Chapter 2
Some Basic Large Sample Theory

1. **Modes of Convergence**
   Convergence in distribution, $\rightarrow_d$
   Convergence in probability, $\rightarrow_p$
   Convergence almost surely, $\rightarrow_{a.s.}$
   Convergence in $r-$th mean, $\rightarrow_r$
   Metrics on probability distributions

2. **Classical Limit Theorems**
   Weak and strong laws of large numbers
   Classical (Lindeberg) CLT
   Liapounov CLT
   Lindeberg-Feller CLT
   Cramér-Wold device; Mann-Wald theorem; Slutsky's theorem
   Delta-method

3. **Examples**
   One-sample $t-$test
   One-sample normal theory test of variance
   Two-sample tests for means
   Linear regression with non-normal errors
   The correlation coefficient
   Pearson's chi-square tests, simple null hypothesis

4. **Replacing $\rightarrow_d$ by $\rightarrow_{a.s.}$**

5. **Empirical Measures and Empirical Processes**
   The empirical distribution function; the uniform empirical process
   The Brownian bridge process
   Finite-dimensional convergence
   A hint of Donsker's Mann-Wald type theorem
   General empirical measures and processes

6. **The Partial Sum Process and Brownian motion**
   Brownian motion
   Existence of Brownian motion and Brownian bridge as continuous processes

7. **Sample Quantiles**

Chapter 2

Some Basic Large Sample Theory

1 Modes of Convergence

Consider a probability space $(\Omega, \mathcal{A}, P)$. For our first three definitions we suppose that X, X_n, $n \geq 1$ are all random variables defined on this one probability space.

Definition 1.1 We say that X_n *converges a.s.* to X, denoted by $X_n \to_{a.s.} X$, if

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

or, equivalently, if, for every $\epsilon > 0$

$$(2) \qquad P(\sup_{m \geq n} |X_m - X| > \epsilon) \to 0 \qquad \text{as } \ n \to \infty\,.$$

Definition 1.2 We say that X_n *converges in probability to* X and write $X_n \to_p X$ if for every $\epsilon > 0$

$$(3) \qquad P(|X_n - X| > \epsilon) \to 0 \qquad \text{as } \ n \to \infty\,.$$

Definition 1.3 Let $0 < r < \infty$. We say that X_n *converges in* $r-th$ *mean* to X, denoted by $X_n \to_r X$, if

$$(4) \qquad E|X_n - X|^r \to 0 \qquad \text{as } \ n \to \infty \ \text{ for functions } \ X_n, X \in L_r(P)\,.$$

Definition 1.4 We say that X_n *converges in distribution* to X, denoted by $X_n \to_d X$, or $F_n \to F$, or $\mathcal{L}(X_n) \to \mathcal{L}(X)$ with $\mathcal{L}$ referring to the the "law" or "distribution", if the distribution functions F_n and F of X_n and X satisfy

$$(5) \qquad F_n(x) \to F(x) \qquad \text{as } \ n \to \infty \qquad \text{for each continuity point } \ x \ \text{ of } F\,.$$

Note that $F_n \equiv 1_{[1/n,\infty)} \to_d 1_{[0,\infty)} \equiv F$ even though $F_n(0) = 0$ does not converge to $1 = F(0)$. The statement $\to_d$ will carry with it the implication that F corresponds to a (proper) probability measure P.

Definition 1.5 A sequence of random variables $\{X_n\}$ is *uniformly integrable* if

$$(6) \qquad \lim_{\lambda \to \infty} \limsup_{n \to \infty} E\left\{|X_n| 1_{[|X_n| \geq \lambda]}\right\} = 0\,.$$

Theorem 1.1 (Convergence implications).
A. If $X_n \to_{a.s.} X$, then $X_n \to_p X$.
B. If $X_n \to_p X$, then $X_{n'} \to_{a.s.} X$ for some subsequence $\{n'\}$.
C. If $X_n \to_r X$, then $X_n \to_p X$.
D. If $X_n \to_p X$ and $|X_n|^r$ is uniformly integrable, then $X_n \to_r X$.
If $X_n \to_p X$ and $\limsup_{n\to} E|X_n|^r \le E|X|^r$, then $X_n \to_r X$.
E. If $X_n \to_r X$ then $X_n \to_{r'} X$ for all $0 < r' \le r$.
F. If $X_n \to_p X$, then $X_n \to_d X$.
G. $X_n \to_p X$ if and only if every subsequence $\{n'\}$ contains a further subsequence $\{n''\}$ for which $X_{n''} \to_{a.s.} X$.

Theorem 1.2 (Vitali's theorem). Suppose that $X_n \in L_r(P)$ where $0 < r < \infty$ and $X_n \to_p X$. Then the following are equivalent:
A. $\{|X_n|^r\}$ are uniformly integrable.
B. $X_n \to_r X$.
C. $E|X_n|^r \to E|X|^r$.
D. $\limsup_n E|X_n|^r \le E|X|^r$.

Before proving the theorems we need a short review of some facts about convex functions and some inequalities. We first briefly review convexity. A real valued function f is *convex* if

$$f(\alpha x + (1-\alpha)y) \le \alpha f(x) + (1-\alpha) f(y) \tag{7}$$

for all x, y and all $0 \le \alpha \le 1$. This holds if and only if

$$f\left(\frac{1}{2}(x+y)\right) \le \frac{1}{2}[f(x) + f(y)] \tag{8}$$

for all x, y provided f is continuous and bounded. Note that (8) holds if and only if

$$f(s) \le \frac{1}{2}[f(s-r) + f(s+r)] \qquad \text{for all } \ r, s\,. \tag{9}$$

Also

$$f''(x) \ge 0 \qquad \text{for all } \ x \ \text{ implies } \ f \ \text{ is convex}\,. \tag{10}$$

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 a prespecified x_0 in the interior of the domain of f; this is called the *supporting hyperplane theorem*.

Definition 1.6 Assuming the following expectations (integrals) exist,

$$\mu \equiv E(X) = \text{ the } \textit{mean} \text{ of } \ X\,. \tag{11}$$
$$\sigma^2 \equiv Var[X] \equiv E(X-\mu)^2 = \text{ the } \textit{variance} \text{ of } \ X\,. \tag{12}$$
$$E(X^k) = \textit{ k-th moment} \text{ of } \ X \text{ for } \ k \ge 1 \text{ an integer}\,. \tag{13}$$
$$E|X|^r = \textit{ r-th absolute moment} \text{ of } \ X \text{ for } r \ge 0\,. \tag{14}$$
$$E(X-\mu)^k = \textit{k- th central moment} \text{ of } \ X\,. \tag{15}$$
$$Cov[X,Y] \equiv E[(X-\mu_X)(Y-\mu_Y)] = \text{ the } \textit{covariance} \text{ of } \ X \text{ and } Y\,. \tag{16}$$

Proposition 1.1 If $E|X|^r < \infty$, then $E|X|^{r'}$ and $E(X^k)$ are finite for all $r' \le r$ and integers $k \le r$.

Proof. Now $|x|^{r'} \le 1 + |x|^r$; and integrability is equivalent to absolute integrability. $\Box$

Proposition 1.2 $Var(X) \equiv \sigma^2 < \infty$ if and only if $E(X^2) < \infty$. In this case $\sigma^2 = E(X^2) - \mu^2$.

Proof. Suppose that $\sigma^2 < \infty$. Then $\sigma^2 + \mu^2 = E(X-\mu)^2 + E(2\mu X - \mu^2) = E(X^2)$. Suppose that $E(X^2) < \infty$. Then $E(X^2) - \mu^2 = E(X^2) - E(2\mu X - \mu^2) = E(X-\mu)^2 = Var[X]$. $\Box$

Proposition 1.3 (c_r-inequality). $E|X+Y|^r \le c_r E|X|^r + c_r E|Y|^r$ where $c_r = 1$ for $0 < r \le 1$ and $c_r = 2^{r-1}$ for $r \ge 1$.

Proof. Case 1: $r \ge 1$. Then $|x|^r$ is a convex function of x; take second derivatives. 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 $\uparrow$ for $x \ge 0$; examine derivatives. Thus

$$\begin{aligned} |x+y|^r - |x|^r &= \int_x^{x+y} r t^{r-1} dt = \int_0^y r(x+s)^{r-1} ds \\ &\le \int_0^y r s^{r-1} ds = |y|^r , \end{aligned}$$

and now take expectations. $\Box$

Proposition 1.4 (Hölder inequality). $E|XY| \le E^{1/r}|X|^r E^{1/s}|Y|^s \equiv \|X\|_r \|Y\|_s$ for $r > 1$ where $1/r + 1/s = 1$ defines s. When the expectations are finite we have equality if and only if there exists A and B not both 0 such that $A|X|^r = B|Y|^s$ a.e.

Proof. The result is trivial if $E|X|^r = 0$ or ∞. Likewise for $E|Y|^s$. So suppose that $E|X|^r > 0$. Now

$$|ab| \le \frac{|a|^r}{r} + \frac{|b|^s}{s}, \qquad \text{as in the figure}.$$

Now let $a = |X|/\|X\|_r$ and $b = |Y|/\|Y\|_s$; and take expectations. Equality holds if and only if $|Y|/\|Y\|_s = (|X|/\|X\|_r)^{1/(1-s)}$ a.e.; if and only if

$$\frac{|Y|^s}{E|Y|^s} = \left(\frac{|X|}{\|X\|_r} \right)^{\frac{s}{s-1}} = \frac{|X|^r}{E|X|^r} \qquad a.e.$$

if and only if there exist $A, B \ne 0$ such that $A|X|^r = B|Y|^s$. This also gives the next inequality as an immediate consequence. $\Box$

Proposition 1.5 (Cauchy-Schwarz inequality). $(E|XY|)^2 \le E(X^2)E(Y^2)$ with equality if and only if there exists A, B not both 0 such that $A|X| = B|Y|$ a.e.

Remark 1.1 Thus for non-degenerate random variables (i.e. non-zero variance) with finite variance we have

$$(17) \qquad -1 \le \rho \le 1$$

where

$$(18) \qquad \rho \equiv Corr[X,Y] \equiv \frac{Cov[X,Y]}{\sqrt{Var[X]Var[Y]}}$$

is called the *correlation* of X and Y. Note that $\rho = 1$ if and only if $X - \mu_X = A(Y - \mu_Y)$ for some $A > 0$ and $\rho = -1$ if and only if $X - \mu_X = A(Y - \mu_Y)$ for some $A < 0$. Thus ρ measures linear dependence, not dependence.

Proposition 1.6 $\log E|X|^r$ is convex in r for $r \ge 0$. It is linear if and only if $|X| = c$ a.e. for some c.

Proof. Let $0 \le r \le s$. Apply the Cauchy-Schwarz inequality to $|X|^{(s-r)/2}$ and $|X|^{(s+r)/2}$ and take logs to get

$$\log E|X|^s \le \frac{1}{2}\left\{\log E|X|^{s-r} + \log E|X|^{s+r}\right\} .$$

□

Proposition 1.7 (Liapunov inequality). Let X be a random variable. Then $E^{1/r}|X|^r$ is $\uparrow$ in r for $r \ge 0$.

Proof. The slope of the chord of $y = \log E|X|^r$ is $\uparrow$ in r by proposition 1.6. That is, $(1/r)\log E|X|^r$ is $\uparrow$ in r. We used $P(\Omega) = 1 < \infty$ to show that $E|X|^{r'} < \infty$ if $E|X|^r < \infty$ for $r' \le r$ in proposition 1.1. □

Exercise 1.1 Let $\mu_r \equiv E|X|^r$. For $r \ge s \ge t \ge 0$ we have $\mu_r^{s-t}\mu_t^{r-s} \ge \mu_s^{r-t}$.

Proposition 1.8 (Minkowski's inequality). For $r \ge 1$ we have $E^{1/r}|X+Y|^r \le E^{1/r}|X|^r + E^{1/r}|Y|^r$.

Proof. It is trivial for $r = 1$. Suppose that $r > 1$. Then for any measure

$$(a) \qquad \begin{aligned} E|X+Y|^r &\le E|X||X+Y|^{r-1} + E|Y||X+Y|^{r-1} \\ &\le \{\|X\|_r + \|Y\|_r\} \, \||X+Y|^{r-1}\|_s \qquad \text{by Hölder's inequality} \\ &= \{\|X\|_r + \|Y\|_r\} \, E^{1/s}|X+Y|^{(r-1)s} \\ &= \{\|X\|_r + \|Y\|_r\} \, E^{1/s}|X+Y|^r . \end{aligned}$$

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

Proposition 1.9 (Basic inequality). Let $g \ge 0$ be an even function which is $\uparrow$ on $[0,\infty)$. Then for all random variables X and for all $\epsilon > 0$

$$(19) \qquad P(|X| \ge \epsilon) \le \frac{Eg(X)}{g(\epsilon)} .$$

Proof. Now

$$
\text{(a)} \qquad \begin{aligned} Eg(X) &= E\{g(X)1_{[|X|\geq\epsilon]}\} + E\{g(X)1_{[|X|<\epsilon]}\} \\ &\geq E\{g(X)1_{[|X|\geq\epsilon]}\} \geq g(\epsilon)E\{1_{[|X|\geq\epsilon]}\} \\ &= g(\epsilon)P(|X|\geq\epsilon) \end{aligned}
$$

as claimed. □

The next two inequalities are immediate corollaries of the basic inequality.

Proposition 1.10 (Markov's inequality).

$$
\text{(20)} \qquad P(|X|\geq\epsilon) \leq \frac{E|X|^r}{\epsilon^r} \qquad \text{for all } \epsilon > 0\,.
$$

Proposition 1.11 (Chebychev's inequality).

$$
\text{(21)} \qquad P(|X-\mu|\geq\epsilon) \leq \frac{Var[X]}{\epsilon^2} \qquad \text{for all } \epsilon > 0\,.
$$

Proposition 1.12 (Jensen's inequality). If g is convex on (a,b) where $-\infty \leq a < b \leq \infty$ and if $P(X\in(a,b)) = 1$ and $E(X)$ is finite (and hence $a < E(X) < b$), then

$$
\text{(22)} \qquad g(EX) \leq Eg(X)\,.
$$

If g is strictly convex, then equality holds in (22) if and only if $X = E(X)$ with probability 1.

Proof. Let l be a supporting hyperplane to g at EX. Then

$$
\begin{aligned} Eg(X) &\geq El(X) \\ &= l(EX) \qquad \text{since } l \text{ is linear and } P(\Omega)=1 \\ &= g(EX)\,. \end{aligned}
$$

Now $g(X)-l(X)\geq 0$. Thus $Eg(X) = El(X)$ if and only if $g(X) = l(X)$ almost surely, if and only if $X = EX$ almost surely. □

Exercise 1.2 For any function $h \in L_2(0,1)$, define a new function Th on $(0,1)$ by $Th(u) = u^{-1}\int_0^u h(s)ds$ for $0 < u \leq 1$. Note that T is an *averaging operator*. use the Cauchy-Schwarz inequality to show that

$$
\int_0^1 \{Th(u)\}^2 du \leq 4\int_0^1 h^2(u)du\,.
$$

Thus $T : L_2(0,1) \to L_2(0,1)$ is a *bounded linear operator* with $\|T\| \leq 2$. [Hint: write $Th(u) = u^{-1}\int_0^u h(s)s^\alpha s^{-\alpha}ds$ for some α.]

Exercise 1.3 Suppose that $X \sim$ Binomial(n,p). Use the basic inequality proposition 1.9 with $g(x) = \exp(rx)$, $r > 0$, to show that for $\epsilon \geq 1$

$$
\text{(23)} \qquad P\left(\frac{X/n}{p} \geq \epsilon\right) \leq \exp(-nph(\epsilon))
$$

where $h(\epsilon) = \epsilon(\log(\epsilon) - 1) + 1$. From this, show that for $\lambda > 0$ we have

$$(24) \qquad P\left(\sqrt{n}\left(\frac{X}{n} - p\right) \geq \lambda\right) \leq \exp\left(-\frac{\lambda^2}{2p}\psi\left(\frac{\lambda}{p\sqrt{n}}\right)\right),$$

where $\psi(x) \equiv 2h(1+x)/x^2$ is monotone decreasing on $[0,\infty)$ with $\psi(0) = 0$ and $\psi(x) \sim 2\log(x)/x$ as $x \to \infty$.

Proof. (Proof of theorem 1.1). A follows easily since, for any fixed $\epsilon > 0$

$$P(|X_n - X| \geq \epsilon) \leq P(\cup_{m \geq n}[|X_m - X| \geq \epsilon]) \to 0\,.$$

To prove B, first note that $X_n \to_p X$ implies that for every $k \geq 1$ there exists an interger n_k such that $P(|X_{n_k} - X| > 1/2^k) < 2^{-k}$; we can assume $n_k \uparrow$ in k; if not, take $n'_k \equiv n_k + k$. Let $A_m \equiv \cup_{k \geq m}[|X_{n_k} - X| > 2^{-k}]$ so that $P(A_m) \leq \sum_{k=m}^\infty 2^{-k} = 2^{-m+1}$. On $A_m^c = \cap_{k \geq m}[|X_{n_k} - X| \leq 2^{-k}|]$, $|X_{n_k} - X| \leq 2^{-k}$ for all $k \geq m$; i.e. on A_m^c, $X_{n_k}(\omega) \to X(\omega)$. Thus $X_{n_k} \to X$ on $A \equiv \cup_{m=1}^\infty A_m^c$, and $P(A^c) = P(\cap_{m=1}^\infty A_m) = \lim_m P(A_m) \leq \lim_m 2^{-m+1} = 0$.

Markov's inequality gives C via $P(|X_n - X| \geq \epsilon) \leq E|X_n - X|^r/\epsilon^r \to 0$. Hölder's inequality with $1/(r/r') + 1/q = 1$ gives E via

$$\begin{aligned} E|X_n - X|^{r'} &\leq \{E|X_n - X|^{r'(r/r')}\}^{r'/r}\{E1^q\}^{1/q} \\ \text{(a)} \qquad &= \{E|X_n - X|^r\}^{r'/r} \to 0\,; \end{aligned}$$

or, alternatively, use Liapunov's inequality.

Vitali's theorem 1.2 gives D.

Consider F. Let $X_n \sim F_n$ and $X \sim F$. Now

$$\begin{aligned} F_n(t) &= P(X_n \leq t) \leq P(X \leq t + \epsilon) + P(|X_n - X| \geq \epsilon) \\ &\leq F(t+\epsilon) + \epsilon \qquad \text{for all } \ n \geq \text{some } N_\epsilon\,. \end{aligned}$$

Also

$$\begin{aligned} F_n(t) &= P(X_n \leq t) \geq P(X \leq t - \epsilon \text{ and } |X_n - X| \leq \epsilon) \equiv P(AB) \\ &\geq P(A) - P(B^c) = F(t-\epsilon) - P(|X_n - X| > \epsilon) \\ &\geq F(t-\epsilon) - \epsilon \qquad \text{for } n \geq \text{ some } N'_\epsilon\,. \end{aligned}$$

Thus

$$\text{(b)} \qquad F(t-\epsilon) - \epsilon \leq \liminf F_n(t) \leq \limsup F_n(t) \leq F(t+\epsilon) + \epsilon\,.$$

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

Half of G follows from B since any $X_{n'} \to_p X$. We turn to the other half. Assume that $X_n \to_p X$ fails. Then there exists $\epsilon_0 > 0$ for which $\delta_0 \equiv \limsup P(|X_n - X| \geq \epsilon_0) > 0$. Thus there exists a subsequence $\{n'\}$ for which $P(|X_{n'} - X| \geq \epsilon_0) \to \delta_0 > 0$. Thus neither $X_{n'}$ nor any further subsequence $X_{n''}$ can $\to_{a.s.} X$. This is a contradiction. Thus $X_n \to_p X$. $\square$

Proof of Vitali's theorem, theorem 1.2: Suppose that A holds: thus $\{|X_n|^r\}_{n\geq 1}$ is uniformly integrable. Now $X_{n'} \to_{a.s.} X$ for some subsequence by Theorem 1.1. Thus $E|X|^r = E(\liminf |X_{n'}|^r) \leq \liminf E|X_{n'}|^r < \infty$ by Fatou's lemma and uniform integrability. Thus $X \in L_r(P)$. Now the

C_r−inequality gives $|X_n - X|^r \le C_r\{|X_n|^r + |X|^r\}$, so that the random variables $|X_n - X|^r$ are uniformly integrable. Thus we can write, for each fixed $\epsilon > 0$,

$$\begin{aligned} E|X_n - X|^r &\le E\{|X_n - X|^r 1_{[|X_n - X|^r \ge \epsilon]} + E\{|X_n - X|^r 1_{[|X_n - X|^r \le \epsilon]} \\ &\le E\{|X_n - X|^r 1_{[|X_n - X|^r \ge \epsilon]} + \epsilon^r \\ &\le \epsilon + \epsilon^r \end{aligned}$$

by the equivalence theorem for uniform integrability by choosing ϵ so small that $\limsup P(|X_n - X| \ge \epsilon) \le \delta_\epsilon$. Thus B holds. B implies C by Theorem 1.1 part E. C implies D trivially. Suppose that D holds. Define f_λ to be the bounded continuous function on $[0, \infty)$ given by

$$f_\lambda(x) = \begin{cases} x^r, & |x|^r \le \lambda \\ 0, & |x|^r \ge \lambda + 1 \\ \lambda^r - \lambda^r(x - \lambda), & \lambda < |x|^r < \lambda + 1. \end{cases}$$

Then

$$\begin{aligned} \limsup_n E\{|X_n|^r 1_{[|X_n|^r > \lambda + 1]}\} &= \limsup_n \left\{E|X_n|^r - E\{|X_n|^r 1_{[|X_n|^r \le \lambda + 1]}\right\} \\ &\le E|X|^r - \liminf E\{|X_n|^r 1_{[|X_n|^r \le \lambda+1]}\} \\ &\le E|X|^r - \liminf E f_\lambda(X_n) \\ &= E|X|^r - E f_\lambda(X) \qquad \text{by the Helly-Bray theorem 2.3.5} \\ &\le E\{|X|^r 1_{[|X|^r \ge \lambda]}\} \\ &\to 0 \ \text{ as } \ \lambda \to \infty \ \text{ since } X \in L_r(P), \end{aligned}$$

thus A holds. $\square$

Some Metrics on Probability Distributions

Suppose that P and Q are two probability measures on some measurable space (or sample space) $(\mathbb{X}, \mathcal{A})$. Let $\mathcal{P}$ denote the collection of all probability distributions on $(\mathbb{X}, \mathcal{A})$.

Definition 1.7 The *total variation* metric d_{TV} on $\mathcal{P}$ is defined by

$$d_{TV}(P, Q) = \sup_{A \in \mathcal{A}} |P(A) - Q(A)|\,. \tag{25}$$

Definition 1.8 The *Hellinger* metric H on $\mathcal{P}$ is defined by

$$H^2(P, Q) = \frac{1}{2}\int |\sqrt{p} - \sqrt{q}|^2 d\mu \tag{26}$$

where p, q are densities with respect to any common dominating measure μ of P and Q (the choice $\mu = P + Q$ always works).

Proposition 1.13 For $P, Q \in \mathcal{P}$, let p and q denote densities with respect to any common dominating measure μ ($\mu = P + Q$ always works). Then

$$\sup_{A \in \mathcal{A}} |P(A) - Q(A)| = \frac{1}{2}\int |p - q| d\mu\,. \tag{27}$$

In other words,

$$d_{TV}(P, Q) = \frac{1}{2}\int |p - q| d\mu\,. \tag{28}$$

Proof. Let $r \equiv p - q$. Note that $0 = \int r d\mu = \int r^+ d\mu - \int r^- d\mu$, so that $\int r^+ d\mu = \int r^- d\mu$, and

$$\int |p-q| d\mu = \int r^+ d\mu + \int r^- d\mu = 2 \int r^+ d\mu\,.$$

Let $B \equiv [p - q \geq 0] = [r \geq 0]$. Then for any measurable set A,

$$\begin{aligned} P(A) - Q(A) &= \int_A p d\mu - \int_A q d\mu = \int_A (p-q) d\mu \\ &= \int_{A\cap B} (p-q) d\mu + \int_{A \cap B^c} (p-q) d\mu \\ \text{(a)} \qquad &\leq \int_A r^+ d\mu \leq \int r^+ d\mu = \frac{1}{2} \int |p-q| d\mu\,. \end{aligned}$$

By a similar argument,

$$\text{(b)} \qquad Q(A) - P(A) \leq \int_A r^- d\mu \leq \int r^- d\mu = \frac{1}{2} \int |p-q| d\mu\,.$$

Combining (a) and (b) yields

$$\text{(c)} \qquad |P(A) - Q(A)| \leq \frac{1}{2} \int |p-q| d\mu\,.$$

On the other hand

$$\text{(d)} \qquad |P(B) - Q(B)| = |\int_B (p-q) d\mu| = \int r^+ d\mu = \frac{1}{2} \int |p-q| d\mu\,.$$

The claimed equality follows immediately from (c) and (d). □

Proposition 1.14 (Scheffé's theorem). Suppose that $\{P_n\}_{n\geq 1}$, and P are probability distributions on a measurable space $(\mathbb{X}, \mathcal{A})$ with corresponding densities $\{p_n\}_{n\geq 1}$, and p with respect to a dominating measure μ, and suppose that $p_n \to p$ almost everywhere with respect to μ. Then

$$\text{(29)} \qquad d_{TV}(P_n, P) \to 0\,.$$

Proof. From the proof of proposition 1.13 it follows that

$$\text{(a)} \qquad d_{TV}(P_n, P) = d_{TV}(P, P_n) = \int r_n^+ d\mu$$

where $r_n^+ = (p - p_n)^+$ satisfies $r_n^+ \to_{a.e.} 0$ and $r_n^+ \leq p$ for all n with $\int p d\mu = 1 < \infty$. The conclusion follows from (a) and the dominated convergence theorem. □

Example 1.1 Suppose that $X_{n,1}, \ldots, X_{n,n}$ are i.i.d. Bernoulli(p_n) random variables with $p_n = \lambda/n$ for some $\lambda > 0$. Then $S_n \equiv \sum_{i=1}^n X_{n,j} \sim$ Binomial(n, p_n) and it is well-known that $S_n \to_d Y \sim$

Poisson(λ). But more is true: for fixed $0 \le k < n$ we have

$$\begin{aligned} p_n(k) &\equiv P(S_n = k) = \binom{n}{k} p_n^k (1-p_n)^{n-k} \\ &= \frac{n!}{(n-k)!}\frac{1}{k!}\left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{n(n-1)\cdots(n-k+1)}{n \cdot n \cdots n} \cdot \frac{\lambda^k}{k!}\left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &\to 1 \cdot \frac{\lambda^k}{k!} e^{-\lambda} = P_\lambda(Y = k). \end{aligned}$$

Thus by Scheffé's theorem, with $\mu =$ counting measure on $\{0, 1, 2, \ldots\}$, $S_n \sim P_n$, $Y \sim P_\lambda$,

$$d_{TV}(P_n, P_\lambda) = \frac{1}{2}\sum_{k=0}^\infty |P(S_n = k) - P_\lambda(Y = k)| \to 0.$$

How fast does this convergence happen? An inequality of Le Cam and Hodges (1960) handles an even more general situation: if $X_{n,1}, \ldots, X_{n,n}$ are independent with $X_{n,j} \sim$ Bernoulli($p_{n,j}$), $j = 1, \ldots, n$. then with $\lambda_n = \sum_{j=1}^n p_{n,j}$,

$$d_{TV}(P_n, P_{\lambda_n}) \le \sum_{j=1}^n p_{n,j}^2.$$

When $p_{n,j} = \lambda/n \equiv p_n$ for all $1 \le j \le n$, the right side in the last display becomes λ^2/n. For further inequalities for Poisson approximation via the methods of Stein and Chen, see Barbour, Holst, and Janson (1992).

Example 1.2 Suppose that $X_r \sim t_r$; thus $X_r = Z/\sqrt{V_r/r}$ where $Z \sim N(0,1)$ and $V_r \sim \chi_r^2$ are independent. Since $V_r/r \stackrel{d}{=} r^{-1}\sum_{j=1}^r Z_j^2$ where $Z_1, \ldots, Z_r$ are i.i.d. $N(0,1)$, we have $V_r/r \to_p 1$, and hence $X_r \to_p Z \sim N(0,1)$. Again, more is true. The density of X_r is given by

$$f_r(x) = \frac{\Gamma((r+1)/2)}{\sqrt{\pi r}\,\Gamma(r/2)} \cdot \frac{1}{\left(1 + \frac{x^2}{r}\right)^{(r+1)/2}} \to \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

by using Stirling's formula and $(1 + b/r)^r \to e^b$ as $r \to \infty$ for fixed $b \in \mathbb{R}$. By Scheffé's theorem with μ being Lebesgue measure on $\mathbb{R}$ it follows that with $X_r \sim P_r$

$$d_{TV}(P_r, N(0,1)) \to 0.$$

How fast does this happen? I claim that $d_{TV}(P_r, N(0,1)) \le .7/r$ for all $r \ge 1$. Reference?

Example 1.3 (Poincaré's lemma). Let $Z_1, Z_2, \ldots$ be independent $N(0,1)$ random variables. Then with $\underline{Z}_n \equiv (Z_1, \ldots, Z_n)^T$, the vector $\underline{Z}_n/\|\underline{Z}_n\|$ has a Uniform distribution on the (surface of) the unit sphere S^{n-1} in R^n, and $\sqrt{n}\underline{Z}_n/\|\underline{Z}_n\|$ has a uniform distribution on $\sqrt{n}S^{n-1}$, the (surface of) the unit sphere in $\mathbb{R}^n$ with radius $\sqrt{n}$. But then $\sqrt{n}\underline{Z}_k/\|\underline{Z}_n\|$ has the same distribution as $\sqrt{n}(Y_1, \ldots, Y_k)$ where $(Y_1, \ldots, Y_n)$ has a uniform distribution on S^{n-1}. Since

$$\frac{\sqrt{n}\underline{Z}_k}{\|\underline{Z}_n\|} = \frac{\underline{Z}_k}{\left(n^{-1}\sum_{i=1}^n Z_i^2\right)^{1/2}} \to_p \underline{Z}_k,$$

by the weak law of large numbers, it follows that if $(Y_1, \ldots, Y_n) \sim$ Uniform(S^{n-1}), then

$$\sqrt{n}\underline{Y}_{n,k} = \sqrt{n}(Y_1, \ldots, Y_k) \rightarrow_d \underline{Z}_k \sim N_k(0, I).$$

Furthermore, by computing the density of $\sqrt{n}\underline{Y}_{n,k}$ it is easily shown that $f_{\sqrt{n}\underline{Y}_{n,k}}(\underline{y}) \rightarrow \phi_k(\underline{y})$ for each fixed $y \in R^k$ where ϕ_k is the density of $\underline{Z}_k$. Thus by Scheffé's theorem, with $\sqrt{n}\underline{Y}_{n,k} \sim Q_{n,k}$ and $\underline{Z}_k \sim P_k$,

$$d_{TV}(Q_{n,k}, P_k) \rightarrow 0 \quad \text{as} \quad n \rightarrow \infty$$

for fixed k (and even for $k = o(n)$). Diaconis and Freedman (1987) show that

$$d_{TV}(Q_{n,k}, P_k) \leq \frac{3(k+3)}{n-k-3} \quad \text{for} \quad 1 \leq k \leq n-4.$$

Exercise 1.4 Show that the Hellinger distance $H(P, Q)$ does not depend on the choice of a dominating measure μ.

Exercise 1.5 Show that

$$H^2(P, Q) = 1 - \int \sqrt{pq}\, d\mu \equiv 1 - \rho(P, Q) \tag{30}$$

where the *Hellinger affinity* $\rho(P, Q)$ satisfies $\rho(P, Q) \leq 1$ with equality if and only if $P = Q$.

Exercise 1.6 Show that

$$d_{TV}(P, Q) = 1 - \int p \wedge q d\mu \equiv 1 - \eta(P, Q) \tag{31}$$

where the *total variation affinity* $\eta(P, Q)$ satisfies $\eta(P, Q) \leq 1$ with equality if and only if $P = Q$.

The Hellinger and total variation metrics are different, but they metrize the same topoplgy on $\mathcal{P}$, as follows from the inequalities in the following proposition.

Proposition 1.15 (Inequalities relating Hellinger and total variation metrics).

$$H^2(P, Q) \leq d_{TV}(P, Q) \leq H(P, Q)\{1 + \rho(P, Q)\}^{1/2} \leq \sqrt{2}H(P, Q)\,. \tag{32}$$

Exercise 1.7 Show that (32) holds.

2 Classical Limit Theorems

We now state some of the classical limit theorems of probability theory which are of frequent use in statistics.

Proposition 2.1 (WLLN). If $X, X_1, \ldots, X_n, \ldots$ are i.i.d. with mean μ (so $E|X| < \infty$ and $\mu = E(X)$, then $\overline{X}_n \rightarrow_p \mu$.

Proposition 2.2 (SLLN). If $X, X_1, \ldots, X_n, \ldots$ are i.i.d with mean μ (so $E|X| < \infty$ and $\mu = E(X)$), then $\overline{X}_n \rightarrow_{a.s.} \mu$.

Proposition 2.3 (CLT). If $X, X_1, \ldots, X_n$ are i.i.d. with mean μ and variance σ^2 (so $E|X|^2 < \infty$), then $\sqrt{n}(\overline{X}_n - \mu) \rightarrow_d N(0, \sigma^2)$.

Proposition 2.4 (Multivariate CLT). If $X, X_1, \ldots, X_n$ are i.i.d. random vectors in R^d with mean $\mu = E(X)$ and covariance matrix $\Sigma = E(X-\mu)(X-\mu)'$ (so $E(X'X) = E\|X\|^2 < \infty$), then $\sqrt{n}(\overline{X} - \mu) \rightarrow_d N_d(0, \Sigma)$.

Proposition 2.5 (Liapunov CLT). Let $X_{n1}, \ldots, X_{nn}$ be row independent random variables with $\mu_{ni} = E(X_{ni})$, $\sigma_{ni}^2 \equiv Var(X_{ni})$, and $\gamma_{ni} \equiv E|X_{ni} - \mu_{ni}|^3 < \infty$. Let $\mu_n \equiv \sum_1^n \mu_{ni}$, $\sigma_n^2 = \sum_{i=1}^n \sigma_{ni}^2$, $\gamma_n \equiv \sum_1^n \gamma_{ni}$. If $\gamma_n / \sigma_n^3 \rightarrow 0$, then $\sum_{i=1}^n (X_{ni} - \mu_{ni})/\sigma_n \rightarrow_d N(0,1)$.

Proposition 2.6 (Lindeberg-Feller CLT). Let X_{ni} be row independent with 0 means and finite variances $\sigma_{ni}^2 \equiv Var(X_{ni})$. Let $S_n \equiv \sum_{i=1}^n X_{ni}$ and $\sigma_n^2 = \sum_{i=1}^n \sigma_{ni}^2$. Then both $S_n/\sigma_n \rightarrow_d N(0,1)$ and $\max\{\sigma_{ni}^2/\sigma_n^2 : 1 \le i \le n\} \rightarrow 0$ if and only if the Lindeberg condition

$$(1) \qquad \frac{1}{\sigma_n^2} \sum_{i=1}^n E\{|X_{ni}|^2 1_{[|X_{ni}| \ge \epsilon \sigma_n]}\} \rightarrow 0 \qquad \text{for all} \quad \epsilon > 0$$

holds.

Proposition 2.7 (The Cramér-Wold device). Random vectors X_n in R^d satisfy $X_n \rightarrow_d X$ if and only if $a'X_n \rightarrow_d a'X$ in R for all $a \in R^d$.

Proposition 2.8 (Continuous mapping or Mann - Wald theorem). Suppose that $g : R^d \rightarrow R$ is continuous a.s. P_X. Then:
A. If $X_n \rightarrow_{a.s.} X$ then $g(X_n) \rightarrow_{a.s.} g(X)$.
B. If $X_n \rightarrow_p X$ then $g(X_n) \rightarrow_p g(X)$.
C. If $X_n \rightarrow_d X$ then $g(X_n) \rightarrow_d g(X)$.

Proposition 2.9 (Slutsky's theorem). Suppose that $A_n \rightarrow_p a$, $B_n \rightarrow_p b$, where a, b are constants, and $X_n \rightarrow_d X$. Then $A_n X_n + B_n \rightarrow_d aX + b$.

Proposition 2.10 (g'-theorem or the delta-method). Suppose that $Z_n \equiv a_n(X_n - b) \rightarrow_d Z$ in R^m where $a_n \rightarrow \infty$, and suppose that $g : R^m \rightarrow R^k$ has a derivative g' at b; here g' is a $k \times m$ matrix. Then

$$(2) \qquad a_n(g(X_n) - g(b)) \rightarrow_d g'(b) Z\,.$$

Definition 2.1 A sequence of random variables $\{X_n\}$ is said to be *bounded in probability*, and we write $X_n = O_p(1)$, if

(3) $$\lim_{\lambda\to\infty} \limsup_{n\to\infty} P(|X_n| \ge \lambda) = 0\,.$$

If $Y_n \to_p 0$, then we write $Y_n = o_p(1)$. For any sequence of non-negative real numbers a_n we write $X_n = O_p(a_n)$ if $X_n/a_n = O_p(1)$, and we write $Y_n = o_p(a_n)$ if $Y_n/a_n = o_p(1)$.

Proposition 2.11 If $X_n \to_d X$, then $X_n = O_p(1)$.

Exercise 2.1 Prove Proposition 2.11.

Exercise 2.2 (a) Show that if $X_n = O_p(1)$ and $Y_n = o_p(1)$, then $X_n Y_n = o_p(1)$.
(b) Show that if $X_n = O_p(a_n)$ and $Y_n = O_p(b_n)$ then $X_n + Y_n = O_p(c_n)$ where $c_n = \max\{a_n, b_n\}$.
(c) Show that if $X_n = O_p(a_n)$ and $Y_n = O_p(b_n)$, then $X_n Y_n = O_p(a_n b_n)$.

Proposition 2.12 (Polya - Cantelli lemma). If $F_n \to_d F$ and F is continuous, then $\|F_n - F\|_\infty \equiv \sup_{-\infty<x<\infty} |F_n(x) - F(x)| \to 0$.

Exercise 2.3 Suppose that $\xi_1, \ldots, \xi_n$ are i.i.d. Uniform$(0,1)$.
(a) Show that $n\xi_{n:1} = n\xi_{(1)} \to_d$ Exponential(1).
(b) What is the joint limiting distribution of $(n\xi_{n:1}, n\xi_{n:2})$?
(c) Can you extend the result of (b) to $(\xi_{n:1}, \ldots, \xi_{n:k})$ for a fixed $k \ge 1$?
(d) How would you extend (c) to the situation with $k_n \to \infty$ as $n \to \infty$?

Exercise 2.4 Suppose that $X_1, \ldots, X_n$ are independent Exponential(λ) random variables with distribution function $F_\lambda(x) = 1 - \exp(-\lambda x)$ for $x \ge 0$.
(a) We expect $X_{n:n}$ to be on the order of $b_n \equiv F_\lambda^{-1}(1 - 1/n)$. Compute this explicitly.
(b) Find a sequence of constants a_n so that $a_n(X_{n:n} - b_n) \to_d$ "something" and find "something".

3 Examples.

Example 3.1 (One-sample $t-$test) Suppose that $X_1, \ldots, X_n$ are i.i.d. with $E(X_1) = \mu$ and $Var(X_1) = \sigma^2$. Consider testing $H : \mu \le \mu_0$ versus $K : \mu > \mu_0$. The normal theory test is "reject H if $T_n \ge t_{n-1,\alpha}$" where

$$T_n \equiv \frac{\sqrt{n}(\overline{X}_n - \mu_0)}{S_n}$$

with $S_n^2 \equiv (n-1)^{-1}\sum_1^n (X_i - \overline{X}_n)^2$ and where $P(t_{n-1} \ge t_{n-1,\alpha}) = \alpha$. We are interested in the behavior of this test when the X_i's are *not normally distributed.*
(a) What if $\mu = \mu_0$ is true? Note that by the Lindeberg central limit theorem $\sqrt{n}(\overline{X}_n - \mu_0) \to_d N(0, \sigma^2)$ when $\mu = \mu_0$ is true, and by the WLLN and Slutsky's theorem

$$S_n^2 = \frac{n}{n-1}\left\{\frac{1}{n}\sum_{i=1}^n (X_i - \mu_0)^2 - (\overline{X} - \mu_0)^2\right\} \to_p 1\left\{\sigma^2 - 0\right\} = \sigma^2.$$

Thus by Slutsky's theorem again, $T_n \to_d Z \sim N(0,1)$ when $\mu = \mu_0$ is true, and we have $P_{\mu_0}(T_n \ge t_{n-1,\alpha}) \to P(Z \ge z_\alpha) = \alpha$.
(b) What if $\mu > \mu_0$ is true, with μ fixed? In this case

$$T_n = \frac{\sqrt{n}(\overline{X}_n - \mu)}{S_n} + \frac{\sqrt{n}(\mu - \mu_0)}{S_n} \to_p Z + \infty/\sigma = \infty,$$

so $P_\mu(T_n > t_{n-1,\alpha}) \to P(Z + \infty > z_\alpha) = 1$.
(c) What if $\mu = \mu_n > \mu_0$ with $\sqrt{n}(\mu_n - \mu_0) \to c > 0$? Then it will usually hold that

$$T_n = \frac{\sqrt{n}(\overline{X}_n - \mu_n)}{S_n} + \frac{\sqrt{n}(\mu_n - \mu_0)}{S_n} \to_d Z + c/\sigma$$

where we may need to apply a Lindeberg-Feller or Liapunov CLT to justify the convergence to normality in the first term. If this holds, then

$$P_{\mu_n}(T_n > t_{n-1,\alpha}) \to P(Z + c/\sigma > z_\alpha) = P(Z > z_\alpha - c/\sigma) = 1 - \Phi(z_\alpha - c/\sigma)$$

gives the limiting power of the test under the local alternatives μ_n. Note that $1 - \Phi(z_\alpha - c/\sigma) > \alpha$ for $c > 0$.

Example 3.2 (One sample normal - theory test of variance) Now suppose that $X_1, \ldots, X_n$ are i.i.d. with $E(X_1) = \mu$, $Var(X_1) = \sigma^2$, and $\mu_4 \equiv E(X_1 - \mu)^4 < \infty$.
(a) Now with $Y_i \equiv (X_i - \mu)^2 \sim (\sigma^2, \mu_4 - \sigma^4)$,

$$\begin{aligned}
\frac{\sqrt{n}(n^{-1}\sum_1^n (X_i - \mu)^2 - \sigma^2)}{\sqrt{2}\sigma^2} &= \frac{\sqrt{n}(\overline{Y}_n - \sigma^2)}{\sqrt{2}\sigma^2} \\
&\to_d \frac{N(0, \mu_4 - \sigma^4)}{\sqrt{2}\sigma^2} \\
&= N\left(0, \frac{\mu_4 - \sigma^4}{2\sigma^4}\right) \\
&= N\left(0, \frac{2\sigma^4 + \mu_4 - 3\sigma^4}{2\sigma^4}\right) \\
&= N(0, 1 + 2^{-1}\gamma_2) \quad \text{with} \quad \gamma_2 \equiv \frac{\mu_4}{\sigma^4} - 3.
\end{aligned}$$

(b) Since

$$n^{-1}\sum_1^n (X_i - \overline{X})^2 = n^{-1}\sum_1^n (X_i - \mu)^2 - (\overline{X} - \mu)^2$$

we have

$$\begin{aligned} \frac{\sqrt{n}(n^{-1}\sum_1^n (X_i - \overline{X}_n)^2 - \sigma^2)}{\sqrt{2}\sigma^2} &= \frac{\sqrt{n}(\overline{Y}_n - \sigma^2)}{\sqrt{2}\sigma^2} - \frac{\sqrt{n}(\overline{X}_n - \mu)(\overline{X}_n - \mu)}{\sqrt{2}\sigma^2} \\ &\to_d N(0, 1 + 2^{-1}\gamma_2) - N(0,1)\cdot 0 = N(0, 1+2^{-1}\gamma_2) \end{aligned}$$

by Slutsky's theorem. Thus with $S_n^2 \equiv (n-1)^{-1}\sum_1^n (X_i - \overline{X}_n)^2$,

$$\frac{\sqrt{n}(S_n^2 - \sigma^2)}{\sqrt{2}\sigma^2} \to_d N(0, 1 + 2^{-1}\gamma_2).$$

Now consider testing $H : \sigma = \sigma_0$ versus $K : \sigma > \sigma_0$. If $X_i \sim N(\mu, \sigma_0^2)$, then $(n-1)S_n^2/\sigma_0^2 \sim \chi^2_{n-1}$ under H, so the usual normal theory test is "reject H if $(n-1)S_n^2/\sigma_0^2 > \chi^2_{n-1,\alpha}$". Then since $\gamma_2(N(\mu, \sigma^2)) = 0$ we have

$$\begin{aligned} \alpha &= P_{\sigma_0, Norm}\left(\frac{(n-1)S_n^2}{\sigma_0^2} > \chi^2_{n-1,\alpha}\right) \\ &= P_{\sigma_0, Norm}\left(\sqrt{\frac{n}{2}}\left(\frac{S_n^2}{\sigma_0^2} - 1\right) > \sqrt{\frac{n}{2}}\left(\frac{\chi^2_{n-1,\alpha}}{n-1} - 1\right)\right) \\ &\to P(Z \ge z_\alpha) = \alpha, \end{aligned}$$

which forces

$$\sqrt{\frac{n}{2}}\left(\frac{\chi^2_{n-1,\alpha}}{n-1} - 1\right) \to z_\alpha.$$

Now suppose we carried out the normal theory test, but the X_i's are *not* normal. Then, under H,

$$\begin{aligned} P_{\sigma_0}\left(\frac{(n-1)S_n^2}{\sigma_0^2} > \chi^2_{n-1,\alpha}\right) &= P_{\sigma_0}\left(\sqrt{\frac{n}{2}}\left(\frac{S_n^2}{\sigma_0^2} - 1\right) > \sqrt{\frac{n}{2}}\left(\frac{\chi^2_{n-1,\alpha}}{n-1} - 1\right)\right) \\ &\to P(N(0, 1 + 2^{-1}\gamma_2) \ge z_\alpha) \ne \alpha \end{aligned}$$

when $\gamma_2 \ne 0$. In general the asymptotic size is smaller than α if $\gamma_2 < 0$, but the asymptotic size is greater than α if $\gamma_2 > 0$.

Example 3.3 (Two-sample tests for means) Suppose that $X_1, \ldots, X_m$ are i.i.d. with mean μ and variance σ^2, and that $Y_1, \ldots, Y_n$ are i.i.d. with mean ν and variance τ^2, independent of the X_j's. If we suppose that $\lambda_N \equiv m/N \equiv m/(m+n) \to \lambda \in [0,1]$, then

$$\begin{aligned} \sqrt{\frac{mn}{N}}\left(\overline{X}_m - \overline{Y}_n - (\mu - \nu)\right) &= \sqrt{\frac{n}{N}}\sqrt{m}(\overline{X}_m - \mu) - \sqrt{\frac{m}{N}}\sqrt{n}(\overline{Y}_n - \nu) \\ &\to_d \sqrt{1-\lambda}Z_1 - \sqrt{\lambda}Z_2 \quad (Z_1, Z_2) \sim N_2(0, \mathrm{diag}(\sigma^2, \tau^2)) \\ &\sim N(0, (1-\lambda)\sigma^2 + \lambda\tau^2). \end{aligned}$$

Thus we see that

$$\begin{aligned} \frac{\overline{X}_m - \overline{Y}_n - (\mu - \nu)}{\sqrt{\frac{S_X^2}{m} + \frac{S_Y^2}{n}}} &= \frac{\sqrt{\frac{mn}{N}}\left(\overline{X}_m - \overline{Y}_n - (\mu - \nu)\right)}{\sqrt{\frac{n}{N}S_X^2 + \frac{m}{N}S_Y^2}} \\ &\to_d \; N(0,1) \end{aligned}$$

by Slutsky's theorem. On the other hand, again by Slutsky's theorem,

$$\begin{aligned} T_{m,n}(\mu,\nu) &\equiv \frac{\sqrt{\frac{mn}{N}}\left(\overline{X}_m - \overline{Y}_n - (\mu - \nu)\right)}{\sqrt{\frac{(m-1)S_X^2 + (n-1)S_Y^2}{N-2}}} \\ &\to_d \frac{N(0,(1-\lambda)\sigma^2 + \lambda\tau^2)}{\sqrt{\lambda\sigma^2 + (1-\lambda)\tau^2}} \\ &= N\left(0, \frac{(1-\lambda)\sigma^2 + \lambda\tau^2}{\lambda\sigma^2 + (1-\lambda)\tau^2}\right) \\ &\neq N(0,1) \end{aligned}$$

unless $\lambda = 1/2$ or $\sigma^2 = \tau^2$.

Since the two-sample $t-$test of $H : \mu \leq \nu$ versus $K : \mu > \nu$ rejects H if $T_{m,n}(0,0) > t_{N-2,\alpha}$, it follows that the test is *not size (or level) robust* against violations of the assumption $\sigma^2 = \tau^2$ when $\lambda \neq 1/2$.

Example 3.4 (Simple linear regression with non-normal errors.) Suppose that $Y_i = \alpha + \beta(x_i - \overline{x}) + \epsilon_i$ for $i = 1, \ldots, n$ where $\overline{x} \equiv n^{-1}\sum_1^n x_i$ and $\epsilon_1, \ldots, \epsilon_n$ are i.i.d. with mean zero and finite variance, $Var(\epsilon_1) = \sigma^2$; the ϵ_i's are *not* assumed to be normally distributed. In matrix form

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \equiv \begin{pmatrix} 1 & x_1 - \overline{x} \\ \vdots & \vdots \\ 1 & x_n - \overline{x} \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \boldsymbol{\epsilon}.$$

The least squares estimators $\hat{\alpha}$ and $\hat{\beta}$ of α and β are given by

$$\hat{\alpha} = \overline{Y}, \qquad \hat{\beta} = \frac{\sum_1^n (x_i - \overline{x})Y_i}{\sum_1^n (x_i - \overline{x})^2}.$$

Claim: if $\max_{1 \leq i \leq n}(x_i - \overline{x})^2 / \sum_1^n (x_i - \overline{x})^2 \to 0$, then

$$(\mathbf{X}^T\mathbf{X})^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \begin{pmatrix} \sqrt{n}(\hat{\alpha} - \alpha) \\ \sqrt{\sum_1^n (x_i - \overline{x})^2}(\hat{\beta} - \beta) \end{pmatrix} \to_d N_2(0, \sigma^2 I_2). \tag{1}$$

Here is a partial proof. Now

$$\sqrt{n}(\hat{\alpha} - \alpha) = \sqrt{n}(\overline{Y} - \alpha) = \sqrt{n}\overline{\epsilon} \to_d N(0, \sigma^2)$$

by the Lindeberg CLT. Thus the first coordinate in (1) converges to the claimed limit marginally. We now use the Lindeberg-Feller CLT to show that the same is true for the second coordinate, and hence the two claimed marginal convergences hold. Note that

$$\sqrt{\sum_1^n (x_i - \overline{x})^2}(\hat{\beta} - \beta) \;=\; \sigma \frac{\sum_1^n (x_i - \overline{x})\epsilon_i}{\sigma\sqrt{\sum_1^n (x_i - \overline{x})^2}} \equiv \sigma \frac{S_n}{\sigma_n}$$

in the context of the Lindeberg-Feller CLT where $X_{n,i} \equiv (x_i - \overline{x})\epsilon_i$ for $i = 1, \ldots, n$, and hence $\mu_{n,i} = EX_{n,i} = 0$, $\sigma_{n,i}^2 = Var(X_{n,i}) = (x_i - \overline{x})^2\sigma^2$, and $\sigma_n^2 = \sigma^2 \sum_1^n (x_i - \overline{x})^2$. Thus we need to verify the condition $LF_n(\delta) \to 0$ for every $\delta > 0$ where

$$LF_n(\delta) \equiv \frac{1}{\sigma_n^2} \sum_1^n E\{|X_{n,i}|^2 1_{[|X_{n,i}| \geq \delta\sigma_n]}\}.$$

But in the present case,

$$\begin{aligned}
LF_n(\delta) &= \frac{1}{\sigma^2 \sum_1^n (x_i - \overline{x})^2} \sum_{i=1}^n (x_i - \overline{x})^2 E\left\{\epsilon_i^2 1\{|x_i - \overline{x}||\epsilon_i| \geq \delta\sigma \sqrt{\sum_1^n (x_i - \overline{x})^2\}}\right\} \\
&= \frac{1}{\sigma^2 \sum_1^n (x_i - \overline{x})^2} \sum_{i=1}^n (x_i - \overline{x})^2 E\left\{\epsilon_1^2 1\{|\epsilon_1| \geq \frac{\delta\sigma}{\sqrt{\frac{|x_i - \overline{x}|^2}{\sum_1^n (x_i - \overline{x})^2}}}\}\right\} \\
&\leq \frac{1}{\sigma^2 \sum_1^n (x_i - \overline{x})^2} \sum_{i=1}^n (x_i - \overline{x})^2 E\left\{\epsilon_1^2 1\left\{|\epsilon_1| \geq \frac{\delta\sigma}{\sqrt{\max_{1 \leq i \leq n} \frac{|x_i - \overline{x}|^2}{\sum_1^n (x_i - \overline{x})^2}}}\right\}\right\} \\
&= \frac{1}{\sigma^2} E\left\{\epsilon_1^2 1\{|\epsilon_1| \geq \delta\sigma / \sqrt{\max_{1 \leq i \leq n} |x_i - \overline{x}|^2 / \sum_1^n (x_i - \overline{x})^2\}}\right\} \\
&\to 0
\end{aligned}$$

for every $\delta > 0$ by the DCT since the integrand converges a.s. to zero by the hypothesis and since $E(\epsilon_1^2) < \infty$, so ϵ_1^2 gives an integrable dominating function. Thus the second coordinate satisfies the claimed marginal convergence. All that remains to be shown is the claimed joint convergence.

Conclusion: the normal theory tests and confidence intervals for α and β have the right asymptotic size and coverage probabilities as long as $\sigma^2 < \infty$ and $\max_{1 \leq i \leq n} |x_i - \overline{x}|^2 / \sum_1^n (x_i - \overline{x})^2 \to 0$.

Example 3.5 (Multiple linear regression with non-normal errors). Insert?

Example 3.6 (The correlation coefficient). Suppose that $(X_1, Y_1)^T, \ldots, (X_n, Y_n)^T$ are i.i.d. with means $E(X_1, Y_1) = (\mu_X, \mu_Y)$, covariance matrix

$$\begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix},$$

and $E|X_1|^4 < \infty$, $E|Y_1|^4 < \infty$. Let

$$S_{XY} \equiv n^{-1} \sum_1^n (X_i - \overline{X})(Y_i - \overline{Y}), \quad S_{XX} \equiv n^{-1} \sum_1^n (X_i - \overline{X})^2, \quad S_{YY} \equiv n^{-1} \sum_1^n (Y_i - \overline{Y})^2,$$

and consider the sample correlation $r \equiv r_n$ defined by

$$r_n \equiv \frac{S_{XY}}{\sqrt{S_{XX} S_{YY}}}.$$

since r_n is invariant with respect to linear transformations of each axis, we may assume without loss of generality that $\mu_X = \mu_Y = 0$ and $\sigma_X^2 = \sigma_Y^2 = 1$; if not replace (X_i, Y_i) by $((X_i - \mu_X)/\sigma_X, (Y_i - \mu_Y)/\sigma_Y)$, $i = 1, \ldots, n$. Note that

$$\begin{aligned} \sqrt{n}\begin{pmatrix} S_{XY} - \rho \\ S_{XX} - 1 \\ S_{YY} - 1 \end{pmatrix} &= \sqrt{n}\begin{pmatrix} \overline{XY} - \rho \\ \overline{X^2} - 1 \\ \overline{Y^2} - 1 \end{pmatrix} + \begin{pmatrix} o_p(1) \\ o_p(1) \\ o_p(1) \end{pmatrix} \\ &\to_d \ \underline{Z} \sim N_3(0, \Sigma) \end{aligned}$$

by the multivariate CLT where

$$\Sigma = \begin{pmatrix} E(X^2Y^2) - \rho^2 & E(X^3Y) - \rho & E(XY^3) - \rho \\ E(X^3Y) - \rho & E(X^4) - 1 & E(X^2Y^2) - 1 \\ E(XY^3) - \rho & E(X^2Y^2) - 1 & E(Y^4) - 1 \end{pmatrix}.$$

Since $g(u, v, w) \equiv u/\sqrt{vw}$ has $\nabla g(\rho, 1, 1) = (1, -\rho/2, -\rho/2)$, it follows by the g' theorem (or delta method) that

$$\sqrt{n}(r_n - \rho) \to_d \nabla g(\rho, 1, 1)\underline{Z} = Z_1 - \frac{\rho}{2}(Z_2 + Z_3).$$

Note that if X_1 and Y_1 are independent, then $\rho = 0$ and the covariance matrix Σ becomes

$$\Sigma = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \mu_{4,X} & 0 \\ 0 & 0 & \mu_{4,Y} \end{pmatrix},$$

and hence $\sqrt{n}(r_n - 0) = \sqrt{n} r_n \to_d N(0, 1)$. Thus under independence the test of $H : \rho = 0$ versus $K : \rho > 0$ that rejects if $\sqrt{n} r_n > z_\alpha$ has asymptotic level α even if the true distribution is not Gaussian (but we have $E|X_1|^4 + E|Y_1|^4 < \infty$). If the true distribution of the (X_i, Y_i) pairs is Normal, then Σ becomes

$$\Sigma = \begin{pmatrix} 1 + \rho^2 & 2\rho & 2\rho \\ 2\rho & 2 & 2\rho^2 \\ 2\rho & 2\rho^2 & 2 \end{pmatrix},$$

and hence

$$\sqrt{n}(r_n - \rho) \to_d N(0, (1, -\rho/2, -\rho/2)\Sigma(1, -\rho/2 - \rho/2)^T) = N(0, (1 - \rho^2)^2).$$

Finally, it is often useful to transform the distribution of r_n (which always has support in $[-1, 1]$) to the whole line $\mathbb{R}$: if $g(x) \equiv 2^{-1}\log((1 + x)/(1 - x))$, then $g'(x) = 1/(1 - x^2)$, and hence, under normality of the (X_i, Y_i)'s,

$$\sqrt{n}(g(r_n) - g(\rho)) \to_d N(0, 1).$$

If we let $Z_n \equiv g(r_n)$ and $\xi \equiv g(\rho)$ (this is sometimes known as *Fisher's Z-transform*), then

$$\sqrt{n-3}\left(Z_n - \xi - \frac{\rho}{2(n-1)}\right) \approx N(0, 1)$$

is an excellent approximation.

Example 3.7 (Chi-square test of a simple null hypothesis). Suppose that $\underline{\Delta}_1, \ldots, \underline{\Delta}_n, \ldots$ are i.i.d. $\text{Multinomial}_k(1, \underline{p})$ random vectors so that $\underline{\Delta}_i$ vector contains only zeros and 1's, but only one 1. Thus

$$\underline{N}_n \equiv \sum_{i=1}^n \underline{\Delta}_i \sim \text{Mult}_k(n, \underline{p}),$$

and $\underline{\hat{p}}_n \equiv n^{-1} N_n \rightarrow_{p,a.s.} \underline{p}$.

Consider testing $H : \underline{p} = \underline{p}_0$ versus $K : \underline{p} \neq \underline{p}_0$. One simple test statistic is the (Pearson) chi-square statistic Q_n defined by

$$Q_n \equiv \sum_{j=1}^k \frac{(N_j - np_{0,j})^2}{np_{0,j}}.$$

To carry out the test based on Q_n we need to know the distribution of Q_n under the null hypothesis either exactly (which is complicated: its depends on k, n, and $\underline{p}_0$), or at least asymptotically which is relative easy. We claim that: when $\underline{p} = \underline{p}_0$,

$$Q_n \rightarrow_d \chi^2_{k-1}.$$

Here is the proof. Set

$$\underline{Z}_n = \left(\frac{N_1 - np_{0,1}}{\sqrt{np_{0,1}}}, \ldots, \frac{N_k - np_{0,k}}{\sqrt{np_{0,k}}} \right)^T \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^n \underline{Y}_i$$

where The $\underline{Y}_i$'s are i.i.d. with $E(\underline{Y}_i) = \underline{0}$, and covariance matrix $\Sigma = I - \sqrt{\underline{p}_0}\sqrt{\underline{p}_0}^T$ where $\sqrt{\underline{p}_0} = (\sqrt{p_{0,1}}, \ldots, \sqrt{p_{0,k}})^T$. Thus

$$\underline{Z}_n \rightarrow_d \underline{Z} \sim N_k(0, \Sigma)$$

by the multivariate CLT. Now $Q_n = \underline{Z}_n^T \underline{Z}_n$ is a continuous function of $\underline{Z}_n$, so $Q_n = \underline{Z}_n^T \underline{Z}_n \rightarrow_d \underline{Z}^T \underline{Z} \equiv Q$ by the Mann-Wald theorem. It remains to show that the distribution of Q is χ^2_{k-1}.

To see this, note that for any orthogonal matrix Γ we have

$$Q = \underline{Z}^T \underline{Z} = (\Gamma \underline{Z})^T (\Gamma \underline{Z}) \equiv \underline{V}^T \underline{V}$$

where $\underline{V} \sim N(0, \Gamma \Sigma \Gamma^T)$. Here is a convenient choice of Γ: choose Γ to be the orthogonal matrix with first row $\sqrt{\underline{p}_0}^T$, and filled out with orthogonal rows. Then

$$\begin{aligned} \Gamma \Sigma \Gamma^T &= \Gamma I \Gamma^T - \Gamma \sqrt{\underline{p}_0} \sqrt{\underline{p}_0}^T \Gamma^T = I - (1, 0, \ldots, 0)^T (1, 0, \ldots, 0) \\ &= \begin{pmatrix} 0 & \underline{0}^T \\ \underline{0} & I_{(k-1)\times(k-1)} \end{pmatrix}. \end{aligned}$$

Thus $V_1 = 0$ with probability 1 and $V_2, \ldots, V_k$ are i.i.d. $N(0,1)$. It follows that $Q = \sum_{j=2}^k V_j^2 \sim \chi^2_{k-1}$. Thus we have

$$P_{\underline{p}_0}(Q_n \geq \chi^2_{k-1,\alpha}) \rightarrow P(\chi^2_{k-1} \geq \chi^2_{k-1,\alpha}) = \alpha.$$

What if $\underline{p} \neq \underline{p}_0$? In this case, since $\underline{\hat{p}}_n \rightarrow_p \underline{p}$, we can write

$$n^{-1} Q_n = \sum_{j=1}^k \frac{(\hat{p}_j - p_{0,j})^2}{p_{0,j}} \rightarrow_p \sum_{j=1}^k \frac{(p_j - p_{0,j})^2}{p_{0,j}} \equiv q > 0.$$

Thus $Q_n \to_p \infty$, and it follows that

$$P_{\underline{p}}(Q_n \geq \chi^2_{k-1,\alpha}) \to 1$$

as $n \to \infty$; i.e. the test is (power) consistent.

What if $\underline{p} = \underline{p}_n$ satisfies $\underline{p}_n = \underline{p}_0 + \underline{c}n^{-1/2}$ where $\underline{1}^T\underline{c} = 0$ and hence $\underline{1}^T\underline{p}_n = 1$ for all n? In this case, by using the Cramér-Wold device together with either the Liapunov CLT or the Lindeberg-Feller CLT,

$$\begin{aligned}
\underline{Z}_n &= \left(\frac{N_1 - np_{0,1}}{\sqrt{np_{0,1}}}, \ldots, \frac{N_k - np_{0,k}}{\sqrt{np_{0,k}}}\right)^T \\
&= \left(\frac{N_1 - np_{n,1}}{\sqrt{np_{0,1}}}, \ldots, \frac{N_k - np_{n,k}}{\sqrt{np_{0,k}}}\right)^T + \mathrm{diag}(1/\sqrt{\underline{p}_0})\sqrt{n}(\underline{p}_n - \underline{p}_0) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^n \underline{Y}_{n,i} + \mathrm{diag}(1/\sqrt{\underline{p}_0})\underline{c} \\
&\to_d \underline{Z} + \mathrm{diag}(1/\sqrt{\underline{p}_0})\underline{c}
\end{aligned}$$

where $\underline{Z} \sim N_k(0, \Sigma)$ with $\Sigma = I - \sqrt{\underline{p}_0}\sqrt{\underline{p}_0}^T$. Thus it follows by the Mann-Wald theorem that under the local alternatives $\underline{p}_n$ we have

$$\begin{aligned}
Q_n &= \underline{Z}_n^T\underline{Z}_n \to_d (\underline{Z} + \mathrm{diag}(1/\sqrt{\underline{p}_0})\underline{c})^T(\underline{Z} + \mathrm{diag}(1/\sqrt{\underline{p}_0})\underline{c}) \\
&= \left(\Gamma(\underline{Z} + \mathrm{diag}(1/\sqrt{\underline{p}_0})\underline{c})\right)^T \left(\Gamma(\underline{Z} + \mathrm{diag}(1/\sqrt{\underline{p}_0})\underline{c})\right) \\
&\equiv \underline{V}^T\underline{V}
\end{aligned}$$

where

$$\underline{V} \sim N_k\left(\Gamma\mathrm{diag}(1/\sqrt{\underline{p}_0})\underline{c}, \begin{pmatrix} 0 & \underline{0}^T \\ \underline{0} & I_{(k-1)\times(k-1)} \end{pmatrix}\right) \equiv N_k\left(\underline{\mu}, \begin{pmatrix} 0 & \underline{0}^T \\ \underline{0} & I_{(k-1)\times(k-1)} \end{pmatrix}\right)$$

Noting that $\underline{\mu}^T\underline{\mu} = \underline{c}^T\mathrm{diag}(1/\underline{p}_0)\underline{c} = \sum_{j=1}^k c_j^2/p_{0,j}$ and

$$\mu_1 = \sqrt{\underline{p}_0}^T\mathrm{diag}(1/\sqrt{\underline{p}_0})\underline{c} = \underline{1}^T\underline{c} = 0,$$

it follows that $\underline{V}^T\underline{V} \sim \chi^2_{k-1}(\delta)$ with $\delta = \sum_1^k c_j^2/p_{0,j}$. We conclude that

$$P_{\underline{p}_n}(Q_n \geq \chi^2_{k-1,\alpha}) \to P(\chi^2_{k-1}(\delta) > \chi^2_{k-1,\alpha}).$$

This leads to approximating the power of the chi-square test based on Q_n by

$$P_{\underline{p}}(Q_n \geq \chi^2_{k-1,\alpha}) \approx P(\chi^2_{k-1}(\delta_n) > \chi^2_{k-1,\alpha})$$

with $\delta_n \equiv \sum_{j=1}^k c_{n,j}^2/p_{0,j}$ where $\underline{c}_n \equiv \sqrt{n}(\underline{p} - \underline{p}_0)$.

Another approach to approximating the power of the chi-square test is based on fixed alternatives as follows. Suppose that $\underline{p} \neq \underline{p}_0$. Then with $q = \sum_{j=1}^k (p_j - p_{0,j})^2/p_{0,j}$,

$$\sqrt{n}\left(n^{-1}Q_n - q\right) = \sum_{j=1}^k \sqrt{n}\left\{\frac{(\hat{p}_j - p_{0,j})^2}{p_{0,j}} - \frac{(p_j - p_{0,j})^2}{p_{0,j}}\right\}$$

$$
\begin{aligned}
&= \sum_{j=1}^{k} \sqrt{n} \frac{\{(\hat{p}_j - p_{0,j}) - (p_j - p_{0,j})\}\{(\hat{p}_j - p_{0,j}) + (p_j - p_{0,j})\}}{p_{0,j}} \\
&= \sum_{j=1}^{k} \frac{\sqrt{n}(\hat{p}_j - p_j)}{\sqrt{p_j}} \left\{ \frac{\{2(p_j - p_{0,j}) + o_p(1)\}\sqrt{p_j}}{p_{0,j}} \right\} \\
&\equiv \sum_{j=1}^{k} B_{n,j} Z_{n,j} \\
&\to_d \underline{b}^T \underline{Z} \sim N(0, \underline{b}^T (I - \sqrt{\underline{p}}\sqrt{\underline{p}}^T)\underline{b}) = N(0, Var_{\underline{p}}(D))
\end{aligned}
$$

where, with $b_j \equiv 2(p_j/p_{0,j} - 1)\sqrt{p_j}$ for $j = 1, \ldots, k$,

$$P(D = b_j/\sqrt{p_j}) = p_j, \quad j = 1, \ldots, k.$$

Here the claimed convergence holds since

$$
\begin{aligned}
&\underline{Z}_n \to_d \underline{Z} \sim N_k(0, I - \sqrt{\underline{p}}\sqrt{\underline{p}}^T), \\
&B_{n,j} = \left\{ 2\left(\frac{p_j}{p_{0,j}} - 1 \right) + o_p(1) \right\} \sqrt{p_j} \to_p \left\{ 2\left(\frac{p_j}{p_{0,j}} - 1 \right) \right\} \sqrt{p_j} = b_j,
\end{aligned}
$$

for $j = 1, \ldots, k$. This suggests the following approximation to the power of the chi-square test:

$$
\begin{aligned}
P_{\underline{p}}(Q_n \geq \chi^2_{k-1,\alpha}) &= P_{\underline{p}}\left(\sqrt{n}\left(\frac{Q_n}{n} - q \right) \geq \sqrt{n}\left(\frac{\chi^2_{k-1,\alpha}}{n} - q \right) \right) \\
&\approx P\left(N(0, Var_{\underline{p}}(D)) \geq \sqrt{n}\left(\frac{\chi^2_{k-1,\alpha}}{n} - q \right) \right) \\
&= P\left(N(0,1) \geq \sqrt{n}\left(\frac{\chi^2_{k-1,\alpha}}{n} - q \right) / \sqrt{Var_{\underline{p}}(D)} \right) \\
&= 1 - \Phi\left(\sqrt{n}\left(\frac{\chi^2_{k-1,\alpha}}{n} - q \right) / \sqrt{Var_{\underline{p}}(D)} \right).
\end{aligned}
$$

4 Skorokhod's Theorem: Replacing $\rightarrow_d$ by $\rightarrow_{a.s.}$

Our goal in this section is to show how we can convert convergence in distribution into the stronger mode of almost sure convergence. This often simplifies proofs and makes them more intuitive.

Definition 4.1 For any distribution function F define F^{-1} by $F^{-1}(t) \equiv \inf\{x : \ F(x) \geq t\}$ for $0 < t < 1$.

Proposition 4.1 F^{-1} is left - continuous.

Proof. To show that F^{-1} is left - continuous, let $0 < t < 1$, and set $z \equiv F^{-1}(t)$. Then $F(z) \geq t$ by the right continuity of F. If F is discontinuous at z, $F^{-1}(t-\epsilon) = z$ for all small $\epsilon > 0$, and hence left continuity holds. If F is continuous at z, then assume F^{-1} is discontinuous from the left at t. Then for all $\epsilon > 0$, $F^{-1}(t-\epsilon) < z - \delta$ for some $\delta > 0$, and hence $F(z-\delta) \geq t - \epsilon$ for all $\epsilon > 0$. hence $F(z-\delta) \geq t$, which implies $F^{-1}(t) \leq z - \delta$, a contradiction. $\Box$

Proposition 4.2 If X has continuous distribution function F, then $F(X) \sim \text{Uniform}(0,1)$. For any distribution function F and any $t \in (0,1)$,

$$P(F(X) \leq t) \leq t$$

with equality if and only if t is in the range of F. Equivalently, $F(F^{-1}(t)) \equiv F \circ F^{-1}(t) \geq t$ for all $0 < t < 1$ with equality if and only if t is in the range of F. Also, $F^{-1} \circ F(x) \leq x$ for all $-\infty < x < \infty$ with strict inequality if and only if $F(x-\epsilon) = F(x)$ for some $\epsilon > 0$. Thus $P(F^{-1} \circ F(X) \neq X) = 0$ where $X \sim F$.

Exercise 4.1 Prove proposition 4.2.

Theorem 4.1 (The inverse transformation). Let $\xi \sim \text{Uniform}(0,1)$ and let $X = F^{-1}(\xi)$. Then for all real x,

(1) $\quad [X \leq x] = [\xi \leq F(x)]\,.$

Thus X has distribution function F.

Proof. Now $\xi \leq F(x)$ implies $X = F^{-1}(\xi) \leq x$ by the definition 4.1 of F^{-1}. If $X = F^{-1}(\xi) \leq x$, then $F(x+\epsilon) \geq \xi$ for all $\epsilon > 0$, so that right continuity of F implies $F(x) \geq \xi$. Thus the claimed event identity holds. $\Box$

Proposition 4.3 (Elementary Skorokhod theorem). Suppose that $X_n \rightarrow_d X_0$. Then there exist random variables X_n^*, $n \geq 0$, all defined on the common probability space $([0,1], \mathcal{B}[0,1], \lambda)$ for which $X_n^* \stackrel{d}{=} X_n$ for every $n \geq 0$ and $X_n^* \rightarrow_{a.s.} X_0^*$.

Proof. Let F_n denote the distribution function of X_n and let

(a) $\quad X_n^* \equiv F_n^{-1}(\xi) \qquad \text{for all} \;\; n \geq 0$

where $\xi \sim \text{Uniform}(0,1)$. Then $X_n^* \stackrel{d}{=} X_n$ for all $n \geq 0$ by theorem 4.1. It remains only to show that $X_n^* \rightarrow_{a.s.} X_0^*$.

Let $t \in (0,1)$ be such that there is at most one value z having $F(z) = t$. (Thus t corresponds to a continuity point of F^{-1}.) Let $z = F^{-1}(t)$. Then $F(x) < t$ for $x < z$. Thus $F_n(x) < t$ for $n \geq N_x$ provided $x < z$ is a continuity point of F. Thus $F_n^{-1}(t) \geq x$ provided $x < z$ is a continuity point of F. Thus $\liminf F_n^{-1}(t) \geq x$ provided $x < z$ is a continuity point of F. Thus $\liminf F_n^{-1}(t) \geq z$ since there are continuity points x that $\uparrow z$.

We also have $F(x) > t$ for $z < x$. Thus $F_n(x) > t$, and hence $F_n^{-1}(t) \leq x$ for $n \geq$ some N_x provided $x > z$ is a continuity point of F. Thus $\limsup_n F_n^{-1}(t) \leq x$ provided $x > z$ is a continuity point of F. Thus $\limsup_n F_n^{-1}(t) \leq z$ since there are continuity points x that $\downarrow z$.

Thus $F_n^{-1}(t) \rightarrow F^{-1}(t)$ for all but a countable number of t's. Since any such set has Lebesgue measure zero, it follows that $X_n^* = F_n^{-1}(\xi) \rightarrow_{a.s.} F^{-1}(\xi) = X_0^*$. □

Proposition 4.4 (Continuous mapping or Mann-Wald theorem). Suppose that $g : R \rightarrow R$ is continuous a.s. P_X. Then:
A. If $X_n \rightarrow_{a.s.} X_0$, then $g(X_n) \rightarrow_{a.s.} g(X_0)$.
B. If $X_n \rightarrow_p X_0$, then $g(X_n) \rightarrow_p g(X_0)$.
C. If $X_n \rightarrow_d X_0$, then $g(X_n) \rightarrow_d g(X_0)$.

Proof. A. Let N_1 be the null set such that $X_n(\omega) \rightarrow X_0(\omega)$ for all $\omega \in N_1^c$, and let N_2 be the null set such that g is continuous at $X_0(\omega)$ for all $\omega \in N_2^c$. Then for $\omega \in N_1^c \cap N_2^c$ we have $g(X_n(\omega)) \rightarrow g(X_0(\omega))$. But $P(N_1 \cup N_2) \leq P(N_1) + P(N_2) = 0 + 0 = 0$, and the convergence asserted in A holds.

B. By theorem 1.1 part G, $X_n \rightarrow_p X_0$ if and only if for every subsequence $\{X_{n'}\}$ there is a further subsequence $\{X_{n''}\} \subset \{X_{n'}\}$ such that $X_{n''} \rightarrow_{a.s.} X_0$. We will apply this to $Y_n = g(X_n)$. Let $Y_{n'} = g(X_{n'})$ be an arbitrary subsequence of $\{Y_n\}$. By part G of theorem 1.1 there exists a subsequence $\{X_{n''}\}$ of $\{X_{n'}\}$ such that $X_{n''} \rightarrow_{a.s.} X_0$. By A we conclude that $Y_{n''} = g(X_{n''}) \rightarrow_{a.s.} g(X_0) = Y_0$. But by part G of theorem 1.1 (in the converse direction) it follows that $Y_n = g(X_n) \rightarrow_p g(X_0) = Y_0$.

C. Replace X_n, X_0 by X_n^*, X_0^* of the Skorokhod theorem, proposition 4.3. Thus

$$\text{(a)} \qquad g(X_n) \stackrel{d}{=} g(X_n^*) \rightarrow_{a.s.} g(X_0^*) \stackrel{d}{=} g(X_0)\,.$$

Since $\rightarrow_{a.s.}$ implies $\rightarrow_p$ which in turn implies $\rightarrow_d$, (a) implies that $g(X_n) \rightarrow_d g(X_0)$. □

Remark 4.1 Proposition 4.4 remains true for random vectors in R^k and, still more generally, for convergence in law (weak convergence) of random elements in a separable metric space. See Billingsley (1986), *Probability and Measure*, page 399, for the first, and Billingsley (1971), *Weak Convergence of Measures: Applications in Probability*, theorem 3.3, page 7, for the second. The original paper is Skorokhod (1956) where the separable metric space case was treated immediately.

Proposition 4.5 (Helly Bray theorem). If $X_n \rightarrow_d X_0$ and g is bounded and continuous (a.s. P_X), then $Eg(X_n) \rightarrow Eg(X_0)$.

Proof. For the random variables X_n^* of proposition 4.3, it follows from A of proposition 4.4 that $g(X_n^*) \rightarrow_{a.s.} g(X_0^*)$. Thus by equality in distribution guaranteed by the construction of proposition 4.3 and the dominated convergence theorem,

$$Eg(X_n) = Eg(X_n^*) \rightarrow Eg(X_0^*) = Eg(X_0)\,.$$

□

Remark 4.2 If $Eg(X_n) \to Eg(X)$ for all bounded continuous functions g, then $X_n \to_d X_0$. (Proof: box in the indicator function $1_{(-\infty,x]}$ by the bounded continuous functions g_+, g_- defined by connecting $(x,1)$ to $(x+\epsilon,0)$ linearly and $(x-\epsilon,1)$ to $(x,0)$ linearly, respectively.) This gives a way of defining $\to_d$ more generally:

Definition 4.2 Suppose that X_n, $n \geq 0$ take values in the complete separable metric space (M,d). Then we say that X_n *converges in law or distribution to* X_0, and we write $X_n \to_d X_0$ or $X_n \Rightarrow X_0$, if

$$Eg(X_n) \to Eg(X_0) \qquad \text{for all} \;\; g \in C_b(M);$$

here $C_b(B)$ denotes the collection of all bounded continuous functions from M to R.

Proposition 4.6 If $X_n \to_d X_0$, then $E|X_0| \leq \liminf_{n\to\infty} E|X_n|$.

Proof. For the random variables X_n^* of proposition 4.3, $X_n^* \to_{a.s.} X_0^*$. It follows from the equality in distribution of proposition 4.3 and Fatou's lemma

$$E|X_0| = E|X_0^*| = E(\liminf_n |X_n^*|) \leq \liminf_n E|X_n^*| = \liminf_n E|X_n|\,.$$

□

Corollary 1 If $X_n \to_d X_0$, then $Var(X_0) \leq \liminf_n Var(X_n)$.

Exercise 4.2 Prove corollary 1. Hint: Note that with $X_n' \stackrel{d}{=} X_n$ for all $n \geq 0$ with X_n' independent of X_n, we have $Var(X_n) = (1/2)E(X_n - X_n')^2$.

Proposition 4.7 (Slutsky's theorem). If $A_n \to_p a$, $B_n \to_p b$, and $Z_n \to_d Z$, then $A_n Z_n + B_n \to_d aZ + b$.

Proof. Now $A_n \to_p a$, $B_n \to_p b$, and $Z_n \to_d Z$ where a, b are constants, implies that $(Z_n, A_n, B_n) \to_d (Z, a, b)$ in R^3. Hence by the R^3 version of Skorokhod's theorem, there exists a sequence $(Z_n^*, A_n^*, B_n^*) \stackrel{d}{=} (Z_n, A_n, B_n)$ such that $(Z_n^*, A_n^*, B_n^*) \to (Z^*, a, b) \stackrel{d}{=} (Z, a, b)$. Hence

(a) $$A_n Z_n + B_n \stackrel{d}{=} A_n^* Z_n^* + B_n^* \to_{a.s.} aZ^* + b \stackrel{d}{=} aZ + b\,.$$

Since $\to_{a.s.}$ implies $\to_p$ which in turn implies $\to_d$, (a) yields the desired conclusion. □

5 Empirical Measures and Empirical Processes

We first introduce the empirical distribution function $\mathbb{G}_n$ and empirical process $\mathbb{U}_n$ of i.i.d. Uniform$(0,1)$ random variables. Suppose that $\xi_1,\ldots,\xi_n,\ldots$ are i.i.d. Uniform$(0,1)$. Their *empirical distribution function* is

$$(1)\qquad \begin{aligned} \mathbb{G}_n(t) &= \frac{1}{n}\sum_{i=1}^n 1_{[0,t]}(\xi_i) \qquad \text{for } 0\le t\le 1\\ &= \frac{\#\{\xi_i\le t,\ i=1,\ldots,n\}}{n}\\ &= \frac{k}{n} \qquad \text{for } \xi_{n:k}\le t<\xi_{n:k+1},\quad k=0,\ldots,n \end{aligned}$$

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. The *uniform empirical process* is defined by

$$(2)\qquad \mathbb{U}_n(t)\equiv\sqrt{n}(\mathbb{G}_n(t)-t) \qquad \text{for } 0\le t\le 1\,.$$

The inverse function $\mathbb{G}_n^{-1}$ of $\mathbb{G}_n$ is the *uniform quantile function*. Thus

$$(3)\qquad \mathbb{G}_n^{-1}(t)=\xi_{n:i} \qquad \text{for } (i-1)/n<t\le i/n,\quad i=1,\ldots,n\,.$$

The *uniform quantile process* $\mathbb{V}_n$ is defined by

$$(4)\qquad \mathbb{V}_n(t)\equiv\sqrt{n}(\mathbb{G}_n^{-1}(t)-t) \qquad \text{for } 0\le t\le 1\,.$$

Note that

$$(5)\qquad n\mathbb{G}_n(t)\sim \text{Binomial}(n,t) \qquad \text{for } 0\le t\le 1\,,$$

so that

$$(6)\qquad \mathbb{U}_n(t) \quad \text{has mean } 0 \quad \text{and variance } t(1-t) \qquad \text{for } 0\le t\le 1\,.$$

In fact

$$(7)\qquad Cov[1_{[0,s]}(\xi_i),1_{[0,t]}(\xi_i)]=s\wedge t-st \qquad \text{for } 0\le s,t\le 1\,.$$

Moreover, applying the multivariate CLT to $(1_{[0,s]}(\xi_i),1_{[0,t]}(\xi_i))$, it is clear that

$$(8)\qquad (\mathbb{U}_n(s),\mathbb{U}_n(t))\rightarrow_d N_2\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}s(1-s) & s\wedge t-st\\ s\wedge t-st & t(1-t)\end{pmatrix}\right) \qquad \text{as } n\to\infty\,,$$

for $0\le s,t\le 1$.

We define $\{\mathbb{U}(t):\ 0\le t\le 1\}$ to be a *Brownian bridge process* if it is a Gaussian process indexed by $t\in[0,1]$ having

$$(9)\qquad E\mathbb{U}(t)=0 \qquad \text{and} \qquad Cov[\mathbb{U}(s),\mathbb{U}(t)]=s\wedge t-st$$

for all $0\le s,t\le 1$. The process $\mathbb{U}$ exists and has sample functions $\mathbb{U}(\cdot,\omega)$ which are continuous for a.e. ω as we will show below. Of course the bivariate result (8) immediately extends to all the finite-dimensional marginal distributions of $\mathbb{U}_n$: for any $k\ge 1$ and any $t_1,\ldots,t_k\in[0,1]$,

$$(10)\qquad (\mathbb{U}_n(t_1),\ldots,\mathbb{U}_n(t_k))\rightarrow_d(\mathbb{U}(t_1),\ldots,\mathbb{U}(t_k))\sim N_k(0,(t_j\wedge t_{j'}-t_jt_{j'}))\,.$$

Thus we have convergence of all the finite-dimensional distributions of $\mathbb{U}_n$ to those of a Brownian bridge process $\mathbb{U}$, and we write

$$(11) \qquad \mathbb{U}_n \to_{f.d.} \mathbb{U} \qquad \text{as} \quad n \to \infty\,.$$

We would like to be able to conclude from (11) that $g(\mathbb{U}_n) \to_d g(\mathbb{U})$ as $n \to \infty$ for various continuous functionals such as $g(x) = \sup_{0\le t\le 1} |x(t)|$ for $x \in D[0,1]$, the space of all right-continuous functions on $[0,1]$ with left-limits. The conclusion (11) is not strong enough to imply this, but (10) can be strengthened to a result that does. The Mann-Wald theorem suggests $g(\mathbb{U}_n) \to_d g(\mathbb{U})$ should be true for "continuous" functionals g, and this raises the question of what metric should be used to define continuous.

The Empirical Process on R

Let $X_1, \ldots, X_n, \ldots$ be i.i.d. F with order statistics $X_{n:1} \le \cdots \le X_{n:n}$. Their *empirical distribution function* $\mathbb{F}_n$ is defined by

$$(12) \qquad \mathbb{F}_n(x) = \frac{1}{n}\sum_{i=1}^n 1_{(-\infty,x]}(X_i) \qquad \text{for} \quad -\infty < x < \infty\,.$$

The *empirical process* is defined to be $\sqrt{n}(\mathbb{F}_n - F)$. It will be very useful to suppose that random variables X_i^*, $i = 1, \ldots, n$, are defined by

$$(13) \qquad X_i^* = F^{-1}(\xi_i) \qquad i = 1, \ldots, n \qquad \text{for the} \;\; \xi_i\text{'s} \;\; \text{of (1)}\,.$$

Theorem 4.1 shows that these X_i^*'s are indeed i.i.d. F. Recall from Theorem 4.1 that also

$$(14) \qquad 1_{[X_i^* \le x]} = 1_{[\xi_i \le F(x)]} \qquad \text{on} \;\; (-\infty, \infty)$$

for these particular X_i^*'s. Thus for these X_i^*'s we have

$$(15) \qquad \mathbb{F}_n^* = \mathbb{G}_n(F) \qquad \text{on} \;\; (-\infty, \infty),$$

and

$$(16) \qquad \sqrt{n}(\mathbb{F}_n^* - F) = \mathbb{U}_n(F) \qquad \text{on} \;\; (-\infty, \infty)\;.$$

Note from (10) and (16) that

$$(17) \qquad \sqrt{n}(\mathbb{F}_n - F) \to_{f.d.} \mathbb{U}(F) \qquad \text{as} \;\; n \to \infty\,.$$

Theorem 5.1 (Glivenko - Cantelli). Let I denote the identity function on $[0,1]$, $I(t) = t$, for $0 \le t \le 1$. Then

$$(18) \qquad \|\mathbb{G}_n - I\|_\infty \equiv \sup_{0\le t\le 1} |\mathbb{G}_n(t) - t| \to_{a.s.} 0$$

and

$$(19) \qquad \|\mathbb{F}_n - F\|_\infty \equiv \sup_{-\infty<x<\infty} |\mathbb{F}_n(x) - F(x)| \to_{a.s.} 0$$

as $n \to \infty$.

Proof. Since

(a) $$\|\mathbb{F}_n - F\|_\infty \stackrel{d}{=} \|\mathbb{F}_n^* - F\|_\infty = \|\mathbb{G}_n(F) - F\|_\infty \le \|\mathbb{G}_n - I\|_\infty\,,$$

where the equality in distribution holds jointly in n and with equality if F is continuous, it suffices to prove the first part.

Fix a large integer M. Then

$$\begin{aligned}
\|\mathbb{G}_n - I\|_\infty &= \max_{1\le j\le M} \sup_{(j-1)/M\le t\le j/M} |\mathbb{G}_n(t) - t| \\
&= \max_{1\le j\le M} \{ \sup_{(j-1)/M\le t\le j/M} (\mathbb{G}_n(t) - t) \vee \sup_{(j-1)/M\le t\le j/M} (t - \mathbb{G}_n(t))\} \\
&\le \max_{1\le j\le M} \{(\mathbb{G}_n(j/M) - (j-1)/M) \ \vee (j/M - \mathbb{G}_n((j-1)/M))\} \\
&\le \max_{1\le j\le M} \{(\mathbb{G}_n(j/M) - j/M) \vee ((j-1)/M - \mathbb{G}_n((j-1)/M))\} + 1/M \\
&\to_{a.s.} 0 + 1/M
\end{aligned}$$

since $\mathbb{G}_n(j/M) \to_{a.s.} j/M$, $j = 1, \ldots, M$. But M was arbitrary; hence $\|\mathbb{G}_n - I\|_\infty \to_{a.s.} 0$. $\Box$

The next natural step is to show that

(20) $$\mathbb{U}_n \Rightarrow \mathbb{U} \qquad \text{as } n\to\infty \qquad \text{in } (D[0,1], \|\cdot\|_\infty)$$

and

(21) $$\sqrt{n}(\mathbb{F}_n - F) \stackrel{d}{=} \sqrt{n}(\mathbb{F}_n^* - F) = \mathbb{U}_n(F) \Rightarrow \mathbb{U}(F) \qquad \text{as } n\to\infty \qquad \text{in } (D(-\infty,\infty), \|\cdot\|_\infty)\,.$$

This is essentially what was proved by Donsker (1952). However, it turned out later that there are measurability difficulties here: $(D[0,1], \|\cdot\|_\infty)$ is an inseparable Banach space, and even though $\mathbb{U}$ takes values in the separable Banach space $(C[0,1], \|\cdot\|_\infty)$, in this case the unfortunate consequence is that $\mathbb{U}_n$ is not a measurable element of $(D[0,1], \|\cdot\|_\infty)$; see Billingsley (1968), Chapter 18. Roughly, the Borel sigma-field is too big. Hence an attractive alternative formulation is one that works around this difficulty essentially by carrying out an explicit Skorokhod construction of uniform empirical processes $\mathbb{U}_n^* \stackrel{d}{=} \mathbb{U}_n$ defined on a common probability space with a Brownian bridge process $\mathbb{U}^*$ and satisfying

(22) $$\|\mathbb{U}_n^* - \mathbb{U}^*\|_\infty = \sup_{0\le t\le 1} |\mathbb{U}_n^*(t) - \mathbb{U}^*(t)| \to_{a.s.} 0\,.$$

This is the content of the following theorem:

Theorem 5.2 There exists a (sequence of) uniform empirical processes $\mathbb{U}_n^*$ corresponding to a triangular array of row independent Uniform$(0,1)$ random variables $\xi_{n1}, \ldots, \xi_{nn}$, $n \ge 1$, and a Brownian bridge process $\mathbb{U}^*$ all defined on a common probability space $(\Omega, \mathcal{A}, P)$, such (22) holds. Thus it follows that

(23) $$\|\sqrt{n}(\mathbb{F}_n^* - F) - \mathbb{U}^*(F)\|_\infty \to_{a.s.} 0 \qquad \text{as } n\to\infty.$$

The convergence in (22) was strengthened in the papers of Komlós, Major, and Tusnády (1975), (1978) as follows: the construction can be carried out so that the convergence in (22) holds with the rate $n^{-1/2}\log n$: there is a construction of $\mathbb{U}_n^*$ and $\mathbb{U}^*$ so that

(24) $$\|\mathbb{U}_n^* - \mathbb{U}^*\|_\infty \le C\frac{\log n}{\sqrt{n}} \qquad \text{a.s.}\,,$$

for some absolute constant C. Moreover, there is a construction of the sequence(s) $\{\mathbb{U}_n^*\}_{n\geq 1}$ and $\mathbb{U}^* = \mathbb{U}^{*n}$ on a common probability space so that the joint in n distributions are correct and

$$\|\mathbb{U}_n^* - \mathbb{U}^*\|_\infty \leq C\frac{(\log n)^2}{\sqrt{n}} \qquad \text{a.s.}\,. \tag{25}$$

In any case, these results have the following corollary:

Corollary 1 (Donsker's theorem). If $g : D[0,1] \to R$ is $\|\cdot\|_\infty$−continuous, then $g(\mathbb{U}_n) \to_d g(\mathbb{U})$.

Here are some examples of this:

Example 5.1 (Kolmogorov's (two-sided) statistic). If F is continuous, then

$$\|\sqrt{n}(\mathbb{F}_n - F)\|_\infty \stackrel{d}{=} \|\mathbb{U}_n(F)\|_\infty = \|\mathbb{U}_n\|_\infty \to_d \|\mathbb{U}\|_\infty\,. \tag{26}$$

It is known, via reflection methods (see Shorack and Wellner (1986), pages 33-42) that

$$P(\|\mathbb{U}\|_\infty > \lambda) = 2\sum_{k=1}^{\infty}(-1)^{k+1}\exp(-2k^2\lambda^2) \qquad \text{for} \;\; \lambda > 0\,. \tag{27}$$

Dvoretzky, Kiefer, and Wolfowitz (1956) showed that

$$P(\|\mathbb{U}_n\|_\infty \geq \lambda) \leq Ce^{-2\lambda^2}$$

for all $\lambda > 0$ and $n \geq 1$ for some constant C. Massart (1990) showed that the inequality in the last display holds with $C = 2$ as conjectured by Birnbaum and McCarty (1958).

Example 5.2 (Kolmogorov's one-sided statistic). If F is continuous, then

$$\|\sqrt{n}(\mathbb{F}_n - F)^+\|_\infty \equiv \sup_{-\infty<x<\infty}\sqrt{n}(\mathbb{F}_n(x) - F(x)) \stackrel{d}{=} \|\mathbb{U}_n^+(F)\|_\infty = \|\mathbb{U}_n^+\|_\infty \to_d \|\mathbb{U}^+\|_\infty\,. \tag{28}$$

It is known (see Shorack and Wellner (1986), pages 37 and 142) that

$$P(\|\mathbb{U}^+\|_\infty > \lambda) = \exp(-2\lambda^2) \qquad \text{for} \;\; \lambda > 0\,. \tag{29}$$

Example 5.3 (Birnbaum's statistic). If F is continuous,

$$\int_{-\infty}^{\infty}\sqrt{n}(\mathbb{F}_n(x) - F(x))dF(x) \stackrel{d}{=} \int_{-\infty}^{\infty}\mathbb{U}_n(F)dF = \int_0^1 \mathbb{U}_n(t)dt \to_d \int_0^1 \mathbb{U}(t)dt\,. \tag{30}$$

Now $\int_0^1 \mathbb{U}(t)dt$ is a linear combination of normal random variables, and hence it has a normal distribution. It has expectation 0 by Fubini's theorem since $E(\mathbb{U}(t)) = 0$ for each fixed t. Furthermore, again by Fubini's theorem,

$$\begin{aligned}
E\left(\int_0^1 \mathbb{U}(t)dt\right)^2 &= E\left(\int_0^1 \mathbb{U}(s)ds\int_0^1 \mathbb{U}(t)dt\right) \\
&= \int_0^1\int_0^1 E\{\mathbb{U}(s)\mathbb{U}(t)\}dsdt \\
&= \int_0^1\int_0^1 (s\wedge t - st)\,dsdt = \frac{1}{12}.
\end{aligned}$$

Hence $\int_0^1 \mathbb{U}(t)dt \sim N(0, 1/12)$.

Example 5.4 (Cramér - von Mises statistic). If F is continuous,

$$\int_{-\infty}^{\infty} \{\sqrt{n}(\mathbb{F}_n(x) - F(x))^2 dF(x) \stackrel{d}{=} \int_{-\infty}^{\infty} \{\mathbb{U}_n(F)\}^2 dF = \int_0^1 \{\mathbb{U}_n(t)\}^2 dt \rightarrow_d \int_0^1 \{\mathbb{U}(t)\}^2 dt \,.$$

In this case it is known that

$$\int_0^1 \{\mathbb{U}(t)\}^2 dt \stackrel{d}{=} \sum_{j=1}^{\infty} \frac{1}{j^2\pi^2} Z_j^2 \,, \tag{31}$$

where the Z_j's are i.i.d. $N(0,1)$, and this distribution has been tabled; see Shorack and Wellner (1986), page 148.

Example 5.5 (Anderson - Darling statistic). If F is continuous,

$$\int_{-\infty}^{\infty} \frac{\{\sqrt{n}(\mathbb{F}_n - F)\}^2}{F(1-F)} dF \stackrel{d}{=} \int_{\infty}^{\infty} \frac{\{\mathbb{U}_n(F)\}^2}{F(1-F)} dF = \int_0^1 \frac{\{\mathbb{U}_n(t)\}^2}{t(1-t)} dt \rightarrow_d \int_0^1 \frac{\{\mathbb{U}(t)\}^2}{t(1-t)} dt \,.$$

It is known in this case that

$$\int_0^1 \frac{\{\mathbb{U}(t)\}^2}{t(1-t)} dt \stackrel{d}{=} \sum_{j=1}^{\infty} \frac{1}{j(j+1)} Z_j^2 \tag{32}$$

where the Z_j's are i.i.d. $N(0,1)$. This distribution has also been tabled; see Shorack and Wellner (1986), page 148.

General Empirical Measures and Processes

Now suppose that $X_1, X_2, \ldots, X_n, \ldots$ are i.i.d. P on the measurable space $(S, \mathcal{S})$. We let $\mathbb{P}_n$ denote the *empirical measure* of the first n of the X_i's:

$$\mathbb{P}_n \equiv \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i} \,; \tag{33}$$

here δ_x denotes the measure with mass 1 at $x \in S$: $\delta_x(B) = 1_B(x)$ for $B \in \mathcal{S}$. Thus for a set $B \in \mathcal{S}$,

$$\mathbb{P}_n(B) = \frac{1}{n} \sum_{i=1}^{n} 1_B(X_i) = \frac{1}{n} \#\{i \le n : \; X_i \in B\} \,. \tag{34}$$

Note that when $S = R$ so that the X_i's are real-valued, and $B = (-\infty, x]$ for $x \in R$, then

$$\mathbb{P}_n(B) = \mathbb{P}_n((-\infty, x]) = \mathbb{F}_n(x) \,, \tag{35}$$

the empirical distribution function of the X_i's at x.

The *empirical process* $\mathbb{G}_n$ is defined by

$$\mathbb{G}_n \equiv \sqrt{n}(\mathbb{P}_n - P) \,. \tag{36}$$

The question is how to "index" $\mathbb{P}_n$ and $\mathbb{G}_n$ as stochastic processes.

Some history: For the case $S = R^d$, the empirical distribution function $\{\mathbb{F}_n(x), \; x \in R^d\}$, is the case obtained by choosing the class of sets to be the lower-left orthants

$$\mathcal{C} = \mathcal{O}_d \equiv \{(-\infty, x] : \; x \in R^d\} \,,$$

and this direction was pursued in some detail through the 1950's and early 1960's. However, with a little thought it becomes clear that many other classes of sets will be of interest in R^d. For example why not consider the empirical process indexed by the class of all rectangles

$$\mathcal{R}_d \equiv \{A = [a_1, b_1] \times \cdots \times [a_d, b_d] :\ a_j, b_j \in R,\ j = 1, \ldots, d\}$$

or the class of all closed balls

$$\mathcal{B}_d \equiv \{B(x, r) :\ x \in R^d,\ r > 0\}$$

where $B(x, r) = \{y \in R^d : |y - x| \le r\}$; or the class of all half-spaces

$$\mathcal{H}_d \equiv \{H(u, t) :\ u \in S^{d-1}, t \in R\}$$

where $H(u, t) \equiv \{y \in R^d :\ \langle y, u\rangle \le t\}$ and $S^{d-1} \equiv \{u \in R^d :\ |u| = 1\}$ denotes the unit sphere in R^d; or the class of all convex sets in R^d

$$\mathcal{C}_d \equiv \{C \subset R^d :\ C \text{ is convex}\}\,?$$

All of these cases correspond to the empirical process indexed by some class of indicator functions

$$\{1_C :\ C \in \mathcal{C}\}$$

for the appropriate choice of $\mathcal{C}$. Thus we can consider the empirical measure and the empirical process as functions on a class of sets $\mathcal{C}$ which map sets $C \in \mathcal{C}$ to the real-valued random variables

$$\mathbb{P}_n(C) \qquad \text{and} \qquad \mathbb{G}_n(C) = \sqrt{n}(\mathbb{P}_n(C) - P(C))\,.$$

Note that for any class of sets $\mathcal{C}$ we have

$$\sup_{C \in \mathcal{C}} \mathbb{P}_n(C) \le 1 < \infty \qquad \text{and} \qquad \sup_{C \in \mathcal{C}} |\sqrt{n}(\mathbb{P}_n(C) - P(C))| \le \sqrt{n} < \infty\,,$$

so we can regard both $\mathbb{P}_n$ and $\mathbb{G}_n$ as elements of the space $l^\infty(\mathcal{C}) \equiv \{x : \mathcal{C} \to R |\ \sup_{C \in \mathcal{C}} |x(C)| < \infty\}$.

More generally still, we can think of indexing the empirical process by a class $\mathcal{F}$ of functions $f : S \to R$. For example, when $S = R^d$ a natural class which might easily arise in applications is the class of functions

$$\mathcal{F} = \{f_t(x) :\ t \in R^d\}$$

where $f_t(x) = |x - t|$. This is already an interesting class of functions when $d = 1$.

For any fixed measureable function $f : S \to$ we will use the notation

$$P(f) = \int f dP, \qquad \mathbb{P}_n(f) = \int f d\mathbb{P}_n = \frac{1}{n} \sum_{i=1}^{n} f(X_i)\,.$$

From the strong law of large numbers it follows that for any fixed function f with $E|f(X)| < \infty$

(37) $\quad \mathbb{P}_n(f) \to_{a.s.} P(f) = Ef(X_1)\,.$

By the central limit theorem (CLT) it follows that for any fixed function f with $E|f(X)|^2 < \infty$

(38) $\quad \mathbb{G}_n(f) = \sqrt{n}(\mathbb{P}_n(f) - P(f)) \to_d \mathbb{G}(f)\,; \sim N(0, Var(f(X_1))$

here $\mathbb{G}$ denotes a $P-$Brownian bridge process: i.e. a mean zero Gaussian process with covariance function

$$(39) \qquad Cov[\mathbb{G}(f), \mathbb{G}(g)] = P(fg) - P(f)P(g)\,.$$

The question of interest is: for what classes $\mathcal{C}$ of subsets of S, $\mathcal{C} \subset \mathcal{A}$ or classes of functions $\mathcal{F}$, can we make these convergences hold uniformly in $C \in \mathcal{C}$, or uniformly in $f \in \mathcal{F}$? These are the kinds of questions with which modern empirical process theory is concerned and can answer.

To state some typical results from this theory, we first need several definitions. If d is a metric on a set $\mathcal{F}$, then we define the *covering numbers of $\mathcal{F}$ with respect to d* as follows:

$$(40) \qquad N(\epsilon, \mathcal{F}, d) \equiv \inf\{k : \text{ there exist } f_1, \ldots, f_k \in \mathcal{F} \text{ such that } \mathcal{F} \subset \cup_{j=1}^k B(f_j, \epsilon)\}\,;$$

here $B(f, \epsilon) \equiv \{g \in \mathcal{F} : d(g, f) \leq \epsilon\}$. Another useful notion is that of a bracket: if $l \leq u$ are two real-valued functions defined on S, then the bracket $[u, l]$ is defined by

$$(41) \qquad [u, l] \equiv \{f : \ l(s) \leq f(s) \leq u(s) \text{ for all } s \in S\}\,.$$

We say that a bracket $[u, l]$ is an $\epsilon-$bracket for the metric d if $d(u, l) \leq \epsilon$. Then the *bracketing covering number* $N_{[]}(\epsilon, \mathcal{F}, d)$ for a set of functions $\mathcal{F}$ is

$$(42) \qquad N_{[]}(\epsilon, \mathcal{F}, d) \equiv \inf\{k : \text{ there exist } \epsilon-\text{brackets } [l_1, u_1], \ldots, [l_k, u_k] \text{ such that } \mathcal{F} \subset \cup_{j=1}^k [l_j, u_j]\}\,.$$

One more bit of notation is needed before stating our theorems: an envelope function F for a class of functions $\mathcal{F}$ is any function satisfying

$$(43) \qquad |f(x)| \leq F(x) \qquad \text{for all } x \in S, \ \text{ and all } \ f \in \mathcal{F}\,.$$

Usually we will take F to be the minimal measurable majorant

$$(44) \qquad F(x) \equiv \left(\sup_{f \in \mathcal{F}} |f(x)|\right)^*\,,$$

where here the $*$ stands for "smallest measurable function above" the quantity in parentheses (which need not be measurable since it is, in general, a supremum over an uncountable collection). [Note that this F *is not a distribution function*!]

Now we can state two generalizations of the Glivenko-Cantelli theorems.

Theorem 5.3 Suppose that $\mathcal{F}$ is a class of functions with finite $L_1(P)-$bracketing numbers: $N_{[]}(\epsilon, \mathcal{F}, L_1(P)) < \infty$ for every $\epsilon > 0$. Then

$$(45) \qquad \|\mathbb{P}_n - P\|_{\mathcal{F}} \equiv \sup_{f \in \mathcal{F}} |\mathbb{P}_n(f) - P(f)| \to_{a.s.} 0\,.$$

Theorem 5.4 Suppose that $\mathcal{F}$ is a class of functions with:
A. An integrable envelope function F: $P(F) < \infty$.
B. The truncated classes $\mathcal{F}_M \equiv \{f 1_{[F \leq M]} : \ f \in \mathcal{F}\}$ satisfy

$$(46) \qquad n^{-1} \log N(\epsilon, \mathcal{F}_M, L_1(\mathbb{P}_n)) \to_{a.s.} 0$$

for every $\epsilon > 0$ and $0 < M < \infty$. Then, if $\mathcal{F}$ is also "suitably measurable",

$$(47) \qquad \|\mathbb{P}_n - P\|_{\mathcal{F}} \equiv \sup_{f \in \mathcal{F}} |\mathbb{P}_n(f) - P(f)| \to_{a.s.} 0\,.$$

Note that the key hypothesis (46) of Theorem 5.4 is clearly satisfied if

$$\sup_Q N(\epsilon, \mathcal{F}_M, L_1(Q)) < \infty \tag{48}$$

for every $\epsilon > 0$ and $M > 0$; here the supremum is over finitely discrete measures Q.

Finally, here are two generalizations of the Donsker theorem.

Theorem 5.5 (Ossiander's uniform CLT). Suppose that $\mathcal{F}$ is a class of functions with $L_2(P)$ bracketing numbers $N_{[]}(\epsilon, \mathcal{F}, L_2(P))$ satisfying

$$\int_0^\infty \sqrt{\log N_{[]}(\epsilon, \mathcal{F}, L_2(P))}\, d\epsilon < \infty\,. \tag{49}$$

Then

$$\mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - P) \Rightarrow \mathbb{G} \qquad \text{in } \ l^\infty(\mathcal{F}) \qquad \text{as } \ n \to \infty\,. \tag{50}$$

Theorem 5.6 (Pollard's uniform CLT). Suppose that $\mathcal{F}$ is a class of functions satisfying:
A. The envelope function F of $\mathcal{F}$ is square integrable: $P(F^2) < \infty$.
B. The uniform covering numbers $\sup_Q \log N(\epsilon\|F\|_{Q,2}, \mathcal{F}, L_2(Q))$ satisfy

$$\int_0^\infty \sqrt{\sup_Q \log N(\epsilon\|F\|_{Q,2}, \mathcal{F}, L_2(Q))}\, d\epsilon < \infty\,. \tag{51}$$

Then

$$\mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - P) \Rightarrow \mathbb{G} \qquad \text{in } \ l^\infty(\mathcal{F}) \qquad \text{as } \ n \to \infty\,. \tag{52}$$

For proofs of Theorems 45 - 5.6, see van der Vaart and Wellner (1996). Treatments of empirical process theory are also given by Dudley (1999, 2014), Van de Geer (1999), and Giné and Nickl (2016).

6 The Partial Sum Process and Brownian Motion

We define $\{\mathbb{S}(t) : 0 \le t \le 1\}$ to be *Brownian motion* if $\mathbb{S}$ is a Gaussian process indexed by $t \in [0,1]$ having

$$(1) \qquad E(\mathbb{S}(t)) = 0 \qquad \text{and} \qquad Cov[\mathbb{S}(s), \mathbb{S}(t)] = s \wedge t$$

for all $0 \le s, t \le 1$. These finite-dimensional distributions are "consistent", and hence a theorem of Kolmogorov shows that the process $\mathbb{S}$ exists. Note that (1) and normality imply that

$$(2) \qquad \mathbb{S} \qquad \text{has stationary independent increments}\,.$$

Exercise 6.1 Suppose that $\mathbb{U}$ is a Brownian bridge and $Z \sim N(0,1)$ is is independent of $\mathbb{U}$. Let

$$(3) \qquad \mathbb{S}(t) \equiv \mathbb{U}(t) + tZ \qquad \text{for} \;\; 0 \le t \le 1$$

is a Brownian motion.

Exercise 6.2 Suppose that $\mathbb{S}$ is a Brownian motion. Show that

$$(4) \qquad \mathbb{U}(t) \equiv \mathbb{S}(t) - t\mathbb{S}(1) \qquad \text{for} \;\; 0 \le t \le 1$$

is a Brownian bridge.

Now suppose that $X_1, \ldots, X_n$ are i.i.d. random variables with mean 0 and and variance 1, and set $X_0 \equiv 0$. We define the partial sum process $\mathbb{S}_n$ by

$$(5) \qquad \mathbb{S}_n(t) \equiv \mathbb{S}_n(k/n) = \frac{1}{\sqrt{n}} \sum_{i=1}^{k} X_i \qquad \text{for} \;\; \frac{k}{n} \le t < \frac{k+1}{n}\,,$$

for $0 \le k < \infty$. Note that

$$\begin{aligned} Cov[\mathbb{S}_n(j/n), \mathbb{S}_n(k/n)] &= \frac{1}{n} \sum_{i=1}^{j} \sum_{i'=1}^{k} Cov[X_i, X_{i'}] \\ &= \frac{1}{n} \sum_{i=1}^{j \wedge k} Var[X_i] = \frac{j \wedge k}{n} \\ &\to s \wedge t \qquad \text{if} \;\; j/n \to s \;\; \text{and} \;\; k/n \to t\,. \end{aligned}$$

Also,

$$(6) \qquad \mathbb{S}_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{[nt]} X_i = \sqrt{\frac{[nt]}{n}} \frac{1}{\sqrt{[nt]}} \sum_{i=1}^{[nt]} X_i \to_d \sqrt{t} N(0,1) \sim N(0,t)\,.$$

for $0 \le t \le 1$ by the CLT and Slutsky's theorem. This suggests that

$$(7) \qquad \mathbb{S}_n \to_{f.d.} \mathbb{S} \qquad \text{as} \;\; n \to \infty\,.$$

This will be verified in exercise 6.3. Much more is true: $\mathbb{S}_n \Rightarrow \mathbb{S}$ as processes in $D[0,1]$, and hence $g(\mathbb{S}_n) \to_d g(\mathbb{S})$ for continuous functionals g.

Exercise 6.3 Show that (7) holds: i.e. for any fixed $t_1, \ldots, t_k \in [0,1]^k$,

$$(\mathbb{S}_n(t_1), \ldots, \mathbb{S}_n(t_k)) \to_d (\mathbb{S}(t_1), \ldots, \mathbb{S}(t_k)) .$$

Existence of Brownian motion and Brownian bridge as continuous processes on $C[0,1]$

The aim of this subsection to convince you that both Brownian motion and Brownian bridge exist as *continuous Gaussian processes* on $[0,1]$, and that we can then extend the definition of Brownian motion to $[0,\infty)$.

Definition 6.1 *Brownian motion* (or *standard Brownian motion*, or a Wiener process) $\mathbb{S}$ is a Gaussian process with continuous sample functions and:
(i) $\mathbb{S}(0) = 0$;
(ii) $E(\mathbb{S}(t)) = 0, \quad 0 \le t \le 1$;
(iii) $E\{\mathbb{S}(s)\mathbb{S}(t)\} = s \wedge t, \quad 0 \le s,t \le 1$.

Definition 6.2 A *Brownian bridge process* $\mathbb{U}$ is a Gaussian process with continuous sample functions and:
(i) $\mathbb{U}(0) = \mathbb{U}(1) = 0$;
(ii) $E(\mathbb{U}(t)) = 0, \quad 0 \le t \le 1$;
(iii) $E\{\mathbb{U}(s)\mathbb{U}(t)\} = s \wedge t - st, \quad 0 \le s,t \le 1$.

Theorem 6.1 Brownian motion $\mathbb{S}$ and Brownian bridge $\mathbb{U}$ exist.

Proof. We first construct a Brownian bridge process $\mathbb{U}$. Let

$$\text{(a)} \qquad h_{00}(t) \equiv h(t) \equiv \begin{cases} t & 0 \le t \le 1/2\,, \\ 1-t & 1/2 \le t \le 1\,, \\ 0 & \text{elsewhere}\,. \end{cases}$$

For $n \ge 1$ let

$$\text{(b)} \qquad h_{nj}(t) \equiv 2^{-n/2} h(2^n t - j), \qquad j = 0, \ldots, 2^n - 1\,.$$

For example, $h_{10}(t) = 2^{-1/2}h(2t)$, $h_{11}(t) = 2^{-1/2}h(2t-1)$, while

$$\begin{aligned} h_{20}(t) &= 2^{-1}h(4t), & h_{21}(t) &= 2^{-1}h(4t-1)\,, \\ h_{22}(t) &= 2^{-1}h(4t-2), & h_{23}(t) &= 2^{-1}h(4t-3)\,. \end{aligned}$$

Note that $|h_{nj}(t)| \le 2^{-n/2}2^{-1}$.

The functions $\{h_{nj} : \ j = 0, \ldots, 2^n - 1, \quad n \ge 0\}$ are called the *Schauder functions*; they are integrals of the orthonormal (with respect to Lebesgue measure on $[0,1]$) family of functions $\{g_{nj} : \ j = 0, \ldots, 2^n - 1, \ \ n \ge 0\}$ called the *Haar functions* defined by

$$\begin{aligned} g_{00}(t) &\equiv g(t) \equiv 2 1_{[0,1/2]}(t) - 1, \\ g_{nj}(t) &\equiv 2^{n/2} g_{00}(2^n t - j)\,, \qquad j = 0, \ldots, 2^n - 1, \ \ n \ge 1\,. \end{aligned}$$

Thus

$$\text{(c)} \qquad \int_0^1 g_{nj}^2(t)dt = 1, \qquad \int_0^1 g_{nj}(t) g_{n'j'}(t)dt = 0 \qquad \text{if} \qquad n \ne n', \ \text{or} \ \ j \ne j'\,,$$

and

$$\text{(d)} \qquad h_{nj}(t) = \int_0^t g_{nj}(s)ds, \qquad 0 \le t \le 1\,.$$

Furthermore, the family $\{g_{nj}\}_{j=0,\,n\ge 0}^{2^n-1} \cup \{g(\cdot/2)\}$ is complete: any $f \in L_2(0,1)$ has an expansion in terms of the g's. In fact the Haar basis is the simplest *wavelet basis* of $L_2(0,1)$, and is the starting point for further developments in the area of wavelets.

Now let $\{Z_{nj}\}_{j=0,\,n\ge 0}^{2^n-1}$ be independent identically distributed $N(0,1)$ random variables; if we wanted, we could construct all these random variables on the probability space $([0,1], \mathcal{B}_{[0,1]}, \lambda)$. Define

$$V_n(t,\omega) = \sum_{j=0}^{2^n-1} Z_{nj}(\omega) h_{nj}(t)\,,$$

$$U_m(t,\omega) = \sum_{n=0}^{m} V_n(t,\omega)\,.$$

For $m > k$

$$\text{(e)} \qquad |U_m(t,\omega) - U_k(t,\omega)| = |\sum_{n=k+1}^{m} V_n(t,\omega)| \le \sum_{n=k+1}^{m} |V_n(t,\omega)|$$

where

$$\text{(f)} \qquad |V_n(t,\omega)| \le \sum_{j=0}^{2^n-1} |Z_{nj}(\omega)||h_{nj}(t)| \le 2^{-(n/2+1)} \max_{0\le j\le 2^n-1} |Z_{nj}(\omega)|$$

since the h_{nj}, $j = 0, \ldots, 2^n - 1$ are $\neq 0$ on disjoint t intervals.

Now $P(Z_{nj} > z) = 1 - \Phi(z) \le z^{-1}\phi(z)$ for $z > 0$ (by "Mill's ratio") so that

$$\text{(g)} \qquad P(|Z_{nj}| \ge 2\sqrt{n}) = 2P((Z_{nj} \ge 2\sqrt{n}) \le \frac{2}{\sqrt{2\pi}}(2\sqrt{n})^{-1}e^{-2n}\,.$$

Hence

$$\text{(h)} \qquad P\left(\max_{0\le j\le 2^n-1} |Z_{nj}| \ge 2\sqrt{n}\right) \le 2^n P(|Z_{00}| \ge 2\sqrt{n}) \le \frac{2^n}{\sqrt{2\pi}} n^{-1/2} e^{-2n}\,;$$

since this is a term of a convergent series, by the Borel-Cantelli lemma $\max_{0\le j\le 2^n-1} |Z_{nj}| \ge 2\sqrt{n}$ occurs infinitely often with probability zero; i.e. except on a null set, for all ω there is an $N = N(\omega)$ such that $\max_{0\le j\le 2^n-1} |X_{nj}(\omega)| < 2\sqrt{n}$ for all $n > N(\omega)$. Hence

$$\text{(i)} \qquad \sup_{0\le t\le 1} |U_m(t) - U_k(t)| \le \sum_{n=k+1}^{m} 2^{-n/2} n^{1/2} \downarrow 0$$

for all $k, m \ge N' \ge N(\omega)$. Thus $U_m(t,\omega)$ converges uniformly as $m \to \infty$ with probability one to the (necessarily continuous) function

$$\text{(j)} \qquad \mathbb{U}(t,\omega) \equiv \sum_{n=0}^{\infty} V_n(t,\omega)\,.$$

Define $\mathbb{U} \equiv 0$ on the exceptional set. Then $\mathbb{U}$ is continuous for all ω.

Now $\{\mathbb{U}(t) : 0 \le t \le 1\}$ is clearly a Gaussian process since it is the sum of Gaussian processes. We now show that $\mathbb{U}$ is in fact a Brownian bridge: by formal calculation (it remains only to justify the interchange of summation and expectation),

$$\begin{aligned}
E\{\mathbb{U}(s)\mathbb{U}(t)\} &= E\left\{\sum_{n=0}^{\infty} V_n(s) \sum_{m=0}^{\infty} V_m(t)\right\} \\
&= \sum_{n=0}^{\infty} E\{V_n(s)V_n(t)\} \\
&= \sum_{n=0}^{\infty} E\left\{\sum_{j=0}^{2^n-1} Z_{nj}\int_0^s g_{nj}d\lambda \sum_{k=0}^{2^n-1} Z_{nk}\int_0^t g_{nk}d\lambda\right\} \\
&= \sum_{n=0}^{\infty}\sum_{j=0}^{2^n-1}\int_0^s g_{nj}d\lambda \int_0^t g_{nj}d\lambda \\
&= \sum_{n=0}^{\infty}\sum_{j=0}^{2^n-1}\int_0^1 1_{[0,s]}g_{nj}d\lambda \int_0^1 1_{[0,t]}g_{nj}d\lambda + st - st \\
&= \int_0^1 1_{[0,s]}(u)1_{[0,t]}(u)du - st \\
&= s\wedge t - st
\end{aligned}$$

where the next to last equality follows from Parseval's identity. Thus $\mathbb{U}$ is Brownian bridge.

Now let Z be one additional $N(0,1)$ random variable independent of all the others used in the construction, and define

(k) $$\mathbb{S}(t) \equiv \mathbb{U}(t) + tZ = \sum_{n=0}^{\infty} V_n(t) + tZ\,.$$

Then $\mathbb{S}$ is also Gaussian with 0 mean and

$$\begin{aligned}
Cov[\mathbb{S}(s),\mathbb{S}(t)] &= Cov[\mathbb{U}(s)+sZ, \mathbb{U}(t)+tZ] \\
&= Cov[\mathbb{U}(s),\mathbb{U}(t)] + stVar(Z) \\
&= s\wedge t - st + st = s\wedge t\,.
\end{aligned}$$

Thus $\mathbb{S}$ is Brownian motion. Since $\mathbb{U}$ has continuous sample paths, so does $\mathbb{S}$. □

Exercise 6.4 Graph the first few g_{nj}'s and h_{nj}'s.

Exercise 6.5 Justify the interchange of expectation and summation used in the proof. [Hint: use the Tonelli part of Fubini's theorem.]

Exercise 6.6 Let $\mathbb{U}$ be a Brownian bridge process. For $0 \le t < \infty$ define a process $\mathbb{B}$ by

(8) $$\mathbb{B}(t) \equiv (1+t)\mathbb{U}\left(\frac{t}{1+t}\right).$$

Show that $\mathbb{B}$ is a Brownian motion process on $[0,\infty)$.

7 Quantiles and Quantile Processes

Let $X_1, \ldots, X_n$ be i.i.d. real-valued random variables with distribution function F, and let $X_{(1)} \le X_{(2)} \le \ldots \le X_{(n)}$ denote the *order statistics*. For $t \in (0,1)$, let

$$(1) \qquad \mathbb{F}_n^{-1}(t) \equiv \inf\{x : \ \mathbb{F}_n(x) \ge t\}$$

so that

$$(2) \qquad \mathbb{F}_n^{-1}(t) = X_{(i)} \qquad \text{for } \ \frac{i-1}{n} < t \le \frac{i}{n}, \quad i = 1, \ldots, n\,.$$

Let $\xi_1, \ldots, \xi_n$ be i.i.d. Uniform$(0,1)$ random variables, and let $0 \le \xi_{(1)} \le \ldots \le \xi_{(n)} \le 1$ denote their order statistics. Thus, with $\mathbb{G}_n^{-1}(t) \equiv \inf\{x : \mathbb{G}_n(x) \ge t\}$,

$$(3) \qquad \mathbb{G}_n^{-1}(t) = \xi_{(i)} \qquad \text{for } \ \frac{i-1}{n} < t \le \frac{i}{n}, \quad i = 1, \ldots, n\,.$$

Now

$$(4) \qquad (X_1^*, \ldots, X_n^*) \equiv (F^{-1}(\xi_1), \ldots, F^{-1}(\xi_n)) \stackrel{d}{=} (X_1, \ldots, X_n),$$

so

$$(5) \qquad (X_{(1)}^*, \ldots, X_{(n)}^*) \equiv (F^{-1}(\xi_{(1)}), \ldots, F^{-1}(\xi_{(n)})) \stackrel{d}{=} (X_{(1)}, \ldots, X_{(n)}).$$

Hence it follows that

$$(6) \qquad \mathbb{F}_n^{-1}(\,\cdot\,) \stackrel{d}{=} F^{-1}(\mathbb{G}_n^{-1}(\,\cdot\,))\,,$$

and to study $\mathbb{F}_n^{-1}$ it suffices to study $\mathbb{G}_n^{-1}$.

Proposition 7.1 The sequence of uniform quantile functions $\mathbb{G}_n^{-1}$ satisfy

$$(7) \qquad \|\mathbb{G}_n^{-1} - I\|_\infty \equiv \sup_{0 \le t \le 1} |\mathbb{G}_n^{-1}(t) - t| = \|\mathbb{G}_n - I\|_\infty \to_{a.s.} 0\,,$$

and hence, if F^{-1} is continuous on $[a,b] \subset [0,1]$, then

$$(8) \qquad \|\mathbb{F}_n^{-1} - F^{-1}\|_a^b \equiv \sup_{a \le t \le b} |\mathbb{F}_n^{-1}(t) - F^{-1}(t)| \to_{a.s.} 0\,.$$

Proof. Note that $\|\mathbb{G}_n^{-1} - I\|_\infty = \|\mathbb{G}_n - I\|_\infty$ by inspection of the graphs. Thus

$$(a) \qquad \|\mathbb{F}_n^{-1} - F^{-1}\|_a^b \stackrel{d}{=} \|F^{-1}(\mathbb{G}_n^{-1}) - F^{-1}(I)\|_a^b \to_{a.s.} 0$$

since F^{-1} is uniformly continuous on $[a,b]$ and $\|\mathbb{G}_n^{-1} - I\|_\infty \to_{a.s.} 0$. $\square$

Definition 7.1 The *uniform quantile process* $\mathbb{V}_n$ is defined by

$$(9) \qquad \mathbb{V}_n \equiv \sqrt{n}(\mathbb{G}_n^{-1} - I)\,.$$

The *general quantile process* is defined by

$$(10) \qquad \sqrt{n}(\mathbb{F}_n^{-1} - F^{-1}) \stackrel{d}{=} \sqrt{n}(F^{-1}(\mathbb{G}_n^{-1}) - F^{-1})\,.$$

Theorem 7.1 The uniform quantile process can be written as

$$(11)\qquad \mathbb{V}_n = -\mathbb{U}_n(\mathbb{G}_n^{-1}) + \sqrt{n}(\mathbb{G}_n \circ \mathbb{G}_n^{-1} - I)\,,$$

and hence for the specially constructed $\mathbb{U}_n$ of theorem 5.2 it follows that, with $\mathbb{V} \equiv -\mathbb{U} \stackrel{d}{=} \mathbb{U}$,

$$(12)\qquad \|\mathbb{V}_n - \mathbb{V}\|_\infty \to_{a.s.} 0 \qquad \text{as} \qquad n \to \infty\,.$$

Proof. First we prove the identity (11):

$$\begin{aligned}
\mathbb{V}_n &= \sqrt{n}(\mathbb{G}_n^{-1} - I) \\
&= \sqrt{n}(\mathbb{G}_n^{-1} - \mathbb{G}_n(\mathbb{G}_n^{-1})) + \sqrt{n}(\mathbb{G}_n(\mathbb{G}_n^{-1}) - I) \\
&= -\mathbb{U}_n(\mathbb{G}_n^{-1}) + \sqrt{n}(\mathbb{G}_n \circ \mathbb{G}_n^{-1} - I)\,.
\end{aligned}$$

Now $\|\mathbb{G}_n^{-1} - I\|_\infty \to_{a.s.} 0$ by proposition 7.1, and

$$\text{(a)}\qquad \|\mathbb{G}_n \circ \mathbb{G}_n^{-1} - I\|_\infty = \sup_{0\le t\le 1} |\mathbb{G}_n(\mathbb{G}_n^{-1}(t)) - t| = \frac{1}{n}\,.$$

Hence

$$\begin{aligned}
\|\mathbb{V}_n - \mathbb{V}\|_\infty &\le \|\mathbb{U}_n(\mathbb{G}_n^{-1}) - \mathbb{U}\|_\infty + \|\sqrt{n}(\mathbb{G}_n(\mathbb{G}_n^{-1}) - I)\|_\infty \\
&\le \|\mathbb{U}_n(\mathbb{G}_n^{-1}) - \mathbb{U}(\mathbb{G}_n^{-1})\|_\infty + \|\mathbb{U}(\mathbb{G}_n^{-1}) - \mathbb{U}\|_\infty + \frac{1}{\sqrt{n}} \\
&\le \|\mathbb{U}_n - \mathbb{U}\|_\infty + \|\mathbb{U}(\mathbb{G}_n^{-1}) - \mathbb{U}\|_\infty + \frac{1}{\sqrt{n}} \\
&\to_{a.s.} 0 + 0 + 0 = 0\,.
\end{aligned}$$

since $\mathbb{U}$ is a continuous (and hence uniformly continuous) function on $[0,1]$. $\square$

Theorem 7.2 Let $Q = F^{-1}$, and suppose that Q is differentiable at $0 < t_1 < \cdots < t_k < 1$. Then

$$(13)\qquad \begin{pmatrix} \sqrt{n}(\mathbb{F}_n^{-1}(t_1) - F^{-1}(t_k)) \\ \cdot \\ \cdot \\ \cdot \\ \sqrt{n}(\mathbb{F}_n^{-1}(t_k) - F^{-1}(t_k)) \end{pmatrix} \to_d \begin{pmatrix} Q'(t_1)\mathbb{V}(t_1) \\ \cdot \\ \cdot \\ \cdot \\ Q'(t_k)\mathbb{V}(t_k) \end{pmatrix} \sim N_k(0, \Sigma)$$

where

$$(14)\qquad \Sigma \equiv (\sigma_{ij}) = \big(Q'(t_i)Q'(t_j)(t_i \wedge t_j - t_i t_j)\big)\,.$$

Moreover, if Q' is nonzero and continuous on $[a,b] \subset [0,1]$, then for any $[c,d] \subset [a,b]$

$$(15)\qquad \|\sqrt{n}(F^{-1}(\mathbb{G}_n^{-1}) - F^{-1}) - Q'\mathbb{V}\|_c^d \to_{a.s.} 0 \qquad \text{as} \;\; n \to \infty\,.$$

Note that $Q'(t) = 1/f(F^{-1}(t))$.

Proof. Suppose that $k=1$ and let $t_1 = t$. Then

$$\begin{aligned}
\sqrt{n}(\mathbb{F}_n^{-1}(t) - F^{-1}(t)) &= \sqrt{n}(Q(\mathbb{G}_n^{-1}(t)) - Q(t)) \\
&= \frac{Q(\mathbb{G}_n^{-1}(t)) - Q(t)}{\mathbb{G}_n^{-1}(t) - t}\sqrt{n}(\mathbb{G}_n^{-1}(t) - t) \\
&\stackrel{d}{=} \frac{Q(\mathbb{G}_n^{-1}(t)) - Q(t)}{\mathbb{G}_n^{-1}(t) - t}\mathbb{V}_n(t) \qquad \text{for the special } \mathbb{V}_n \text{ process} \\
&\to_{a.s.} Q'(t)\mathbb{V}(t) \sim N(0, (Q'(t))^2 t(1-t)).
\end{aligned}$$

Similarly

$$\text{(a)} \qquad \begin{pmatrix} \sqrt{n}(\mathbb{F}_n^{-1}(t_1) - F^{-1}(t_k)) \\ \cdot \\ \cdot \\ \cdot \\ \sqrt{n}(\mathbb{F}_n^{-1}(t_k) - F^{-1}(t_k)) \end{pmatrix} \stackrel{d}{=} \sqrt{n}\begin{pmatrix} Q(\mathbb{G}_n^{-1}(t_1)) - Q(t_1) \\ \cdot \\ \cdot \\ \cdot \\ Q(\mathbb{G}_n^{-1}(t_1)) - Q(t_1) \end{pmatrix} \to_{a.s.} \begin{pmatrix} Q'(t_1)\mathbb{V}(t_1) \\ \cdot \\ \cdot \\ \cdot \\ Q'(t_k)\mathbb{V}(t_k) \end{pmatrix}.$$

□

Theorem 7.3 (Bahadur representation of quantile processes). The uniform quantile process can be written as

$$\text{(16)} \qquad \mathbb{V}_n = -\mathbb{U}_n + o_p(1)$$

where the $o_p(1)$ term is uniform in $0 \le t \le 1$; i.e.

$$\text{(17)} \qquad \|\mathbb{V}_n + \mathbb{U}_n\|_\infty \to_p 0.$$

In fact

$$\text{(18)} \qquad \limsup_{n\to\infty} \frac{n^{1/4}\|\mathbb{V}_n + \mathbb{U}_n\|_\infty}{\sqrt{b_n(\log n)}} = \frac{1}{\sqrt{2}} \qquad \text{a.s.}$$

where $b_n \equiv \sqrt{2\log\log n}$. Moreover, if $Q'(t)$ exists, then

$$\text{(19)} \qquad \sqrt{n}(\mathbb{F}_n^{-1}(t) - F^{-1}(t)) = -\,Q'(t)\sqrt{n}(\mathbb{F}_n(F^{-1}(t)) - t) + o_p(1).$$

Corollary 1 (Asymptotic normality of the $t-$th quantile). Suppose that $Q = F^{-1}$ is differentiable at $t \in (0,1)$. Then

$$\text{(20)} \qquad \sqrt{n}(\mathbb{F}_n^{-1}(t) - F^{-1}(t)) \to_d Q'(t)N(0, t(1-t)) = N\left(0, \frac{t(1-t)}{f^2(F^{-1}(t))}\right).$$

Corollary 2 (Asymptotic normality of a linear combination of order statistics). Suppose that J is bounded and continuous a.e. F^{-1}, and suppose that $E(X^2) < \infty$. Let

$$\text{(21)} \qquad T_n \equiv \frac{1}{n}\sum_{i=1}^n J\left(\frac{i}{n+1}\right)X_{(i)}, \qquad \mu = \int_0^1 J(u)F^{-1}(u)du.$$

Then

$$\text{(22)} \qquad \sqrt{n}(T_n - \mu) \to_d \int_0^1 J\mathbb{V}dF^{-1} \sim N(0, \sigma^2(J,F))$$

where

$$\text{(23)} \qquad \sigma^2(J,F) = \int_0^1\int_0^1 J(s)J(t)(s\wedge t - st)dF^{-1}(s)dF^{-1}(t).$$