8), it suggests that

(9) $\quad \langle M\rangle(x) = \int_{-\infty}^{x} [1 - \Delta A(y)]\, dA(y).$

Thus (note (13.5.12) and (5)), the process

$$M^2(x) - \langle M\rangle(x) \qquad \text{has} \qquad \mathrm{E}\{d[M^2(x) - \langle M\rangle(x)]|\mathcal{A}_{x-}\} = 0,$$

which suggests that, provided that each $\mathrm{E}\, M^2(x) < \infty$,

(10) $\quad M^2(x) - \langle M\rangle(x) \quad \text{for } x \in R \qquad$ is a 0-mean mg with respect to the $\mathcal{A}_x$.

Summary Starting with a martingale $M(\cdot)$ having all $\mathrm{E}M^2(x) < \infty$, it seems that

(11) $\quad \langle M\rangle(\cdot)$ is the predictable variation of the submartingale $\{M^2(x), \mathcal{A}_x\}_{t\in R}$.

That is, $\langle M\rangle(\cdot)$ is the $\nearrow$, ≥ 0, and $\mathcal{A}_{x-}$-measurable process whose existence is guaranteed by the Doob–Meyer decomposition theorem (theorem 13.**??** below). [Note that we "discovered" this without using said Doob–Meyer theorem; that theorem will merely guarantee that our heuristic guess-and-verify approach (assuming that we can make it rigorous) gives us the "right answer." This is typical.]

Consider the martingale transform

(12) $\quad W(x) \equiv \int_{-\infty}^{x} H(y)\, dM(y), \qquad$ where $H(x,\cdot)$ is $\mathcal{A}_{x-}$-measurable for all x.

Then $\mathrm{E}\{dW(x)|\mathcal{A}_x\} = \mathrm{E}\{H(x)dM(x)|\mathcal{A}_{x-}\} = H(x)\mathrm{E}\{dM(x)|\mathcal{A}_{x-}\} = 0$ by (1), so

(13) $\quad \{W(x), \mathcal{A}_x\}_{x\in R}$ is a martingale, $\qquad$ provided that each $\quad \mathrm{E}|W(x)| < \infty$.

Moreover, its predictable variation is given by

$$\begin{aligned} d\langle W\rangle(x) &= \mathrm{E}\{[dW(x)]^2|\mathcal{A}_{x-}) && \text{appealing directly to (6) this time} \\ &= \mathrm{E}\{[H(x)\, dM(x)]^2|\mathcal{A}_{x-}\} \\ &= H^2(x)\, \mathrm{E}\{[dM(x)]^2|\mathcal{A}_{x-}\} && \text{since } H(x) \text{ is } \mathcal{A}_{x-}\text{-measurable} \\ &= H^2(x)\, d\langle M\rangle(x) && \text{by (6),} \end{aligned}$$

suggesting that

(14) $\quad \langle W\rangle(x) = \int_{-\infty}^{x} H^2\, d\langle M\rangle = \int_{-\infty}^{x} H^2(y)\, [1 - \Delta A(y)]\, dA(y).$

This also suggests, provided that each $\mathrm{E}W^2(x) < \infty$, that

(15) $\quad \{L(x) \equiv W^2(x) - \langle W\rangle(x), \mathcal{A}_x\}_{x\in R}, \qquad$ is a 0-mean mg.

Processes $H(\cdot)$ that are $\mathcal{A}_{x-}$-measurable satisfy $H(x) = \mathrm{E}\{H(x)|\mathcal{A}_{x-}\}$, and so $H(x)$ can be determined by averaging $H(\cdot)$ over the past; such an H is thus called *predictable*. The martingale transform statement (12) can be summarized as

(16) $\quad \int_{-\infty}^{x} [\text{predictable}]\, d\, [\text{martingale}] = [\text{martingale}],$

provided that expectations exist.

Suppose now that we have a sequence of martingales M_n whose increments satisfy a type of Lindeberg condition; this suggests that any limiting process $M(\cdot)$ ought to be a normal process. From the martingale condition we hope that

$$\mathrm{Cov}[M(y) - M(x), M(x)] = \lim_n \mathrm{E}\{[M_n(y) - M_n(x)]\, M_n(x)\}$$

$$= \lim_n \mathrm{E}\{M_n(x)\, \mathrm{E}\{M_n(x,y]|\mathcal{A}_x\}\} = \lim_n \mathrm{E}\{M_n(x) \cdot 0\} = 0;$$

and for a normal process $M(\cdot)$, uncorrelated increments also mean independent increments. The variance process of $M(\cdot)$ should be $\mathrm{E}M^2(x) = \lim_n \mathrm{E}M_n^2(x) = \lim_n \mathrm{E}\langle M_n\rangle(x)$ by (6). So it seems reasonable to hope that [recall (12.1.15)]

(17) $\quad M_n \to_d M \cong \mathbb{S}(V) \qquad \text{on } (D_R, \mathcal{D}_R, \rho_\infty) \qquad \text{as } n \to \infty$

for a Brownian motion $\mathbb{S}$, provided that

(18) the increments of M_n satisfy a type of Lindeberg condition,

and provided that (note (9))

(19) $\quad \langle M_n\rangle(x) \to_p [\text{some } V(x)] \qquad \text{as } n \to \infty, \text{ for each } x \in R,$

where

(20) $\quad V$ is $\nearrow$ and right continuous with $V(-\infty) = 0$.

As noted above,

(21) it often holds that $\quad V(x) = \lim_n \mathrm{E}\langle M_n\rangle(x) = \lim_n \mathrm{E}\, M_n^2(x) = \mathrm{E}\, M^2(x).$

Of course, the original martingales M_n need to be square integrable. This "quasi theorem" is roughly Rebolledo's CLT.

One other bit of heuristics seems in order. Suppose now that we have several counting processes $N_i(x)$ and that we perform the above calculations and determine martingales $M_i(x) = N_i(x) - A_i(x)$ with $\langle M_i\rangle(x) = \int_{-\infty}^{x} [1 - \Delta A_i]\, dA_i$. Now, for $\mathcal{A}_{x-}$-measurable functions $c_i(\cdot)$,

(22) $\quad \mathbb{M}_n(x) \equiv \sum_{i=1}^n c_i(x)\, M_i(x) \qquad \text{is also a martingale.}$

We note from (6) that

$$d\langle\mathbb{M}_n\rangle(x) = \mathrm{E}\{[d\mathbb{M}_n(x)]^2|\mathcal{A}_{x-}\}$$

$$= \sum_{i=1}^n c_i^2(x)\ \mathrm{E}\{[dM_i(x)]^2|\mathcal{A}_{x-}\}$$

$$+ \sum\sum_{i\neq j} c_i(x)\, c_j(x)\ \mathrm{E}\{[dM_i(x)]\,[dM_j(x)]|\mathcal{A}_{x-}\}$$

(23) $\quad = \sum_{i=1}^n c_i^2(x)\, \langle M_i\rangle(x),$

provided that the

(24) $\quad M_i(x,y] \quad \text{and} \quad M_j(x,y] \quad$ are uncorrelated, given $\mathcal{A}_{x-}$.

In fact, conditions under which all of the previous heuristics are actually true are given below. Even without these, we can use these heuristics as the first step in a guess-and-verify approach.

The Guess-and-Verify Approach in a Single Sample IID as F

Example 8.1 (Single-sample martingale) Let $\tau_o \equiv F^{-1}(1)$. Suppose that

(25) $\quad N_i(x) \equiv 1_{[X_i \le x]}$ for all real x,

for $X_1, \ldots, X_n$ iid F on R. Let

$$\mathcal{A}_x \equiv \sigma[1_{[X_i \le y]} : y \le x, 1 \le i \le n] = \sigma[1_{[X_i > y]} : -\infty \le y \le x, 1 \le i \le n].$$

Then N_i is a counting process with

$$\mathrm{E}\{dN_i(x)|\mathcal{A}_{x-}\} = P(dN_i(x) = 1|N_i(x-) = 0)$$
$$= dA_i(x) \equiv 1_{[X_i \ge x]}\, d\Lambda(x),$$

where $d\Lambda(x) \equiv [1 - F_-(x)]^{-1}\, dF(x)$ with $\Delta\Lambda(x) = [\Delta F(x)]/[1 - F_-(x)] \le 1$, so

(26) $\quad M_i(x) \equiv N_i(x) - A_i(x) = N_i(x) - \int_{-\infty}^{x} 1_{[X_i \ge y]}\, d\Lambda(y)$ for all real x

satisfies (as verified in exercise 13.1.4)

(27) $\quad \{M_i(x), \mathcal{A}_x\}_{x \in R}$ is a 0-mean mg.

The predictable variation process is (letting $\int_{-\infty}^{x} \equiv \int_{(-\infty,x]}$)

$$\langle M_i \rangle(x) = \int_{-\infty}^{x} [1 - \Delta A_i(y)]\, dA_i(y)$$
$$= \int_{-\infty}^{x} [1 - 1_{[X_i \ge y]} \times \Delta\Lambda(y)] \times 1_{[X_i \ge y]}\, d\Lambda(y)$$

(28) $\quad = \int_{-\infty}^{x} 1_{[X_i \ge y]} \times [1 - \Delta\Lambda(y)]\, d\Lambda(y)$ for all real x.

Thus,

$$\mathrm{Cov}[M_i(x), M_i(y)] = \mathrm{E}\, M_i^2(x \wedge y) \qquad \text{since } M_i \text{ is a mg}$$
$$= \mathrm{E}\, \langle M_i \rangle(x \wedge y) = \mathrm{E} \int_{-\infty}^{x \wedge y} 1_{[X_i \ge t]} \times [1 - \Delta\Lambda(t)]\, d\Lambda(t)$$
$$= \int_{-\infty}^{x \wedge y} \mathrm{E}(1_{[X_i \ge t]} \times [1 - \Delta\Lambda(t)]\, d\Lambda(t)$$
$$= \int_{-\infty}^{x \wedge y} [1 - \Delta\Lambda]\, dF \qquad \text{for all real } x$$

(29) $\quad = V(x \wedge y),$

where

(30) $\quad V(x) \equiv \int_{-\infty}^{x} [1 - \Delta\Lambda]\, dF = \int_{-\infty}^{x} [(1 - F)/(1 - F_-)]\, dF\,.$

Since the sum of martingales is also a martingale, we have

(31) $\quad \mathbb{M}_n(x) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} M_i(x)$ is a 0-mean mg on R with respect to the $\mathcal{A}_x$

(32) $\quad = \sqrt{n}\, \mathbb{F}_n(x) - \int_{-\infty}^{x} \sqrt{n}\, [1 - \mathbb{F}_{n-}]\, d\Lambda$

(33) $\quad = \mathbb{U}_n(F(x)) + \int_{-\infty}^{x} \mathbb{U}_n(F_-)\, d\Lambda.$

Moreover, (23) tells us that

$$\langle \mathbb{M}_n \rangle(x) = \int_{-\infty}^{x} [1 - \mathbb{F}_{n-}] \times [1 - \Delta\Lambda]\, d\Lambda$$

$$\to_p \int_{-\infty}^{x} [1 - F_-] \times [1 - \Delta\Lambda] \times [1 - F_-]^{-1}\, dF$$

(34) $$= \int_{-\infty}^{x} [1 - \Delta\Lambda]\, dF = V(x) \quad \text{for } V \text{ as in (30)} \qquad \text{for all real } x.$$

Our heuristic Rebolledo CLT thus suggests that for a Brownian motion $\mathbb{S}$,

(35) $$\mathbb{M}_n \to_d \mathbb{S}(V) \qquad \text{on } D_{[0,\infty)}, \mathcal{D}_{[0,\infty)}, \rho_\infty);$$

see (12.1.15) for ρ_∞.

It is clear from theorem 12.10.1 that a Skorokhod embedding version $\tilde{\mathbb{M}}_n$ of the original process $\mathbb{M}_n$ should (and will) satisfy

(36) $$\|\tilde{\mathbb{M}}_n - \mathbb{M}\,\| \to_p 0 \qquad \text{as } n \to \infty,$$

where (note exercise 8.1 below)

(37) $$\mathbb{M} \equiv \mathbb{U}(F) + \int_{-\infty}^{\cdot} \mathbb{U}(F_-)\, d\Lambda \cong \mathbb{S}(V) \qquad \text{for all real } x. \ \square$$

Exercise 8.1 (a) Computing means and covariances to suggest that $\mathbb{M} \cong \mathbb{S}(V)$.
(b) Verify that $(\mathbb{Z}_n \equiv \mathbb{M}_n^2 - \langle \mathbb{M}_n \rangle, \mathcal{A}_x)_{x \in \bar{R}}$ is a u.i. 0-mean mg, and identify $\mathbb{Z}_n(\infty)$.

The Random Censorship Model

Example 8.2 (Random Censorship) Suppose now that $X_1, \ldots, X_n$ are iid nondegenerate survival times with df F on $[0,\infty)$. However, we are not able to observe the X_i's due to the following random censoring. Let $Y_1, \ldots, Y_n$ be iid censoring times whose df G is an arbitrary df on $[0,\infty]$. Suppose also that the Y_i's are independent of the X_i's. All that is observable is

(38) $$Z_i \equiv X_i \wedge Y_i \qquad \text{and} \qquad \delta_i \equiv 1_{[X_i \le Y_i]} \qquad \text{for } i = 1, \ldots, n.$$

Let $\mathcal{A}_x \equiv \sigma[\{1_{[Z_i \le y] \cap [\delta=1]}, 1_{[Z_i \le y] \cap [\delta=0]} : y \le x, 1 \le i \le n\}]$. Let the ordered values be denoted by $Z_{n:1} \le \cdots \le Z_{n:n}$ (we agree that in a tied group of $Z_{n:i}$'s, those with a δ of 1 are given smaller subscripts than those with a δ of 0), and let $\delta_{n:i}$ correspond to $Z_{n:i}$. Now, $Z_1, \ldots, Z_n$ are iid H, where $1 - H = (1-F)(1-G)$, and we let $\mathbb{H}_n$ denote the empirical df of the Z_i's. Thus,

(39) $$\mathbb{H}_n(t) \equiv \frac{1}{n} \sum_{i=1}^{n} 1_{[Z_i \le t]} \qquad \text{and} \qquad 1 - H = (1 - F)\,(1 - G)\,.$$

Let $\mathbb{H}_n^{uc}$ and H^{uc} denote the empirical df and the true df of the uncensored rvs; so

(40) $$\mathbb{H}_n^{uc}(t) \equiv \tfrac{1}{n} \textstyle\sum_{i=1}^{n} 1_{[Z_i \le t,\ \delta_i = 1]} \qquad \text{and}$$
$$H^{uc}(t) \equiv \mathrm{E} 1_{[Z_i \le t,\ \delta_i = 1]} = \int_{[0,t]} (1 - G_-)\, dF\,.$$

The basic counting process here is

(41) $$N_i(t) \equiv 1_{[Z_i \le t]}\, \delta_i \qquad \text{on } [0,\infty).$$

In analogy with exercise 13.1.4, it can be shown (see exercise 8.2 below) that, for the cumulative hazard function defined by

$$\Lambda(t) \equiv \int_0^t \frac{1}{1-F_-}\, dF = \int_0^t \frac{1-G_-}{1-H_-}\, dF \qquad \text{on } [0,\infty), \tag{42}$$

the compensator must satisfy

$$dA_i(t) \equiv \mathrm{E}\{dN_i(t)|\mathcal{A}_{t-}\} = 1_{[Z_i \ge t]}\, \frac{1-G_-(t)}{1-H_-(t)}\, dF(t) = 1_{[Z_i \ge t]}\, d\Lambda(t). \tag{43}$$

This defines the basic martingale on $[0,\infty)$ to be

$$\begin{aligned} \mathbb{M}_{1i}(t) &\equiv N_i(t) - A_i(t) = 1_{[Z_i \le t]}\, \delta_i - \textstyle\int_0^t 1_{[Z_i \ge s]}\, d\Lambda(s) \\ &= 1_{[Z_i \le t]}\, \delta_i - \Lambda(Z_i \wedge t). \end{aligned} \tag{44}$$

The predictable variation process should be

$$\langle \mathbb{M}_{1i} \rangle(t) = \textstyle\int_{[0,t]} 1_{[Z_i \ge s]} \times [1 - \Delta\Lambda(s)]\, d\Lambda(s) \qquad \text{on } [0,\infty). \tag{45}$$

Now, observe (with $\int_0^t \equiv \int_{[0,t]}$) that

$$\begin{aligned} \mathbb{M}_n(t) &\equiv \tfrac{1}{\sqrt{n}} \textstyle\sum_{i=1}^n \mathbb{M}_{1i}(t) = \sqrt{n}\,[\mathbb{H}_n^{uc}(t) - \int_0^t (1-\mathbb{H}_{n-})\, d\Lambda] \\ &= \sqrt{n}\,[\mathbb{H}_n^{uc}(t) - H^{uc}(t)] + \int_0^t \frac{\sqrt{n}\,[\mathbb{H}_{n-} - H_-]}{1-H_-}\, dH^{uc} \end{aligned} \tag{46}$$

$$= \mathbb{E}_n^{uc}(t) + \int_0^t \frac{\mathbb{E}_{n-}}{1-H_-}\, dH^{uc}, \qquad \text{on } [0,\infty), \tag{47}$$

where we have defined

$$\mathbb{E}_n^{uc} \equiv \sqrt{n}\,[\mathbb{H}_n^{uc} - H^{uc}] \qquad \text{and} \qquad \mathbb{E}_n \equiv \sqrt{n}\,[\mathbb{H}_n - H] \qquad \text{on } [0,\infty). \tag{48}$$

We would expect that a Skorokhod version $\tilde{\mathbb{M}}_n$ of $\mathbb{M}_n$ would satisfy

$$\|\tilde{\mathbb{M}}_n - \mathbb{M}\| \to_p 0 \qquad \text{as } n \to \infty, \tag{49}$$

where for appropriately defined $\mathbb{E}^{uc}$ and $\mathbb{E}$,

$$\mathbb{M} \equiv \mathbb{E}^{uc} + \int_0^{\cdot} \frac{\mathbb{E}}{1-H_-}\, dH^{uc} \qquad \text{on } [0,\infty). \tag{50}$$

Also,

$$\langle \mathbb{M}_n \rangle(t) = \textstyle\int_0^t [1 - \mathbb{H}_{n-}(s)] \times [1 - \Delta\Lambda(s)]\, d\Lambda(s) \qquad \text{on } [0,\infty), \tag{51}$$

and

$$V(t) \equiv \mathrm{E}\langle \mathbb{M}_n \rangle(t) = \textstyle\int_0^t (1-G_-)\,(1-\Delta\Lambda)\, dF = \int_0^t (1-H_-) \times [1-\Delta\Lambda]\, d\Lambda. \tag{52}$$

As would now be expected, it can be shown (see (12.1.15) for ρ_∞) that

$$\mathbb{M} \cong \mathbb{S}(V) \qquad \text{on } (D_{[0,\infty)}, \mathcal{D}_{[0,\infty)}, \rho_\infty). \tag{53}$$

The *cumulative hazard function* $\Lambda(\cdot)$ and the Aalen–Nelson *cumulative hazard function estimator* $\hat{\Lambda}_n(\cdot)$ are defined by

$$\Lambda(t) \equiv \int_0^\infty \frac{1}{1-F_-}\, dF = \int_0^t \frac{1-G_-}{1-H_-}\, dF \quad \text{and} \quad \hat{\Lambda}_n(t) \equiv \int_0^t \frac{1}{1-\mathbb{H}_{n-}}\, d\,\mathbb{H}_n^{uc} \tag{54}$$

for all $t \in [0,\infty)$. This is motivated by the deterministic halves of (39) and (40). We use this to form an estimator of the df F, that is, a type of instantaneous life table. The Kaplan–Meier *product-limit estimator* of the survival function $1-F$ is defined by

(55) $$1 - \hat{\mathbb{F}}_n(t) \equiv \prod_{s\le t}[1-\Delta\hat{\Lambda}_n(s)] = \prod_{Z_{n:i}\le t}[1-1/(n-i+1)]^{\delta_{n:i}} \quad \text{on } [0,\infty).$$

Fundamental Processes We define

(56) $$\mathbb{X}_n \equiv \sqrt{n}\,[\hat{\mathbb{F}}_n - F], \qquad \mathbb{B}_n \equiv \sqrt{n}\,[\hat{\Lambda}_n - \Lambda], \qquad \mathbb{Z}_n \equiv \sqrt{n}\,\frac{[\hat{\mathbb{F}}_n - F]}{1-F}$$

on $[0,\infty)$; and, with

(57) $$T \equiv Z_{n:n} \qquad \text{and} \qquad J_n(\cdot) \equiv 1_{[0,T]}(\cdot) = 1_{[0,Z_{n:n}]}(\cdot) = 1_{[\mathbb{H}_{n-}(\cdot)<1]}$$

(where $J_n(\cdot)$ is predictable), we further define $\tau_o \equiv H^{-1}(1)$ and

(58) $$\mathbb{X}_n^T \equiv \mathbb{X}_n(T\wedge\cdot), \qquad \mathbb{B}_n^T \equiv \mathbb{B}_n(T\wedge\cdot), \qquad \mathbb{Z}_n^T \equiv \mathbb{Z}_n(T\wedge\cdot) \qquad \text{on } [0,\infty).$$

Note that for $t\in[0,\infty)$, (54) and (46) give

$$\mathbb{B}_n^T(t) = \sqrt{n}\,[\int_0^{T\wedge t}\tfrac{1}{1-\mathbb{H}_{n-}}\,d\,\mathbb{H}_n^{uc} - \int_0^t d\Lambda]$$

$$= \sqrt{n}\int_0^{T\wedge t}\tfrac{1}{1-\mathbb{H}_{n-}}\,d\,[\mathbb{H}_n^{uc} - \int_0^t(1-\mathbb{H}_{n-})\,d\Lambda]$$

(59) $$= \int_0^t \frac{J_n}{1-\mathbb{H}_{n-}}\,d\,\mathbb{M}_n = \int_0^t[\text{predictable}]\,d[\text{martingale}] = [\text{martingale}](t)\,.\square$$

Exercise 8.2 Suggest that on $[0,\infty)$ we have

(60) $$\langle B_n^T\rangle = \int_0^t \frac{J_n}{[1-\mathbb{H}_{N-}]}[1-\Delta\Lambda]\,d\Lambda \to C(t) \equiv \int_0^t \frac{1}{1-H_-}[1-\Delta\Lambda]\,d\Lambda.$$

Exercise 8.3 (a) Verify (43).
(b) For $0\le s,t<\infty$, evaluate $\operatorname{Cov}[\mathbb{E}_n^{uc}(s),\mathbb{E}_n(t)]$.

Exercise 8.4 Use integration by parts to show that

(61) $$\mathbb{Z}_n^T(t) = \frac{\mathbb{X}_n(T\wedge t)}{1-F(T\wedge t)} = \int_0^t 1_{[0,Z_{n:n}]}\,\frac{1-\hat{\mathbb{F}}_{n-}}{1-F}\,\frac{1}{1-\mathbb{H}_{n-}}\,d\mathbb{M}_n \qquad \text{on } [0,\infty)$$

(62) $$= \int_0^t[\text{predictable}]\,d[\text{martingale}] = [\text{martingale}](t)\,.$$

Use the above heuristics to suggest (it can be proved by either mg or empirical process methods) that

(63) $$\langle\mathbb{Z}_n\rangle^T(t) = \int_0^t 1_{[0,Z_{n:n}]}\,\big[\tfrac{1-\hat{\mathbb{F}}_{n-}}{1-F}\big]^2\,\frac{1}{(1-\mathbb{H}_{n-})^2}\,d\langle\mathbb{M}\rangle_n \qquad \text{on } [0,\infty)$$

$$= \int_0^t 1_{[0,Z_{n:n}]}\,\big[\tfrac{1-\hat{\mathbb{F}}_{n-}}{1-F}\big]^2\,\frac{1}{1-\mathbb{H}_{n-}}\times[1-\Delta\Lambda]\,d\Lambda$$

(64) $$\to D(t) \equiv \int_0^t \frac{1}{1-H_-}\,\frac{1}{1-\Delta\Lambda}\,d\Lambda \qquad \text{on } [0,\infty)$$

(which is a ≥ 0, right continuous, and $\nearrow$ function), since

(65) $$[(1-F_-)/(1-F)]^2 = [1/(1-\Delta\Lambda)]^2.$$

These facts suggest that (for the $D(\cdot)$ of (64))

(66) $\quad \mathbb{Z}_n^T \to_d \mathbb{S}(D) \qquad \text{and} \qquad \mathbb{X}_n^T \to_d (1-F)\,\mathbb{S}(D) \quad \text{on } (D_{[0,\infty)}, \mathcal{D}_{[0,\infty)}, \rho_{[\infty)}).$

We now prove the corresponding result for $\mathbb{B}_n^T$.

Theorem 8.1 Let F denote an arbitrary df F on $[0,\infty)$. Then for C as in (60),

(67) $\quad \mathbb{B}_n^T \to_d \mathbb{S}(C) \qquad \text{on } (D_{[0,\infty)}, \mathcal{D}_{[0,\infty)}, \rho_{[\infty)}).$

Proof. We will show that for all $0 \le r \le s \le t < \tau_o$, we have

(68) $\quad \mathrm{E}\{\, \mathbb{B}_n^T(r,s]^{3/2}\ \mathbb{B}_n^T(s,t]^{3/2}\,\} \le 35\ C(r,s]^{3/4}\ C(s,t]^{3/4} \qquad \text{for } C \text{ as in (60)},$

so Chentsov's theorem gives relative compactness. We leave convergence of the finite-dimensional distributions to an exercise below. Define $Y_t \equiv \mathbb{B}_n^T$, and note that Y_t is of the form

(a) $\quad Y_t = \int_{[0,t]} H\,dM \qquad \text{for } H \text{ predictable and } M \text{ a mg;}$

in fact,

(b) $\quad H \equiv J_n/(1-\mathbb{H}_{n-}) \qquad \text{and} \qquad M \equiv \mathbb{M}_n\,.$

We will require two basic results. It holds that

(c) $\quad \mathrm{E}\{Y(s,t]^2\,|\mathcal{A}_s\} \le 2\,\beta_n\,C(s,t] \qquad\qquad \text{for all } 0 \le s \le t < \infty,$

where $C(\cdot)$ as defined in (60) is ≥ 0, right continuous, and $\nearrow$, and where

(d) $\quad \beta_n \equiv \|\,(1-H_-)/(1-\mathbb{H}_{n-})\,\|_0^{Z_{n:n}} \qquad \text{has} \qquad \mathrm{E}\beta_n^k \le 1 + 2\,k\,\Gamma(k+2).$

These will be established below. However, we first use them to give a proof of (67).

Now, for all $0 \le r \le s \le t < \tau_o$ we have

$$\mathrm{E}\{\,|Y(r,s]\,|^{3/2}\ |Y(s,t]\,|^{3/2}\,\} = \mathrm{E}\{\,|Y(r,s]\,|^{3/2}\ \mathrm{E}\{|Y(s,t]\,|^{3/2}\,|\mathcal{A}_s\}\,\}$$

(e) $\quad \le \mathrm{E}\{\,|Y(r,s]\,|^{3/2}\ \mathrm{E}\{Y(s,t]^2\,|\mathcal{A}_s\}^{3/4}\,\} \qquad \text{by conditional Liapunov}$

(f) $\quad \le \mathrm{E}\{\,|Y(r,s]\,|^{3/2}\,(2\,\beta_n\,C(s,t])^{3/4}\} \qquad \text{by (c)}$

(g) $\quad \le 2^{3/4}\ \ C(s,t]^{3/4}\ \mathrm{E}\,\{\,(Y(r,s]^2)^{3/4}\ \ \beta_n^{3/4}\}$

$\quad \le 2^{3/4}\ \ (\mathrm{E}\beta_n^3)^{1/4}\ C(s,t]^{3/4}\ \ (\mathrm{E}\{Y(r,s]^2\})^{3/4} \qquad \text{by Hölder's inequality}$

$\quad \le 2^{3/4}\ \ (\mathrm{E}\beta_n^3)^{1/4}\ C(s,t]^{3/4}\ \ (\mathrm{E}\{\mathrm{E}\{Y(r,s]^2\,|\mathcal{A}_r\}\})^{3/4}$

(h) $\quad \le 2^{3/4}\ (\mathrm{E}\beta_n^3\,)^{1/4}\ \ C(s,t]^{\,3/4}\ \ (\mathrm{E}\{\,2\,\beta_n\,C(r,s]\})^{3/4} \qquad \text{by (c)}$

$$= 2^{3/2}\,(\mathrm{E}\beta_n^3\,)^{1/4}\,(\mathrm{E}\beta_n)^{3/4}\;\; C(r,s]^{3/4}\,C(s,t]^{3/4}$$

(i) $\leq 2^{3/2}\,(\mathrm{E}\beta_n^3\,)^{1/2}\;\; C(r,s]^{3/4}\,C(s,t]^{3/4}$ by Liapunov's inequality

(j) $\leq 35\,C(r,s]^{3/4}\,C(s,t]^{3/4}$ by (d),

as claimed in (66).

We now establish (c). Now,

$$\mathrm{E}\{\,Y(s,t]^2\,|\mathcal{A}_s\,\} = \mathrm{E}\{\,Y_t^2 - 2Y_sY_t + Y_s^2\,|\mathcal{A}_s\,\}$$

(k) $= \mathrm{E}\{\,Y_t^2 - Y_s^2\,|\mathcal{A}_s\,\}$ since $\{Y_t, \mathcal{A}_t\}_{[t\geq 0]}$ is a mg

$$= \mathrm{E}\{\,Y_t^2 - \langle Y\rangle_t\,|\mathcal{A}_s\,\} - (Y_s^2 - \langle Y\rangle_s) + \mathrm{E}\{\langle Y\rangle_t - \langle Y\rangle_s|\mathcal{A}_s\}$$

$= \mathrm{E}\{\,\langle Y\rangle_t - \langle Y\rangle_s|\mathcal{A}_s\}$ since $\{Y_t^2 - \langle Y\rangle_t, \mathcal{A}_t\}_{[t\geq 0]}$ is a mg

(l) $= \mathrm{E}\{\int_{(s,t]} H^2\,d\langle \mathbb{M}_n\rangle\,|\mathcal{A}_s\}$ so far, holding very generally

(m) $= \mathrm{E}\{\int_{(s.t]} \frac{J_n}{(1-\mathbb{H}_{n-})^2}\,(1-\mathbb{H}_{n-}) \times [1-\Delta\Lambda]\;\, d\Lambda|\mathcal{A}_s\}$ using (58) and (50)

(n) $= \int_{(s,t]} \mathrm{E}\{\frac{n\,J_n(v)}{n\,[1-\mathbb{H}_{n-}(v)]}\,|\mathcal{A}_s\,\} \times [1-\Delta\Lambda(v)]\,d\Lambda(v)$ by Fubini

(o) $= J_n(s)\int_{(s,t]} n\,\mathrm{E}\{\frac{1_{[Z(v)>0]}}{Z(v)}\} \times [1-\Delta\Lambda(v)]\;\, d\Lambda(v)$

where $Z(v) \cong \mathrm{Binomial}(n[1-\mathbb{H}_{n-}(s)], [(1-H_-(v)]/[1-H_-(s)])$

(p) $\leq J_n(s)\displaystyle\int_{(s,t]} \frac{2n}{n[1-\mathbb{H}_{n-}(s)]}\,\frac{1-H_-(s)}{1-H_-(v)} \times [1-\Delta\Lambda(v)]\,d\Lambda(v)$

by (68) below

$$= 2J_n(s)\,\frac{1-H_-(s)}{1-\mathbb{H}_{n-}(s)}\;\int_{(s,t]} \frac{1}{1-H_-} \times [1-\Delta\Lambda]\,d\Lambda$$

(q) $\leq 2\;\beta_n\;C(s,t]$,

as claimed in (c). Step (o) used the fact that

(69) $\mathrm{E}\{Z^{-1}\,1_{[Z>0]}\} \leq 2/(mp)$ when $Z \cong \mathrm{Binomial}(m,p)$.

This is true, since $Z^{-1}1_{[Z>0]} \leq 2/(Z+1)$ and

$$\mathrm{E}(Z+1)^{-1} = \textstyle\sum_{k=0}^m \frac{1}{k+1}\,\frac{m!}{k!(m-k)!}p^k q^{m-k}$$

$$= \tfrac{1}{(m+1)p}\,\textstyle\sum_{k+1=1}^{m+1} \frac{(m+1)!}{(k+1)!((m+1)-(k+1))!}\;p^{k+1}q^{(m+1)-(k+1)}$$

(r) $\leq 1/((m+1)p) \leq 1/(mp)$.

We now turn to (d). Now, by (6.4.13) and (12.11.14) we have

$$\mathrm{E}\beta_n^k = \int_0^\infty P(\beta_n \geq x) k x^{k-1}\, dx \leq 1 + \int_1^\infty 2x^2 e^{-x} k x^{k-1}\, dx$$

(s) $\qquad \leq 1 + 2k\,\Gamma(k+2) \qquad$ for general k

(t) $\qquad \leq 145 \qquad$ for $k = 3$,

as claimed in (d). □

Exercise 8.5 Complete the proof of theorem 8.1, by showing convergence of the finite-dimensional distributions.

Exercise 8.6 Show that (69) may be extended to give

(70) $\quad \mathrm{E}\{\frac{1}{Z^k}\, 1_{[Z>0]}\} \leq k(k+1)/(mp)^k \qquad$ when $Z \cong \text{Binomial}(m, p)$, and $k \geq 1$.

Remark 8.1 At this point, three sections of the 1*st* Edition have been entirely omitted. The first two contained a general form of the Doob–Meyer decomposition for continuous parameter martingales followed by a treatment of martingales of the form $\int H\, d\mathbb{M} = \int [\,\text{predictable}\,]\, \mathrm{d}[\,\text{martingale}\,]$. The final omitted section treated the basic censored data martingale. Together, these form the continuous analog to the current section 13.5 and section 13.8. □

9 CLTs for Dependent RVs

Let $\{X_{nk} : k = 1, 2, \ldots$ and $n = 1, 2, \ldots\}$ be an array of rvs on a basic probability space $(\Omega, \mathcal{A}, P)$. For each n we suppose that $X_{n1}, X_{n2}, \ldots$ are adapted to an $\nearrow$ sequence of σ-fields $\mathcal{A}_{n0} \subset \mathcal{A}_{n1} \subset \mathcal{A}_{n2} \subset \cdots \subset \mathcal{A}$. For each n we suppose that κ_n in an integer-valued stopping time with respect to these $(\mathcal{A}_{nk})_{k=0}^{\infty}$. We now introduce

(1)
$$P_{k-1}(\cdot) \equiv P(\cdot\,|\mathcal{A}_{n,k-1}),$$
$$\mathrm{E}_{k-1}(\cdot) \equiv \mathrm{E}(\cdot\,|\mathcal{A}_{n,k-1}), \qquad \text{and} \qquad \mathrm{Var}_{k-1}[\cdot] \equiv \mathrm{Var}[\cdot\,|\mathcal{A}_{n,k-1}].$$

Our interest is in the sum

(2) $\quad S_n \equiv \sum_{k=1}^{\kappa_n} X_{nk}\,.$

We will have reason to consider

(3) $\quad X'_{nk} \equiv X_{nk}\,1_{[\,|X_{nk}|\le 1]} \qquad \text{and} \qquad X''_{nk} \equiv X_{nk}\,1_{[\,|X_{nk}|>1]}\,.$

What follows is the most basic CLT in this monograph. For row-independent rvs and $\kappa_n \equiv n$ it reduces to the asymptotic normality condition of theorem **??.??**, with $c = 1$. The second theorem implies the first, and is very much in the spirit of the Lindeberg theorem.

Theorem 9.1 (Basic dependent CLT) Conclude that $S_n \to_d N(0,1)$ if

(4)
$$\sum_{k=1}^{\kappa_n} P_{k-1}(|X_{nk}| \ge \epsilon) \to_p 0 \qquad \text{for all } \epsilon > 0$$
$$(\text{equivalently,} \quad \max_{1\le k\le \kappa_n} |X_{nk}| \to_p 0),$$

(5) $\quad \sum_{k=1}^{\kappa_n} \mathrm{E}_{k-1}(X'_{nk}) \to_p 0 \qquad$ (i.e., partial sums of the X_{nk} are nearly a mg),

(6) $\quad \sum_{k=1}^{\kappa_n} \mathrm{Var}_{k-1}[X'_{nk}] \to_p 1\,.$

Theorem 9.2 Conclude that $S_n \to_d N(0,1)$ if

(7) $\quad \sum_{k=1}^{\kappa_n} \mathrm{E}_{k-1}(X_{nk}^2\,1_{[\,|X_{nk}|\ge\epsilon]}) \to_p 0 \qquad$ for all $\epsilon > 0$,

(8) $\quad \sum_{k=1}^{\kappa_n} \mathrm{E}_{k-1}(X_{nk}) \to_p 0\,,$

(9) $\quad \sum_{k=1}^{\kappa_n} \mathrm{Var}_{k-1}[X_{nk}] \to_p 1\,.$

Comments If one replaces (5) by

(10) $\quad \sum_{k=1}^{\kappa_n} |\,\mathrm{E}_{k-1}(X'_{nk})|^2 \to_p 1\,,$

then (6) may be replaced by

(11) $\quad \sum_{k=1}^{\kappa_n} \mathrm{E}_{k-1}({X'_{nk}}^2) \to_p 1\,.$

If (4) holds, then (11) is equivalent to

(12) $\quad \sum_{k=1}^{\kappa_n} X_{nk}^2 \to_p 1\,.$

If $X_{n1}, X_{n2}, \ldots$ are mg differences, then (10) and (5) are implied by

(13) $\sum_{k=1}^{\kappa_n} |\, \mathrm{E}_{k-1}(X_{nk}\, 1_{[\,|X_{nk}|>1]})| \to_p 0\,.$

Moreover, (13) is implied by either (7) or

(14) $\mathrm{E}(\max_{1\le k\le \kappa_n} |X_{nk}|) \to_p 0\,.$

We now summarize these last claims.

Theorem 9.3 (MG CLT) Suppose the $X_{n1}, X_{n2}, \ldots$ are mg differences, for each n. Then $S_n \to_d N(0,1)$, provided that any of the following occurs:

(15) Conditions (7) and (9) hold,

(16) Conditions (12) and (14) hold,

(17) Conditions (4) , (12), and (13) hold.

Chapter 14

Convergence in Law on Metric Spaces

1 Convergence in Distribution on Metric Spaces

Many results for convergence in distribution generalize to probability measures on a general metric space (M, d) equipped with its Borel σ-field $\mathcal{M}_d$. We call such measures *Borel measures.* Instead of using convergence of dfs to define convergence in distribution (or law), we use directly the Helly–Bray idea embodied in the $\to_d$ equivalences of theorem 9.1.4, and that theorem is here extended to general metric spaces. Skorokhod's construction is also generalized to complete and separable metric spaces. Section 12.1 gave specific information on the two very important metric spaces that gave rise to $(C, \mathcal{C})$ and $(D, \mathcal{D})$. In section 14.2 the dual bounded Lipschitz metric is introduced, along with Hellinger, Prohorov, and total variation metrics. These are useful on function spaces.

Definition 1.1 (Convergence in distribution) If $\{P_n : n \geq 1\}$ and P are probability measures on $(M, d, \mathcal{M}_d)$ satisfying

$$(1) \qquad \int g\, dP_n \to \int g\, dP \qquad \text{for all } g \in C_b(M)$$

[where $C_b(M) \equiv$ {all bounded and d-continuous functions g from M to R}, and $C_{bu}(M)$ denotes those functions that are additionally d-uniformly continuous], then we say that P_n *converges in distribution* (or *law*) to P, or that P_n *converges weakly* to P; and we write $P_n \to_d P$ or $P_n \to_L P$. Similarly, if X_n, X are random elements in M for which

$$(2) \qquad \mathrm{E}\, g(X_n) \to \mathrm{E}\, g(X) \qquad \text{for all } g \in C_b(M),$$

then we write $X_n \to_d X$, $X_n \to_{\mathcal{L}} X$ or $\mathcal{L}(X_n) \to \mathcal{L}(X)$.

Theorem 1.1 (Portmanteau theorem; Billingsley) For probability measures $\{P_n : n \geq 1\}$ and P on any metric space $(M, d, \mathcal{M}_d)$ the following are equivalent:

(3) $\quad P_n \to_d P \qquad$ [i.e., $\int g\, dP_n \to \int g\, dP$ for all $g \in C_b(M)$].

(4) $\quad \int g\, dP_n \to \int g\, dP \qquad$ for all $g \in C_{bu}(M)$.

(5) $\quad \overline{\lim}\, P_n(B) \leq P(B) \qquad$ for all closed sets $B \in \mathcal{M}_d$.

(6) $\quad \underline{\lim}\, P_n(B) \geq P(B) \qquad$ for all open sets $B \in \mathcal{M}_d$.

(7) $\quad \lim P_n(B) = P(B) \qquad$ for all P-continuity sets $B \in \mathcal{M}_d$.

Proof. Clearly, (3) implies (4).

Consider (4) implies (5): Suppose that (4) holds and that B is closed. Let $\epsilon > 0$. Then for integral m large enough, the set $B_m \equiv \{x : d(x, B) < 1/m\}$ satisfies

(a) $\quad P(B_m) \leq P(B) + \epsilon,$

since $B_m \searrow B$ as $m \to \infty$. Let $g_m(x) \equiv \psi(md(x, B)) = \max\{0, (1 - m\, d(x, B))\}$, where $\psi(t)$ is equal to 1, $1 - t$, 0 according as t has $t \leq 0$, $0 \leq t \leq 1$, $1 \leq t$. Then

(b) $\quad 1_B \leq g_m \leq 1_{B_m},$

and for each $m \geq 1$, g_m is Lipschitz and uniformly continuous. Hence, by (4) and also (a) and (b),

(c) $\quad \overline{\lim}_n\, P_n(B) \leq \lim_n \int g_m\, dP_n = \int g_m\, dP \leq P(B_m) \leq P(B) + \epsilon.$

Since $\epsilon > 0$ was arbitrary, (5) follows.

Equivalence of (5) and (6) follows easily by taking complements.

Consider (5) implies (3): Suppose that $g \in C_b(M)$ and that (5) holds. Now, transform g linearly so that $0 \leq g(x) \leq 1$. Fix $k \geq 1$, and define the closed set

(d) $\quad B_j \equiv \{x \in M : j/k \leq g(x)\} \qquad$ for $\quad j = 0, \ldots, k+1.$

Then it follows that

$$\sum_{j=1}^{k+1} \frac{j-1}{k}\, P(x : \tfrac{j-1}{k} \leq g(x) < \tfrac{j}{k})$$

(e) $$\leq \int g\, dP < \sum_{j=1}^{k+1} \frac{j}{k}\, P(x : \tfrac{j-1}{k} \leq g(x) < \tfrac{j}{k}).$$

Rewriting the sum on the right side and summing by parts gives

(f) $\quad \sum_{j=1}^{k} (j/k)\; [P(B_{j-1}) - P(B_j)] = (1/k) + (1/k) \sum_{j=1}^{k} P(B_j),$

which together with a similar summation by parts on the left side yields

(g) $\quad (1/k) \sum_{j=1}^{k} P(B_j) \leq \int g\, dP \leq (1/k) + (1/k) \sum_{j=1}^{k} P(B_j).$

Applying the right side of (g) to P_n, and then using (5) for the closed sets B_j, and then applying the left side of (g) to P gives

$$\overline{\lim}_n \int g\, dP_n \le \overline{\lim}_n \left[\tfrac{1}{k} + \tfrac{1}{k}\sum_{j=1}^k P_n(B_j)\right]$$

(h) $$\le \left[\tfrac{1}{k} + \tfrac{1}{k}\sum_1^k P(B_j)\right] \le \tfrac{1}{k} + \int g\, dP\ .$$

Letting $k \to \infty$ in (h) yields

(i) $$\overline{\lim}_n \int g\, dP_n \le \int g\, dP.$$

Applying (i) to (the nontransformed) $-g$ yields

(j) $$\underline{\lim}_n \int g\, dP_n \ge \int g\, dP.$$

Combining (i) and (j) gives (3).

Consider (5) implies (7): With B^0 the interior of any set $B \in \mathcal{M}$ and $\bar{B}$ its closure, (5) and (6) give

(k) $$P(B^0) \le \underline{\lim}\, P_n(B^0) \le \underline{\lim}\, P_n(B) \le \overline{\lim}\, P_n(B) \le \overline{\lim}\, P_n(\bar{B}) \le P(\bar{B}).$$

If B is a P-continuity set, then $P(\partial B) = 0$ and $P(\bar{B}) = P(B^0)$, so the extreme terms in (k) are equal; thus $\lim P_n(B) = P(B)$, as required by (7).

Consider (7) implies (5): Since $\partial\{x : d(x,B) \le \delta\} \subset \{x : d(x,B) = \delta\}$, the boundaries are disjoint for different $\delta > 0$, and hence at most countably many of them can have positive P-measure. Therefore, for some sequence $\delta_k \to 0$, the sets $B_k \equiv \{x : d(x,B) < \delta_k\}$ are P-continuity sets and $B_k \searrow B$ if B is closed. It follows from $B \subset B_k$ and then (7)

(l) $$\overline{\lim}\, P_n(B) \le \overline{\lim}\, P_n(B_k) = P(B_k).$$

Then (5) follows from the monotone property of P, since $B_k \searrow B$ as $k \to \infty$. □

Proposition 1.1 $P_n \to_d P$ if and only if each subsequence $\{P_{n'}\}$ contains a further subsequence $\{P_{n''}\}$ such that $P_{n''} \to_d P$.

Proof. This is easy from definition 1.1 (and the fact that a sequence of real numbers has $x_n \to x$ if and only if each $\{x_{n'}\}$ contains a further subsequence $\{x_{n''}\}$ such that $x_{n''} \to x$), as in the corollary to Helly's selection in theorem 9.1.3. □

Theorem 1.2 (Slutsky's theorem) Suppose that X_n, Y_n are random elements taking values in a separable metric space $(M, d, \mathcal{M}_d)$, both defined on some Ω_n.

(a) Show that $d(X_n, Y_n)$ is a rv whenever (M, d) is separable.

(b) If $X_n \to_d X$ and $d(X_n, Y_n) \to_p 0$, then $Y_n \to_d X$.

Proof. For a closed set B and $\delta > 0$ let $B_\delta \equiv \{x : d(x, B) < \delta\}$. Then

(a) $P(Y_n \in B) = P(Y_n \in B,\, d(X_n, Y_n) < \delta) + P(Y_n \in B,\, d(X_n, Y_n) \ge \delta)$

(b) $\le P(X_n \in \bar{B}_\delta) + P(d(X_n, Y_n) \ge \delta)$.

The second term on the right side of (a) goes to zero, since $d(X_n, Y_n) \to_p 0$ (note the following exercise). Then $X_n \to_d X$ gives

(c) $\overline{\lim}\, P(Y_n \in B) \le \overline{\lim}\, P(X_n \in \bar{B}_\delta) \le P(X \in \bar{B}_\delta)$, for every $\delta > 0$,

via the portmanteau theorem (5) applied to the X_n's. Then $\bar{B}_\delta \searrow B$ as $\delta \searrow 0$, since B is closed, so $P(X \in \bar{B}_\delta) \searrow P(X \in B)$. Thus $\overline{\lim}\, P(Y_n \in B) \le P(X \in B)$; thus $Y_n \to_d X$ follows from applying the portmanteau theorem to the Y_n's. □

Exercise 1.1 Prove theorem 1.2(a). (Recall proposition 2.2.4.)

Theorem 1.3 (Continuous mapping theorem) Let $X_n \to_d X$ on $(M, d, \mathcal{M}_d)$, and suppose $g : M \to \bar{M}$ (where $(\bar{M}, \bar{d})$ is another metric space) is continuous a.s. with respect to $P \equiv P_X$ (that is, $P(X \in C_g) = 1$ for the continuity set C_g of g). Then, necessarily, $g(X_n) \to_d g(X)$.

Proof. We simply note that this is essentially no different from the proof of the Mann–Wald theorem. (Only now we apply the general Skorokhod construction of the following theorem, instead of the elementary Skorokhod theorem.) [This proof, however, requires that the metric space be complete and separable. The next exercise asks the reader to provide a general proof.] □

Exercise 1.2 Prove the continuous mapping theorem above by appeal to (5) and then proposition 2.2.4 (that the discontinuity set of a transformation between metric spaces is always measurable). This proof requires neither completeness nor separability! (And this is actually a much more elementary proof.)

Skorokhod's Construction

We earlier established the elementary form of Skorokhod's theorem: If random variables $X_n \to_d X_0$, then there exist random variables $\{Y_n : n \ge 0\}$ defined on a common probability space satisfying $Y_n \cong X_n$ for all $n \ge 0$ and $Y_n \to_{a.s.} Y_0$. That proof relied on the inverse transformation. We now turn to the extension of the elementary Skorokhod theorem from R to a complete and separable metric space (M, d), whose open sets generate the Borel σ-field. [The first step in the proof will be to establish a preliminary result for just one P on $(M, d, \mathcal{M}_d)$.]

Theorem 1.4 (Skorokhod construction) Suppose $(M, d, \mathcal{M}_d)$ is a complete and separable metric space and the measures $\{P_n : n \ge 0\}$ satisfy $P_n \to_d P_0$. Then there exist random elements $\{X_n : n \ge 0\}$ taking values in the space M (thus $X_n(\omega) \in M$ for all $\omega \in \Omega$) and all defined on the common probability space $(\Omega, \mathcal{A}, P) \equiv ([0,1], \mathcal{B}([0,1]), \text{Lebesgue})$, with $X_n \cong P_n$ and satisfying

(8) $d(X_n(\omega), X_0(\omega)) \to 0$ for each $\omega \in \Omega$.

Proposition 1.2 Suppose P is a probability measure on $(M, d, \mathcal{M}_d)$. Then there is a random element X defined on $(\Omega, \mathcal{A}, P) = ([0,1], \mathcal{B}[0,1], \text{Lebesgue})$ and taking values in M that has distribution P.

Proof. For each k, decompose M into a countable number of disjoint sets $A_{k1}, A_{k2}, \ldots$ whose diameter is less than $1/k$. Then arrange it so that $\mathcal{A}_{k+1}$ refines $\mathcal{A}_k \equiv \{A_{k1}, A_{k2}, \ldots\}$. Make a corresponding decomposition of the unit interval as $\mathcal{I}_k \equiv \{I_{k1}, I_{k2}, \ldots\}$ where the subintervals I_{kj} satisfy $P(A_{kj}) = \lambda(I_{kj})$ and where the decomposition $\mathcal{I}_{k+1}$ refines $\mathcal{I}_k$.

Let x_{kj} be a point in A_{kj}, and define

(a) $\quad X_k(\omega) = x_{kj} \qquad \text{if } \omega \in I_{kj} \subset [0,1].$

Since $\{X_k(\omega), X_{k+1}(\omega), \ldots\} \subset$ (some one A_{kj}), its diameter is bounded by $1/k$. Thus, $\{X_k(\omega)\}$ is Cauchy for each ω, $\lim_k X_k(\omega) \equiv X(\omega)$ exists, and

(9) $\quad d(X(\omega), X_k(\omega)) \le 1/k \to 0 \qquad \text{as } k \to \infty.$

For a given set B, write $\sum^* \equiv \sum_{\{j: A_{jk} \cap B \neq \emptyset\}}$, and similarly for unions of sets. Then

$$P(X_k \in B) \le P(X_k \in \textstyle\bigcup^* A_{kj}) = \sum^* P(X_k \in A_{kj})$$

$$= \textstyle\sum^* \lambda(I_{kj}) = \sum^* P(A_{kj})$$

(b) $$\le P(\overline{B^{1/k}}),$$

where $B^\delta \equiv \{x : d(x, B) \equiv \inf_{y \in B} d(x,y) < \delta\}$. For a closed set B we have $\bigcap_{k=1}^\infty \overline{B^{1/k}} = B$, and so

(c) $\quad \overline{\lim}_k P(X_k \in B) \le \overline{\lim}_k P(X_k \in \overline{B^{1/k}}) \le P(B) \qquad \text{for closed sets } B,$

and hence the distribution of X_k converges to P by (5) of the portmanteau theorem. It follows from Slutsky's theorem (with $Y_k = X$ for all k) that $X \cong P$. □

Proof. Consider Skorokhod's theorem. First construct the decompositions $\mathcal{A}_k$ of the proof of the previous proposition, but now do it in a way that makes each A_{kj} a P-continuity set. Because $\partial\{y : d(x,y) < \delta\} \subset \{y : d(y,x) = \delta\}$, the spheres about x are P-continuity sets for all but countably many radii; hence M can be covered by countably many P-continuity sets all with diameter at most $1/k$. The usual disjointification procedure preserves P-continuity because $\partial(B \cap C) \subset (\partial B) \cup (\partial C)$.

Consider the decompositions $\mathcal{I}_k$ as before, and, for each n, construct successively finer partitions $\mathcal{I}_k^{(n)} = \{I_{k1}^{(n)}, I_{k2}^{(n)}, \ldots\}$ with $\lambda(I_{kj}^{(n)}) = P_n(A_{kj})$. Inductively arrange the indexing so that $I_{ki}^{(n)} < I_{kj}^{(n)}$ if and only if $I_{ki} < I_{kj}$; here $I < J$ for intervals I and J means that the right endpoint of I does not exceed the left endpoint of J. In other words, we ensure that for each k the families $\mathcal{I}_k, \mathcal{I}_k^{(1)}, \mathcal{I}_k^{(2)}, \ldots$ are ordered similarly.

Define X_k as before, where $x_{kj} \in A_{kj}$, and define

(a) $\qquad X_k^{(n)}(\omega) = x_{kj} \qquad \text{if } \omega \in I_{kj}^{(n)}.$

Again $X_k(\omega)$ converges to an $X(\omega)$ satisfying (9), and $X_k^{(n)}(\omega)$ converges to an $X^{(n)}(\omega)$ satisfying

(b) $\qquad d(X^{(n)}(\omega), X_k^{(n)}(\omega)) \le 1/k \to 0 \qquad \text{as } k \to \infty.$

And again X has distribution P and $X^{(n)}$ has distribution P_n.

Since $\sum_j [P(A_{kj}) - P_n(A_{kj})] = 0$, it follows that

(c) $\qquad \sum_j |\lambda(I_{kj}) - \lambda(I_{kj}^{(n)})| = \sum_j |P(A_{kj}) - P_n(A_{kj})|$

$\qquad\qquad = 2 \sum_j'' [P(A_{kj}) - P_n(A_{kj})]$

(d) $\qquad\qquad = 2 \sum_j [P(A_{kj}) - P_n(A_{kj})]^+,$

where the next to the last sum extends over those j for which the summand is positive. Each summand goes to 0 as $n \to \infty$ because the A_{kj} are P-continuity sets, and it follows by the DCT (with dominating function identically equal to the constant function with value 1) that

(e) $\qquad \lim_n \sum_j |\lambda(I_{kj}) - \lambda(I_{kj}^{(n)})| = 0.$

Fix k and j_0, let α and α_n be the left endpoints of I_{kj_0} and $I_{kj_0}^{(n)}$, respectively, and let $\sum'$ indicate summation over the set for which $I_{kj} < I_{kj_0}$, which is the same as the set for which $I_{kj}^{(n)} < I_{kj_0}^{(n)}$. Then (d) implies

(f) $\qquad \alpha = \sum_j' \lambda(I_{kj}) = \lim_n \sum_j' \lambda(I_{kj}^{(n)}) = \lim_n \alpha_n\,.$

Similarly, the right endpoint of the interval $I_{kj}^{(n)}$ converges as $n \to \infty$ to the right endpoint of the interval I_{kj}.

Hence, if ω is interior to I_{kj} (which now fixes k and j), then ω lies in $I_{kj}^{(n)}$ for all n large enough, so that $X_k(n)(\omega) = x_{kj} = X_k(\omega)$ for all $n \ge$ (some $n_{k,j,\omega}$), andthe conclusions (9) and (b) give

(g) $\qquad d(X(\omega), X^{(n)}(\omega)) \le 2/k \qquad \text{for all } m \ge n_{k,j,\omega}.$

Thus, if ω is not an endpoint of any I_{kj}, then for each k we have that (g) holds for all sufficiently large n. In other words, $\lim_n X^{(n)}(\omega) = X(\omega)$ if ω is not in the set of endpoints of the I_{kj}. This last set, being countable, has Lebesgue measure 0; thus if $X^{(n)}(\omega)$ is redefined as $X(\omega)$ on this set, $X^{(n)}$ still has distribution P_n and there is now convergence for all ω. This completes the proof, with $X^{(n)}$ denoting X_n and with X denoting X_0. □

Exercise 1.3 Recall the partial sum process $\mathbb{S}_n = \{\mathbb{S}_n(t) : t \ge 0\}$ defined earlier. Now, $\mathbb{S}_n \to_d \mathbb{S}$ by Donsker's theorem (theorem 12.8.3), where $\mathbb{S}$ is a Brownian motion process in $M = C[0,1]$. Consider the following four functions:

(a) $g(x) = \sup_{0 \le t \le 1} x(t)$,
(b) $g(x) = \int_0^1 x(t)\,dt$,
(c) $g(x) = \lambda(\{t \in [0,1] : x(t) > 0\})$, where λ denotes Lebesgue measure,
(d) $g(x) = \inf\{t > 0 : x(t) = b\}$, with $b > 0$ fixed.

For each of these real-valued functionals g of $x \in C[0,1]$, find the discontinuity set D_g of g. [If we can show that the P measure of these discontinuity sets is zero, where P denotes the measure of $\mathbb{S}$ on $C[0,1]$, then it follows immediately from the continuous mapping theorem that $g(\mathbb{S}_n) \to_d g(\mathbb{S})$.]

Exercise 1.4 Let $S_0 \equiv 0, S_1 \equiv X_1, S_k \equiv X_1 + \cdots + X_k$ for $k \ge 1$ be the partial sums of the iid $(0,1)$ rvs that go into the definition of $\mathbb{S}_n$. Represent $n^{-3/2} \sum_{k=1}^n |S_k|$ as a functional $g(\mathbb{S}_n)$. Is the resulting g continuous?

Tightness and Relative Compactness

The notion of tightness comes into play in a crucial way in the general theory of convergence in distribution on a metric space (M,d), since there now are more ways to "leave" the space than simply for mass to drift off to infinity.

Definition 1.2 (Tightness) Let $\mathcal{P}_0$ denote a collection of probability measures on $(M, d, \mathcal{M}_d)$. Then $\mathcal{P}_0$ is *tight* (or *uniformly tight*) if and only if for every $\epsilon > 0$ there is a compact set $K_\epsilon \subset M$ with

(10) $\quad P(K_\epsilon) > 1 - \epsilon \qquad$ for all $P \in \mathcal{P}_0$.

Definition 1.3 (Sequential compactness, or relative compactness) Let $\mathcal{P}_0$ be a family of probability measures on $(M, d, \mathcal{M}_d)$. We call $\mathcal{P}_0$ *relatively compact* (or *sequentially compact*) if every sequence $\{P_n\} \subset \mathcal{P}_0$ contains a weakly convergent subsequence. That is, every sequence $\{P_n\} \subset \mathcal{P}_0$ contains a subsequence $\{P_{n'}\}$ with $P_{n'} \to_d$ (some probability Q) (not necessarily in $\mathcal{P}_0$).

Proposition 1.3 Let $(M, d, \mathcal{M}_d)$ be a separable metric space.
(a) If $\{P_n\}_{n=1}^\infty$ is relatively compact with limit set $\{P\}$, then $P_n \to_d P$.
(b) If $P_n \to_d P$, then $\{P_n\}_{n=1}^\infty$ is relatively compact.
(c) We have thus related convergence in distribution to relative compactness.

Proof. See proposition 1.1 for both(a) and (b). (That is, we have merely rephrased things we already know.) □

Theorem 1.5 (Prohorov) Let $\mathcal{P}_0$ denote a collection of probability measures on the metric space $(M, d, \mathcal{M}_d)$.
(a) If $\mathcal{P}_0$ is tight, then it is relatively compact.
(b) Suppose that $(M, d, \mathcal{M}_d)$ is separable and complete. If the collection $\mathcal{P}_0$ is relatively compact, then it is tight.
(c) We have thus related relative compactness to tightness, at least on complete and separable metric spaces.

Proof. A full proof progresses from $M = R_k$ to R_∞, to sigma compact M, and finally to general M, at each step using the next simpler case. We present only the proof of (a) for the case $M = R_k$.

If $\{P_n\}$ is any sequence in $\mathcal{P}_0$, then Helly's selection theorem implies that the corresponding sequence of dfs $\{F_n\}$ defined by $F_n(\mathsf{x}) \equiv P_n(-\infty, \mathsf{x}]$ contains a subsequence $\{F_{n'}\}$ satisfying

(p) $\quad F_{n'}(\mathsf{x}) \to F(\mathsf{x}) \qquad$ for all $\mathsf{x} \in C_F$,

where the sub df F is continuous from above. Now, there is a measure P on R_k such that $P(\mathsf{a}, \mathsf{b}]$ equals F differenced around the vertices of the k-dimensional rectangle $(\mathsf{a}, \mathsf{b}]$. Now $P_{n'} \to_d P$ will follow if we can show that $P(R_k) = 1$.

Given $\epsilon > 0$, choose $K \subset R_k$ compact with $P_{n'}(K) > 1 - \epsilon$ for all n'; this is possible by tightness of $\mathcal{P}_0$. Now choose a, b in R_k such that $K \subset (\mathsf{a}, \mathsf{b}]$ and all 2^k vertices of $(\mathsf{a}, \mathsf{b}]$ are continuity points of F (we can do this because at most a countable number of parallel $(k-1)$-dimensional hyperplanes can possibly have positive P-measure. Since $P_{n'}(\mathsf{a}, \mathsf{b}]$ equals $F_{n'}$ differenced around $(\mathsf{a}, \mathsf{b}]$, (a) yields $P_{n'}(\mathsf{a}, \mathsf{b}] \ge P_{n'}(K) \ge 1 - \epsilon$, so $P(\mathsf{a}, \mathsf{b}] \ge 1 - \epsilon$. Since ϵ was arbitrary, $P(R_k) = 1$. Hence P is a probability measure, $P_{n'} \to_d P$, and $\mathcal{P}_0$ is relatively compact. □

Convergence in Distribution on $(D, \mathcal{D})$

We phrase results carefully in this subsection, so as to mention primarily the familiar metric $\|\cdot\|$ on D (while limiting mention of the contrived metric d of exercise 12.1.4 for which (D, d) is both complete and separable with $\mathcal{D}_d = \mathcal{D}$). Recall that $\mathcal{D}$ denotes the finite-dimensional σ-field.

Theorem 1.6 (Criterion for $\to_d$ on $(D, \mathcal{D})$; Chentsov) Let each X_n denote a process on $(D, \mathcal{D})$. Suppose that for some $a > \frac{1}{2}$ and $b > 0$ the increments of the X_n processes satisfy

(11) $\quad \mathrm{E}\,|X_n(r,s]\,X_n(s,t]\,|^b \le [\,\mu_n(r,s] \times \mu_n(s,t]\,]^a \qquad$ for all $0 \le r \le s \le t \le 1$

for some finite measure μ_n on the Borel subsets of $[0,1]$. (Often, $a = 1$ and $b = 2$, and note the relaxation of this condition in the remark below.) Suppose that μ is a continuous measure on the Borel subsets of $[0,1]$, and that either

(12) $\quad \mu_n(s,t] \le \mu(s,t] \qquad$ for all $0 \le s \le t \le 1$ and for all $n \ge 1 \qquad$ or

(13) $\quad \mu_n/\mu_n([0,1]) \to_d \mu/\mu([0,1]) \qquad$ as $n \to \infty$.

(a) Then for the metric d of exercise 12.1.4 we have that

(14) $\quad \{X_n : n \ge 1\}$ is relatively compact on (D, d).

(b) Especially, if we further have $X_n \to_{f.d.}$ (some X) and if $P(X \in C) = 1$), then

(15) $\quad g(X_n) \to_d g(X)$ for all $\mathcal{D}$-measurable and a.s. $\|\cdot\|$-continuous $g : D \to R$.

Remark 1.1 For processes like the partial sum process $\mathbb{S}_n$ the condition (11) is troublesome; but since $\mathbb{S}_n$ is constant on the intervals $[(i-1)/n, i/n)$, it should be enough to verify (11) for r, s, t restricted to be of the form i/n. We now make this rough idea precise.

For $m \geq 1$ we let $T_m \equiv \{0 \equiv t_{m0} < t_{m1} < \cdots < t_{mk_m} \equiv 1\}$, and measure the coarseness of this partition by defining $\text{mesh}(T_m) \equiv \max\{t_{mi} - t_{m,i-1} : 1 \leq i \leq k_m\}$. Let $x \in D$ and let $A_m(x)$ denote the function in D that equals $x(t_{mi})$ at t_{mi} for each $0 \leq i \leq k_m$ and that is constant in between these points. We agree to call $A_m(x)$ the T_m-*approximation* of x. Suppose now that one of the two conditions (12) or (13) holds, that $X_n = A_n(X_n)$ with $\text{mesh}(T_n) \to 0$ as $n \to \infty$, and that (11) holds for all r, s, t in T_n. Then both (14) and (15) hold. [That $X_n = A_n(X_n)$ means that X_n is equal to its own T_n-approximation; and this clearly holds for $\mathbb{S}_n$ when all $t_{ni} = i/n$.] □

"Proof." By analogy with Helly's selection theorem, it should suffice to show the tightness. Thus, for each $\epsilon > 0$ we must exhibit a compact set K_ϵ of functions on $[0, 1]$ for which $P(X_n \in K_\epsilon) > 1 - \epsilon$ for all n. According to Arzelà's theorem (see exercise **??**.**??**), a compact set of functions consists of a uniformly bounded set of functions that have a uniform bound on their "wiggliness." A complex and delicate argument (slightly in the spirit of Kolmogorov's inequality) based on (6) can be given to bound this "wiggliness." Since the details are long and hard and are only used for the present theorem, they will be skipped. See Billingsley (1968). □

Exercise 1.5 (Prohorov) For any real-valued function on $[0, 1]$ we now define the *modulus of continuity* $\omega_\delta(x)$ of x by $\omega_\delta(x) \equiv \max\{|x_t - x_s| : |t - s| \leq \delta\}$ for each $\delta > 0$. Let $X, X_1, X_2, \ldots$ denote processes on $(C, \mathcal{C})$. Then $X_n \to_d X$ on $(C, \|\cdot\|)$ if and only if both $X_n \to_{f.d.} X$ and $\lim_{\delta \to 0} \limsup_{n\to\infty} P(\omega_\delta(X_n) > \epsilon) = 0$ for all $\epsilon > 0$. (The modulus of continuity also measures the "wiggliness" of the process, and Prohorov's condition implies that the processes are "not too wiggly.")

Exercise 1.6 (Doob) Use theorem 1.6 to establish that $g(\mathbb{U}_n) \to_d g(\mathbb{U})$ for all $\mathcal{D}$-measurable and a.s. $\|\cdot\|$-continuous functionals g on D.

Exercise 1.7 (Donsker) Use theorem 1.6 to establish that $g(\mathbb{S}_n) \to_d g(\mathbb{S})$ for all $\mathcal{D}$-measurable and a.s. $\|\cdot\|$-continuous functionals g on D.

Exercise 1.8 (Prohorov) Linearize $\mathbb{S}_n$ between the i/n-points so as to make it a process on $(C, \|\cdot\|)$, and then use exercise 1.3 to show that this linearized process converges in distribution to Brownian motion.

2 Metrics for Convergence in Distribution

Definition 2.1 (Prohorov metric) [The Lévy metric of exercise **??**.1.5 extends in a nice way to give a metric for $\to_d$ more generally.] For any Borel set $B \in \mathcal{M}_d$ and $\epsilon > 0$, define

$$B^\epsilon \equiv \{y \in M : d(x,y) < \epsilon \quad \text{for some } x \in B\}.$$

Let P, Q be two probability measures on $(M, d, \mathcal{M}_d)$. If we set

(1) $$\rho(P,Q) \equiv \inf\{\epsilon : Q(B) \le P(B^\epsilon) + \epsilon \quad \text{for all } B \in \mathcal{M}_d\},$$

then ρ is the *Prohorov metric* (see exercise 2.1). (We note that this definition is not formed in a symmetric fashion.)

Definition 2.2 (Dudley metric) (i) Label as $BL(M,d)$ the set of all real-valued functions g on the metric space (M,d) that are bounded and Lipschitz (in the sense that both of the quantities

(2) $$\|g\|_\infty \equiv \sup_{x \in M} |g(x)| \qquad \text{and} \qquad \|g\|_L \equiv \sup_{x \ne y} [|f(x) - f(y)|/d(x,y)]$$

are finite). For functions g in $BL(M,d)$ we define

(3) $$\|g\|_{BL} \equiv \|g\|_L + \|g\|_\infty,$$

and so $BL(M,d) = \{g : \|g\|_{BL} < \infty\}$.
(ii) Now let P, Q be two probability measures on $(M, \mathcal{M}_d)$, and set

(4) $$\beta(P,Q) \equiv \sup\{|\int g\,dP - \int g\,dQ| : \|g\|_{BL} \le 1\}.$$

Then β is called the *dual bounded Lipschitz distance* (or *Dudley distance*) between the probability distributions P and Q.

Proposition 2.1 Let $\mathcal{P} \equiv \{$all probability distributions P on $(M, \mathcal{M})$ $\}$. Then both ρ and β are metrics on $\mathcal{P}$. (Equation (12) below will show that $\mathcal{M}_\rho = \mathcal{M}_\beta$, which we abbreviate here as $\mathcal{M}$.)

Exercise 2.1 Prove the previous proposition.

The following theorem says that both ρ and β metrize $\to_d$ on $(M, d, \mathcal{M}_d)$ just as the Lévy distance L metrized the convergence in distribution $\to_d$ of dfs on R_1.

Theorem 2.1 (Metrizing $\to_d$; Dudley) For any separable metric space (M,d) and probability measures $\{P_n : n \ge 1\}$ and P on the Borel σ-field $\mathcal{M}_d$, the following are equivalent conditions:

(5) $$P_n \to_d P,$$

(6) $$\int g\,dP_n \to \int g\,dP \qquad \text{for all } g \in BL(M,d),$$

(7) $$\beta(P_n, P) \to 0,$$

(8) $$\rho(P_n, P) \to 0.$$

Theorem 2.2 (Ulam) For (M,d) complete and separable, each single P on $(M,\mathcal{M}_d)$ is tight.

Proof. Let $\epsilon > 0$. By the separability of M, for each $m \geq 1$ there is a sequence $A_{m1}, A_{m2}, \ldots$ of open $1/m$ spheres covering M. Choose i_m such that $P(\bigcup_{i\leq i_m} A_{mi}) > 1-\epsilon/2^m$. Now, the set $B \equiv \bigcap_{m=1}^{\infty}\bigcup_{i\leq i_m} A_{mi}$ is totally bounded in M, meaning that for each $\epsilon > 0$ it has a finite ϵ-net (that is, a set of points x_k with $d(x,x_k) < \epsilon$ for some x_k, for each $x \in B$). [Nice trick!] By completeness of M, $K \equiv \bar{B}$ is complete and is also compact; see exercises **??**.**??**(c) and **??**.**??**(e). Since

(a) $\quad P(K^c) = P(\bar{B}^c) \leq P(B^c) \leq \sum_{m=1}^{\infty} P([\bigcup_{i\leq i_m} A_{mi}]^c) < \sum_1^{\infty} \epsilon/2^m = \epsilon,$

the conclusion follows. □

Proof. We now prove theorem 2.1, since Ulam's theorem is in hand, but only under the additional assumption that M is complete. Clearly, (5) implies (6). We will now show that (6) implies (7). By Ulam's theorem, for any $\epsilon > 0$ we can choose K compact so that $P(K) > 1-\epsilon$. Let $f|K$ denote f restricted to K. Now, the set of functions $\mathcal{F} \equiv \{f|K : \|f\|_{BL} \leq 1\}$ forms a $\|\cdot\|$-totally bounded subset of the functions $\mathcal{C}_b(K)$ (by the Arzelà theorem of exercise **??**.**??**(a)). Thus, for every $\epsilon > 0$ there is some finite $k \equiv k_\epsilon$ and functions $f_1,\ldots,f_k \in \mathcal{F}$ such that for any $f \in \mathcal{F}$ there is an f_j with

(a) $\quad \sup_{x\in K} |f(x) - f_j(x)| \leq \epsilon; \qquad \text{moreover,} \qquad \sup_{x\in K^\epsilon} |f(x)-f_j(x)| \leq 3\epsilon,$

since f and f_j are in $\mathcal{F}$ (note the second half of (2)). Let

$$g(x) \equiv \max\{0, (1 - d(x,K)/\epsilon)\}\,;$$

then $g \in BL(M,d)$ and $1_K \leq g \leq 1_{K^\epsilon}$. Thus, $\int g\, dP_n \to \int g\, dP$, so that for n large enough we have

(b) $\quad P_n(K^\epsilon) \geq \int g\, dP_n > \int g\, dP - \epsilon \geq \int 1_K\, dP - \epsilon = P(K) - \epsilon > 1-2\epsilon.$

Hence, for any $f \in \mathcal{F}$ we have from (a), (2), (b), and $P(K) > 1-\epsilon$ that

$$|\textstyle\int f\, d(P_n - P)| = |\int (f-f_j)\, d(P_n-P) + \int f_j\, d(P_n-P)|$$

$$\leq |\textstyle\int (f-f_j)\, dP_n| + |\int (f-f_j)\, dP| + |\int f_j\, d(P_n-P)|$$

(c) $\quad \leq (3\epsilon + 2\times 2\epsilon) + (\epsilon + 2\times\epsilon) + |\int f_j\, d(P_n - P)| \leq 11\epsilon$

for n chosen large enough. Hence (7) holds.

We next show that (7) implies (8). Suppose a Borel set B and an $\epsilon > 0$ are given. We let $f_\epsilon(x) \equiv \max\{0,(1-d(x,B)/\epsilon)\}$. Then $f_\epsilon \in BL(M,d)$, $\|f_\epsilon\|_{BL} \leq 1+\epsilon^{-1}$, and $1_B \leq f_\epsilon \leq 1_{B^\epsilon}$. Therefore, for any P and Q on M we have from (4) that

(d) $\quad Q(B) \leq \int f_\epsilon\, dQ \leq \int f_\epsilon\, dP + (1+\epsilon^{-1})\,\beta(P,Q) \leq P(B^\epsilon) + (1+\epsilon^{-1})\,\beta(P,Q),$

and it follows that

(e) $\rho(P,Q) \le \max\{\epsilon, (1+\epsilon^{-1})\,\beta(P,Q)\}$.

Hence, if $\beta(P,Q) < \epsilon^2$, then $\rho(P,Q) < \epsilon + \epsilon^2 < 2\epsilon$. Hence, for all P, Q we have $\rho(P,Q) \le 2\sqrt{\beta(P,Q)}$. Thus, (7) implies (8). [It is also possible to establish the inequality $\beta(P,Q)/2 \le \rho(P,Q)$. This would give

(f) $\beta(P,Q)/2 \le \rho(P,Q) \le 2\sqrt{\beta(P,Q)}$,

showing that ρ and β are equivalent metrics (see (12) and exercise 2.3 below).]

Finally, we will show that (8) implies (5). Suppose (8) holds. Let B denote a P-continuity set, and let $\epsilon > 0$. Then for $0 < \delta < \epsilon$ with δ small enough, we have $P(B^\delta \setminus B) < \epsilon$ and $P((B^c)^\delta \setminus B^c) < \epsilon$. Then for n large enough we have

(g) $P_n(B) \le P(B^\delta) + \delta \le P(B) + 2\epsilon$

and also

(h) $P_n(B^c) \le P((B^c)^\delta) + \delta \le P(B^c) + 2\epsilon$.

Combining these yields

(i) $|P_n(B) - P(B)| \le 2\epsilon$,

and hence $P_n(B) \to P(B)$. By the portmanteau theorem, this yields (5). □

Strassen's Coupling Theorem

Suppose $P_n \to_d P_0$ on a separable metric space (M, d). Then theorem 2.1 gives $\rho(P_n, P_0) \to 0$, for Prohorov's metric ρ, while Skorokhod's theorem gives existence of random elements X_n on a common $(\Omega, \mathcal{A}, P)$ satisfying $d(X_n, X_0) \to_{a.s.} 0$. Claiming less than is true, $d(X_n, X_0) \to_p 0$, or $P(d(X_n, X_0) \ge \epsilon) \to 0$ as $n \to \infty$. We are then naturally led to ask how rapidly this convergence occurs. It turns out that this is essentially determined by $\rho(P_n, P)$. Alternatively, the following theorem can be used to bound $\rho(P_n, P)$, provided that a rate is available regarding Skorokhod.

Theorem 2.3 (Strassen) Suppose that P and Q are measures on the Borel sets of a separable metric space (M, d). Then $\rho(P, Q) < \epsilon$ if and only if there exist X and Y defined on a common probability space with $X \cong P$ and $Y \cong Q$ and coupled closely enough that $P(d(X,Y) \ge \epsilon) \le \epsilon$.

Proof. See Dudley (1976, section 18). □

Some Additional Metrics

As shown in theorem 2.1, both the Prohorov metric ρ and also the dual-bounded Lipschitz metric β metrize weak convergence ($\to_d$). Other stronger metrics are also available and often useful.

Definition 2.3 (Total variation metric) For probability measures P and Q on the measurable space $(M, \mathcal{M}_d)$, let

(9) $$d_{TV}(P,Q) \equiv 2 \sup\{|P(A) - Q(A)| : A \in \mathcal{M}_d\};$$

d_{TV} is called the *total variation metric.*

Proposition 2.2 The total variation metric d_{TV} is equal to

(10) $$d_{TV}(P,Q) = \int |f-g|\, d\mu = 2[1 - \int f \wedge g\, d\mu],$$

where $f = dP/d\mu$, $g = dQ/d\mu$, and μ is any measure dominating both P and Q (for example, $P + Q$).

Proof. Note that $|f - g| = (f \vee g) - (f \wedge g) = (f + g) - 2(f \wedge g)$. □

Definition 2.4 (Hellinger metric) For probabilities P and Q on $(M, \mathcal{M}_d)$, let

(11) $$d_H^2(P,Q) \equiv \int [f^{1/2} - g^{1/2}]^2\, d\mu = 2[1 - \int \sqrt{f\,g}\, d\mu],$$

where $f = dP/d\mu$, $g = dQ/d\mu$, and μ is any measure dominating both P and Q (for example, $P + Q$); then d_H is called the *Hellinger metric.*

Exercise 2.2 d_H does not depend on the choice of μ.

Here is a theorem relating these metrics and the Prohorov and bounded Lipschitz metrics. The inequalities in (12) show that ρ and β induce identical topologies. The inequalities in (13) show that the total variation metric d_{TV} and the Hellinger metric d_H induce identical topologies. Moreover, (14) shows that the β and ρ topology is the finer topology (with more open sets and admitting more continuous functions). [Recall exercise **??.??** dealing with equivalent metrics.]

Theorem 2.4 (Inequalities among the metrics) (a) For P and Q probability measures on $(M, \mathcal{M}_d)$, the following inequalities necessarily hold:

(12) $$\beta(P,Q)/2 \le \rho(P,Q) \le 2\sqrt{\beta(P,Q)},$$

(13) $$d_H^2(P,Q) \le d_{TV}(P,Q) \le d_H(P,Q)\,\{4 - d_H^2(P,Q)\}^{1/2},$$

(14) $$\rho(P,Q) \le d_{TV}(P,Q).$$

(b) For dfs F, G on R (or R_k) we have the following:

(15) $$L(F,G) \le \rho(F,G) \le d_{TV}(F,G),$$

(16) $$L(F,G) \le d_K(F,G) \le d_{TV}(F,G),$$

where $d_K(F,G) \equiv \|F - G\|_\infty \equiv \sup_x |F(x) - G(x)|$ is the Kolmogorov distance.

Exercise 2.3 Prove the first inequality in (12).

Exercise 2.4 Prove (13). [Hint. To prove the left inequality, establish the inequality $\int \sqrt{f\,g}\,d\mu \geq \int f \wedge g\,d\mu$ and use the second equality in (11). To show the right inequality, write $|f-g| = |\sqrt{f} - \sqrt{g}\,|\,|\sqrt{f} + \sqrt{g}\,|$.

Exercise 2.5 Prove (14).

Exercise 2.6 (Statistical interpretation of the d_{TV} metric) Consider testing P versus Q. Find the test that minimizes the sum of the error probabilities, and show that the minimum sum of errors is $\|P \wedge Q\| \equiv \int f \wedge g\,d\mu$. Note that P and Q are orthogonal if and only if $\|P - Q\| \equiv d_{TV}(P, Q) = 2$ if and only if $\|P \wedge Q\| = 0$ if and only if $\int \sqrt{f\,g}\,d\mu \equiv \int \sqrt{dP\,dQ} = 0$.

Chapter 15

Asymptotics via Empirical Processes

0 Introduction

In section 15.1 we rederive the usual CLT for iid samples from any distribution in the domain of attraction of the normal distribution, but now using empirical process methods. We then obtain a corresponding result for the trimmed mean that is valid for samples from any df in *any* domain of attraction, provided only that the number of observations trimmed from each tail grows to infinity so slowly that the fraction of observations trimmed from each tail goes to zero. When the qfs of all distributions in a class are bounded by an appropriate envelope qf, then all of these convergence results are uniform across the class of qfs. This uniformity allows random trimming. In section 15.3 *complete* analogs are presented for L-statistics. In section 15.2 similar results are derived for linear rank tests and permutation tests, and a uniform studentized CLT is given for sampling from a finite population. Also, two very interesting ways of creating normality are discussed in this section.

All of the results are based on the empirical process construction of section 12.10 combined with the quantile method inequalities from sections C.5–C.6. We will frequently obtain conclusions of the form $T_n \to_p Z$ for a special construction version of an interesting statistic T_n and its limiting normal rv Z. Or, we may even obtain $\sup_{\mathcal{K}} |T_n - Z(K)| \to_p 0$ for a family of normal rvs $Z(K)$ indexed by the qf K in a class $\mathcal{K}$. These strong $\to_p$ conclusions are only true for the special construction versions of the statistic T_n. However, $T_n \to_p Z$ for the special construction T_n implies that $T_n \to_d Z$ for *any* version of T_n. In like fashion, $\sup_{\mathcal{K}} |T_n - Z(K)| \to_p 0$ for a special construction version of T_n implies that the rate at which $T_n \to_d Z$ is uniform across the entire class of qfs in $\mathcal{K}$ for *any* version of the statistic T_n.

1 Ťrimmed and W̃insorized Means

Notation 1.1 Let $X_{n1}, \ldots, X_{nn}$ be iid with df F. Let $X_{n:1} \le \cdots \le X_{n:n}$ denote their order statistics, with empirical qf $\mathbb{K}_n$. Define $\mu \equiv \mu_K \equiv \int_0^1 K(t)\,dt = \mathrm{E}K(\xi)$ and $\sigma^2 \equiv \sigma_K^2 \equiv \mathrm{Var}[K(\xi)]$, when these exist. We also let $\bar{X}_n = \sum_1^n X_{ni}/n$ denote the sample mean and $S_n^2 = \sum_1^n (X_{ni} - \bar{X}_n)^2/n$ denote the "sample variance". For trimming numbers $k_n \wedge k_n^{'} \ge 1$ and for $a_n \equiv k_n/n$ and $a_n' \equiv k_n'/n$, we let $\tilde{K}_n(\cdot)$ denote $K(\cdot)$ Winsorized outside (a_n, a_n') (see notation 6.5.1) and define

$$
(0)\qquad
\begin{aligned}
&\check{\mu}_n \equiv \check{\mu}_K(a_n, a_n') \equiv \textstyle\int_{a_n}^{1-a_n'} K(t)\,dt = \mathrm{E}\{K(\xi) \times 1_{(a_n, 1-a_n')}(\xi)\}\,,\\
&\tilde{\sigma}_n^2 \equiv \tilde{\sigma}_K^2(a_n, a_n') \equiv \textstyle\int_0^1 \int_0^1 [s \wedge t - st]\, d\tilde{K}_n(s)\, d\tilde{K}_n(t) = \mathrm{Var}[\tilde{K}_n(\xi)]\,,\\
&\check{X}_n \equiv \check{X}_n(a_n, a_n') \equiv \tfrac{1}{n} \textstyle\sum_{i=k_n+1}^{n-k_n'} X_{n:i} = \check{\mu}_{\mathbb{K}_n}(a_n, a_n'),\\
&\tilde{S}_n^2 \equiv \tilde{\sigma}_{\mathbb{K}_n}^2(a_n, a_n').
\end{aligned}
$$

Here $\tilde{S}_n \equiv \tilde{\sigma}_{\mathbb{K}_n}(a_n, a_n')$ denotes the *sample* $(a_n, a_n^{'})$*-Winsorized standard deviation*, and the rv $\check{X}_n = \check{\mu}_{\mathbb{K}_n}(a_n, a_n')$ is being called the *sample* (a_n, a_n')*-truncated mean*, while $\breve{X}_n \equiv \check{X}_n/(1 - a_n - a_n')$ denotes the *sample* (a_n, a_n')*-trimmed mean.* Also, $\breve{\mu}_n \equiv \check{\mu}_n/(1 - a_n - a_n')$ is the population trimmed mean, and $\breve{\sigma}_n \equiv \tilde{\sigma}_n/(1 - a_n - a_n')$. Now, $\check{X}_n$ is our vehicle of convenience for studying the trimmed mean $\breve{X}_n$, since

$$
(1)\qquad \sqrt{n}\,(\check{X}_n - \check{\mu}_n)/\tilde{\sigma}_n = \breve{Z}_n \equiv \sqrt{n}\,(\breve{X}_n - \breve{\mu}_n)/\breve{\sigma}_n\,.
$$

[Said another way, $\breve{X}_n$ has statistical meaning, while $\check{X}_n$ does not; but $\check{X}_n$ is much more convenient to work with notationally and probabilistically.] For $k_n \wedge k_n' \ge 1$ we *always* have (integrating Brownian bridge $\mathbb{U}$, which for each fixed ω is just a continuous function)

$$
(2)\qquad Z(\tilde{K}_n) \equiv -\int_{a_n}^{1-a_n'} \mathbb{U}\, dK/\tilde{\sigma}_n = -\int_0^1 \mathbb{U}\, d\tilde{K}_n/\tilde{\sigma}_n \cong N(0,1).
$$

Generally, $\tilde{K}_{t,t}(\cdot)$ denotes $K(\cdot)$ Winsorized outside $(t, 1-t)$, and $\tilde{\sigma}^2(t) \equiv \mathrm{Var}[\tilde{K}_{t,t}(\xi)]$. Recall that $R_n =_a S_n$ denotes that $R_n - S_n \to_p 0$. □

Convention We now specify that *throughout this chapter* $X_{ni} \equiv K(\xi_{ni}) = F^{-1}(\xi_{ni})$, for $1 \le i \le n$, for the ξ_{ni} described in notation 1.2 below. (Recall the sentence above (6.4.3) noting that this representation of rvs allows alternative methods of proof.)

Note The conclusions in (4), and (6) below are true *only* for these particular rvs $F^{-1}(\xi_{ni})$ that are iid $F(\cdot)$, but the implied $\to_d$ conclusion is true for *any* iid rvs $X_{n1}, \ldots, X_{nn}$ having df $F(\cdot)$.

Theorem 1.1 (The CLT) Let $K \in \mathcal{D}(\text{Normal})$. Equivalently, suppose

$$
(3)\qquad t\,[K_+^2(ct) \vee K^2(1-ct)]\,/\,\tilde{\sigma}^2(t) \to 0 \qquad \text{as } t \to 0, \qquad \text{for each fixed } c > 0.
$$

(Note proposition 10.6.1 regarding (3).) Define $k_n \equiv k'_n \equiv 1$, $\check{\mu}_n \equiv \mu_K(1/n, 1/n)$ and $\tilde{\sigma}_n \equiv \tilde{\sigma}_K(1/n, 1/n)$, and let $\tilde{K}_n(\cdot)$ denote $K(\cdot)$ Winsorized outside $(1/n, 1-1/n)$. Then iid rvs $X_{n1}, \ldots, X_{nn}$ with qf $K(\cdot)$ satisfy (as also shown in theorem 10.6.1)

$$(4) \qquad Z_n = \sqrt{n}\,(\bar{X}_n - \check{\mu}_n)/\tilde{\sigma}_n =_a Z(\tilde{K}_n) \cong N(0,1) \qquad \text{and} \qquad S_n/\tilde{\sigma}_n \to_p 1\,.$$

Notation 1.2 For the following theorem, suppose the integers k_n and k'_n satisfy

$$(5) \qquad (k_n \wedge k'_n) \to \infty, \qquad (a_n \vee a'_n) \to 0, \qquad \text{and} \qquad a'_n/a_n \to 1.$$

Of course, $\tilde{K}_n(\cdot)$ now denotes $K(\cdot)$ Winsorized outside $(a_n, 1-a'_n)$. □

Theorem 1.2 (CLT for trimmed means) Suppose the qf $K(\cdot)$ is such that the partial variance $\tilde{\sigma}^2(t) \equiv \mathrm{Var}[\tilde{K}_{t,t}(\xi)]$ is in one of the regularly varying classes $\mathcal{R}_{-\beta}$, for some $\beta \ge 0$. [This union of all the $\mathcal{R}_{-\beta}$ classes was labeled as $\tilde{\mathcal{D}}$ in the definition (C.5.33) $\mathcal{R}_0 = \mathcal{D}(\text{Normal})$. Note (C.5.4).] If (5) holds, then

$$(6) \qquad \check{Z}_n = \sqrt{n}\,(\check{X}_n - \check{\mu}_n)/\tilde{\sigma}_n =_a Z(\tilde{K}_n) \cong N(0,1) \qquad \text{and} \qquad \tilde{S}_n/\tilde{\sigma}_n \to_p 1.$$

If $\beta = 0$, we can weaken (5) to $a_n \vee a'_n \to 0$ when $0 < \underline{\lim}\, a_n/a'_n \le \overline{\lim}\, a_n/a'_n < \infty$.

Corollary 1 (Trimming fixed fractions) Suppose $a_n \to a$ and $a'_n \to a'$ for $0 < a < 1 - a' < 1$. Then (4) holds for any qf $K(\cdot)$ continuous at both a and $1 - a'$.

Notation 1.3 Be clear that we are working on the specific probability space $(\Omega, \mathcal{A}, P)$ of theorem 12.10.3 on which are defined a fixed Brownian bridge $\mathbb{U}$ and a triangular array of rvs whose nth row members $\xi_{n1}, \ldots, \xi_{nn}$ are iid Uniform(0,1) with order statistics $0 \le \xi_{n:1} \le \cdots \le \xi_{n:n} \le 1$, empirical df $\mathbb{G}_n$, and empirical process $\mathbb{U}_n = \sqrt{n}\,[\mathbb{G}_n - I]$ that not only satisfy $\|\mathbb{U}_n = \mathbb{U}\| \to_p 0$, but in fact, for each fixed $0 \le \nu < \frac{1}{4}$, satisfy

$$(7) \qquad \Delta_{\nu n} \equiv \left\| \frac{n^{\nu}\,(\mathbb{U}_n - \mathbb{U})}{[I \wedge (1-I)]^{(1/2)-\nu}} \right\|_{1/n}^{1-1/n} = O_p(1)\,.$$

For any $\xi_{n1}, \ldots, \xi_{nn}$ (and hence for this realization also), for each fixed $0 < \nu < 1$,

$$(8) \qquad \Delta^0_{\nu n} \equiv \left\| \frac{n^{\nu}\,(\mathbb{G}_n - I)}{[I \wedge (1-I)]^{1-\nu}} \right\|_0^1 = O_p(1).$$ □

Remark 1.1 We prove these theorems in such a way that the uniformity available is apparent. To see the full range of uniformity possible, consult the 1*st* Edition of this text, where this topic is pursued much more completely. Random k_n and k'_n, (useful to applied statisticians) are also considered therein. □

Proofs. Integration by parts (with $\int_{(d,c]} \equiv -\int_{(c,d]}$ if $c < d$ in (b), (c), etc. below) [a.s., dK puts no mass at any $\xi_{n:i}$] yields

$$\check{Z}_n = \sqrt{n}\,(\check{X}_n - \check{\mu}_n)/\tilde{\sigma}_n$$

$$(a) \qquad = \sqrt{n}\{\textstyle\int_{[\xi_{n:k_n+1}, \xi_{n:n-k'_n}]} K\, d\mathbb{G}_n - \int_{(a_n, 1-a'_n)} K(t)\, dt\}/\tilde{\sigma}_n$$

(b) $$= \sqrt{n}\,\{-K\mathbb{G}_n\big|_{\xi_{n:k_n+1}-} - \int_{[\xi_{n:k_n+1},\xi_{n:n-k'_n}]}\mathbb{G}_n\,dK + K\mathbb{G}_n\big|_{\xi_{n:n-k'_n}+}\}/\tilde{\sigma}_n$$
$$+ \sqrt{n}\,\{tK\big|_{a_n+} + \int_{(a_n,1-a'_n)} t\,dK - tK\big|_{1-a'_n-}\}/\tilde{\sigma}_n$$

(c) $$=_{a.s.}\ -\int_{(a_n,1-a'_n)}\mathbb{U}\,d\tilde{K}_n/\tilde{\sigma}_n - \int_{(a_n,1-a'_n)}[\mathbb{U}_n-\mathbb{U}]\,d\tilde{K}_n/\tilde{\sigma}_n$$
$$-\int_{(\xi_{n:k_n+1},a_n]}\sqrt{n}\,[\mathbb{G}_n(t)-a_n]\,dK/\tilde{\sigma}_n$$
$$-\int_{[1-a'_n,\xi_{n:n-k'_n})}\sqrt{n}\,[\mathbb{G}_n(t)-(1-a'_n)]\,dK/\tilde{\sigma}_n$$

(9) $$\equiv Z(\tilde{K}_n) + \gamma_{1n} + \gamma_{2n} + \gamma'_{2n}$$
$$\{+\,\gamma_{3n} + \gamma'_{3n} \text{ when replacing } \check{X}_n - \check{\mu}_n \text{ by } \bar{X}_n - \check{\mu}_n \text{ during theorem 1.1}\},$$

where $\gamma_{3n} \equiv K(\xi_{n:1})/\sqrt{n}\,\tilde{\sigma}_n$ and $\gamma'_{3n} \equiv K(\xi_{n:n})/\sqrt{n}\,\tilde{\sigma}_n$ are the extreme summands of $\check{Z}_n$. [We require the added terms γ_{3n} and γ'_{3n} only for the theorem 1.1 identity, because in this case only we have used the "non-natural for the identity" value $\check{\mu}(1/n,1/n)$ in the theorem statement. The "natural value" $\mu = \check{\mu}(0,0)$ would have caused trouble in the proof. Note the definition of $k_n = k'_n = 1$ in theorem 1.1.]

We begin an examination of the various γ-terms. Concerning γ_{1n}, we note that

(10) $$|\gamma_{1n}| \le \Delta_{\nu n} \times \int_{a_n}^{1-a'_n} n^{-\nu}\,[t(1-t)]^{1/2-\nu}\,dK(t)/\,\tilde{\sigma}_n \equiv \Delta_{\nu n} \times M_{\nu n}(K),$$

and it is this sort of *factorization* into a random term times a deterministic term that is key to the proof. The randomness refers only to the Uniform$(0,1)$ distribution, and the only reference to $K(\cdot)$ is in the deterministic term (to which inequality (C.6.1) applies). In particular, $\Delta_{\nu n} = O_p(1)$ by (7). Then (C.6.3) gives (the second term appearing in (d) is the γ_n of (C.6.3), and the γ'_n term in (d) is the symmetric term from the right tail)

(d) $$M_{\nu n}(K) \le \frac{(3/\sqrt{\nu}\,)}{[(k_n \vee r)\wedge(k'_n \vee r')]^{\nu}} + \frac{\sqrt{r}\,|K_+(a_n)-K_+(r/n)|}{\sqrt{n}\,\tilde{\sigma}_n}\,1_{[a_n<r/n]} + \gamma'_n.$$

Note with regard to γ_{2n} that monotonicity of $\mathbb{G}_n(\cdot)$ implies that

(e) the integrand of γ_{2n} is uniformly bounded by $|\mathbb{U}_n(a_n)| = \sqrt{a_n}\,E_n$,

where it is trivially true that $E_n \equiv |\mathbb{U}_n(a_n)/\sqrt{a_n}\,| = O_p(1)$. We also let $1_{n\epsilon}$ denote the indicator function of an event, having probability exceeding $1-\epsilon$, on which (with $\lambda \equiv \lambda_\epsilon$ small enough) we can specify our choice of one of the following:

(f)
- (i) $\lambda\,a_n \le \xi_{n:k_n} \le \xi_{n:k_n+1} \le a_n/\lambda$.
- (ii) $\xi_{n:k_n+1}$ lies between $a_n \mp \lambda\sqrt{k_n}/n = a_n(1 \mp \lambda/\sqrt{k_n}\,)$.
- (iii) Bound $\xi_{n:k_n+1}$ by its lower and upper $\epsilon/2$-quantiles $t^{\mp}_{n\epsilon}$.

[The bounds in (f) are derived as follows: For (ii), the Chebyshev (second moment) inequality with beta rv moments of $\xi_{n:k_n+1}$. Or for (i), the Markov (first moment) inequality, or from the in probability linear bounds of inequality 12.11.2. Or for choice (iii), the exact distribution of $\xi_{n:k_n+1}$.]

Consider theorem 1.1, when $a_n = a'_n = 1/n$. Now, γ_{1n} is controlled via (d) (at the left end) solely by $|K_+(a_n)|/\sqrt{n}\,\tilde{\sigma}_n$ (just choose r first to be very large in the first term in the bound of (d)). To summarize, $\gamma_{1n} \to_p 0$ whenever

$$(11) \qquad \{\sqrt{a_n}\,|K_+(a_n)| + \sqrt{a'_n}\,|K(1-a'_n)|\,\}/\tilde{\sigma}_n \to 0\,.$$

For γ_{2n} and γ_{3n} we apply (f)(i) in choosing the $1_{n\epsilon}$ in (f) to get (when $K_+(0) < 0$)

$$(g) \qquad |\gamma_{2n}| \times 1_{n\epsilon} \le E_n \times \left|\sqrt{a_n}\int_{(\lambda a_n, a_n/\lambda)} dK/\tilde{\sigma}_n\right| \le 2E_n \times |\sqrt{a_n}\,K(\lambda a_n)/\tilde{\sigma}_n|\,,$$

$$(h) \qquad |\gamma_{3n}| \times 1_{n\epsilon} \le 2 \times |K_+(\lambda a_n)|\,/(\sqrt{n}\,\tilde{\sigma}_n) \le 2 \times |\sqrt{a_n}\,K(\lambda a_n)/\tilde{\sigma}_n|\,.$$

Thus the CLT claim of (4) holds, provided only (adding the symmetric condition in the right tail)

$$(12) \qquad \{\sqrt{a_n}\,|K_+(\lambda a_n)| + \sqrt{a'_n}\,|K((1-\lambda a'_n))|\,\}/\tilde{\sigma}_n \to 0$$

for each $0 < \lambda \le 1$, and it holds uniformly in any class of qfs in which (12) holds uniformly. But (12) follows from (3) (or any of the equivalent conditions (10.6.32), or (10.6.16), or (10.6.28), or (10.6.21), or (C.2.6), or (C.2.15), or (C.1.53)—for example. Thus the normality in theorem 1.1 is established again—using a fundamentally different proof from that in section 10.6. [Note that (the crude upper bound) condition (12) implies (11), (g), and (h).]

Consider theorem 1.2. The first term in (d) converges to 0, and the other two terms that appear in (d) equal zero. Thus, again $\gamma_{1n} \to_p 0$ whenever $k_n \wedge k'_n \to \infty$ (but now, it converges *uniformly* in *all* qfs). In the present context the γ_{3n} term always equals 0. Thus, *only* γ_{2n} must be shown to be negligible; but we must now be much more careful than the crude bound (12). Now, using (f)(ii) in the definition of $1_{n\epsilon}$ in (f), we obtain

$$(i) \qquad |\gamma_{2n}| \times 1_{n\epsilon} \le E_n \times \left|\sqrt{a_n}\int_{I_{\lambda n}} dK/\tilde{\sigma}_n\right|,$$

where $I_{\lambda n} \equiv (a_n(1-\lambda/\sqrt{k_n}\,), a_n(1+\lambda/\sqrt{k_n}\,))$. Thus the CLT claim of (6) holds, provided that (with symmetric requirements in the right tail)

$$(13) \qquad \sqrt{a_n}\left|K(a_n(1 \mp \lambda/\sqrt{k_n}\,)) - K_+(a_n)\right|/\tilde{\sigma}_n \equiv \sqrt{a_n}\int_{I^{\mp}_{\lambda n}} dK/\tilde{\sigma}_n \to 0$$

for each $\lambda > 0$. Thus, normality holds uniformly in any class $\mathcal{K}_u$ in which both (13) and its right tail analogue hold uniformly; call such a class $\mathcal{K}_u$ a *uniformity class*.

Summary (so far) Whenever $k_n \wedge k'_n \to \infty$,

$$(14) \qquad \sup_{K \in \mathcal{K}_u}\left|\sqrt{n}\,[\check{X}_n(a_n, a'_n) - \check{\mu}_n]/\tilde{\sigma}_n - Z(\tilde{K}_n)\right| \to_p 0$$

for any class $\mathcal{K}_u$ in which both (13) and its right tail analogue hold uniformly. (Also, we may replace $\tilde{\sigma}_n$ by $\tilde{S}_n$ in (14) under this same requirement (as was be shown in the variance estimation proof given in the 1*st* Edition)).

Now, (13) does hold for a fixed K whenever both K is in any $\mathcal{R}_{-\beta}$ and the trimming numbers of (5) are employed (appeal to theorem C.5.1), and this gives theorem 1.2. (Two uniformity classes $\mathcal{K}_u$ are exhibited in theorem C.5.2.) Aside from variance estimation, the proofs of theorems 1.1 and 1.2 are complete. The 1*st* Edition carefully considers both variance estimation and the uniformity of the asymptotic normality, via making (12) hold uniformly. □

Remark 1.2 The class $\tilde{\mathcal{D}}$ of qfs K having $\tilde{\sigma}^2(\cdot) \in \mathcal{R}_{-\beta}$ for some $\beta \geq 0$ is strictly bigger than the stable laws; the class of stable laws also require that

$$v^+(t)/v^-(t) \to (\text{some } (1-p)/p) \in [0,1],$$

so that both the contributions from the two extreme tails {(from 1 to k_n) and (from $n - k_n' + 1$ to n)} can be balanced. (Recall the De Haan result in exercise C.4.2.) But we do not need to balance them; we threw them away. □

Exercise 1.1 Verify that $\sup\{\mathbb{U}_n(t)|/\sqrt{a_n} : c\, a_n \leq t \leq a_n/c\} = O_p(1)$.

Exercise 1.2* (Rossberg) Suppose $0 \leq \xi_{n:1} \leq \cdots \leq \xi_{n:n}$ are the order statistics of the first n of the infinite sequence of independent Uniform(0, 1) rvs $\xi_1, \xi_2, \ldots$. Let $\alpha_n \equiv \{j : 1 \leq j \leq k_n\}$, $\delta_n \equiv \{j : k_n + 1 \leq j \leq l_n\}$, $\beta_n \equiv \{j : l_n + 1 \leq j \leq n - l_n'\}$, $\delta_n' \equiv \{j : n - l_n' + 1 \leq j \leq n - k_n'\}$, and $\alpha_n' \equiv \{j : n - k_n' + 1 \leq j \leq n\}$. Show that the collections of rvs $\{\xi_{n:j} : j \in \alpha_n\}$, $\{\xi_{n:j} : j \in \beta_n\}$, and $\{\xi_{n:j} : j \in \alpha_n'\}$ are asymptotically independent when $k_n \nearrow \infty$ with $a_n \equiv k_n/n \to 0$ and $l_n \equiv k_n^{1+2\nu}$ for $0 < \nu < \frac{1}{4}$ (with analogous conditions in the right tail).

Remark 1.3 Consider asymptotic normality of the sample mean. Let U_n, V_n, W_n, V_n', and U_n' denote the sum of those $X_{n:j} = K(\xi_{n:j})$ for which j is in α_n, δ_n, β_n, δ_n', and α_n', respectively. The previous exercise shows that W_n, U_n, and U_n' are asymptotically independent. Exercise C.6.1 shows that V_n and V_n' are *always* asymptotically negligible. We saw in appendix C that the condition $\tilde{\sigma}^2(\cdot) \in \mathcal{L}$ is also necessary for (4). Since the vast middle is "nearly always" normal, one needs to determine what is happening only in the extreme tails (as the mid-tails were always negligible). This asymptotic independence is at the heart of the very general asymptotics for $\bar{X}_n$ found in S. Csörgő, Haeusler, and Mason (1989). □

Exercise 1.3 (Winsorized mean) Let $\tilde{Z}_n \equiv \sqrt{n}\,(\tilde{X}_n - \tilde{\mu}_n)/\tilde{\sigma}_n$ for the sample mean $\tilde{X}_n$ of the Winsorized sample $\tilde{X}_{n1}, \ldots, \tilde{X}_{nn}$. For the $N(0,1)$ rv $Z(\tilde{K}_n)$ of (2), the identity (9) becomes

$$\tilde{Z}_n = Z(\tilde{K}_n) + \gamma_{1n} + \tilde{\gamma}_{2n} + \tilde{\gamma}_{2n}',$$

$$\text{where} \quad \tilde{\gamma}_{2n} \equiv \sqrt{n} \int_{I_n} \mathbb{G}_n \, dK/\tilde{\sigma}_n \quad \text{and} \quad \tilde{\gamma}_{2n}' \equiv \sqrt{n} \int_{I_n'} \mathbb{G}_n \, dK/\tilde{\sigma}_n$$

and γ_{1n} is as before. Here I_n is equal to $[\xi_{n:k_n+1}, a_n]$ or $(a_n, \xi_{n:k_n+1})$ according as $\xi_{n:k_n+1} < a_n$ or $\xi_{n:k_n+1} \geq a_n$, and I_n' is equal to $[1-a_n', \xi_{n:n-k_n'}]$ or $(\xi_{n:n-k_n'}, 1-a_n')$ according as $1 - \xi_{n:n-k_n'} < a_n'$ or $1 - \xi_{n:n-k_n'} \geq a_n'$. Show that this quantity $\tilde{\gamma}_{2n}$ for the Winsorized mean essentially exceeds the γ_{2n} of the trimmed mean proof by the factor $\sqrt{k_n}$. This is just enough to prevent analogues of the theorems for trimmed means that we proved in this chapter.

(a) Prove what you can for the Winsorized mean.

(b) Graph a typical K on $(0,1)$. Locate a_n and $\xi_{n:k_n+1}$ near the 0 endpoint (and suppose $K(0) < 0$). Obtain the natural graphical upper bounds on the magnitudes of γ_{2n} and $\tilde{\gamma}_{2n}$, and note how the second quantity is inherently larger than the first (pictorially, a "trapezoid" versus a "triangle").

Exercise 1.4 (Uniform WLLN) Suppose the qf $K_0(\cdot)$ is of order one, in that

$$(15)\qquad t\,(\,|K_0(t)| + |K_0(1-t)|\,) \to 0 \qquad \text{as } t \to 0\,.$$

Then one can claim the uniform WLLN

$$(16)\qquad \sup_{K \in \mathcal{K}_0} \left| \bar{X}_n - \int_{1/n}^{1-1/n} K(t)\,dt \right| \to_p 0,$$

when $\mathcal{K}_0 \equiv \{K : |K| \le |K_0| \text{ on } (0, a_0] \cup [1-a_0, 1)\}$, for some $0 < a_0 \le \frac{1}{2}$.

2 Linear Rank Statistics and Finite Sampling

Example 2.1 (Linear rank statistics) Consider the $\mathbb{R}_N$ process of section 12.10, with the same notation and assumptions as were made there. Thus (for c_{Ni}'s with mean $\bar{c}_N = 0$, standard deviation $\sigma_{cN}^2 = 1$, and standardized fourth central moment $\overline{c_N^4}/\sigma_{cN}^4 \le M < \infty$ for all N) we define (for the antiranks $\mathsf{D} \equiv (D_{N1}, \ldots, D_{NN})$)

$$(1) \qquad \mathbb{R}_N(t) \equiv \frac{1}{\sqrt{N}} \sum_{i=1}^{[(N+1)t]} \frac{c_{ND_{Ni}} - c_{N\cdot}}{\sigma_{cN}} = \frac{1}{\sqrt{N}} \sum_{i=1}^{[(N+1)t]} c_{ND_{Ni}} \quad \text{on } [0,1],$$

and this process satisfies $\ddot{\Delta}_{\nu N} = O_p(1)$, as in (12.10.35). The known constants $\mathsf{c}_N \equiv (c_{N1}, \ldots, c_{NN})'$ are called *regression constants*. Let $\mathsf{a}_N \equiv (a_{N1}, \ldots, a_{NN})'$ specify known *scores*. Let $a_{N\cdot}$, $\sigma_{aN}^2 > 0$, and $\mu_4(\mathsf{a}_N)$ denote their mean, variance, and fourth central moment. The class of *simple linear rank statistics* is defined by

$$(2) \qquad T_N \equiv T_N(\mathsf{a}_N) \equiv \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{a_{Ni} - a_{N\cdot}}{\sigma_{aN}} \, c_{ND_{Ni}} \,.$$

Now, $\mathrm{E}\, T_N = 0$ and $\mathrm{Var}[T_N] = \frac{N}{N-1}$ by exercise 2.1 below. Assume (for convenience only) that the scores are ordered as

$$a_{N1} \le \cdots \le a_{NN} \,,$$

and we define an $\nearrow$ left-continuous qf K_N on $[0,1]$ by

$$(3) \qquad K_N(t) = a_{Ni} - a_{N\cdot} \qquad \text{for } \ (i-1)/N < t \le i/N, \ \text{and } 1 \le i \le N,$$

with $K_N(0) \equiv a_{N1}$. Note that

$$(4) \qquad T_N \equiv \int_0^1 K_N \, d\,\mathbb{R}_N/\sigma_{aN} = -\int_0^1 \mathbb{R}_N \, dK_N/\sigma_{aN} \,.$$

The basic probability space and rvs are as defined in notation 15.1.3. Let

$$(5) \qquad Z_N \equiv Z(\mathsf{a}_N) \equiv -\int_0^1 \mathbb{W} \, dK_N/\sigma_{aN} \,.$$

Clearly, Z_N is a normal rv with mean 0, since it is just a linear combination of the jointly normal rvs $\mathbb{W}(1/N), \ldots, \mathbb{W}(1-1/N)$. Fubini's theorem gives

$$\mathrm{Var}[Z_N] = \mathrm{Var}\left[-\textstyle\int_0^1 \mathbb{W}\, dK_N/\sigma_{aN}\right] = \textstyle\int_0^1\int_0^1 [s \wedge t - st]\, dK_N(s)\, dK_N(t)/\sigma_{aN}^2$$

$$(6) \qquad = \mathrm{Var}[K_N(\xi)]/\sigma_{aN}^2 = 1.$$

We will also consider the sum of independent rvs given by

$$(7) \qquad \begin{aligned} T_N^0 &\equiv T_N^0(\mathsf{a}_N) \equiv \tfrac{-1}{\sigma_{aN}} \textstyle\int_0^1 \mathbb{W}_N \, dK_N = \tfrac{1}{\sigma_{aN}} \textstyle\int_0^1 K_N \, d\mathbb{W}_N \\ &= \tfrac{1}{\sqrt{N}} \textstyle\sum_{i=1}^n \left[\, c_{Ni} K_N(\xi_{Ni}) \,\right] / \, \sigma_{aN} \cong (0,1). \end{aligned}$$

We next show that $T_N =_a T_N^0 =_a Z_N \cong N(0,1)$ under rather mild conditions. $\square$

Exercise 2.1 Show that $\mathrm{E}\, T_N = 0$ and $\mathrm{Var}[T_N] = N/(N-1)$.

Definition 2.1 ($\mathcal{D}(\mathsf{a}_N)$-negligibility) Call such a_N a *negligible* array if

(8) $$\mathcal{D}(\mathsf{a}_N) \equiv \frac{\max_{1\le i\le N} |a_{Ni} - a_{N\cdot}|}{\sqrt{N}\,\sigma_{aN}} \le \ (\text{some } \epsilon_N) \searrow 0\,.$$

We let $\mathcal{A}$ denote any collection of such arrays that uses the same ϵ_N's. When the a_{Ni} are random, we call them

p-negligible if $\mathcal{D}(\mathsf{a}_N) \to_p 0$ and *a.s.-negligible* if $\mathcal{D}(\mathsf{a}_N) \to_{\text{a.s.}} 0$.

Theorem 2.1 (Uniform CLT for linear rank statistics) Suppose that the regression constants satisfy $\overline{c_N^4} \le M < \infty$ for all N. Let $\mathcal{A}$ be a collection of uniformly negligible arrays, in that $\epsilon_N \equiv \sup_{\mathcal{A}} \mathcal{D}(\mathsf{a}_N) \to 0$. Then

(9) $$\sup_{\mathcal{A}} |T_N(\mathsf{a}_N) - Z(\mathsf{a}_N)| \to_p 0 \quad \text{and} \quad \sup_{\mathcal{A}} |T_N^0(\mathsf{a}_N) - Z(\mathsf{a}_N)| \to_p 0\,.$$

Proof. Fix $0 \le \nu < \frac{1}{4}$. Now, $\gamma(\mathsf{a}_N) \equiv T_N - Z_N$ satisfies

(a) $$|\gamma(\mathsf{a}_N)| = |T_N - Z_N| = |-\textstyle\int_{1/(N+1)}^{N/(N+1)} (\mathbb{R}_N - \mathbb{W})\, dK_N/\sigma_{aN}|$$

(b) $$\le \left\| \frac{N^\nu(\mathbb{R}_N - \mathbb{W})}{[I(1-I)]^{1/2-\nu}} \right\|_{1/(N+1)}^{N/(N+1)} \times \int_{1/(N+1)}^{N/(N+1)} N^{-\nu}[t(1-t)]^{1/2-\nu}\, dK_N(t)/\sigma_{aN}$$

(10) $$\equiv \ddot{\Delta}_{\nu N} \times M_{\nu N}(\tilde{K}_N),$$

where $\ddot{\Delta}_{\nu N} = O_p(1)$ by (12.10.35). Thus (**??.??.??**) gives

$$|T_N - Z_N| \le \ddot{\Delta}_{\nu N}$$
$$\times \left\{ \frac{\sqrt{9/\nu}}{r^\nu} + \sqrt{r}\,\frac{|K_N(1/(N+1)) - a_{N\cdot}|}{\sqrt{N}\sigma_{aN}} + \sqrt{r}\,\frac{|K_N(N/(N+1)) - a_{N\cdot}|}{\sqrt{N}\sigma_{aN}} \right\}$$

(c) $$\le \ddot{\Delta}_{\nu N} \times \{\sqrt{9/\nu}\,/r^\nu + 2\sqrt{r}\,[\max_{1\le i\le N} |a_{Ni} - a_{N\cdot}|/\sqrt{N}\,\sigma_{aN}]\}$$

(11) $$\le \ddot{\Delta}_{\nu N} \times \{\sqrt{9/\nu}\,/r^\nu + 2\sqrt{r}\,\mathcal{D}(\mathsf{a}_N)\}.$$

By first choosing r large and then letting $N \to \infty$, we see that

(d) $$|T_N - Z_N| = O_p(1) \times o(1) \to_p 0\,.$$

Note that we have separated the randomness from properties of the a_{Ni}'s, so that the convergence is uniform over $\mathcal{A}$. [The rate for T_N in (9) depends on the sequence $\epsilon_N \to 0$, and on the c_{Ni}'s only through the M_ϵ's in the statement $P(\ddot{\Delta}_{\nu N} \ge M_\epsilon) \le \epsilon$ for all N.] Since (see (12.10.34))

(e) $$|T_N^0 - Z_N| = |\textstyle\int_{1/(N+1)}^{N/(N+1)} (\mathbb{W}_N - \mathbb{W})\, dK_N/\sigma_{aN}| \le \dot{\Delta}_{\nu N} \times M_{\nu N}(\tilde{K}_N),$$

comparing (e) with (a) and (10) shows that the proof is the same in this case.

[One can also allow random regression constants c_N and scores a_N that are independent of the antiranks f0or which $\overline{c_N^4} = O_p(1)$ and $\sup_{\mathcal{A}} \mathcal{D}(\mathsf{a}_N) \to_p 0$.] □

Example 2.2 (Creating normality) The following are known, and intuitive.
(A) *(Using normal regression constants)* When $\overline{\lim}\ \overline{c_N^4} \le \infty$, present methods give

(12) $\quad T_N \to_d N(0,1) \quad$ if and only if $\quad$ either $\mathcal{D}(\mathsf{a}_N) \to 0$ or $c_{ND_{N1}} \to_d N(0,1)$.

Result 1 Thus, with absolutely no hypotheses,

(13) $\quad$ the choice $c_{Ni} \equiv \Phi^{-1}(i/(N+1))$
always gives $T_N \to_d N(0,1)$ for *every* choice of a_{Ni}'s.

(B) *(Winsorizing a finite sample)* Let $\tilde{a}_{N\cdot}$, $\tilde{\sigma}_{aN}$, $\tilde{T}_N$, and $\tilde{Z}_N$ be defined as before, but now based on the (k_N, k'_N)-*Winsorized population* $\tilde{\vec{a}}_N$ consisting of

(14) $\quad a_{N,k_N+1}, \ldots, a_{N,k_N+1};\ a_{N,k_N+1}, \ldots, a_{N,N-k'_N};\ a_{N,N-k'_N}, \ldots, a_{N,N-k'_N}.$

Of course, theorem 2.1 also applies to $\tilde{T}_N$. But note that now

$$\begin{aligned}
\mathcal{D}(\tilde{\mathsf{a}}_N) &= \max_{1\le i\le N} |\tilde{a}_{Ni} - \tilde{a}_{N\cdot}|/\sqrt{N}\,\tilde{\sigma}_{aN} \\
&= \left[\, |a_{N,k_N+1} - \tilde{a}_{N\cdot}| \vee |a_{N,N-k'_N} - \tilde{a}_{N\cdot}| \right] / (\sqrt{N}\,\tilde{\sigma}_{aN}) \\
&\le \frac{|a_{N,k_N+1} - \tilde{a}_{N\cdot}| \vee |a_{N,N-k'_N} - \tilde{a}_{N\cdot}|}{\left\{(k_N+1)(a_{N,k_N+1} - \tilde{a}_{N\cdot})^2 + (k'_N+1)(a_{N,N-k'_N} - \tilde{a}_{N\cdot})^2\right\}^{1/2}} \\
&\le 1/\sqrt{(k_N \wedge k'_N)+1}, \qquad \text{provided only that } a_{N,k_N+1} < a_{N,N-k'_N}. \qquad (15)
\end{aligned}$$

Thus $\mathcal{D}(\tilde{\mathsf{a}}_N) \to 0$ whenever

(16) $\quad k_N \wedge k'_N \to \infty,\ $ and $a_{N,k_N+1} < a_{N,N-k'_N}$ for all N sufficiently large.

Result 2 Suppose $\overline{c_N^4} \le M < \infty$ for all N. For fixed $k_N \wedge k'_N \to \infty$, we have

(17) $\quad \sup_{\mathcal{A}} \left\{ \left| \tilde{T}_N(\mathsf{a}_N) - \tilde{Z}(\mathsf{a}_N) \right| : \text{all arrays in } \mathcal{A} \text{ with } a_{N,k_N+1} < a_{N,N-k'_N} \right\} \to_p 0.$

Summary Asymptotic normality is guaranteed by Winsorizing a number that slowly increases to infinity, provided only that we do not collapse the whole sample. □

Exercise 2.2 Argue heuristically why (13) should be true. [See Shorack(1996).]

Example 2.3 (Permutation statistics) Suppose $X_1, \ldots, X_N$ are iid rvs with nondegenerate df F on $(\Omega, \mathcal{A}, P)$. Then let $\mathsf{X}_N \equiv (X_1, \ldots, X_N)'$ denote the full population of observed values, having order statistics $X_{N:1} \le \cdots \le X_{N:N}$, antiranks $(D_{N1}, \ldots, D_{NN})$, sample mean $\bar{X}_N$, sample variance S_N^2, empirical df $\mathbb{F}_N$, and empirical qf $\mathbb{K}_N \equiv \mathbb{F}_N^{-1}$. Let $0 \le k_N < N - k'_N \le N$, and let $\tilde{\mathsf{X}}_N$ denote the (k_N, k'_N)-Winsorized population $\tilde{X}_{N:1} \le \cdots \le \tilde{X}_{N:N}$ (as in (14)) whose parameters are the Winsorized mean $\tilde{T}_N$, the Winsorized variance $\tilde{S}_N^2$, and empirical qf $\tilde{\mathbb{K}}_N$. We note that

(18) $\quad X_{N:1} \le \cdots \le X_{N:N}$ and $(D_{N1}, \ldots, D_{NN})$ are independent rvs,
if tied X_i's are randomly assigned their ranks.

We also recall (from theorem 10.7.1, or from exercise 8.4.20) that

(19) $\mathcal{D}(\mathsf{X}_N) \to_{a.s.} 0$ if and only if $0 < \mathrm{Var}[X] < \infty$.

(20) $\mathcal{D}(\mathsf{X}_N) \to_p 0$ if and only if $F \in D(\text{Normal})$.

Moreover, from (16),

(21) $\mathcal{D}(\tilde{\mathsf{X}}_N) \to_{a.s.} 0$ [making $\tilde{\mathsf{X}}_N$ a.s. negligible] for a.e. $X_1, X_2, \ldots$,

provided only that

(22) $k_N \wedge k'_N \to_{a.s.} \infty$ and $\underline{\lim}\,(X_{N:N-k'_N} - X_{N:k_N+1}) > 0$ a.s.

for k_N and k'_N that are either fixed integer sequences or integer-valued rvs that are independent of the antiranks D_N. Condition (22) *necessarily holds* if

(23) F is *any* nondegenerate *df*, and if $k_N \wedge k'_N \to \infty$ while $(k_N \vee k'_N)/N \to 0$.

[We shall maintain the order statistics (which are on $(\Omega, \mathcal{A}, P)$), but we can replace the independent antiranks by a realization (on some $(\Omega^*, \mathcal{A}^*, P^*)$ independent of $(\Omega, \mathcal{A}, P)$) for which $\ddot{\Delta}^*_{\nu N} = O_p(1)$ on $(\Omega^*, \mathcal{A}^*, P^*)$ (from (12.10.25)) for some Brownian bridge $\mathbb{W}$. This is possible whenever $\overline{c^4_N} \le M < \infty$ for all N (see theorem 12.10.3).]

By a *permutation statistic* we mean a rv of the form

(24) $$T_N \equiv T_N(\mathsf{X}_N) \equiv \frac{1}{\sqrt{N}} \sum_{i=1}^N c_{Ni}\, X_i / S_N = \frac{1}{\sqrt{N}} \sum_{i=1}^N c_{N D_{Ni}}\, X_{N:i}/S_N\,,$$

with $S^2_N = \sigma^2_{XN} = \sum_1^N (X_k - \bar{X}_N)^2/N$ and with a $\mathsf{c}_N \equiv (c_{N1}, \ldots, c_{NN})'$ population that is standardized. (Note that the distribution of T_N is unaltered by using this different realization of the antiranks.) □

Theorem 2.2 (Permutation tests) If $\ddot{\Delta}^*_{\nu N} = O_p(1)$ (as when $\overline{\lim}\ \overline{c^4_N} < \infty$), then the asymptotic normality

$$T_N =_a Z(\mathsf{X}_N) \cong N(0,1)$$

(25) $$\text{holds} \begin{cases} \text{on } (\Omega^*, \mathcal{A}^*, P^*) \text{ for a.e. } X_1,\ X_2,\ \ldots \text{ if } c_{Ni} = \Phi^{-1}(i/N+1), \\ \text{on } (\Omega^*, \mathcal{A}^*, P^*) \text{ for a.e. } X_1, X_2, \ldots \text{ if } 0 < \mathrm{Var}[X] < \infty, \\ \text{on } (\Omega \times \Omega^*, \mathcal{A} \times \mathcal{A}^*, P \times P^*) \qquad \text{if } F \in D(\text{Normal}). \end{cases}$$

[The convergence is uniform over classes $\mathcal{F}$ of dfs F in which $\mathcal{D}(\mathsf{X}_N) \to_p$ or $_{a.s.}$ 0 uniformly.] Also, whenever (22) (or (23)) holds a.e. we have

(26) $\tilde{T}_N =_a Z(\tilde{\mathsf{X}}_N) \cong N(0,1)$ on $(\Omega^*, \mathcal{A}^*, P^*)$ for a.e. $X_1, X_2, \ldots$.

Similar results hold for $T_N(\mathsf{Y}_N)$ and $\tilde{T}_N(\mathsf{Y}_N)$, where $\mathsf{Y}_N \equiv (Y_{N1}, \ldots, Y_{NN})'$ with

(27) $Y_{Ni} \equiv \hat{g}_N(X_i)$ for any function $\hat{g}_N(\cdot)$ independent of the antiranks D_N.

Proof. Equation (11) now becomes (recall (10) for $M_{\nu N}(\cdot)$)

(a) $|\gamma(\mathsf{X}_N)| \equiv |T_N - Z(\mathsf{X}_N)| = |-\int_0^1 (\mathbb{R}_N - \mathbb{W})\, d\mathbb{K}_N/S_N| \le \ddot{\Delta}^*_{\nu N} \times M_{\nu N}(\mathbb{K}_N)$

(28) $\le \ddot{\Delta}^*_{\nu N} \times \{\sqrt{9/\nu}/r^{\nu} + 2\sqrt{r}\,\mathcal{D}(\mathsf{X}_N)\}$

(b) $= O_p(1) \times o(1) \to_p 0,$ as in (25),

using either (19), (20) on subsequences, or the proof of (13) on $\mathcal{D}(\mathsf{X}_N)$. So, (25) holds. Likewise, conclusion (26) holds by using (22) to apply (15) to $\mathcal{D}(\tilde{\mathsf{X}}_N)$. □

Sampling from Finite Populations

Example 2.4 (Simple random sampling) Let $X_1, \ldots, X_n$ be a random sample without replacement from an $\mathsf{a}_N \equiv (a_{N1}, \ldots, a_{NN})'$ population. As usual, let $\bar{X}_n$ and $S_n^2 \equiv \overline{X_n^2} - \bar{X}_n^2$ denote the sample mean and "sample variance." Suppose that $a_{N1} \le \cdots \le a_{NN}$, that $n \equiv n_N$, and that the $\mathcal{D}(\mathsf{a}_N)$ of (8) satisfy both

(29) $0 < \underline{\lim}\, n/N \le \overline{\lim}\, n/N < 1$ and

(30) $\sup_{\mathcal{A}} \mathcal{D}(\mathsf{a}_N) = \sup_{\mathcal{A}} \{ \max_{1\le i\le N} |a_{Ni} - a_{N\cdot}|/\sqrt{N}\,\sigma_{aN}\} \to 0\,.$

Prior to normalizing, the c_{Ni}'s consist of n values of 1 and $m \equiv N - n$ values of 0, with $c_{N\cdot} = -n/N$ and $\sigma_{cN}^2 = mn/N^2$. After normalizing,

$$\overline{c_N^4} = (m^3 + n^3)/(mnN) \le 2(m \vee n)/(m \wedge n)\,.$$

Thus (29) implies that all $\overline{c_N^4} \le$ (some M) $< \infty$. Since $c_{ND_{N1}} \to_d N(0,1)$ clearly fails, (12) shows that $T_N \to_d N(0,1)$ if and only if $\mathcal{D}(\mathsf{a}_N) \to 0$. The limiting rv (as in (5)) will be $Z_N \equiv Z(\mathsf{a}_N) \equiv -\int_0^1 \mathbb{W}\, dK_N/\sigma_{aN} \cong N(0,1)$. Now define

(31) $T_N \equiv T_N(\mathsf{a}_N) \equiv \dfrac{\sqrt{n}\,(\bar{X}_n - a_{N\cdot})}{\sigma_{aN}\sqrt{1 - n/N}} = -\int_0^1 \mathbb{R}_N\, dK_N/\sigma_{aN} \cong (0, N/(N-1))\,,$

(32) $\hat{T}_N \equiv \hat{T}_N(\mathsf{a}_N) \equiv \dfrac{\sqrt{n}\,(\bar{X}_n - a_{N\cdot})}{S_n\sqrt{1 - n/N}}.$ □

Theorem 2.3 (Simple random sampling) Suppose (29) holds, and suppose the arrays $\mathcal{A}$ are uniformly negligible with $\sup_{\mathcal{A}} \mathcal{D}(\mathsf{a}_N) \to 0$ (as in (30)). Then

(33) $\sup_{\mathcal{A}} |T_N(\mathsf{a}_N) - Z(\mathsf{a}_N)| \to_p 0,$

(34) $\sup_{\mathcal{A}} |\frac{1}{\sigma_{aN}} S_n - 1| \to_p 0,$

(35) $\hat{T}_N(\mathsf{a}_N) - Z(\mathsf{a}_N) \to_p 0.$ In fact, $\sup_{\mathcal{A}} |\hat{T}_N(\mathsf{a}_N) - Z(\mathsf{a}_N)| \to_p 0$

if (36) also holds. That is, the uniform convergence conclusion in (35) holds if

(36) $\sup_N \int_{1/(N+1)}^{N/(N+1)} g(t)\, d\frac{K_N(t) - a_{N\cdot}}{\sigma_{aN}} < \infty$

for $g(t) = b(t)\,[t \wedge (1-t)]^{1/2}$ and $b(t) = b(1-t) = 1 \vee [2\log_2 1/t]^{1/2}$ for $t \in [0, 1/2]$.

Exercise 2.3 (a) Show that (30) and (36) both hold whenever the qfs $K_N(\cdot)$ have a uniformly bounded $2+\delta$ moment for any $\delta > 0$.
(b) Devise a "logarithmic moment" that will suffice.

Proof. Now, (33) follows from (9). Consider (34). Let $d_N \equiv \sqrt{m/nN}$. Simple algebra (start on the right) gives

(a) $$\frac{S_n^2 - \sigma_{aN}^2}{\sigma_{aN}^2} = \left\{ \frac{\frac{1}{n}\sum_1^n (X_i - a_{N\cdot})^2 - \sigma_{aN}^2}{\sigma_{aN}^2} \right\} - \left\{ \frac{\bar{X}_n - a_{N\cdot}}{\sigma_{aN}} \right\}^2 \equiv I_{2n} - I_{1n}^2 \,.$$

Using Chebyshev's inequality with the finite sampling variance of (A.1.9) yields

(b) $$P(|I_{1n}| \ge \epsilon) \le \mathrm{Var}[\bar{X}_n]/\epsilon^2\sigma_{aN}^2 = [1 - \tfrac{n-1}{N-1}]/n\epsilon^2 \to 0.$$

Letting $Y_i \equiv (X_i - a_{N\cdot})^2$, we use (A.1.9) again for

(c) $$P(|I_{2n}| \ge \epsilon) = P(|\bar{Y}_n - \mu_Y| \ge \epsilon\,\sigma_{aN}^2) \le \mathrm{Var}[\bar{Y}_n]/\epsilon^2\sigma_{aN}^4$$

(d) $$\le \frac{\sigma_Y^2}{n\,\epsilon^2\,\sigma_{aN}^4}\left[1 - \frac{n-1}{N-1}\right] \le \frac{\mathrm{E}(X - a_{N\cdot})^4}{n\,\epsilon^2\,\sigma_{aN}^4}\,\frac{m}{N-1}$$

$$= \frac{\sum_1^N (X_i - a_{N\cdot})^4/N}{n\,\epsilon^2\,\sigma_{aN}^4}\,\frac{m}{N-1} \le \mathcal{D}^2(\mathsf{a}_N)\,\frac{\sigma_{aN}^2}{\sigma_{aN}^2}\,\frac{2m}{\epsilon^2\,n}$$

(e) $$\le \mathcal{D}^2(\mathsf{a}_N)\,\frac{2m}{\epsilon^2\,n} \to 0 \qquad \text{by (29).}$$

Thus (34) holds. Then (33) and (34) gives the first claim in (35) via Slutsky's theorem. (Note that (29) uniformly bounds the ratio $\sqrt{m/n}$.) The second claim made in (35) will now follow from the identity

(f) $$(\hat{T}_N - Z_N) = (T_N - Z_N) \times \{[(\sigma_{aN}/S_n) - 1] + 1\} + Z_N \times [(\sigma_{aN}/S_n) - 1],$$

provided we show that

(37) $$\sup_N |Z(\mathsf{a}_N)| = O_p(1).$$

The proof that (36) implies (37) is found in the 1*st* Edition. □

Remark 2.1 At this point in the 1*st* Edition, the next section was used to rederive the bootstrap results of Chapter 10 using the present methods instead. □

3 L-Statistics

Let $K \equiv F^{-1}$, and define $X_{ni} \equiv K(\xi_{ni})$, for $1 \le i \le n$, in terms of the Uniform $(0,1)$ rvs $\xi_{n1}, \ldots, \xi_{nn}$ of notation 15.1.3. Then $X_{n1}, \ldots, X_{nn}$ are iid F, and we let $X_{n:1} \le \cdots \le X_{n:n}$ denote the order statistics. Suppose the statistician specifies a known $\nearrow$ and left-continuous function h, known constants $c_{n1}, \ldots, c_{nn}$, and known integers $0 \le k_n < n - k'_n \le n$. We wish to establish the asymptotic normality of the *trimmed* L-*statistic*

$$(1) \qquad L_n \equiv L_n(k_n, k'_n) \equiv \frac{1}{n} \sum_{i=k_n+1}^{n-k'_n} c_{ni}\, h(X_{n:i}) = \frac{1}{n} \sum_{i=k_n+1}^{n-k'_n} c_{ni}\, H(\xi_{n:i}),$$

where $H \equiv h(F^{-1}) = h(K)$ is also $\nearrow$ and left continuous. [Other useful cases such as $h(x) = x^2$ are dealt with by considering $(H^-)^2$ and $(H^+)^2$ separately, and then adding the results. Here $H^- \equiv -H \cdot 1_{[H \le 0]}$ and $H^+ \equiv H \cdot 1_{[H \ge 0]}$ denote the negative and positive parts of H. Thus there is no theoretical loss in now assuming that $h(X) = X$ and $H = F^{-1}$.]

Now, $\mathbb{G}_n$ and $\mathbb{U}_n$ denote the empirical df and the empirical process of those specially constructed ξ_{ni}'s of notation 15.1.2, whose empirical process $\mathbb{U}_n$ converges pathwise to the Brownian bridge $\mathbb{U}$ in the manner described. This will figure heavily in our proofs and in $=_a$ claims, but not in any $\to_d$ claims.

We need a centering constant μ_n for L_n. We define

$$(2) \qquad J_n(t) = c_{ni} \qquad \text{for } (i-1)/n < t < i/n \ \text{ and } 1 \le i \le n,$$

where the value of J_n at the i/n points is totally inconsequential. Suppose that J_n "converges" to J in some sense. Define $a_n \equiv k_n/n$, $a'_n \equiv k'_n/n$ as before, and then define centering constants

$$(3) \qquad \mu_n \equiv \int_{a_n}^{1-a'_n} J_n(t) H(t)\, dt \qquad \text{and} \qquad \mu_n^0 \equiv \int_{a_n}^{1-a'_n} J(t) H(t)\, dt$$

with $\mu^0 \equiv \mu_n^0(0,0) = \int_0^1 J(t)H(t)\,dt$. Note that $\mu_n = \sum_{i=k_n+1}^{n-k'_n} c_{ni} \int_{(i-1)/n}^{i/n} H(t)\,dt$, which means that $k_n = 0$ and $c_{n1} > 0$ (that $k'_n = 0$ and $c_{nn} > 0$) entails the added requirement that $\mathrm{E}H^-(\xi)$ be finite (that $\mathrm{E}H^+(\xi)$ be finite), for $\xi \cong$ Uniform (0,1). [Our main interest is in μ_n, while μ_n^0 is secondary; μ_n is *the* data analysis constant, while μ_n^0 is just a constant for theory.]

It is convenient to assume that on $(0,1)$

$$(4) \qquad J_n \ge 0, \quad J \ge 0 \text{ is continuous}, \quad \text{and } H \text{ is } \nearrow \text{ and left continuous.}$$

[More generally, we can apply our results for two separate J functions and then just subtract the results.] Now specify $a. \in (0,1)$ to satisfy $H_+(a.) = 0$, and define

$$(5) \qquad K(t) \equiv \int_{(a.,t)} J(s)\, dH(s),$$

where $\int_{(a.,t)} \equiv -\int_{[t,a.]}$. (But set $a. = 0$ if $H(\cdot) \ge 0$, and use $\int_{[0,t)}$ in (5); and set $a. = 1$ if $H(\cdot) \le 0$, and use $\int_{[t,1]}$ in (5).) Thus (in case (5))

$$(6) \qquad K \text{ is } \nearrow \text{ and left continuous on } (0,1) \text{ with } \quad K_+(a.) = 0,$$

where K_+ is the right-continuous version, and $\Delta K \equiv K_+ - K$. [Since $L_n - \mu_n$ is invariant under vertical shift, there is actually no *theoretical* loss in also assuming as we did above that H satisfies $H_+(a.) = 0$.] Since K is a qf, the *unobservable* rvs

(7) $$Y_{ni} \equiv K(\xi_{ni}), \text{ for } 1 \le i \le n, \text{ are iid with qf } K, \text{ and let } \bar{Y}_n \equiv \frac{1}{n}\sum_{i=1}^n Y_{ni}\,.$$

The most historically important case obtains when

(8) $$\sigma^2 \equiv \mathrm{Var}[K(\xi)] \in (0,\infty), \qquad \mu \equiv \mathrm{E}K(\xi), \qquad \text{and} \qquad k_n = k_n' = 0.$$

In this case we would *desire to show* that (on the same probability space where the special ξ_{ni}'s above are defined) for some $N(0,1)$ rv that we will denote by Z_K (or alternatively, and suggestively, we will also denote by $\int_0^1 K\, d\mathbb{U}/\sigma$) we have

(9) $$\sqrt{n}\,[L_n(0,0) - \mu_n(0,0)]/\sigma =_a \sqrt{n}\,(\bar{Y}_n - \mu)/\sigma =_a Z_K \equiv \int_0^1 K\, d\mathbb{U}/\sigma \cong N(0,1).$$

We would also like a consistent estimator of σ, and we might want to be able to replace $\mu_n(0,0)$ by $\mu^0 = \int_0^1 J(t)H(t)\, dt$. Of course, J_n will have to approximate J sufficiently closely.

Let $\tilde{K}_n$ denote K Winsorized outside $(a_n, 1 - a_n')$, and define the *unobservable* Winsorized rvs

(10) $$\tilde{Y}_{ni} \equiv \tilde{K}_n(\xi_{ni}) \qquad \text{for } 1 \le i \le n.$$

Then $\tilde{Y}_{n1}, \ldots, \tilde{Y}_{nn}$ are iid with qf $\tilde{K}_n$ and mean $\tilde{\mu}_n$ and variance $\tilde{\sigma}_n^2$ given by

(11) $$\tilde{\mu}_n \equiv \mathrm{E}\tilde{Y}_{ni} = \int_0^1 \tilde{K}_n(t)\, dt \quad \text{and} \quad \tilde{\sigma}_n^2 \equiv \mathrm{Var}[\tilde{Y}_{ni}] = \mathrm{Var}[\tilde{K}_n(\xi)]\,.$$

(We can allow only $k_n = 0$ or $k_n' = 0$ if the variance double integral is finite.) Let

(12) $$\tilde{Y}_{n\cdot} \equiv \tilde{Y}_n(a_n, a_n') \equiv \frac{1}{n}\sum_{i=1}^n \tilde{Y}_{ni} = \int_0^1 \tilde{K}_n\, d\mathbb{G}_n\,.$$

In this case it is our *desire to show* that

(13) $$\sqrt{n}\,(L_n - \mu_n)/\tilde{\sigma}_n =_a \sqrt{n}\,(\tilde{Y}_{n\cdot} - \tilde{\mu}_n)/\tilde{\sigma}_n = \int_0^1 \tilde{K}_n\, d\mathbb{U}_n/\tilde{\sigma}_n = -\int_0^1 \mathbb{U}_n\, d\tilde{K}_n/\tilde{\sigma}_n$$

(14) $$=_a Z_K(a_n, a_n') \equiv \int_0^1 \tilde{K}_n\, d\mathbb{U}/\tilde{\sigma}_n \cong N(0,1)\,.$$

We also seek an appropriate estimator of $\tilde{\sigma}_n$, and we may want to be able to replace μ_n by μ_n^0. Whenever $k_n \wedge k_n' \ge 1$, we always define the symbol $\int_0^1 \tilde{K}_n\, d\mathbb{U}$ to mean $-\int_0^1 \mathbb{U}\, d\tilde{K}_n$ (that is, pathwise integration for each ω) for *all* qfs.

Make *throughout* without further comment the rather modest assumptions that a_n and a_n' satisfy $\liminf(1 - a_n - a_n') > 0$ and $\liminf \tilde{\sigma}_n = \liminf \sigma_K(a_n, a_n') > 0$ for the df F or F_0 under consideration at the particular moment. The first says that we will deal with averaging, rather than just "quantiles" (though we could easily have added in fixed quantiles had we chosen to do so). The second says that the statistician has at least enough insight to avoid removing all the variation.

The now state the two most elementary theorems about L-statistics found in the 1*st* Edition. The results found there establish uniform convergence to normality over large classes of dfs, and they present studentized versions of such results. This is a very complete treatment of L-statistics. Roughly, suppose the finite sample score function J_n function is sufficiently close to a limiting score function J, as defined in the 1*st* Edition. Then any asymptotic normality theorem for the mean (whether trimmed or untrimmed) of a sample from the df K defined in (5) is also true for the corresponding L-statistic of (1) based on samples from the df F.

Theorem 3.1 (CLT for L-Statistics) Suppose the score function $J(\cdot)$ of (4) is approximated "sufficiently closely" by a sequence $J_n(\cdot)$. Let the statistic L_n in (1) be untrimmed. Suppose also that $Y \equiv K(\xi) \cong (\mu, \sigma^2)$ with $\sigma^2 \in (0, \infty)$ for the K of (5). Let $\tilde{\mu}_n$ and $\tilde{\sigma}_n$ be as in (11) (with $a_n = a'_n = 1/n$, for the sake of the proof). Let μ_n be as in (3). Then (in the context of notation 15.1.3)

(15) $$\sqrt{n}\,(L_n - \mu_n)/\sigma =_a \sqrt{n}\,(\bar{Y}_n - \mu)/\sigma =_a Z_n(K) \equiv \int_{1/n}^{1-1/n} \mathbb{U}\, dK/\tilde{\sigma}_n \cong N(0,1)\,.$$

(Moreover, $V_n/\sigma \to_p 1$ for an estimator V_n^2 of σ^2 presented in the 1*st* Edition.)

Theorem 3.2 (CLT for trimmed L-statistics) Suppose that $J(\cdot)$ as in (4) is approximated "sufficiently closely" by a sequence $J_n(\cdot)$. Suppose the statistician protects himself by specifying trimming numbers k_n and k'_n for which $k_n \wedge k'_n \to \infty$, while $a_n \vee a'_n \to 0$ with $a_n/a'_n \to 1$, and suppose that K is in the statistical domain of attraction $\tilde{\mathcal{D}}$ (recall (C.5.33), (15.1.3), and proposition 10.6.1). Then (in the context of notation 15.1.3)

(16) $$\sqrt{n}\,(L_n - \mu_n)/\tilde{\sigma}_n =_a \sqrt{n}\,(\tilde{Y}_{n\cdot} - \tilde{\mu}_n)/\tilde{\sigma}_n$$

$$= -\int_{a_n}^{1-a'_n} \mathbb{U}_n\, dK/\tilde{\sigma}_n = -\int_0^1 \mathbb{U}_n\, d\tilde{K}_n/\tilde{\sigma}_n,$$

(17) $$=_a Z_n \equiv Z_K(a_n, a'_n) \equiv -\int_{a_n}^{1-a'_n} \mathbb{U}\, dK/\tilde{\sigma}_n = -\int_0^1 \mathbb{U}\, d\tilde{K}_n/\tilde{\sigma}_n \cong N(0,1)\,.$$

(Moreover, $\tilde{V}_n/\tilde{\sigma}_n \to_p 1$ for an estimator $\tilde{V}_n^2$ of $\tilde{\sigma}_n^2$ presented in the 1*st* Edition.) [If $K \in \mathcal{D}$(Normal), only $(k_n \wedge k'_n) \geq 1$ is required and $a_n/a'_n \to 1$ may be omitted.]

Appendix A

Special Distributions

1 Elementary Probability

Independent Bernoulli Trials

If $P(X = 1) = p = 1 - P(X = 0)$, then X is said to be a *Bernoulli*(p) rv. We refer to the event $[X = 1]$ as "success," and $[X = 0]$ as "failure." Let $X_1, \ldots, X_n$ be iid Bernoulli(p), and let $T_n \equiv X_1 + \cdots + X_n$ denote the number of successes in n independent Bernoulli(p) trials. Now,

$$P(X_i = x_i \text{ for } 1 \leq i \leq n) = p^{\sum_1^n x_i}(1-p)^{n - \sum_1^n x_i} \qquad \text{if all } x_i \text{ equal 0 or 1;}$$

this formula gives the joint distribution of $X_1, \ldots, X_n$. From this we obtain

$$(1) \qquad P(T_n = k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad \text{for } 0 \leq k \leq n,$$

since each of the $\binom{n}{k}$ different possibilities that place k of the 1's in specific positions in an n-vector containing k outcomes 1 and $n - k$ outcomes 0 has probability $p^k(1-p)^{n-k}$, from the earlier display. We denote this by writing $T_n \cong$ *Binomial*(n, p) when (1) holds. Note that Binomial($1, p$) is the same as Bernoulli(p).

Let $X_1, X_2, \ldots$ be iid Bernoulli(p); call this a *Bernoulli*(p) *process.* Interesting rvs include $Y_1 \equiv W_1 \equiv \min\{n : T_n = 1\}$. Since we can rewrite the event $[Y_1 = k] = [X_1 = \cdots = X_{k-1} = 0, X_k = 1]$, we have

$$(2) \qquad P(Y_1 = k) = (1-p)^{k-1} p \qquad \text{for } k = 1, 2, \ldots.$$

We write $Y_1 \cong$ *GeometricT*(p). Now let $W_m \equiv \min\{n : T_n = m\}$. We call W_m the *waiting time* to the mth success; W_m counts the number of *turns* until the mth success. We let $Y_m \equiv W_m - W_{m-1}$ for $m \geq 1$, with $W_0 \equiv 0$, and we call the Y_m's the *interarrival times.* Note that $[W_m = k] = [T_{k-1} = m - 1 \text{ and } X_k = 1]$. Hence

$$(3) \qquad P(W_m = k) = \binom{k-1}{m-1} p^m (1-p)^{k-m} \qquad \text{for } k = m, m+1, \ldots.$$

We write $W_m \cong$ *Negative Binomial Turns*(m, p) $\equiv$ *NegBiT*(m, p). [We agree that *NegBiF*(m, p) denotes the distribution of $W_m - m$, and that this "F" connotes "failures"; the rv $W_m - m$ counts the number of *failures* prior to the mth success.]

Exercise 1.1 Explain why $Y_1, Y_2, \ldots$ are iid GeometricT(p).

Since the number of successes in the first n_1+n_2 trials is the same as the number of successes in the first n_1 trials plus the number of successes in the next n_2 trials, it is clear that

$$(4) \qquad T_1+T_2 \cong \text{Binomial}(n_1+n_2, p) \qquad \text{for independent rvs } T_i \cong \text{Binomial}(n_i, p).$$

Likewise, waiting for m_1 successes and then waiting for m_2 more successes is the same as waiting for $m_1 + m_2$ successes in the first place. Hence,

$$(5) \qquad W_1+W_2 \cong \text{NegBiT}(m_1+m_2, p) \quad \text{for independent rvs } W_i \cong \text{NegBiT}(m_i, p).$$

Urn Models

Suppose an urn contains N balls that are identical, except that M bear the number 1 and $N - M$ bear the number 0. Thoroughly mix the balls in the urn. Draw one ball at random. Let X_1 denote the number on the ball drawn. Then $X_1 \cong$ Bernoulli(p) with $p \equiv M/N$. Now replace the ball in the urn, thoroughly mix, and draw at random a second ball with number X_2. Continue the process. This is the *sampling with replacement* scheme. Then $T_n \equiv X_1 + \cdots + X_n \cong \text{Binomial}(n, p)$, where $p = M/N$ represents the probability of success in n independent Bernoulli(p) trials.

Suppose now that the same scheme is repeated, except that the balls are not replaced. In this *sampling without replacement* scheme $X_1, \ldots, X_n$ are dependent Bernoulli(p) rvs with $p = M/N$. Also,

$$(6) \qquad P(T_n = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}, \qquad \text{provided that the value } k \text{ is possible.}$$

We write $T_n \cong$ *Hypergeometric*$(M, N - M; n)$.

Suppose now that sampling is done without replacement, but the N balls in the urn bear the numbers $a_1, \ldots, a_N$. Let $X_1, \ldots, X_n$ denote the numbers on the first n balls drawn, and let $T_n \equiv X_1 + \cdots + X_n$. We call this the general *finite sampling model*. Call $\bar{a} \equiv \sum_1^N a_i/N$ the *population mean* and $\sigma_a^2 \equiv \sum_1^N (a_i - \bar{a})^2/N$ the *population variance*. Note that $X_i \cong (\bar{a}, \sigma_a^2)$ for all $1 \le i \le n$, since we now assume $n \le N$. From (6.3.4), we have

$$(7) \qquad 0 = \text{Var}[\textstyle\sum_1^N X_i] = N\,\text{Var}[X_1] + N(N-1)\text{Cov}[X_1, X_2],$$

with the 0 valid, since $\sum_1^N X_i$ is a constant. Solving (7) yields

$$(8) \qquad \text{Cov}[X_1, X_2] = -\sigma_a^2/(N-1).$$

As in (7), and using (8), $\text{Var}[T_n] = n\sigma_a^2 - n(n-1)\sigma_a^2/(N-1)$. Thus

$$(9) \qquad \text{Var}[T_n/n] = \frac{1}{n}\,\sigma_a^2\left[1 - \frac{n-1}{N-1}\right],$$

where $[1 - (n-1)/(N-1)]$ is called the *correction factor for finite population sampling.*

Exercise 1.2 Verify (8) and (9).

Exercise 1.3 Suppose that $T_1 \cong \text{Binomial}(m, p)$ and $T_2 \cong \text{Binomial}(n, p)$ are independent. Then the conditional distribution of T_1 given that $T_1 + T_2 = k$ is $\text{Hypergeometric}(k, m + n - k; m)$.

The Poisson Process

Suppose now that $X_{n1}, \ldots, X_{nn}$ are iid Bernoulli(p_n), where $np_n \to \lambda$ as $n \to \infty$. Let $T_n \equiv X_{n1} + \cdots + X_{nn}$, so that $T_n \cong \text{Binomial}(n, p_n)$. Simple calculations give

$$(10) \qquad P(T_n = k) \to \lambda^k e^{-\lambda}/k! \qquad \text{for } k = 0, 1, \ldots .$$

When $P(T = k) = \lambda^k e^{-\lambda}/k!$ for $k = 0, 1, \ldots$, we write $T \cong \textit{Poisson}(\lambda)$.

This is now used to model a Geiger counter experiment. A radioactive source with large half-life is placed near a Geiger counter. Let $\mathbb{N}(t)$ denote the number of particles registered by time t. We will say that $\{\mathbb{N}(t) : t \geq 0\}$ is a *Poisson process*. (Do note that our treatment is purely informal.) Physical considerations lead us to believe that the increments $\mathbb{N}(t_1), \mathbb{N}(t_1, t_2], \ldots, \mathbb{N}(t_{k-1}, t_k]$ should be independent rvs; here, the *increment* $\mathbb{N}(t_{i-1}, t_i] \equiv \mathbb{N}(t_i) - \mathbb{N}(t_{i-1})$ is the number of particle counts across the interval $(t_{i-1}, t_i]$. We say that $\mathbb{N}$ has *independent increments*. Let us now define

$$(11) \qquad \nu \equiv E\,\mathbb{N}(1) \equiv \text{ [the } \textit{intensity} \text{ of the process]}.$$

Let M denote the number of radioactive particles in our source, and let X_i equal 1 or 0 depending on whether or not the ith particle registers by time $t = 1$. It seems possible to assume that $X_1, \ldots, X_M$ are iid Bernoulli. Since $\mathbb{N}(1) = X_1 + \cdots + X_M$ has mean $\nu = E\,\mathbb{N}(1) = M\,EX_1$, this leads to $\mathbb{N}(1) \cong \text{Binomial}(M, \nu/M)$. By the first paragraph of this section, $\mathbb{N}(1)$ is thus approximately a Poisson(ν) rv. We now alter our point of view slightly, and agree that we will *use this approximation as our model*. Thus $\mathbb{N}(1)$ is a Poisson(ν) rv. Since M is huge, the accuracy should be superb. Because of the stationary and independent increments we thus have

$$(12) \qquad \mathbb{N}(s, t] \equiv \mathbb{N}(t) - \mathbb{N}(s) \cong \text{Poisson}(\nu(t - s)) \qquad \text{for all } 0 \leq s \leq t, \qquad \text{and}$$

$$(13) \qquad \mathbb{N} \text{ has independent increments.}$$

Agree also that $\mathbb{N}(0) \equiv 0$. (This is actually enough to rigorously specify a Poisson process.) Let $Y_1 \equiv W_1 \equiv \inf\{t : \mathbb{N}(t) = 1\}$. Since

$$(14) \qquad [Y_1 > t] = [\mathbb{N}(t) < 1] = [\mathbb{N}(t) = 0],$$

we see that $1 - F_{Y_1}(t) = P(Y_1 > t) = P(\mathbb{N}(t) = 0) = e^{-\nu t}$ by (12). Thus Y_1 has df $1 - \exp(-\nu t)$ for $t \geq 0$ and density

$$(15) \qquad f_{Y_1}(t) = \nu e^{-\nu t} \qquad \text{for } t \geq 0;$$

we write $Y_1 \cong \textit{Exponential}(\nu)$. Now let $W_m \equiv \inf\{t : N(t) = m\}$; we call W_m the mth *waiting time*. We call $Y_m \equiv W_m - W_{m-1}$, $m \geq 1$, the mth *interarrival time*.

In light of the physical properties of our Geiger counter model, and using (13), it seems reasonable that

(16) $\quad Y_1, Y_2, \ldots$ are iid Exponential(ν) rvs.

Our assumption of the previous sentence could be expressed as follows:

(17) $\quad Y_1$ and $\mathbb{N}_1(t) \equiv \mathbb{N}(Y_1, Y_1 + t] = \mathbb{N}(Y_1 + t) - \mathbb{N}(Y_1)$ are independent, $\mathbb{N}_1$ is again a Poisson process, with intensity ν.

We will call this the *strong Markov property* of the Poisson process. Additionally,

(18) $\quad [W_m > t] = [\mathbb{N}(t) < m]$,

so that $1 - F_{W_m}(t) = P(W_m > t) = \sum_{k=0}^{m-1} (\nu t)^k e^{-\nu t}/k!$; the derivate of this expression telescopes, and shows that W_m has density

(19) $\quad f_{W_m}(t) = \nu^m t^{m-1} e^{-\nu t}/\Gamma(m) \qquad$ for $t \geq 0$.

We write $W_m \cong$ *Gamma*(m, ν). Since waiting for m_1 counts and then waiting for m_2 more counts is the same as waiting for $m_1 + m_2$ counts in the first place,

(20) $\quad Z_1 + Z_2 \cong \text{Gamma}(m_1 + m_2, \nu) \qquad$ for independent $Z_i \cong \text{Gamma}(m_i, \nu)$.

It is true that (19) is a density for any real number $m > 0$, and the property (20) still holds for all positive m_i's.

Exercise 1.4 Verify (10), that Binomial$(n, p_n) \to$ Poisson(λ) as $np_n \to \lambda$.

Exercise 1.5 Verify (19), that F_{W_n} has derivative f_{W_n}.

Exercise 1.6 Verify that (20) holds for arbitrary real $m_i > 0$.

Exercise 1.7 If $X \cong$ Poisson(ν_1) and $Y \cong$Poisson(ν_2), then the conditional distribution of X given that $X + Y = n$ is Binomial$(n, \nu_1/(\nu_1 + \nu_2))$.

Exercise 1.8 Use Kolmogorov's extension theorem to show that a Poisson process $\mathbb{N}$ exists on $(R_{[0,\infty)}, \mathcal{B}_{[0,\infty)})$. Then apply the smoother realizations theorem 5.4.2.

Location and Scale

If $a > 0$, then

$$F_{aZ+b}(x) = P(aZ + b \leq x) = P(Z \leq (x - b)/a) = F_Z((x - b)/a)$$

holds for any $F_Z(\cdot)$. Thus for any density $f_Z(\cdot)$, the rv $aZ + b$ has density

(21) $\quad f_{aZ+b}(x) = \frac{1}{a} f_Z(\frac{x-b}{a}) \qquad$ for $-\infty < x < \infty$.

Normal Distributions

Suppose the rv Z has density

(22) $\quad f_Z(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2}) \qquad \text{for } -\infty < x < \infty;$

then Z is said to be a *standard normal rv.* So the rv $X \equiv \mu + \sigma Z \cong (\mu, \sigma^2)$ has density

(23) $\quad \frac{1}{\sqrt{2\pi}\,\sigma} \exp(-\frac{1}{2}(\frac{x-\mu}{\sigma})^2) \qquad \text{for } -\infty < x < \infty,$

and we write $X \cong$ *Normal*(μ, σ^2), or just $X \cong N(\mu, \sigma^2)$.

Exercise 1.9 Show that the formula $f_Z(\cdot)$ of (22) is a density. Then show that this density has mean 0 and variance 1. [Transform to polar coordinates to compute $(\int f_Z(x)\,dx)^2 = 1$.]

The importance of the normal distribution derives from the following theorem. Recall that if $X_1, \ldots, X_n$ are iid (μ, σ^2), then $\sqrt{n}\,(\bar{X} - \mu)/\sigma \cong (0,1)$ for the sample average $\bar{X}_n \equiv (X_1 + \cdots + X_n)/n$. This is only a statement about moments. But much more is true. The powerful result we now state will be proved in chapter **??**. We will use it in the meantime for motivational purposes.

Theorem 1.1 (Classical CLT) If $X_1, \ldots, X_n$ are iid (μ, σ^2), then

(24) $\quad \sqrt{n}\,(\bar{X}_n - \mu) \to_d N(0, \sigma^2) \qquad \text{as } n \to \infty.$

Let $\sigma > 0$. Then the Z_n below is *asymptotically normal*, in that

(25) $\quad Z_n \equiv \sqrt{n}\,(\bar{X}_n - \mu)/\sigma \to_d N(0,1) \qquad \text{as } n \to \infty.$

Suppose that Z is $N(0,1)$. Then

(26) $\quad F_{Z^2}(x) = P(Z^2 \le x) = P(-\sqrt{x} \le Z \le \sqrt{x}\,) = F_Z(\sqrt{x}\,) - F_Z(-\sqrt{x}\,);$

thus Z^2 has density

(27) $$f_{Z^2}(x) = \frac{1}{2\sqrt{x}}\,[f_Z(\sqrt{x}\,) - f_Z(-\sqrt{x}\,)] \qquad \text{for } x \ge 0.$$

[Note that formula (27) is true for any density $f_Z(\cdot)$.] Plugging into (27) for this Z shows that

(28) $\quad f_{Z^2}(x) = (2\pi x)^{-1/2} \exp(-x/2) \qquad \text{for } x \ge 0;$

this is called the Chisquare(1) distribution. Note that Chisquare(1) is the same as Gamma$(\frac{1}{2}, \frac{1}{2})$. Thus (20) establishes that

(29) $$\text{if } X_1, \ldots, X_m \text{ are iid } N(0,1), \text{ then } \sum_{i=1}^m X_i^2 \cong \text{Chisquare}(m),$$

where *Chisquare*$(m) \equiv$ Gamma$(\frac{m}{2}, \frac{1}{2})$.

Uniform and Related Distributions

Write $X \cong$ *Uniform*(a,b) if

(30) $\quad f_X(x) = \frac{1}{(b-a)} 1_{[a,b]}(x) = \frac{1}{(b-a)} \qquad$ on $[a,b]$.

By far the most important special case is Uniform$(0,1)$. A generalization of this is the Beta(c,d) family. We write $X \cong$ *Beta*(c,d) if

(31) $\quad f_X(x) = \frac{1}{\beta(c,d)}\, x^{c-1}(1-x)^{d-1}\, 1_{[0,1]}(x) = \frac{1}{\beta(c,d)}\, x^{c-1}(1-x)^{d-1} \qquad$ on $[0,1]$,

where $\beta(c,d) \equiv \Gamma(c)\Gamma(d)/\Gamma(c+d)$. Here, $b > 0$ and $c > 0$ are required.

Suppose that $\xi_1, \ldots, \xi_n$ are iid Uniform$(0,1)$. Let $0 \le \xi_{n:1} \le \cdots \le \xi_{n:n} \le 1$ denote the ordered values of the ξ_i's; we call the $\xi_{n:i}$'s the *uniform order statistics*. It seems intuitive that $\xi_{n:i}$ equals x if $(i-1)$ of the ξ_i's fall in $[0,x)$, 1 of the ξ_i's is equal to x, and $n-i$ of the ξ_i's fall in $(x,1]$. There are $n!/[(i-1)!(n-i)!]$ such designations of the ξ_{ni}'s, and for each such designation the "chance" of the rv's falling in the correct parts of $[0,1]$ is $x^{i-1}(1\cdot dx)(1-x)^{n-i}$. Thus

(32) $\quad f_{\xi_{n:i}}(x) = \dfrac{n!}{(i-1)!(n-1)!} x^{i-1}(1-x)^{n-i} 1_{[0,1]}(x),$ or $\xi_{n:i} \cong$ Beta$(i, n-i+1)$.

Exercise 1.10 Give a rigorous derivation of (32) by computing $1 - F_{\xi_{n:i}}(x)$ and then differentiating it.

Exercise 1.11 Choose a point at random on the surface of the unit sphere (with probability proportional to area). Let Θ denote the longitude and Φ denote the latitude (relative to some fixed axes) of the point so chosen. Determine the joint density of Θ and Φ.

The Cauchy Distribution

Write $X \cong$ *Cauchy*(b,a) if

(33) $\quad f_X(x) = 1/\{a\pi[1+(x-b)^2/a^2]\} \qquad$ on $(-\infty,\infty)$.

By far the most important special case is Cauchy$(0,1)$; we then say simply that $X \cong$ *Cauchy*, and its density is given by $1/[\pi(1+x^2)]$ on $(-\infty,\infty)$. Verify that $E|X| = \infty$. We will see below that if $X_1,\ldots,X_n$ are iid Cauchy, then the sample average $\bar{X}_n \equiv (X_1+\cdots+X_n)/n \cong$ Cauchy. These two facts make the Cauchy ideal for many counterexamples.

Double Exponential and Logistic Distributions

We say $X \cong$ *Double Exponential*(b,a) when $(X-b)/a$ has density $\frac{1}{2}\exp(-|x|)$ on the line. We say $X \cong$ *Logistic*(b,a) when $(X-b)/a$ has density $e^x/(1+e^x)^2 = 1/(e^{-x/2}+e^{x/2})^2$ on the line.

Exercise 1.12 Now, $X \equiv F^{-1}(\xi)$ has df F by the inverse transformation. So, compute F^{-1} for the Logistic$(0,1)$ and the Double Exponential$(0,1)$ distributions.

Rademacher Random Variables and Symmetrization

Many problems become simpler if the problem is symmetrized. One way of accomplishing this is by the appropriate introduction of Rademacher rvs. We say that ϵ is a *Rademacher rv* if $P(\epsilon = 1) = P(\epsilon = -1) = \frac{1}{2}$. Thus $\epsilon \cong 2$ Bernoulli$(\frac{1}{2}) - 1$.

We say that X is a *symmetric rv* if $X \cong -X$. If X and X' are iid, then $X^s \equiv (X - X') \cong (X' - X) = -(X - X') = -X^s$; hence X^s is a symmetric rv.

Exercise 1.13 If X is a symmetric rv independent of the Rademacher rv ϵ, then $X \cong \epsilon X$ always holds.

The Multinomial Distribution

Suppose that $B_1 + \cdots + B_k = R$ for Borel sets $B_i \in \mathcal{B}$; recall that we call this a *partition* of R. Let $Y_1, \ldots, Y_n$ be iid rvs on $(\Omega, \mathcal{A}, P)$. Let $\mathsf{X}_i \equiv (X_{i1}, \ldots, X_{ik})$ $\equiv (1_{B_1}(Y_i), \ldots, 1_{B_k}(Y_i))$ for $1 \le i \le n$, and set

$$\text{(34)} \qquad \mathsf{T} \equiv (T_1, \ldots, T_k)' = (\sum_{i=1}^n X_{i1}, \ldots, \sum_{i=1}^n X_{ik}) = (\sum_{i=1}^n 1_{B_1}(Y_i), \ldots, \sum_{i=1}^n 1_{B_k}(Y_i)).$$

Note that $X_{1j}, \ldots, X_{nj}$ are iid Bernoulli(p_j) with $p_j \equiv P(Y_i \in B_j)$, and thus $T_j \cong$ Binomial(n, p_j) (marginally). Note that $T_1, \ldots, T_n$ are dependent rvs. The joint distribution of $(T_1, \ldots, T_n)'$ is called the *Multinomial*(n, p) distribution. We now derive it. The number of ways to designate n_1 of the Y_i's to fall in $B_1, \ldots, B_k$, and n_k of the Y_i's to fall in B_k is the *multinomial coefficient*

$$\text{(35)} \qquad \binom{n}{n_1 \ldots n_k} \equiv \frac{n!}{n_1! \cdots n_k!}, \qquad \text{where } n_1 + \cdots + n_k = n.$$

Each such designation occurs with probability $\prod_1^k p_i^{n_i}$. Hence for each possible n,

$$\text{(36)} \qquad P(\mathsf{T} = \mathsf{n}) \equiv P(T_1 = n_1, \ldots, T_k = n_k) = \binom{n}{n_1 \ldots n_k} p_1^{n_1} \cdots p_k^{n_k}.$$

It is now a trivial calculation that

$$\text{(37)} \qquad \mathrm{Cov}[X_{ij}, X_{il}] = \mathrm{E}1_{B_j}(Y_i)1_{B_l}(Y_i) - \mathrm{E}1_{B_j}(Y_i)\mathrm{E}1_{B_l}(Y_i) = -p_j p_l \qquad \text{if } j \ne l.$$

Thus

$$\text{(38)} \qquad \mathrm{Cov}[T_j, T_l] = -n\, p_j\, p_l \qquad \text{for all } j \ne l.$$

Thus (with D_{p} a diagonal matrix having each $d_{ii} = p_i$)

$$\text{(39)} \qquad \begin{pmatrix} T_1 \\ \cdot \\ \cdot \\ \cdot \\ T_k \end{pmatrix} \cong n \left(\begin{pmatrix} p_1 \\ \cdot \\ \cdot \\ \cdot \\ p_k \end{pmatrix}, \begin{pmatrix} p_1(1-p_1) & & -p_1 p_k \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ -p_k p_1 & & p_k(1-p_k) \end{pmatrix} \right) = n(\, \mathsf{p}\, , [D_{\mathsf{p}} - \mathsf{p}\,\mathsf{p}'\,]\,).$$

Assorted Facts

Stirling's Formula For all $n > 1$ we have

(40) $\quad n! = e^{a_n}\, n^{n+1/2}\, e^{-n}\sqrt{2\pi}, \qquad \text{where } 1/(12n+1) < a_n < 1/(12n).$

Eulers's Constant

(41) $\quad \sum_{i=1}^{n} 1/i - \log n \uparrow \gamma \equiv 0.577215664901533\ldots .$

Exercise 1.14 (An added touch) If $\sum_1^\infty a_n < \infty$, there exists a $c_n \uparrow \infty$ such that $\sum_1^\infty c_n a_n < \infty$.

Elementary Conditional Probability

One defines the *conditional probability* of the event A given that the event B has occurred via $P(A|B) \equiv P(AB)/P(B)$ when $P(B) \neq 0$. One then calls A and B *independent* if $P(A|B) = P(A)$, because the probability of A is then unaffected by whether or not B occurred. Thus both of the following statements hold:

(42) Definition: $P(A|B) \equiv P(AB)/P(B)$
leads to Theorem: $P(AB) = P(B)P(A|B)$.

(43) Definition: Independence means $P(A|B) = P(A)$
leads to Theorem: $P(AB) = P(A)P(B)$ if A and B are independent.

The big advantage of computation of $P(A|B)$ via the theorem of (42) is that one can often *revisualize* $P(A|B)$ in the context of a much simpler problem. Thus the probability of drawing two Reds when drawing at random without replacement from an urn containing 6 Reds and 4 Whites is $P(R_1R_2) = P(R_1)P(R_2|R_1) = (6/10) \times (5/9)$, where we revisualized to an urn containing 5 Reds and 4 Whites to compute $P(R_2|R_1) = 5/9$. [Had we used sampling with replacement, our answer would have been $(6/10) \times (6/10)$ via (43).] [In the next exercise, revisualization works superbly to trivialize the problem.]

Exercise 1.15 (Craps, according to Hoyle) (a) The "shooter" rolls two dice, and obtains a total (called the "point"). If "point" equals "seven" or "eleven," the game is over and "shooter" wins. If point equals "two" or "twelve," the game is over and "shooter" loses. Otherwise, the game continues. It is now a race between "point" and "seven." If "point" comes first, the "shooter" wins; otherwise, he loses. Determine the probability that the "shooter" wins in the game of craps.
[When trying to "convert" a "point" of "ten" (say), we can revisualize and say that on the turn on which the game ends the dice will be showing either one of the 3 tens or one of the 6 sevens, and the probability of this conversion is clearly $3/(3+6)$.]
(b) (The Las Vegas game) The above game is favorable to the "shooter." Thus the version played in Las Vegas has different rules. Specifically, a "three" on the first roll of the two dice is also an immediate loss for "shooter." Determine the probability that "shooter" wins the Las Vegas version of craps.

2 Distribution Theory for Statistics

Convolution

If X and Y are independent rvs on $(\Omega, \mathcal{A}, P)$, then

$$F_{X+Y}(z) = P(X+Y \le z) = \iint_{x+y\le z} dF_X(x)\, dF_Y(y)$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{z-x} dF_Y(y)\, dF_X(x)$$

$$(1) \qquad = \int_{-\infty}^{\infty} F_Y(z-x)\, dF_X(x) \equiv F_X * F_Y(z)$$

is a formula, called the *convolution formula*, for F_{X+Y} in terms of F_X and F_Y (the symbol $*$ defined here stands for "convolution"). In case Y has density f_Y with respect to Lebesgue measure, then so does $X+Y$. In fact, since

$$\int_{-\infty}^{z}\int_{-\infty}^{\infty} f_Y(y-x)\, dF_X(x)\, dy = \int_{-\infty}^{\infty}[\int_{-\infty}^{z} f_Y(y-x)\, dy]\, dF_X(x)$$

$$= \int_{-\infty}^{\infty} F_Y(z-x)\, dF_X(x) = F_{X+Y}(z),$$

we see that $X+Y$ has a density given by

$$(2) \qquad f_{X+Y}(z) = \int_{-\infty}^{\infty} f_Y(z-x)\, dF_X(x).$$

In case both X and Y have densities, we further note that

$$(3) \qquad f_{X+Y}(z) = \int_{-\infty}^{\infty} f_Y(z-x) f_X(x)\, dx \equiv f_Y * f_X(z).$$

Exercise 2.1 Use (2) to show that for X and Y independent:
(i) $X \cong N(\mu_1, \sigma_1^2)$ and $Y \cong N(\mu_2, \sigma_2^2)$ implies $X+Y \cong N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$.
(ii) $X \cong$ Cauchy$(0, a_1)$ and $Y \cong$ Cauchy$(0, a_2)$ has $X+Y \cong$ Cauchy$(0, a_1+a_2)$.
(iii) $X \cong$ Gamma(r_1, θ) and $Y \cong$ Gamma(r_2, θ) has $X+Y \cong$ Gamma(r_1+r_2, θ).

Exercise 2.2 (i) If $X_1, \ldots, X_n$ are iid $N(0,1)$, then the normed sample average necessarily satisfies $(X_1 + \cdots + X_n)/\sqrt{n} \cong N(0,1)$.
(ii) If $X_1, \ldots, X_n$ are iid Cauchy$(0,1)$, then $(X_1 + \cdots + X_n)/n \cong$ Cauchy$(0,1)$.

If X and Y are independent rvs taking values in $0, 1, 2, \ldots$, then clearly

$$(4) \qquad P(X+Y=k) = \sum_{i=0}^{k} P(X=i)P(Y=k-i) \qquad \text{for } k = 0, 1, 2, \ldots.$$

Exercise 2.3 Use (3) to show that for X and Y independent:

$$X \cong \text{ Poisson}(\lambda_1) \text{ and } Y \cong \text{ Poisson}(\lambda_2) \text{ has } X+Y \cong \text{ Poisson}(\lambda_1+\lambda_2)\,.$$

A fundamental problem in probability theory is to determine constants b_n and $a_n > 0$ for which iid rvs $X_1, \ldots, X_n, \ldots$ satisfy

$$(5) \qquad (X_1 + \cdots + X_n - b_n)/a_n \to_d G, \qquad \text{as } n \to \infty$$

for some nondegenerate df G. Exercise 2.2 gives us two examples of such convergence; each was derived via the convolution formula. Except in certain special cases, such as exercises 2.1 – 2.3, the various convolution formulas are too difficult to deal with directly. For this reason we need to develop a more oblique, but ultimately more convenient, approach if we are to solve problems of the form (5). This is taken up in chapters 11, 13, 14, and 15.

Other Formulas

Exercise 2.4 Suppose that X and Y are independent with $P(Y \geq 0) = 1$. Show that products and quotients of these rvs satisfy

$$(6) \qquad F_{XY}(z) \equiv P(XY \leq z) = \int_0^\infty F_X(z/y)\, dF_Y(y) \qquad \text{for all } z,$$

$$(7) \qquad F_{X/Y}(z) \equiv P(X/Y \leq z) = \int_0^\infty F_X(zy)\, dF_Y(y) \qquad \text{for all } z.$$

If F_X has a density f_X, then changing the order of integration above shows that F_{XY} and $F_{X/Y}$ have densities given by

$$(8) \qquad f_{XY}(z) = \int_0^\infty y^{-1}\, f_X(z/y)\, dF_Y(y) \qquad \text{for all } z,$$

$$(9) \qquad f_{X/Y}(z) = \int_0^\infty y\, f_X(yz)\, dF_Y(y) \qquad \text{for all } z.$$

Exercise 2.5 Let $Z \cong N(0,1)$, $U \cong \chi^2_m$, and $V \cong \chi^2_n$ be independent.
(a) Establish these classically important results:

$$(10) \qquad \frac{Z}{\sqrt{U/m}} \cong \text{ Student's } t_m\,.$$

$$(11) \qquad \frac{U/m}{V/n} \cong \text{ Snedecor's } F_{m,n}\,.$$

$$(12) \qquad \frac{U}{U+V} \cong \text{ Beta}(m/2, n/2)\,.$$

Here

$$(13) \qquad f_{t_m}(x) \equiv \frac{\Gamma((m+n)/2)}{\sqrt{\pi\, m}\,\Gamma(m/2)} \frac{1}{(1+x^2/m)^{(m+n)/2}} \qquad \text{for } -\infty < x < \infty,$$

$$(14) \qquad f_{F_{m,n}}(x) \equiv \frac{\Gamma((m+n)/2)}{\Gamma(m/2)\Gamma(n/2)} \frac{(m/n)^{m/2} x^{m/2-1}}{(1+mx/n)^{(m+n)/2}} \qquad \text{for } 0 < x < \infty.$$

(b) Compute the kth moment of each of these three distributions.

Exercise 2.6 If $Y_1, \ldots, Y_{n+1}$ are iid Exponential(θ), then

(15) $\quad (Y_1 + \cdots + Y_i)/(Y_1 + \cdots + Y_{n+1}) \cong \text{ Beta}(i, n-i+1).$

Exercise 2.7 Let $X_1, \ldots, X_n$ be iid N(μ, σ^2).
(a) Show that $W_n \equiv \sqrt{n}\,(\bar{X}_n - \mu)/\,\sigma \cong \text{N}(0,1)$.
(b) Show that $(n-1)S_n^2/\,\sigma^2 \equiv \sum_1^n (X_k - \bar{X}_n)^2/\,\sigma^2 \cong \chi^2_{n-1}$.
(c) Show that W_n and S_n^2 are independent rvs.
(d) Show that $T_n \equiv \sqrt{n}\,(\bar{X}_n - \mu)/\,S_n \cong$ Student's t_{n-1}.
[Hint. Let $\Gamma \equiv |[\,\gamma_{ij}\,]|$ be an orthogonal matrix with all $\gamma_{1j} = 1/\sqrt{n}$. Now let $\vec{Z} \equiv \Gamma\,(\vec{X} - \mu\,\vec{1}\,)/\,\sigma$. This yields iid N(0,1) rvs $Z_1, \ldots, Z_n$, with $W_n = Z_1 \cong \text{N}(0,1)$ and $(n-1)S_n^2 = \sum_2^n Z_k^2 \cong \chi^2_{n-1}$. Apply exercise 2.5.]

Statistical Confidence Intervals

Example 2.1 Suppose we model the performances of n independent repetitions $X_1, \ldots, X_n$ of an experiment as iid $N(\mu, \sigma^2)$ rvs. The previous exercise shows that $\sqrt{n}\,(\bar{X}_n - \mu)/\sigma$ is a $N(0,1)$ rv independent of the sample variance estimator $S_n^2 \equiv \sum_1^n (X_k - \bar{X}_n)^2/(n-1)$ of σ^2, and that $S_n/\sigma \cong \{\chi^2_{n-1}/(n-1)\}^{1/2}$. Thus

(16) $\quad T_n \equiv \sqrt{n}\,[\bar{X}_n - \mu]/S_n \cong \mathcal{T}_{n-1} \equiv \text{Student } t_{n-1}.$

Specify $t_{p/2}$ such that $P(-t_{p/2} \le \mathcal{T}_{n-1} \le t_{p/2}) = 1-p$; perhaps $p = .05$. Then with the "large" probability of $1 - p = .95$ we have

(17) $\quad 1 - p = P(-t_{p/2} \le T_n \le t_{p/2}) = P(-t_{p/2} \le \sqrt{n}\,[\bar{X}_n - \mu]/S_n \le t_{p/2})$

(18) $\quad\quad = P(\mu - t_{p/2}\,S_n/\sqrt{n} \le \bar{X}_n \le \mu + t_{p/2}\,S_n/\sqrt{n} \le \bar{X}_n)$

(19) $\quad\quad = P(\bar{X}_n - t_{p/2}\,S_n/\sqrt{n} \le \mu \le \bar{X}_n + t_{p/2}\,S_n/\sqrt{n})\,.$

That is:

(20) The random interval $\bar{X}_n \pm t_{p/2}\,S_n/\sqrt{n}$
will contain the unknown value of μ
an average of $(1-p) \times 100\%$ of the time.

So when we apply this to the data values $x_1, \ldots, x_n$, we can have $(1-p) \times 100\%$ confidence that the interval $\bar{x}_n \pm t_{p/2}\,s_n/\sqrt{n}$ did enclose the true (but unknown) value of μ. We say that

(21) $\quad \bar{X}_n \pm t_{p/2}\,S_n/\sqrt{n}$ provides a $(1-p) \times 100\%$ *confidence interval*

for the unknown mean μ. Or we say that

(22) $\quad \bar{x}_n \pm t_{p/2}\,s_n/\sqrt{n}$ provides a $(1-p) \times 100\%$ *numerical confidence interval*

for the unknown mean μ. There is a probability of $1-p$ (or a $(1-p) \times 100\%$ chance) that the former *will contain* the unknown value of μ when the X-experiment is repeated n times. There is a $(1-p) \times 100\%$ confidence (or degree of belief) that the latter *did contain* the unknown value of μ after the X-experiment was repeated n times giving the actual data values $x_1, \ldots, x_n$. We call $t_{p/2}\,s_n/\sqrt{n}$ the *numerical margin for error* exhibited by our experiment. □

Transformations of Random Variables

Exercise 2.8 Suppose X has density $f_X(\cdot)$ with respect to Lebesgue measure $\lambda_n(\cdot)$ on n-dimensional Euclidean space R_n.
(a) Let $Y \equiv AX$ denote a linear transformation with A a nonsingular matrix. The Jacobian of this linear transformation is

$$J \equiv \Big|\Big[\frac{\partial(\text{old})}{\partial(\text{new})}\Big]\Big| \equiv \Big|\Big[\frac{\partial x_i}{\partial y_j}\Big]\Big| = A^{-1}, \qquad \text{with} \qquad |J|^+ = |A^{-1}|^+ = 1/|A|^+.$$

Verify that the rv Y has a density $f_Y(\cdot)$ with respect to Lebesgue measure that is given by $f_Y(y) = f_X(A^{-1}y)/\,|A|^+$ on R_n.
(b) Suppose now that X has density $f_X(\cdot)$ with respect to Lebesgue measure on a region R_X in R_n. Suppose the 1-to-1 transformation $Y \equiv g(X)$ from R_X to the region $R_Y \equiv g(R_X)$ has a nonsingular Jacobian with continuous elements at each point of the region. Show that Y has a density given by

$$f_Y(y) = f_X(g^{-1}(y)) \times |[\partial(\text{old})/\partial(\text{new})]|^+.$$

(Any "nice" transformation is locally linear.)

Exercise 2.9 Suppose that $U \equiv XY$ and $V \equiv X/Y$ for rvs having joint density $f_{XY}(\cdot,\cdot)$ on the region where $x > 0$ and $y > 0$. The inverse transformation is $X = \sqrt{UV}$ and $Y = \sqrt{U/V}$ with a "nice" Jacobian that is equal to $2v$. Thus the joint density of U, V is

$$f_{UV}(u,v) = \tfrac{1}{2v}\, f_{XY}(\sqrt{uv}, \sqrt{u/v}\,)) \quad \text{on the appropriate } (u,v)\text{-region},$$

provided that the transformation is 1-to-1. Now evaluate $f_{UV}(\cdot,\cdot)$ and $f_V(\cdot)$ in the following cases.
(a) X and Y are independent Exponential(1).
(b) X and Y are independent with density $1/(xy)^2$ on $x, y \geq 1$. Evaluate $f_U(\cdot)$.
(c) X and Y are independent $\text{N}(0,1)$. [Note that this transformation is not 1-1.]
(d) $X \cong \text{N}(0,1)$ and $Y \cong \text{Uniform}(0,1)$ are independent.
[This exercise demonstrates vividly the important role played by the regions R_X and R_Y.]

3 Linear Algebra Applications

Notation 3.1 (Mean vector and covariance matrix) Let $X \equiv (X_1, \ldots, X_n)'$ be a rv. Then $\mathrm{E}(X) \equiv \mu \equiv (\mu_1, \ldots, \mu_n)'$, where $\mu_i \equiv \mathrm{E}(X_i)$ is called the *mean vector*. And $\Sigma \equiv |[\,\sigma_{ij}\,]| \equiv |[\,\mathrm{Cov}[X_i, X_j]\,]|$ is called the *covariance matrix*. (By the Cauchy–Schwarz inequality, both of μ and Σ are well-defined provided that each of $\sigma_{ii} \equiv \mathrm{Var}[X_i] \equiv \mathrm{Cov}[X_i, X_i]$ is finite.) □

Definition 3.1 (Linear algebra) We will operate on n-dimensional space R_n with $n \times n$ matrices and $n \times 1$ vectors.
(i) A matrix Γ with column vectors γ_i (that is, $\Gamma = [\gamma_1, \ldots, \gamma_n]$) is called *orthogonal* if $\Gamma'\,\Gamma = I$. [Thus $\gamma_i'\,\gamma_j$ equals 1 or 0 according as $i = j$ or $i \neq j$; when $\gamma_i'\,\gamma_j = 0$ we say that these vectors are orthogonal, and we write $\gamma_i \perp \gamma_j$.] Under the orthogonal transformation of R_n onto itself defined by $y = \Gamma x$, the image of each γ_i is the standardized basis vector $e_i \equiv (0, \ldots, 0, 1, 0, \ldots, 0)'$ with the 1 in the ith slot.
(ii) Call a symmetric matrix A *positive definite* (written $A > 0$) if $x'Ax > 0$ for all vectors $x \neq 0$. Call it *nonnegative definite* (written $A \geq 0$) if $x'Ax \geq 0$ for all vectors $x \neq 0$.
(iii) If A is symmetric and *idempotent* (that is, if $AA = A$), then A is called a *projection matrix* (the symbol P is often used for a projection matrix).
(iv) Let D_a be the *diagonal matrix* with $d_{ii} = a_i$ (and $d_{ij} = 0$ for all $i \neq j$).
(v) Let $\mathcal{R}[A]$ denote the *column space* of A; that is, it is the set of all vectors that can be written as linear combinations of the column vectors of A.
(vi) Call $x'Ax = \sum_{j=1}^n \sum_{i=1}^n x_i\, a_{ij}\, x_j$ a *quadratic form* in the vector x.

What follows is the statistician's main result from linear algebra. We simply state it, then interpret it geometrically in discussion 3.1, and then put it into a very useful format in discussion 3.2.

Theorem 3.1 (Principal axes theorem) Let A denote an arbitrary real and symmetric matrix of rank r.
(a) There exists an orthogonal matrix $\Gamma \equiv [\gamma_1, \ldots, \gamma_n]$ and a diagonal matrix D for which we have the representation

$$(1) \qquad A = \Gamma\, D\, \Gamma' \qquad \text{and/or} \qquad \Gamma'\, A\, \Gamma = D \qquad \text{with } \mathrm{rank}(D) = r.$$

The γ_i are called *eigenvectors*, while the corresponding d_{ii} are called *eigenvalues*. (See (39) below for further comments.)
(b) If $A > 0$ ($A \geq 0$), then all $d_{ii} > 0$ ($d_{ii} \geq 0$).
We can specify Γ such that $d_{11} \geq \cdots \geq d_{rr} > 0 = d_{r+1,r+1} = \cdots = d_{nn}$.
(c) If P is a projection matrix, then all $d_{ii} = 1$ or 0. Moreover, we must have $r \equiv \mathrm{rank}(A) = \mathrm{tr}(D) = \mathrm{tr}(A) = \sum_1^n a_{ii}$.

Discussion 3.1 (Spectral decomposition) Consider a projection matrix P of rank r. Then the transformation $y = P\,x$ can be broken down as

$$(2) \qquad Px = \Gamma\, D\, \Gamma' x = [\gamma_1, \ldots, \gamma_n]\, D\, [\gamma_1, \ldots, \gamma_n]' x = \textstyle\sum_1^n d_{ii}\, (\gamma_i'\, x)\, \gamma_i\,,$$

where $(\gamma_i' x)\gamma_i$ is the projection of x onto γ_i in the direction of γ_i, and where this term is present when $d_{ii} = 1$ and is absent when $d_{ii} = 0$. Also, $Px \perp (I - P)x$, where the transformation

(3) $$(I - P)x = \sum_{i=1}^{n} [1 - d_{ii}]\,(\gamma_i' x)\,\gamma_i$$

projects onto $\mathcal{R}[\gamma_{r+1}, \ldots, \gamma_n] = \mathcal{R}^{\perp}[\gamma_1, \ldots, \gamma_r\,]$. Finally,

(4) $$Px = \sum_{i=1}^{r} d_{ii}\,(\gamma_i\,\gamma_i'\,)\,x = [\,\sum_{i=1}^{r} P_i\,]\,x$$

with $P_i \equiv \gamma_i\,\gamma_i'$. This is called the *spectral decomposition* of the transformation $y = Px$. □

Exercise 3.1 (a) Show that for compatible matrices B and C,

(5) $$\text{tr}(BC) = \text{tr}(CB) \quad \text{and} \quad \text{rank}(BC) \le \text{rank}(B) \wedge \text{rank}(C),$$

giving $\text{rank}(A\,\Gamma) = \text{rank}(A)$ above.
(b) Prove theorem 3.1(b)(c) using theorem 3.1(a).
(c) Show that $\mathcal{R}[A] = \mathcal{R}[AA'\,]$ and $\mathcal{R}[A'\,] = \mathcal{R}[A'A]$.

Proposition 3.1 (Properties of E(·)) (a) It holds that

(6) $$\text{E}(A\,X\,B + C) = A\,\text{E}(X)\,B + C \quad \text{and} \quad \text{Cov}[A\,X, B\,Y] = A\,\text{Cov}[X, Y]\,B'\,.$$

(b) Any covariance matrix $\Sigma_X \equiv |[\,\text{Cov}[X_i, X_j]\,]|$ satisfies $\Sigma_X \ge 0$.

Exercise 3.2 Prove proposition 3.1.

Discussion 3.2 (Versions of Σ^- and $\Sigma^{-1/2}$) Let $X \cong (\mu, \Sigma)$. According to the principal axes theorem, we can make the decomposition (for any orthogonal matrix Δ whatsoever)

(7) $$\begin{aligned}\Sigma = \Gamma\,D\,\Gamma' &= \left[\Gamma \begin{bmatrix} D^{1/2} & 0 \\ 0 & 0 \end{bmatrix} \Delta'\right] \left[\Delta \begin{bmatrix} D^{1/2} & 0 \\ 0 & 0 \end{bmatrix} \Gamma'\right] \\ &= \left[\Gamma \begin{bmatrix} D^{1/2} & 0 \\ 0 & 0 \end{bmatrix}\right] \left[\begin{bmatrix} D^{1/2} & 0 \\ 0 & 0 \end{bmatrix} \Gamma'\right] = (\,\Gamma\,D^{1/2}\,)(\,D^{1/2}\,\Gamma'\,) \equiv AA'\,,\end{aligned}$$

where $D^{1/2}$ has the numbers $d_{ii}^{1/2}$ on its diagonal and where A is $n \times k$. The presence of $\Delta'\,\Delta$ (which equals I) shows that this decomposition is not unique. Continuing on gives

(8) $$\Sigma = \Gamma\,D\,\Gamma' = (\Gamma\,D\,\Gamma')\,(\Gamma\,D\,\Gamma') \equiv \Sigma^{1/2}\,\Sigma^{1/2},$$

(9) $$\Sigma^{-1/2} \equiv \Gamma\,D^{-1/2}\,\Gamma', \qquad \text{where} \qquad d_{ii}^{-1/2} \equiv \begin{cases} 0 & \text{if } d_{ii} = 0, \\ \frac{1}{\sqrt{d_{ii}}} & \text{if } d_{ii} > 0, \end{cases}$$

(10) $$\Sigma^- \equiv \Gamma\,D^-\,\Gamma', \qquad \text{where} \qquad D^- \equiv D^{-1/2}\,D^{-1/2}.$$

Note that

$$\text{(11)} \qquad \Sigma\,\Sigma^-\,\Sigma = \Sigma \qquad \text{and} \qquad \Sigma^{1/2}\,\Sigma^{-1/2}\,\Sigma^{1/2} = \Sigma^{1/2}\,,$$

$$\text{(12)} \qquad \Sigma^{1/2}\,\Sigma^{-1/2} = \begin{bmatrix} I_k & 0 \\ 0 & 0 \end{bmatrix} = \Sigma^{-1/2}\,\Sigma^{1/2} \quad \text{and} \quad \Sigma\,\Sigma^- = \begin{bmatrix} I_k & 0 \\ 0 & 0 \end{bmatrix} = \Sigma^-\,\Sigma\,.$$

These last two results are in keeping with the definition of generalized inverses. □

Recall that the *generalized inverse* B^- of the matrix B is defined to be any matrix B^- that satisfies $B\,B^-\,B = B$. A generalized inverse always exists. It has the following interpretation. Fix the matrix B and the vector c. Then

$$\text{(13)} \qquad B\,\beta = c \quad (\text{with any } c \in \mathcal{R}[B]\,) \qquad \text{has the solution} \quad \hat{\beta} = B^-\,c\,.$$

(It is clear that such a solution does always exist, for a fixed c.) Suppose such a B^- exists, in general; which we accept, and will use freely. Then replace c in (13) by each column of B, and see that such a B^- must necessarily satisfy $BB^-B = B$.

Theorem 3.2 (Properties of covariance matrices)
(a) The following results are equivalent for real matrices:

(14) Σ is the covariance matrix of some rv Y.

(15) Σ is symmetric and nonnegative definite.

(16) There exists an $n \times n$ matrix A such that $\Sigma = AA'$. (Recall (7) for A.)

(b) The matrix Σ_X is positive definite (that is, $\Sigma_X > 0$) if and only a vector $c \neq 0$ and a constant b do not exist for which $c'\,X = b$ a.s.

Proof. Now, (14) implies (15): Σ is symmetric, since $\mathrm{E}Y_iY_j = \mathrm{E}Y_jY_i$. Also, $a'\Sigma a = \mathrm{Var}[a'Y] \geq 0$ for all vectors a, so that $\Sigma \geq 0$.
Also, (15) implies (16): Just recall (7).
Also, (16) implies (14): Let $X \equiv (X_1, \ldots, X_n)'$, where $X_1, \ldots, X_n$ are independent $N(0,1)$. Let $Y \equiv AX$. Then Y has covariance matrix $\Sigma = AA'$ by (6). □

Exercise 3.3 Prove theorem 3.2(b).

Exercise 3.4 Let $X \cong (\theta, \Sigma)$ and let B be symmetric.
(a) $\mathrm{E}\{(X-b)'\,B\,(X-b)\} = \mathrm{tr}(B\,\Sigma) + (\theta-b)'\,B\,(\theta-b)$.
(b) If $\Sigma = \sigma^2\,I$, then $\mathrm{tr}(B\,\Sigma) = \sigma^2\ \mathrm{tr}(B) = \sigma^2 \sum_{i=1}^n b_{ii}$.
(c) If $\Sigma = \sigma^2\,I$ and B is idempotent, then $\mathrm{tr}(B\,\Sigma) = \sigma^2\ \mathrm{tr}(B) = \sigma^2\ \mathrm{rank}(B)$.

Exercise 3.5 For symmetric A there exists an upper (or lower) triangular matrix H for which $A = H\,H'$. If $A > 0$ (or $A \geq 0$), we may suppose that all $h_{ii} > 0$ (or that all $h_{ii} \geq 0$).

Discussion 3.3 (Best linear predictor and multiple correlation) Consider the partitioned random vector

$$(17)\qquad \begin{bmatrix} Y_0 \\ Y \end{bmatrix} \cong \left[\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{00} & \sigma_0' \\ \sigma_0 & \Sigma \end{bmatrix} \right] \qquad \text{with } |\Sigma| \neq 0 .$$

The *best linear predictor* of Y_0 based on Y is

$$(18)\qquad \alpha_0' \, Y \equiv \sigma_0' \, \Sigma^{-1} \, Y \qquad (\text{or } \alpha_0 \equiv \Sigma^{-1} \, \sigma_0) ,$$

where "best" is in the sense that

$$(19)\qquad \mathrm{Var}[\, Y_0 - \beta' \, Y \,] \geq \mathrm{Var}[\, Y_0 - \alpha_0' \, Y \,] = \sigma_{00} - \sigma_0' \, \Sigma^{-1} \, \sigma_0 \qquad \text{for all } \beta.$$

In parallel with this,

$$(20)\qquad \mathrm{Corr}[\, Y_0, \beta' \, Y \,] \geq \mathrm{Corr}[\, Y_0, \alpha_0' \, Y \,] \qquad \text{for all } \beta.$$

The maximized value of the correlation (that is, the *multiple correlation coefficient*) is given by

$$(21)\qquad \rho_{0\cdot 1,\ldots,n} \equiv \mathrm{Corr}[\, Y_0, \alpha_0' \, Y \,] = \sqrt{\frac{\sigma_0' \, \Sigma^{-1} \, \sigma_0}{\sigma_{00}}} ,$$

and the variance of the best linear predictor is also easily seen to equal

$$(22)\qquad \mathrm{Var}[\, Y_0 - \alpha_0' \, Y \,] = \sigma_{00} - \sigma_0' \, \Sigma^{-1} \, \sigma_0 = \sigma_{00} \, (1 - \rho^2_{0\cdot 1,\ldots,n}) .$$

[Proof. The first holds, since

$$\begin{aligned} \mathrm{Var}[\, Y_0 - \beta' \, Y \,] &= \mathrm{Var}[\, (Y_0 - \alpha_0' \, Y) + (\alpha_0' - \beta) Y \,] \\ &= \mathrm{Var}[\, Y_0 - \alpha_0' \, Y \,] + 2 \cdot 0 + (\alpha_0 - \beta)' \, \Sigma \, (\alpha_0 - \beta) \\ &\geq \mathrm{Var}[\, Y_0 - \alpha_0' \, Y \,] . \end{aligned}$$

The second holds, since

$$\mathrm{Corr}^2 \, [\, Y_0, \beta \, Y \,] = \frac{(\beta' \, \sigma_0)^2}{\sigma_{00} \;\; \beta' \, \Sigma \, \beta} \leq \frac{\sigma_0' \, \Sigma^{-1} \, \sigma_0}{\sigma_{00}} ,$$

with equality only at $\beta = c \, \Sigma^{-1} \, \sigma_0$ (as follows from application of Cauchy–Schwarz).]

Simple linear regression model We now want the best linear predictor of Y based on X. The conditional distribution of Y given that $X = x$ is given by

$$(23)\qquad Y \Big| X = x \;\cong\; \left(\mu_Y + \frac{\sigma_{XY}}{\sigma_X^2} \, (x - \mu_X), \; \sigma_Y^2 - \frac{\sigma_{XY}^2}{\sigma_X^2} \right) = (\alpha + \beta \, x, \, \sigma_\epsilon^2) ,$$

expressing the moments in terms of

$$(24)\qquad \rho \equiv \frac{\sigma_{XY}}{\sigma_X \, \sigma_Y} , \quad \beta \equiv \rho \, \frac{\sigma_Y}{\sigma_X} , \quad \alpha \equiv \mu_Y - \beta \, \mu_X , \quad \sigma_\epsilon^2 \equiv \sigma_Y^2 \, (1 - \rho^2) .$$

This leads directly to the *simple linear regression model* that conditionally on $X = x$ the observations Y_i satisfy

$$(25)\qquad Y_i = \alpha + \beta \, x_i + \epsilon_i \qquad \text{where } \epsilon_i \cong (0, \sigma_\epsilon^2) \text{ are iid}$$

with

$$(26)\qquad \beta = \rho \, \frac{\sigma_Y}{\sigma_X} , \quad \sigma_\epsilon^2 \equiv \sigma_Y^2 \, (1 - \rho^2) , \quad \alpha \equiv \mu_Y - \mu_X \, \beta . \qquad \square$$

Discussion 3.4 (Conditional moments and projections) Suppose that

$$(27) \qquad Y = \begin{bmatrix} Y^{(1)} \\ Y^{(2)} \end{bmatrix} \cong \left[\begin{bmatrix} \mu^{(1)} \\ \mu^{(2)} \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right] .$$

Then the moments of the conditional distribution of $Y^{(1)}$ given that $Y^{(2)} = y^{(2)}$ are summarized in

$$(28) \qquad Y^{(1)} \Big| Y^{(2)} = y^{(2)} \quad \cong \quad (\mu^{(1)} + \Sigma_{12}\,\Sigma_{22}^{-1}\,(y^{(2)} - \mu^{(2)}),\ \Sigma_{11} - \Sigma_{12}\,\Sigma_{22}^{-1}\,\Sigma_{21}).$$

To see this, just define

$$(29) \qquad Z \equiv \begin{bmatrix} Z^{(1)} \\ Z^{(2)} \end{bmatrix} \equiv \begin{bmatrix} (Y^{(1)} - \mu^{(1)}) - \Sigma_{12}\,\Sigma_{22}^{-1}\,(Y^{(2)} - \mu^{(2)}) \\ Y^{(2)} - \mu^{(2)} \end{bmatrix} .$$

It is a minor calculation that

$$(30) \qquad Z \cong [\vec{0}, \Sigma_Z] \equiv \left[\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_{Z,11} & \Sigma_{Z,12} \\ \Sigma_{Z,21} & \Sigma_{Z,22} \end{bmatrix} \right] = \left[\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} - \Sigma_{12}\,\Sigma_{22}^{-1}\,\Sigma_{21} & 0 \\ 0 & \Sigma_{22} \end{bmatrix} \right].$$

The exercises will show that

$$(31) \qquad |\Sigma_Z| = |\Sigma_{22}|\,|\Sigma_{11} - \Sigma_{12}\,\Sigma_{22}^{-1}\,\Sigma_{21}| .$$

[Proof. It is straightforward to compute $\Sigma_{Z,12} = \Sigma_{12} - \Sigma_{12}\,\Sigma_{22}^{-1}\,\Sigma_{22} = 0$. Trivially, we have $\Sigma_{Z,22} = \Sigma_{22}$. Then

$$(a) \qquad \Sigma_{Z,11} = \Sigma_{12} - 2\Sigma_{12}\,\Sigma_{22}^{-1}\,\Sigma_{22} + \Sigma_{12}\,\Sigma_{22}^{-1}\,\Sigma_{22}\,\Sigma_{22}^{-1}\,\Sigma_{21}$$

$$(b) \qquad = \Sigma_{12} - \Sigma_{12}\,\Sigma_{22}^{-1}\,\Sigma_{21}.$$

Since $Y^{(1)} = \mu^{(1)} + Z^{(1)} + \Sigma_{12}\,\Sigma_{22}^{-1}\,Z^{(2)}$ with $Y^{(2)} - \mu^{(2)} = Z^{(2)}$, conditionally

$$(c) \qquad Y^{(1)} \Big| Y^{(2)} = y^{(2)} \quad \cong \quad \mu^{(1)} + Z^{(1)} + \Sigma_{12}\,\Sigma_{22}^{-1}\,z^{(2)}$$

$$\cong (\mu^{(1)} + \Sigma_{12}\,\Sigma_{22}^{-1}\,z^{(2)}, \Sigma_{Z,11})$$

$$= (\mu^{(1)} + \Sigma_{12}\,\Sigma_{22}^{-1}\,(y^{(2)} - \mu^{(2)}), \Sigma_{Z,11}) ,$$

as required.] [See exercise 3.7 below for (31).] □

Exercise 3.6 Consider the rvs $Z^{(1)}$ and $Z^{(2)}$ in (29). Suppose $\mu^{(1)} = 0$ and $\mu^{(2)} = 0$. Let $\mathcal{H}_1^0$ and $\mathcal{H}_2^0$ denote the Hilbert spaces generated by the rv subsets $Y^{(1)} \equiv (Y_1, \ldots, Y_k)$ and $Y^{(2)} \equiv (Y_{k+1}, \ldots, Y_n)$, respectively. Show that $Z^{(1)}$ is the projection of $Y^{(1)}$ into the Hilbert space $(\mathcal{H}_2^0)^{\perp}$. (See (**??**.**??**.**??**).)

Discussion 3.5 (Partitioned matrices) Let

$$(32) \qquad A \equiv \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} , \qquad \text{and write} \qquad A^{-1} \equiv \begin{bmatrix} A^{11} & A^{12} \\ A^{21} & A^{22} \end{bmatrix}$$

when the inverse exists. We agree that A_{11} is $k \times k$.

Exercise 3.7 (a) If $|A_{22}| \neq 0$, show that $|A| = |A_{22}|\,|A_{11} - A_{12}\,A_{22}^{-1}\,A_{21}|$.
(b) If $|A| \neq 0$, show that $|A + x\,y'| = |A|\,(1 + y'\,A^{-1}\,x)$ for all vectors x, y.
[Hint. Appeal to

$$\begin{bmatrix} C & 0 \\ E & D \end{bmatrix} = |C|\,|D| \qquad \text{and work with} \qquad B \equiv \begin{bmatrix} I & 0 \\ -A_{12}\,A_{22}^{-1} & I \end{bmatrix}$$

for appropriate choices.]

Exercise 3.8 (a) Show that for a symmetric A having $|A_{11}| \neq 0$ and $|A_{22}| \neq 0$:

$$A^{11} = (A_{11} - A_{12}\,A_{22}^{-1}\,A_{21})^{-1} \quad \text{and} \quad A^{12} = -A_{11}^{-1}\,A_{12}\,A^{22}. \tag{33}$$

$$A^{22} = (A_{22} - A_{21}\,A_{11}^{-1}\,A_{12})^{-1} \quad \text{and} \quad A^{21} = -A_{22}^{-1}\,A_{21}\,A^{11}. \tag{34}$$

[Hint. Start multiplying the partitioned form of $A\,A^{-1} = I$.]
(b) Obtain analogous formulas from $A^{-1}\,A = I$.
(c) Show that

$$A^{11}\,A_{11} + A^{12}\,A_{21} = I \quad \text{and} \quad A_{11}\,A^{12} + A_{12}\,A^{22} = 0. \tag{35}$$

□

Exercise 3.9 Show that for symmetric A,

$$\frac{\partial}{\partial\beta}\,[\beta'\,A\,\beta] = 2\,A\,\beta. \tag{36}$$

Discussion 3.6 (Simultaneous decomposition) For a real symmetric matrix A that is nonnegative definite (that is, $A \geq 0$) we wrote

$$A = \Gamma\,D\,\Gamma' \qquad \text{and} \qquad \Gamma'\,A\,\Gamma = D \tag{37}$$

with $d_{11} \geq \cdots \geq d_{rr} > 0$.
(A) We note that

$$|A - \lambda\,I| = |\Gamma|\,|D - \lambda\,I|\,|\Gamma'| = |D - \lambda\,I| = 0 \tag{38}$$

all have the same solutions $d_{11}, \ldots, d_{rr}, 0$, and thus $d_{11}, \ldots, d_{rr}$ are indeed the nonzero eigenvalues of A. Moreover, (37) gives

$$A\,\Gamma = \Gamma\,D \quad \text{or} \quad A\,\gamma_i = d_{ii}\,\gamma_i \quad \text{for } 1 \leq i \leq r, \tag{39}$$

so that $\gamma_1, \ldots, \gamma_r$ are the corresponding eigenvectors.
(B) Suppose $A > 0$ and $B \geq 0$. Then

$$\begin{array}{l} |B - \lambda\,A| = 0, \quad |A^{-1/2}\,B\,A^{-1/2} - \lambda\,I| = 0, \quad |A^{-1}\,B - \lambda\,I| = 0 \\ \text{all have the same solutions } \lambda. \end{array} \tag{40}$$

[Just note that $|B - \lambda\,A| = |A^{1/2}|\,|A^{-1/2}\,B\,A^{-1/2} - \lambda\,I|\,|A^{1/2}|$ and $|A^{-1}|\,|B - \lambda\,A| = |A^{-1}\,B - \lambda\,I|$.] Writing $A^{-1/2}\,B\,A^{-1/2} = \Delta\,D_\theta\,\Delta'$ with Δ orthogonal gives

$$B = (A^{1/2}\,\Delta)\,D_\theta\,(\Delta'\,A^{1/2}) \quad \text{and} \quad A = (A^{1/2}\,\Delta)\,(\Delta'\,A^{1/2}). \tag{41}$$

This last formula is called the *simultaneous decomposition* of A and B. □

Discussion 3.7 (a) (Cauchy–Schwarz) For all vectors x, y:

(42) $(x' y)^2 \le \|x\|^2 \, \|y\|^2,$

with equality (for $y \neq 0$) if and only if $x = cy$ for some constant c.
(b) For any real symmetric matrix $A > 0$

(43) $$\max_{a \neq 0} \frac{a' A\, a}{a'\, a} = d_{11}(A)$$

(as follows immediately from (1), with $d_{11}(A)$ the largest eigenvalue of A).
(c) Let $A > 0$, and fix $C \ge 0$ and $b \neq 0$. Then

(44) $(x'y)^2 \le (x' A\, x)\,(y' A^{-1}\, y),$

with equality (when $y \neq 0$) if and only if $x = c\, A^{-1}\, y$ for some c. Also,

(45) $$\min_{a' 1 = 1} \{a' A\, a\} = 1/(1' A^{-1}\, 1),$$

with equality only at $a_0 \equiv A^{-1}\, 1/(1' A^{-1}\, 1)$. Also,

(46) $$\max_{a \neq 0} \frac{(a'\, b)^2}{a' A\, a} = b' A^{-1}\, b,$$

with equality only when $a =$ (some c) $A^{-1}\, b$. Also,

$$\max_{a \neq 0} \frac{a' C\, a}{a' A\, a} = \max_{a \neq 0} \frac{(a' A^{1/2})\,(A^{-1/2}\, C\, A^{-1/2})\,(A^{1/2}\, a)}{(a' A^{1/2})\,(A^{1/2}\, a)}$$

(47) $$= d_{11}(A^{-1/2}\, C A^{-1/2}) = d_{11}(C\, A^{-1})\,.$$

Here, $(a' C\, a) = (a'\, b)^2$ is an important special case (already solved via (46)).
(d) Let $A > 0$, let $B^{k \times n}$ have rank$(B) = k$, and let $b^{k \times 1} \neq 0$. Then

(48) $$\min_{B\, a = b} \{a' A\, a\} = b' \,[B\, A^{-1}\, B'\,]^{-1}\, b, \text{ is achieved at } a_0 \equiv B' \,[B\, A^{-1}\, B'\,]^{-1} b\,. \square$$

Exercise 3.10 Prove (42)–(47) (the equality in (42) needs some attention). [The harder (48) is proven below.]

Proof. Consider (48). Now,

$$a' A\, a \ge \frac{(a'\, y)^2}{y' A^{-1}\, y} \qquad \text{for all } y \neq 0$$

(a) $$= \frac{[a' B' \,[B\, A^{-1}\, B'\,]^{-1}\, b\,]^2}{b' \,[B\, A^{-1}\, B'\,]^{-1} (B\, A^{-1}\, B'\,) [B\, A^{-1}\, B'\,]^{-1}\, b} \qquad \text{if } y \equiv B' \,[B\, A^{-1}\, B'\,]^{-1} b$$

(b) $$= \frac{[b' \,[B\, A^{-1}\, B'\,]^{-1}\, b\,]^2}{b' \,[B\, A^{-1}\, B'\,]^{-1}\, b} \qquad \text{for all } a, \text{ using } B\, a = b$$

$$= b' \,[B\, A^{-1}\, B'\,]^{-1}\, b$$

yielding a bound not depending on a, which proves (48). $\square$

Discussion 3.8 (General Linear Model) Consider the general linear model

(49) $Y = X\beta + \epsilon \equiv \theta + \epsilon,$

where $X^{n\times p}$ is a matrix of known constants, where $\beta^{p\times 1}$ is a vector of unknown parameters, and where the rv $\epsilon^{n\times 1} \cong (\vec{0}, \sigma^2 I)$ with σ^2 unknown. Recall that

(50) $\mathcal{R}[X] \equiv$ (the *column space* of X) $= \{y : y = Xa$ with any vector $a\}$

is a vector space (of rank r, say). Noting that $\theta = X\beta \in \mathcal{R}[X]$, the *least squares estimator* (or LSE) of θ is defined to be that value $\hat{\theta}$ in $\mathcal{R}[X]$ that minimizes

(51) $\|\epsilon\|^2 = \|Y - X\beta\|^2 = \|Y - \theta\|^2 .$

This minimization clearly occurs when $\hat{\theta}$ is the projection of Y onto $\mathcal{R}[X]$; so

(52) $\hat{\theta} =$ (the unique projection of Y onto $\mathcal{R}[X]$) $=$ (the unique LSE $\hat{\theta}$ of θ).

We note that $\hat{\beta}$ need not be unique, since

(53) any $\hat{\beta}$ for which $X\hat{\beta} = \hat{\theta}$ gives this same LSE $\hat{\theta}$.

Since $(Y - \hat{\theta})$ is $\perp$ to $\mathcal{R}[X]$, it must be that $\hat{\theta}$ and $\hat{\beta}$ satisfies the *normal equations*

(54) $X'(Y - \hat{\theta}) = \vec{0}, \qquad$ or equivalently $\qquad X'X\hat{\beta} = X'Y.$

Conversely, suppose $\hat{\beta}$ satisfies the normal equations. Then $X\hat{\beta} \in \mathcal{R}[X]$ with $X'(Y - X\hat{\beta}) = \vec{0}$, showing that $(Y - X\hat{\beta}) \in \mathcal{R}[X]^\perp$; and thus $X\hat{\beta}$ and $(Y - X\hat{\beta})$ must be the projections of Y onto the spaces $\mathcal{R}[X]$ and $\mathcal{R}[X]^\perp$, respectively. Thus

(55) $X\hat{\beta} = \hat{\theta} \equiv$ (the unique projection of Y onto $\mathcal{R}[X]$) iff $X'X\hat{\beta} = X'Y.$

Let $\Omega \equiv \mathcal{R}[X]$, let $r \equiv \text{rank}(X)$, and let $P_\Omega \equiv$ (the projection matrix onto Ω). We will next prove that:

(56) $P_\Omega = X(X'X)^- X' \equiv$ (the *hat matrix*), with P_Ω and $(I - P_\Omega)$ idempotent.

(57) $\hat{\theta} \equiv \hat{Y} \equiv$ (the *fitted value*) $= [X(X'X)^- X']\, Y = P_\Omega Y.$

(58) $\hat{\epsilon} \equiv$ (the *residuals*) $\equiv (Y - \hat{Y}) = [I - P_\Omega]\, Y = P_{\Omega^\perp} Y.$

(59) $\text{E}\,\hat{\theta} = \theta, \quad \text{E}\,\|\hat{\epsilon}\|^2 = \text{E}\,\|Y - \hat{Y}\|^2 = (n - r)\sigma^2 \quad$ and $\quad \text{Cov}[\hat{\theta} - \theta, Y - \hat{\theta}] = \vec{0}.$

Call $RSS \equiv \|\hat{\epsilon}\|^2 = \hat{\epsilon}'\hat{\epsilon} = \|Y - \hat{Y}\|^2 = Y'\,[I - P_\Omega]\, Y$ the *residual sum of squares*.

Proof. Let $B \equiv X'X$, with $\mathcal{R}[B] = \mathcal{R}[X']$ (by exercise 3.1(c)); and then define $c \equiv X'Y \in \mathcal{R}[X']$. Then (by (13)) the projection of Y onto Ω is given by $P_\Omega Y = \hat{\theta} = X\hat{\beta} = XB^- c = X(X'X)^- X'Y$; so $P_\Omega = X(X'X)^- X'$. Also, $\text{E}(\hat{\theta}) = \text{E}(P_\Omega Y) = P_\Omega\, \text{E}\, Y = P_\Omega\, \theta = \theta$. Finally, the residuals satisfy $\|Y - \hat{Y}\|^2 = (Y - \hat{Y})'(Y - \hat{Y}) = Y'(I - P_\Omega)'(I - P_\Omega)Y = Y'(I - P_\Omega)Y$, with expectation (given by exercise 3.4) $\sigma^2\,\text{tr}(I - P_\Omega) + \theta'(I - P_\Omega)\theta = \sigma^2\,\text{tr}(I - P_\Omega) + \theta'\,\vec{0} = (n - r)\sigma^2$.

When $\text{rank}(X) = p$, then $Xb = \hat{\theta}$ has a *unique* solution $\hat{\beta}$. Moreover,

(60) $\hat{\beta} = (X'X)^{-1}X'Y$

is the unique solution of the normal equations ($\hat{\beta}$ is now called the LSE of β), and

(61) $\text{E}\,\hat{\beta} = (X'X)^{-1}X'X\beta = \beta$, so that $\hat{\beta}$ is an *unbiased estimator* of β, and

(62) $\Sigma_{\hat{\beta}} = (X'X)^{-1}X'(\sigma^2 I)X(X'X)^{-1} = \sigma^2\,(X'X)^{-1}$, with

(63) $\text{E}\,S^2 = \sigma^2$ for $S^2 \equiv \|Y - \hat{Y}\,\|^2/(n-p)$, so that S^2 is unbiased for σ^2.

We thus say that β is *identifiable* (and *estimable*) in this *full rank case* when $\text{rank}(X) = p$ (that is, when X is non-singular, or $|X| \neq 0$). □

Exercise 3.11 (Gauss–Markov) Let $Y = X\beta + \epsilon = \theta + \epsilon$ (as in (49)) with the rv $\epsilon \cong (0, \sigma^2 I)$ and with $\text{rank}(X) = r$. Consider some $c'\theta$ $(= c'X\beta)$. Show that among the class of all linear unbiased estimators of $c'\theta$, the estimator $c'\hat{\theta}$ is the unique one having minimum variance (so, it is *best*). Determine its variance.

Exercise 3.12 (Distribution theory under normality) (a) Let $Y = X\beta + \epsilon = \theta + \epsilon$ (as in (49)) with $\epsilon \cong N(0, \sigma^2 I)$, and with $r \equiv \text{rank}(X)$. Show that

(64) $(\hat{\beta} - \beta)'(X'X)(\hat{\beta} - \beta) = \|X(\hat{\beta} - \beta)\|^2 = \|\,\hat{\theta} - \theta\,\|^2 \cong \sigma^2\,\text{Chisquare}_r$.

(65) $\|\hat{\epsilon}\|^2 = \|Y - \hat{Y}\,\|^2 = \|Y - \hat{\theta}\,\|^2 \cong \sigma^2\,\text{Chisquare}_{n-r}$, or

(66) $RSS/(n-r) = \|Y - \hat{Y}\,\|^2/(n-r) \cong \sigma^2\,\text{Chisquare}_{n-r}/(n-r)$, and

(67) $\hat{\theta}$ and $\hat{\epsilon} = (Y - \hat{Y}) = (Y - \hat{\theta}\,)$ are uncorrelated.

When $r \equiv \text{rank}(X) = p$ (so that $\hat{\beta}$ is unique, identifiable, and estimable), show that

(68) $\hat{\beta} \cong N(\beta, \sigma^2\,(X'X)^{-1})$ and

(69) $\hat{\beta}$ is independent of $\hat{\epsilon} = (Y - \hat{Y}) = (Y - \hat{\theta}\,)$, and hence of $\|Y - \hat{\theta}\,\|^2$ also.

(b) Suppose instead that $\epsilon \cong (\vec{0}, V)$, with $\text{rank}(V) = n$. Define $Z \equiv V^{-1/2}Y$, and show that this Z satisfies the linear model equation $Z = X^*\beta + \epsilon^*$ where $X^* \equiv V^{-1/2}X$ and $\epsilon^* \cong (0, \sigma^2\,I)$. So, analogs of all the formulas in (a) are trivial.

Exercise 3.13 (Alternative minimization in the general linear model) In the context of the model $Y = X\beta + \epsilon = \theta + \epsilon$ (as in (49)) we now let $\tilde{\theta}$ denote that $\theta \in \mathcal{R}[X]$ that minimizes (for some positive definite covariance matrix M)

(70) $\|\epsilon\,\|_M^2 \equiv \epsilon'\,M\,\epsilon = \|Y - \theta\,\|_M^2 = \|Y - X\beta\,\|_M^2 = (Y - X\beta)'M(Y - X\beta)$

(instead of minimizing $\|\epsilon\,\|^2$ (as in (49))). Show that $(Y - \tilde{\theta}) \perp_M \mathcal{R}[X]$, and so this resulting weighted LSE $\tilde{\theta} = X\tilde{\beta}$ must satisfy the *weighted normal equations*

(71) $X'M\tilde{\theta} = X'MY$, (equivalently, $X'MX\tilde{\beta} = X'MY$).

Summary: $X\tilde{\beta} = \tilde{\theta}$ if and only if $\tilde{\beta}$ satisfies the weighted normal equations. Also,

(72) $\tilde{\beta} = (X'MX)^-(X'MY)$ does satisfy the weighted normal equations and

(73) $\tilde{\theta} = X\tilde{\beta} = [X(X'MX)^-X'M]\,Y \equiv P_\Omega Y$ projects Y onto $\Omega \equiv \mathcal{R}[X]$.

Exercise 3.14 (Minimum variance unbiased linear estimators) (a) Let $X_1, \ldots, X_n$ be uncorrelated with common mean μ and common finite variance σ^2. All linear estimators $T \equiv \sum_1^n a_i X_i$ having $\sum_1^n a_i = 1$ are unbiased estimators of μ (that is, $\mathrm{E}T = \mu$). Show that the choice with all $a_i = 1/n$ has minimum variance within this class of linear unbiased estimators.
(b) Determine the minimum variance unbiased linear estimator of the common mean μ when the variances are $\sigma^2/c_1, \ldots, \sigma_n^2/c_n$, with the c_k being known constants.

4 The Multivariate Normal Distribution

Definition 4.1 (Jointly normal) Call $Y = (Y_1, \ldots, Y_n)'$ *jointly normal with* 0 *means* if there exist iid $N(0,1)$ rvs $X_1, \ldots, X_k$ and an $n \times k$ matrix A of known constants for which $Y = AX$. [We again write Y in this section, rather than $\vec{Y}$, when the context seems clear.] Note that the $n \times n$ *covariance matrix* $\Sigma_Y \equiv \Sigma$ of the random vector Y is

(1) $\Sigma \equiv \Sigma_Y = \mathrm{E}\,YY' = \mathrm{E}AXX'A' = AA'.$

The covariance matrix of X is the $k \times k$ identity matrix I_k. We will write $X \cong N(0, I_k)$, and we will write $Y \cong N(0, \Sigma)$. Then write $Y \cong N(\mu, \Sigma)$ if $Y - \mu \cong N(0, \Sigma)$. Call Y *multivariate normal* with mean vector μ and covariances matrix Σ. Call Y *nondegenerate* when $|\Sigma| \neq 0$ (that is, the determinant of Σ is not equal to 0). Say that $Y_1, \ldots, Y_n$ are *linearly independent* if (rank Σ) $= n$. Of course, this means that

(2) Y is nondegenerate if and only if $\operatorname{rank}(A) = n$.

Now, Σ is symmetric. Also $a\Sigma a' = \operatorname{Var}[aY] \geq 0$ for all vectors a. When $a\Sigma a' \geq 0$ for all vectors a, the symmetric matrix Σ is called *nonnegative definite*, and one writes $\Sigma \geq 0$.

Theorem 4.1 (Densities) If $Y \cong N(0, \Sigma)$ is nondegenerate, then Y has density (with respect to Lebesgue measure on R_n) given by

(3) $$f_Y(y) = \frac{1}{(2\pi)^{n/2}\,|\Sigma|^{1/2}} \exp(-y'\Sigma^{-1}y/2) \qquad \text{for all } y \in R_n.$$

[Note that each possible normal distribution is completely determined by μ and Σ.]

Proof. Now, $Y = XA$, where $AA' = \Sigma$, $(\operatorname{rank} A) = n$, $|A| \neq 0$, $X \cong N(0, I_k)$. It is trivial that

(a) $P(X \in B_n) = \int 1_{B_n} f_X(x)dx$ with $f_X(x) \equiv (2\pi)^{-n/2}\exp(-x'x/2)$.

Thus $X = A^{-1}Y$ gives

$$P(Y \in B_n) = P(AX \in B_n) = P(X \in A^{-1}B_n) = \int 1_{A^{-1}B_n}(x) f_X(x)\,dx$$

$$= \int 1_{A^{-1}B_n}(A^{-1}y) f_X(A^{-1}y) \left|\frac{\partial x}{\partial y}\right|^+ dy$$

$$= \int 1_{B_n}(y)(2\pi)^{-n/2}\exp(-(A^{-1}y)'(A^{-1}y)/2) \left|\frac{\partial x}{\partial y}\right|^+ dy$$

(b) $$= \int_{B_n} (2\pi)^{-n/2}|\Sigma|^{-1/2}\exp(-y'\Sigma^{-1}y/2)\,dy,$$

since $(A^{-1})'(A^{-1}) = (AA')^{-1} = \Sigma^{-1}$ and

(c) $$\left|\frac{\partial x}{\partial y}\right|^+ = |A^{-1}|^+ = \sqrt{|A'^{-1}||A^{-1}|} = \sqrt{|\Sigma^{-1}|} = 1/\sqrt{|\Sigma|}.$$

This is the required statement. □

Theorem 4.2 (Characteristic functions and representations)
(a) If we are given a random vector $Y = A^{n\times k} X^{k\times 1}$ where $X \cong N(0, I_k)$, we have

(4) $$\phi_Y(t) \equiv \mathrm{E}\, e^{it'Y} = \exp(-t'\Sigma t/2)$$

with $\Sigma \equiv AA'$ and $\mathrm{rank}(\Sigma) = \mathrm{rank}(A)$.
(b) If Y has characteristic function $\phi_Y(t) \equiv \mathrm{E}e^{itY} = \exp(-t'\Sigma t/2)$ with $\Sigma \geq 0$ of rank k, then

(5) $$Y \cong A^{n\times k} X^{k\times 1} \qquad \text{with (rank } A) = k \text{ and } X \cong N(0, I).$$

(Thus the number of independent rvs X_i's needed is equal to the rank of A.)

Proof. Our proof will use the fact that the characteristic function ϕ_Y of any rv Y is unique (as will be shown below in chapter 13.) [When a density function does not exist, one can use this characteristic function for many of the same purposes.] We observe that

$$\phi_Y(t) = \mathrm{E}\exp(it'AX) = \mathrm{E}\ \exp(i(A't)'X)$$

(a) $$= \exp(-(A't)'(A't)/2) \qquad \text{since } \mathrm{E}\, e^{itX_j} = \exp(-t^2/2)$$

by example 9.3.2 below

(b) $$= \exp(-t'(AA')t/2).$$

The converse follows from (A.3.7). □

Even when a multivariate normal rv Y does not have a density, the characteristic function can often be manipulated to establish a desired result.

Theorem 4.3 (Marginals, independence, and linear combinations) Suppose that

$$Y = (Y_1, \ldots, Y_k, Y_{k+1}, \ldots, Y_n)' \cong N(0, \Sigma) \qquad \text{with} \qquad \Sigma \equiv \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$

(i) The marginal covariance matrix of $(Y_1, \ldots, Y_k)'$ is the $k \times k$ matrix Σ_{11}, and

(6) $$(Y_1, \ldots, Y_k)' \cong N(0, \Sigma_{11}).$$

(ii) If $\Sigma_{12} = 0$, then $(Y_1, \ldots, Y_k)'$ and $(Y_{k+1}, \ldots, Y_n)'$ are independent.
(iii) If (Y_1, Y_2) is a jointly normal rv, then Y_1 and Y_2 are independent if and only if they have the zero covariance $\mathrm{Cov}[Y_1, Y_2] = 0$.
(iv) Linear combinations of normals are normal.

Proof. (i) Use the first k coordinates of the representation $Y = AX$.
(ii) Use the fact that one can factor

$$\phi_Y(t) = \exp(-\frac{1}{2}\, t' \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{pmatrix} t).$$

(iii) Just apply (ii), as the other direction is trivial.
(iv) $Z^{m\times 1} \equiv B^{m\times n} Y^{n\times 1} = B(AX) = (BA)X$. □

Theorem 4.4 (Conditional distributions) If

(7) $$Y = \begin{bmatrix} Y^{(1)} \\ Y^{(2)} \end{bmatrix} \cong N\left[\begin{bmatrix} \mu^{(1)} \\ \mu^{(2)} \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}\right],$$

then

(8) $$Y^{(1)}\Big|Y^{(2)} = y^{(2)} \;\cong\; N(\mu^{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(y^{(2)} - \mu^{(2)}),\; \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}).$$

Note that

(9) $$|\Sigma| = |\Sigma_{22}||\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}|.$$

Proof. The vector

(10) $$Z \equiv \begin{bmatrix} Z^{(1)} \\ Z^{(2)} \end{bmatrix} \equiv \begin{bmatrix} (Y^{(1)} - \mu^{(1)}) - \Sigma_{12}\Sigma_{22}^{-1}(Y^{(2)} - \mu^{(2)}) \\ Y^{(2)} - \mu^{(2)} \end{bmatrix}$$

is just a linear combination of the Y_i's, and so it is normal. We need only verify the means and variances. But we did this in discussion A.3.4. □

Exercise 4.1 Show that (Y_1, Y_2) can have normal marginals without being jointly normal. [Hint. Consider starting with a joint $N(0, I)$ density on R_2 and move mass in a symmetric fashion to make the joint distribution nonnormal, but still keeping the marginals normal.]

Quadratic Forms

Exercise 4.2 Let $Y^{n\times 1} \cong N(0, I)$, and suppose that A is symmetric and of rank r. Then $Y'AY \cong \chi_r^2$ if and only if A is a projection matrix (that is, $A^2 = A$).

Exercise 4.3 Let $Y^{n\times 1} \cong N(0, I)$. Suppose that A and B are symmetric and both $Y'AY$ and $Y'BY$ have chisquare distributions. Show that $Y'AY$ and $Y'BY$ are independent if and only if $AB = 0$.

Exercise 4.4 Suppose A and B are $n \times n$ projection matrices with ranks r_A and r_B, and suppose $AB = 0$ and $I - A - B \geq 0$. Then:
(a) $I - A$ is a projection matrix of rank $n - r_A$.
(b) $I - A - B$ is a projection matrix of rank $n - r_A - r_B$.

Exercise 4.5 Suppose $Y^{n\times 1} \cong N(0, \Sigma)$, and let A be an arbitrary symmetric matrix of rank r. Show that $Y'AY \cong \chi_r^2$ if and only if $A\Sigma A = A$.

The Multivariate CLT

The following result is theorem 10.1.3, but we also list it here for convenient referral.

Theorem 4.5 Suppose that the random vectors $X_1, \ldots, X_n$ are iid (μ, Σ). Then

(11) $$\sqrt{n}\,(\bar{X}_n - \mu) \to_d N(0, \Sigma) \qquad \text{as } n \to \infty.$$

Normal Processes

To specify a normal process, we must specify consistent distributions (in the sense of Kolmogorov's consistency theorem). But μ and Σ completely specify $N(\mu, \Sigma)$, while the marginals of $N(\mu, \Sigma)$ are $N(\mu^{(1)}, \Sigma_{11})$. Thus a normal process exists, provided only that the mean value function $\mu(\cdot)$ on I and the covariance function $\mathrm{Cov}(\cdot, \cdot)$ on $I \times I$ are well-defined and are such that $\mathrm{Cov}(\cdot, \cdot)$ is *nonnegative definite* (meaning that every n-dimensional covariance matrix formed from it is nonnegative definite).

We call $\{\mathbb{S}(t) : 0 \le t < \infty\}$ a *Brownian motion* if $\mathbb{S}$ is a normal process having

(12) $\quad \mathrm{E}\mathbb{S}(t) = 0 \qquad \text{and} \qquad \mathrm{Cov}[\mathbb{S}(s), \mathbb{S}(t)] = s \wedge t \qquad \text{for all } s, t \ge 0.$

Since this covariance function is nonnegative definite, a version of the process $\mathbb{S}$ exists on $(R_{[0,\infty)}, \mathcal{B}_{[0,\infty)})$ by the Kolmogorov consistency condition. Then

(13) $\quad \mathbb{U}(t) \equiv -[\mathbb{S}(t) - t\,\mathbb{S}(1)]$ for all $0 \le t \le 1$ is called *Brownian bridge*.

It is a normal process on $(R_{[0,1]}, \mathcal{B}_{[0,1]})$ for which

(14) $\quad \mathrm{E}\mathbb{U}(t) = 0 \qquad \text{and} \qquad \mathrm{Cov}[\mathbb{U}(s), \mathbb{U}(t)] = s \wedge t - st \qquad \text{for all } 0 \le s, t \le 1.$

References

I would like to use this discussion of the literature to say a very heartfelt "Thank you!" to a number of people who have figured prominently in my professional life. Especially, I want to thank my professors Fred Andrews (University of Oregon), Donald Truax (University of Oregon), and Lincoln Moses (Stanford University), whose voluntary efforts on my behalf had far-reaching consequences on most aspects of my life. I shall offer some thoughts on my own personal history as well as the subject matter of this book. My view is strongly affected by how I came to learn about these things. Others have undoubtedly had different experiences.

Measure theory
This text begins with five chapters devoted to measure theory. Halmos (1950) has had a major influence on what future books on measure theory and real analysis would contain and how they would present the subject. Other books on measure theory and real analysis that I have found to be especially useful include Royden (1963), Hewitt and Stromberg (1965), Rudin (1966), and the nicely simplified presentation of Bartle (1966). Many theorems in this introductory part are to some degree recognizable from several of these sources (and/or from the other sources listed in the probability section below). Certainly, Halmos's book was a popular one while I was getting my M.S. degree in mathematics at the University of Oregon, 1960–1962. My own introduction to "real and abstract analysis" came from a beautiful course taught by Karl Stromberg. Later, Edwin Hewitt was a colleague at the University of Washington. So it is a real pleasure for me to cite their work at various points. Lou Ward taught the topology course that I took at Oregon. He gave us a list of theorems, and we had to come up with proofs and present them. That was the most fun I ever had in the classroom. A good deal of appendix B reflects what I learned in his course. Kelly (1955), Copson (1968), and Housain (1977) are useful published sources. Watching over the Oregon graduate students in mathematics was Andrew Moursand, chairman. He really cared about all of us, and I owe him my thanks.

Probability
Loève's (1977–78, originally 1955) presentation has been a very influential work on probability, certainly from the pedagogical point of view. To me, it refines and specializes much general analysis to probability theory, and then treats a broad part of this subject. Clearly, many learned probability from his text. Also, many seem to follow notational conventions used in his book. But I was rather late in learning from it. My original training was at Stanford from lectures that became Chung (1974), and those lectures also reflected Chung's efforts regarding translation of Gnedenko and Kolmogorov (1954). I truly enjoyed Chung's course, and his book. Breiman's (1968) style coincided most closely with my own. I particularly liked his treatment of partial sum and empirical processes, as one would suspect from my own research. I have sometimes used his text as a "permanent reference" to stand beside my own notes in my courses on probability theory. My choice of notation has been most influenced by Loève and Breiman. Feller (1966) has a different flavor from most probability texts, and it includes various interesting approaches not found

elsewhere. And it is informative on rates of approximation. Billingsley (1968) created some excitement and spawned much interesting work, and a bit of that is included here. Doob's (1954) work on martingales has had a huge influence on the subject. I find Meyer (1966) and Hall and Heyde (1980) particularly significant. Lectures by Tom Fleming that led to Fleming and Harrington (1991) sparked part of my martingale presentation here. Whittaker and Watson (1963) is still a superb source for the gamma function. Lehmann (1959) has greatly influenced my view of conditional probability and expectation. This brings me back to the University of Oregon, and to Fred Andrews. Fred "recruited me to statistics" and then taught a year-long course out of Lehmann's book (even though I was the only student), and he was one of those who lined me up for a National Science Foundation fellowship that made it possible for me to go to Stanford University. Don Truax also figured heavily in this. He cared about me, and I learned a lot from him. Thank you both!

The scene shifts southward. My years at Stanford were very fruitful, and I met some fine people. Ingram Olkin is fun and a good teacher, and he went out of his way to be helpful to me. The multivariate topics in appendix A represent things I learned from him. Lincoln Moses was my thesis advisor. This relationship grew out of a course in nonparametric statistics that I took from him. One of the topics in his course was Charles Stein's approach to the central limit theorem. Lin spoke on it for three days, even though he had to leave a couple of well-acknowledged gaps in his presentation—because he believed it was good work. That gave me a profound respect for him as a teacher. The topic caught my imagination, and chapters 11 and 17 reflect this. Lin was also my assigned advisor when I arrived at Stanford. His second sentence to me was, "OK, Shorack, what's important to you in life"? My answer had a lot to do with the geography of the Pacific Northwest. Two months before I graduated he responded on my behalf to a University of Washington opening. Wow!

At Washington I had a chance to teach courses in probability and statistics. And I learned a lot from my association with Ron Pyke, and later with Jon Wellner. The presentations in parts of chapters 12 and 14 reflect this to varying degrees. Fritz Scholz got me started on gamma approximations in the central limit theorem. Likewise, work with David Mason on quantile functions, embedding, and trimmed means is reflected in parts of chapters 6 and appendix C. I offer them all my thanks.

Obviously, I also owe a huge debt to "the literature" in regard to all these topics, and I will list some of those sources below. However, this is a textbook. It is not a research monograph. My emphasis is on presentation, not attribution. Often, my citation concerns where I learned something rather than who did it originally. And in some areas (especially, chapters 6 and 12) I have offered only a sampling of the citations that I could have given. Moreover, I have often chosen to cite a book instead of an original paper. My own work is cited "too heavily" because it is written in the same style and notation as this book.

The bibliography contains quite a number of other books on probability theory, and many are very good books. But it is the ones listed above that have had the most influence on me. I hope that the reader will find that my book also has a somewhat different flavor—a statistical flavor. That flavor will be enhanced if you think of chapters 16 and 17 and the first appendix of the original 2000 Edition as part of the total package.

Special thanks to Chari Boehnke, Roger and Heather Shorack, the Michael Boehnke family, the Barbara Aalund family, Kathleen Shorack, David Mason, Michael Perlman, Fritz and Roberta Scholz, Jon Wellner, the Jan Beirlant family, Piet Groeneboom, Frits Ruymgaart, Derek Dohn, and Pauline Reed for enabling me to write this book.

Thanks Offered to Publishers
I would like to thank several publishers for allowing me to use material in this text that originated elsewhere in my body of work. The citations made here are to publications referenced below. Various parts of Chapter 12, a few small pieces throughout the text, and much of the current Section 13.11 appeared in Shorack and Wellner (1986) (published by John Wiley & Sons). (Regarding the 1*st* Edition: The trimmed means example of Section 16.1 owes a substantial debt to Shorack (1997a) (published by Gordon and Breach Publishers), the trimmed mean and bootstrap examples of Sections 16.2 and 16.3 are largely from Shorack (1998) (published by Elsevier Science), while the L-statistics example of Section 16.4 is largely from Shorack (1997b) (published by Elsevier Science).) The inequalities from these publications upon which the applications are based are found at various points in the current Chapter 6 and Appendix B. The author appreciates the generosity of these publishers.

References

CHAPTERS 1–5 and APPENDIX B

Ash, R. (1972) *Real Analysis and Probability.* Academic Press, New York.

Bartle, R. (1966) *The Elements of Integration.* John Wiley & Sons, New York.

Breiman, L. (1968) *Probability.* Addision-Wesley, Reading, MA.

Chow, Y. and H. Teicher (1997) *Probability Theory: Independence, Interchangeability, Martingales* 3rd ed. Springer-Verlag, New York.

Cohn, D. (1980) *Measure Theory.* Birkhäuser, Boston.

Copson, E. (1968) *Metric Spaces.* Cambridge University Press, Cambridge.

Halmos, P. (1950) *Measure Theory.* Van Nostrand, Princeton, NJ.

Hewitt, E. and K. Stromberg (1965) *Real and Abstract Analysis.* Springer-Verlag, New York.

Husain, T. (1977) *Topology and Maps.* Plenum Press, New York.

Loève, M. (1977–78) *Probability Theory.* Springer-Verlag, New York.

Meyer, P. (1966) *Probability and Potentials.* Blaisdell, Waltham, MA.

Royden, H. (1967) *Real Analysis.* Macmillan, New York.

Rudin, W. (1966) *Real and Complex Analysis.* McGraw-Hill, New York.

CHAPTERS 6, 12, 15 and Appendix C

Billingsley, P. (1968) *Convergence of Probability Measures.* John Wiley & Sons, New York.

Breiman, L. (1968) *Probability.* Addision-Wesley, Reading, MA.

Chung, K. (1974) *A Course in Probability Theory* 2nd ed. Academic Press, New York.

Cramér, H. and M. Leadbetter (1967) *Stationary and Related Stochastic Processes.* John Wiley & Sons, New York.

Csörgő, M., S. Csörgő, L. Horváth, and D. Mason (1986) What portion of the sample makes a partial sum asymptotically stable or normal? *Probab. Theory Related Fields* **72**, 1–16.

Csörgő, M., S. Csörgő, L. Horváth, and D. Mason (1986) Weighted empirical and quantile processes. *Ann. Probab.* **14**, 31–85.

Csörgő, S., E. Haeusler, and D. Mason (1988) The asymptotic distribution of trimmed sums. *Ann. Probab.* **16**, 672–699. **Csörgő**, S., E. Haeusler, and D. Mason (1989) A probabilistic approach to the asymptotic distribution of sums of independent identically distributed random variables. *Adv. in Appl. Math.* **9**, 259–333.

Doob, J. (1949) Heuristic approach to the Kolmogorov–Smirnov theorems. *Ann. Math. Statist.* **20**, 393–403.

Freedman, D. (1971) *Brownian Motion and Diffusion.* Holden-Day, San Fran.

Haeusler, E. and D. Mason (1989) A law of the iterated logarithm for modulus trimming. *Colloquia Math. Soc. J. Bolyai,* **57**, *Limit Theorems in Probability and Statistics.*

Mason, D. (2001) An exponential inequality for a weighted approximation to the uniform empirical process with applications. *Institute of Mathematical Statistics: Lecture Notes-Monograph Series* **36**, 477-498.

Mason, D. and G. Shorack (1990) Necessary and sufficient conditions for asymptotic normality of trimmed L-statistics. *J. Statist. Plan. Inf.* **25**, 111–139.

Mason, D. and G. Shorack (1992) Necessary and sufficient conditions for asymptotic normality of L-statistics. *Ann. Probab.* **20**, 1779–1804.

Meyer, P. (1966) *Probability and Potentials.* Blaisdell, Waltham, MA.

O'Brien, R. (1980) A limit theorem for sample maxima and heavy branches in Galton–Watson trees. *J. Appl. Prob.* **17** 539–545.

Ross, S. (1997) *An Introduction to Probability Models*, 6th ed., Academic Press, San Diego.

Shorack, G. (1972) Functions of order statistics. *Ann. Math. Statist.* **43**, 412–427.

Shorack, G. (1998) The asymptotic normality condition. Univ. Washington Dept. of Statistics Technical Reports 323, 336, 337, and 338.

Shorack, G. (1991) Embedding the finite sampling process at a rate. *Ann. Probability* **19**, 826–842.

Shorack, G. (1996) Linear rank statistics, finite sampling, permutation tests and Winsorizing. *Ann. Statist.* **24**, 1371–1385.

Shorack, G. (1997a) Inequalities for quantile functions with a uniform studentized CLT that includes trimming. *Nonparametric Statist.* **8**, 307–335.

Shorack, G. (1997b) Uniform CLT, WLLN, LIL and bootstrapping in a data analytic approach to trimmed L-statistics. *J. Statist. Plan. Inf.* **60**, 1–44.

Shorack, G. (1998) Applications of weighted approximations via quantile inequalities. *Asymptotic Methods in Probability and Statistics, A Volume in Honor of Miklós Csörgő.* B. Szyszkowicz, ed., 151–167, Elsevier, Amsterdam.
Shorack, G. and J. Wellner (1986) *Empirical Processes with Applications to Statistics.* John Wiley & Sons, New York.

CHAPTER 7

Breiman, L. (1968) *Probability.* Addision-Wesley, Reading, MA.
Lehmann, E. (1959) *Testing Statistical Hypotheses.* John Wiley & Sons, New York.

CHAPTER 8

Breiman, L. (1968) *Probability.* Addision-Wesley, Reading, MA.
Chow, Y. and H. Teicher (1997) *Probability Theory. Independence, Interchangeability, Martingales* 3rd ed. Springer-Verlag, New York.
Chung, K. (1974) *A Course in Probability Theory* 2nd ed. Academic Press, New York.
Feller, W. (1966) *An Introduction to Probability Theory and Its Applications.* Vol. 2. John Wiley & Sons, New York.
Giné, E., F. Gótze, and D. Mason (1998) When is the student t-statistic asymptotically standard normal? *Ann. Probab.* **25**, 1514–1531.
Gnedenko, B. and A. Kolmogorov (1954) *Limit Distributions for Sums of Independent Random Variables.* Addison-Wesley, Cambridge, Mass.
Hoffman-Jorgensen, J. (1974) Sums of independent Banach space valued random variables. *Studia Math.* **52**, 159–186.
Kallenberg, O. (1997) *Foundations of Modern Probability.* Springer-Verlag, New York.
Loève, M. (1977–78) *Probability Theory.* Springer-Verlag, New York.
Resnick, S. (1999) *A Probability Path.* Birkhäuser, Boston.
Strassen, V. (1967). Almost sure behavior of sums of independent random variables and martingales. *Proceeding of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* **2**, 315–343, University of California Press, Berkeley, CA.

CHAPTERS 9–11 and Appendix A

Ahlfors, L. (1979) *Complex Analysis* McGraw-Hill, Inc., New York.
Anderson, T. (1984) *Introduction to Multivariate Statistical Analysis.* John Wiley & Sons., New York.
Bickel, P. and K. Doksum (1977) *Mathematical Statistics: Basic Ideas and Special Topics.* Holden-Day, San Francisco.
Bhattacharya, R. and R. Rao (1976) *Normal Approximation and Asymptotic Expansions.* John Wiley & Sons., New York.
Billingsley, P. (1986) *Probability and Measure.* John Wiley & Sons, New York.
Breiman, L. (1968) *Probability.* Addision-Wesley, Reading, MA.
Csörgő, S. and D. Mason (1989) Bootstrapping empirical functions. *Ann. Statist.* **17**, 1447–1471.

Durrett, R. (1996). *Probability: Theory and Examples.* Wadsworth, Belmont, CA.
Feller, W. (1966) *An Introduction to Probability Theory and Its Applications.* Vol. 2. John Wiley & Sons, New York. **Galambos**, J. (1995) *Advanced Probability Theory* 2nd ed. Marcel Dekker, New York.
Giné, E., F. Götze, and D. Mason (1997) When is the student t-statistic asymptotically standard normal? *Ann. Probab.* **25** 1514–1531.
Giné, E. and J. Zinn (1990) Necessary conditions for the bootstrap of the mean. *Ann. Statist.* **17**, 684–691.
Gnedenko, B. and A. Kolmogorov (1954) *Limit Distributions for Sums of Independent Random Variables.* Addison-Wesley, Cambridge, Mass.
Hall, P. (1983) Chi squared approximations to the distribution of a sum of independent random variables. *Ann. Statist.* **11**, 1028–1036.
Hall, P. (1990) Asymptotic properties of the bootstrap for heavy tailed distributions. *Ann. Statist.* **18**, 1342–1360.
Johnson, N., A. Kotz, and N. Balakrishnan (1994) *Continuous Univariate Distributions* Volumes 1 and 2. John Wiley & Sons, New York.
Kendall, M. and A. Stuart (1977) *The Advanced Theory of Statistics* Vol. 2. Macmillan, New York.
Lehmann, E. and G. Casella (1998) *Theory of Point Estimation* 3rd ed. Springer-Verlag, New York.
Lévy, P. (1937) *Théorie de l'Addition des Variables Aléatoires.* Bautier-Villars, Paris.
Loève, M. (1977–78) *Probability Theory.* Springer-Verlag, New York.
Mason, D. and G. Shorack (1992) Necessary and sufficient conditions for asymptotic normality of L-statistics. *Ann. Probab.* **20** 1799–1804.
Petrov, V. (1975) *Sums of Independent Random Variables.* Springer-Verlag, New York.
Rao, C. (1965) *Linear Statistical Inference and Its Applications* 2nd ed. John Wiley & Sons, New York.
Shiryayev, A. (1984) *Probability.* Springer-Verlag, New York.

CHAPTER 13

Andersen, P., O. Borgan, R. Gill, and N. Keiding (1993) *Statistical Models Based on Counting Processes.* Springer-Verlag, New York.
Breiman, L. (1968) *Probability.* Addision-Wesley, Reading, MA.
Doob, J. (1953) *Stochastic Processes.* John Wiley & Sons, New York.
Fleming, T. and D. Harrington (1991) *Counting Processes and Survival Analysis.* John Wiley & Sons, New York.
Hall, P. and C. Heyde (1980) *Martingale Limit Theory and Its Application.* Academic Press, New York.
Liptser, R. and A. Shiryaev (1977) *Statistics of Random Processes I.* Springer-Verlag, New York.
Meyer, P. (1966) *Probability and Potentials.* Blaisdell, Waltham, MA.
Williams, D. (1991) *Probability with Martingales.* Cambridge University Press, Cambridge.

Chapter 14

Billingsley, P. (1968) *Convergence of Probability Measures.* John Wiley & Sons, New York.

Dudley, R. (1976) *Convergence of laws on metric spaces, with a view to statistical testing.* Matematisk Institut Aarhus Universitet.

Dudley, R. (1989) *Real Analysis and Probability.* Wadsworth & Brooks/Cole, Pacific Grove, CA.

Skorokhod, A. (1956) Limit theorems for stochastic processes. *Theor. Probab. Appls.* **1**, 261–290.

Index