EMPIRICAL PROCESSES

If $X_1, X_2, \ldots, X_n$ are independent identically distributed random variables (RVs) with values in a measurable space ($\mathbf{X}$, $\mathcal{B}$) and with common probability measure P on $\mathbf{X}$, the *empirical measure* or *empirical distribution* $\mathbb{P}_n$ of $(X_1, \ldots, X_n)$ is the measure which puts mass $1/n$ at each $X_i, i = 1, \ldots, n$:

$$\mathbb{P}_n = n^{-1}(\delta_{X_1} + \cdots + \delta_{X_n}), \tag{1}$$

where $\delta_x(A) = 1$ if $x \in A$, 0 if $x \notin A$, $A \in \mathcal{B}$. Thus $n\mathbb{P}_n(A)$ is simply the number of X_i's in A for any set $A \in \mathcal{B}$. The *empirical process* G_n, defined for each $n \geqslant 1$ by

$$G_n = n^{1/2}(\mathbb{P}_n - P), \tag{2}$$

may be viewed as a stochastic process* indexed (a) by some class of sets $\mathcal{C} \subset \mathcal{B}$,

$$G_n(C) = n^{1/2}(\mathbb{P}_n(C) - P(C)), \qquad C \in \mathcal{C}, \tag{3}$$

or (b) by some class of functions $\mathcal{F}$ from $\mathbf{X}$ to the real line R^1,

$$\begin{aligned} G_n(f) &= \int_{\mathbf{X}} f d\{n^{1/2}(\mathbb{P}_n - P)\} \\ &= n^{1/2} \int_{\mathbf{X}} f(x)\{\mathbb{P}_n(dx) - P(dx)\}, \\ &\qquad\qquad f \in \mathcal{F}. \end{aligned} \tag{4}$$

Frequently, in applications of interest the observations $X_1, \ldots, X_n$ are dependent, or nonidentically distributed, or perhaps both. In such cases we will continue to speak of the empirical measure $\mathbb{P}_n$ and empirical process G_n, perhaps with P replaced in (2) by an appropriate average measure.

In the classical case of real-valued random variables, $\mathcal{X} = R^1$, the class of sets $\mathcal{C} = \{(-\infty, x] : x \in R^1\}$ in (3), or the class $\mathcal{F} = \{1_{(-\infty,x]} : x \in R^1\}$ of indicator functions in (4) [where $1_A(x) = 1$ if $x \in A$, 0 if $x \notin A$], yields the usual *empirical distribution function* $\mathbf{F}_n$ given by

$$\begin{aligned} \mathbf{F}_n(x) &= \mathbb{P}_n(-\infty, x] \\ &= n^{-1}\{\text{number of } i \leqslant n \\ &\qquad \text{with } X_i \leqslant x\}, \end{aligned} \tag{5}$$

and the *empirical process*

$$G_n(x) = n^{1/2}(\mathbf{F}_n(x) - F(x)) \tag{6}$$

indexed by $x \in R^1$.

The subject of empirical processes is concerned with the large- and small-sample properties of the processes $\mathbb{P}_n$ and G_n, methods for studying these processes, and with the use of these properties and methods to treat systematically the extremely large number of statistics which may be viewed as functions of the empirical measure $\mathbb{P}_n$ or of the empirical process G_n. Much of the motivation for the study of $\mathbb{P}_n, G_n$, and functions thereof comes both historically and in current work from the desirability and attractiveness of nonparametric or distribution-free* statistical methods, methods which have proved to be of interest in a wide variety of problems, ranging from rank* and goodness-of-fit* tests, to density estimation*, clustering and classification*, and survival analysis*. The study of empirical processes also has strong connections with the related probabilistic topics of weak convergence and invariance principles*, as will be seen in the course of this article.

For any fixed set $C \in \mathcal{B}, n\mathbb{P}_n(C)$ is simply a binomial RV with parameters n and $P(C)$. Hence, by the classical weak law of large numbers, central limit theorem*, and law of the iterated logarithm*, respectively, as $n \to \infty$,

$$\mathbb{P}_n(C) \xrightarrow{\text{p}} P(C), \tag{7}$$

$$\begin{aligned} G_n(C) = n^{1/2}(\mathbb{P}_n(C) - P(C)) &\xrightarrow{\text{d}} G_P(C) \\ &\sim N(0, P(C)(1 - P(C))), \end{aligned} \tag{8}$$

and

$$\begin{aligned} &\limsup_{n\to\infty} \frac{|G_n(C)|}{(2\log\log n)^{1/2}} \\ &\quad = \limsup_{n\to\infty} \frac{n^{1/2}|\mathbb{P}_n(C) - P(C)|}{(2\log\log n)^{1/2}} \\ &\quad = [P(C)(1 - P(C))]^{1/2} \text{ a.s.} \end{aligned} \tag{9}$$

(where we write "$\sim$" for "is distributed as", $N(\mu, \sigma^2)$ denotes the "normal" or Gaussian distribution with mean μ and variance σ^2, "$\xrightarrow{p}$" denotes convergence in probability, and "$\xrightarrow{d}$" denotes convergence in distribution or law; *see* CONVERGENCE OF SEQUENCES OF RANDOM VARIABLES). A large part of the theory of empirical processes is concerned with strengthened versions of (7) to (9), versions of these convergence results that hold simultaneously (i.e., uniformly) for all sets C in some collection $\mathscr{C}$. *See* GLIVENKO–CANTELLI THEOREMS for such uniform extensions of (7) and applications thereof. We will concentrate in this article on various uniform analogs of (8) (sometimes called Donsker theorems or functional central limit theorems), and of (9) (which we call Strassen-Finkelstein log-log laws or functional laws of the iterated logarithm). In the same way that Glivenko-Cantelli theorems serve as tools for establishing the consistency of various estimators or statistics, uniform versions of (8) and (9) serve as tools for establishing convergence in distribution (often asymptotic normality*), or laws of the iterated logarithm, respectively, for those estimators and statistics expressible in terms of $\mathbb{P}_n$ or G_n. We have chosen to concentrate this exposition on the large-sample theory* of empirical processes since so little is known concerning finite sample sizes beyond the classical one-dimensional case of real-valued RVs; for useful summaries of finite-sample size results in the one-dimensional case, see Durbin [1], Niederhausen [4] and the references given there.

The article has been divided into the following four sections:

1. The one-dimensional case
2. More general sample spaces and index sets
3. Dependent or nonidentically distributed observations
4. Miscellaneous topics

Many topics have been omitted or are mentioned only briefly. For a nice survey of earlier work and a helpful exposition of weak convergence issues, see Pyke [3]. For a recent comprehensive review of the i.i.d. case, see Gaenssler and Stute [2].

ONE-DIMENSIONAL CASE

Here we focus on the classical empirical distribution function $\mathbf{F}_n$ and empirical process G_n of real-valued random variables given in (5) and (6). In this one-dimensional situation, a significant further simplification is possible by use of the fundamental transformations of nonparametric statistics (the probability integral transformation* and the inverse probability integral transformation or quantile transformation):

$$\text{If } F \text{ is continuous and } X \sim F, \text{ then } F(X) \sim \text{ uniform}(0,1); \tag{10}$$

$$\text{If } U \sim \text{ uniform } (0,1), \text{ then } F^{-1}(U) \equiv X \sim F \text{ for an arbitrary df } F, \tag{11}$$

where F^{-1} is the left-continuous inverse of $F, F^{-1}(u) = \inf\{x : F(x) \geqslant u\}$. Thus, letting $U_1, U_2, \ldots, U_n$ be independent, identically distributed (i.i.d.) uniform (0, 1) RVs with distribution function $I(t) = t$ on [0, 1], empirical distribution function Γ_n, and corresponding uniform empirical process $\mathbf{U}_n$ defined by

$$\mathbf{U}_n(t) = n^{1/2}(\Gamma_n(t) - t), \quad 0 \leqslant t \leqslant 1,$$

it follows from (10) and (11) that

$$G_n \circ F^{-1} \stackrel{d}{=} \mathbf{U}_n \quad \text{if } F \text{ is continuous}, \tag{12}$$

and

$$G_n \stackrel{d}{=} \mathbf{U}_n \circ F \quad \text{for arbitrary} F, \tag{13}$$

where "$\stackrel{d}{=}$" means equal in distribution or law (so that the two processes are probabilistically equivalent), and "$\circ$" denotes functional composition, $f \circ g(t) = f(g(t))$. By virtue of (12) and (13), we may restrict attention to the uniform empirical process $\mathbf{U}_n$ throughout most of the remainder of this section.

The random function Γ_n is a nondecreasing, right-continuous step function equal to 0 at $t = 0$ and 1 at $t = 1$, which increases by jumps of size $1/n$ at the *order statistics** $0 \leqslant U_{n:1} \leqslant \cdots \leqslant U_{n:n} \leqslant 1$. The random function or process $\mathbf{U}_n$ equals 0 at both $t = 0$ and 1, decreases linearly between successive order statistics with slope $-n^{1/2}$, and jumps upward at the order statistics with jumps of size $n^{-1/2}$. Both Γ_n and $\mathbf{U}_n$ take values in $D = D[0,1]$, the set of functions on [0, 1] which are right continuous and have left limits.

Donsker's Theorem; Weak Convergence of $\mathbf{U}_N$

Convergence in distribution of specific functions of the process $\mathbf{U}_n$ was first treated by Cramér [16], von Mises [65], Kolmogorov [39], and Smirnov [61,62] in the course of investigations of the now well-known Cramér-von Mises* and Kolmogorov goodness-of-fit statistics. A general unified approach to the large-sample theory of statistics such as these did not emerge until Doob [28] gave his heuristic approach to the Kolmogorov-Smirnov limit theorems. Doob's approach was to note that (a) $\mathbf{U}_n$ is a zero-mean stochastic process on [0, 1] with covariance function

$$\operatorname{cov}[\mathbf{U}_n(s), \mathbf{U}_n(t)] = \min(s,t) - st \quad \text{for all } 0 \leqslant s,t \leqslant 1; \tag{14}$$

(b) by a simple application of the multivariate central limit theorem*, all the finite-dimensional joint df's of $\mathbf{U}_n$ converge to the corresponding normal df's which are the finite-dimensional joint df's of a mean-zero Gaussian process* $\mathbf{U}$ on [0, 1] with covariance as in (14), called a *Brownian Bridge process*; and (c) hence, for any real-valued "continuous" function g of $\mathbf{U}_n$ it should follow that

$$g(\mathbf{U}_n) \xrightarrow{\mathrm{d}} g(\mathbf{U}) \quad \text{as } n \to \infty \tag{15}$$

in the ordinary sense of convergence in distribution of RVs. For example, for the Kolmogorov statistic $g(u) = \sup_{0 \leqslant t \leqslant 1} |u(t)| \equiv ||u||$, and Doob [28] showed that the limiting distribution of $g(\mathbf{U}_n) = ||\mathbf{U}_n|| = n^{1/2}||\Gamma_n - I||$, obtained earlier by Kolmogorov [39], is exactly that of $g(\mathbf{U}) = ||\mathbf{U}||$.

A precise formulation of Doob's heuristic approach requires a careful definition of the idea of *weak convergence* of a sequence of stochastic processes*, a notion which extends the more familiar concept of convergence in distribution of random variables or random vector. Donsker [27] succeeded in justifying Doob's [28] heuristic approach, and this in combination with related work on invariance principles by Erdös and Kac [32,33], Donsker [26], and others led to the development of a general theory of *weak convergence* of stochastic processes (and their associated probability laws) by Prohorov [44] and Skorokhod [59]. This theory has been very clearly presented and further developed in an exemplary monograph by Billingsley [5]; see also Billingsley [6].

Unfortunately for the theory of empirical processes, the space $D = D[0,1]$ in which $\mathbf{U}_n$ takes its values is inseparable when considered as a metric space with the supremum or uniform metric $||\cdot||$ (i.e., $||f - g|| \equiv \sup_{0 \leqslant t \leqslant 1} |f(t) - g(t)|$), as pointed out by Chibisov [14]; see Billingsley [5, Sec. 18]. This lack of separability creates certain technical difficulties in the weak convergence theory of $\mathbf{U}_n$ and has led to a number of different approaches to the study of its weak convergence: Skorokhod [59] introduced a metric d with which D becomes separable (see Billingsley [5, Sec. 14]), while Dudley [69], Pyke and Shorack [48], and Pyke [47] give different definitions of weak convergence. These difficulties are largely technical in nature, however. Here we follow Pyke and Shorack [48] and Pyke [47] and say that $\mathbf{U}_n \Rightarrow \mathbf{U}$ ("$\mathbf{U}_n$ converges weakly to $\mathbf{U}$") if $g(\mathbf{U}_n) \xrightarrow{\mathrm{d}} g(\mathbf{U})$ for all $||\cdot||$-continuous real-valued functions g of $\mathbf{U}_n$ for which $g(\mathbf{U}_n)(n \geqslant 1)$ and $g(\mathbf{U})$ are (measurable) RVs. With this definition we have:

Theorem 1. (Donsker [27]). $\mathbf{U}_n \Rightarrow \mathbf{U}$ on $(D, ||\cdot||)$.

The importance of Theorem 1 for applications in statistics is that the limiting distribution of any statistic that can be expressed as $g(\mathbf{U}_n)$ for some $||\cdot||$-continuous measurable function g is that of $g(\mathbf{U})$. For example:

$$||\mathbf{U}_n|| \xrightarrow{\mathrm{d}} ||\mathbf{U}|| \qquad [g(u) = ||u||]; \tag{16}$$

$$\int_0^1 [\mathbf{U}_n(t)]^2 dt \xrightarrow{\mathrm{d}} \int_0^1 [\mathbf{U}(t)]^2 dt \quad \left[g(u) = \int_0^1 (u(t))^2 dt \right]; \tag{17}$$

and

$$n^{1/2}\left(\overline{U}_n - \frac{1}{2}\right) = -\int_0^1 \mathbf{U}_n(t)dt \xrightarrow{\mathrm{d}} -\int_0^1 \mathbf{U}(t)dt \sim N(0, \tfrac{1}{12})$$
$$\left[g(u) = -\int_0^1 u(t)dt\right] \quad (18)$$

as $n \to \infty$. Of course, the distribution function of the RV $g(\mathbf{U})$ must be computed in order to complete the program. For linear functions of $\mathbf{U}_n$ [as in (18)], and hence of $\mathbf{U}$, this is easy: under appropriate integrability conditions, linear functions of the Gaussian process $\mathbf{U}$ have normal distributions with easily computed variances. In general, evaluation of the distribution of $g(\mathbf{U})$ is not an easy task, but well-developed tools are available for quadratic and supremum-type functionals as illustrated by Doob [28] and Darling [24]; see also Sähler [51] and Durbin [30]. More sophisticated applications making essential use of the identities (12) and (13) may be found, for example, in Pyke and Shorack [48] (rank statistics); Shorack [53,54] (linear combinations of order statistics, quantile*, and spacings processes); and Bolthausen [7], Pollard [43], and Boos [8] (minimum distance estimators* and tests). The weak convergence approach, in combination with the device of almost surely convergent versions of weakly convergent processes (to be discussed in the section "Almost Surely Convergent Constructions; Strong Approximations"), has become a key tool in the modern statistical workshop.

Iterated Logarithm Laws

Following a pattern similar to that outlined above, iterated logarithm laws for specified functions of the process $\mathbf{U}_n$ were established by Smirnov [62], Chung [15], Cassels [12], and others in connection with investigations of particular goodness-of-fit statistics, especially $||\mathbf{U}_n||$ and $||\mathbf{U}_n^+||$:

$$\limsup_{n\to\infty} \frac{||\mathbf{U}_n||}{(2\log\log n)^{1/2}} = \limsup_{n\to\infty} \frac{n^{1/2}||\Gamma_n - I||}{(2\log\log n)^{1/2}} = \frac{1}{2} \text{ a.s.} \quad (19)$$

A general law of the iterated logarithm result for the uniform empirical process $\mathbf{U}_n$ comparable to Donsker's theorem emerged in the light of work on almost surely convergent constructions or embeddings of partial sum processes in Brownian motion by Skorokhod [60] and the "invariance principle for the law of the iterated logarithm" by Strassen [63]. Let $\mathcal{U}^0$ be the set of functions on [0, 1] which are absolutely continuous with respect to Lebesgue measure, equal to 0 at 0 and 1, and whose derivatives have L_2-norm no larger than 1; alternatively, $\mathcal{U}^0$ is simply the unit ball of the reproducing kernel Hilbert space with kernel given by the covariance function (14) of the Brownian bridge process $\mathbf{U}$.

Theorem 2. (Finkelstein [34]). With probability 1 every subsequence of

$$\left\{\frac{\mathbf{U}_n}{(2\log\log n)^{1/2}} : n \geqslant 3\right\}$$

has a uniformly convergent subsequence, and the set of limit functions is precisely $\mathcal{U}^0$.

In a way completely parallel to the applications of Donsker's theorem given in the preceding section, Finkelstein's theorem yields laws of the iterated logarithm for $||\cdot||$-continuous functions g of $\mathbf{U}_n/b_n$, where $b_n = (2\log\log n)^{1/2}$:

$$\limsup_{n\to\infty} g(\mathbf{U}_n/b_n) = \sup\{g(u) : u \in \mathcal{U}^0\} \text{ a.s.,} \quad (20)$$

where the problem of evaluating the supremum on the right side for specific functions g may be thought of as analogous to the problem of finding the distribution of $g(\mathbf{U})$ in the case of weak convergence. For example, in parallel to (16) to (18), Finkelstein's theorem yields

$$\limsup_{n\to\infty} ||\mathbf{U}_n/b_n|| = \sup\{||u|| : u \in \mathcal{U}^0\} = \frac{1}{2} \text{ a.s.} \quad (21)$$

with equality when $u(t) = \min(t, 1 - t)$, $0 \leqslant t \leqslant 1$;

$$\limsup_{n\to\infty} \int_0^1 [\mathbf{U}_n(t)/b_n]^2 dt = \sup\left\{\int_0^1 (u(t))^2 dt : u \in \mathcal{U}^0\right\} = 1/\pi^2 \text{ a.s.} \quad (22)$$

with equality when $u(t) = (2^{1/2}/\pi)\sin(\pi t)$, $0 \leqslant t \leqslant 1$; and

$$\limsup_{n\to\infty} \int_0^1 (\mathbf{U}_n(t)/b_n) dt = \sup\left\{\int_0^1 u(t) dt : u \in \mathcal{U}^0\right\} = 1/\sqrt{12}^{1/2} \text{ a.s.} \quad (23)$$

with equality when $u(t) = \sqrt{12}\int_0^1 (\min(s, t) - st) ds = \sqrt{3}t(1 - t)$.

Almost Surely Convergent Constructions; Strong Approximations

In Skorokhod's [59] paper the basis for a different and very fruitful approach to weak convergence was already in evidence: that of replacing weak convergence by almost sure (a.s.) convergence. *See also* HUNGARIAN CONSTRUCTIONS OF EMPIRICAL PROCESSES.

Theorem 3. (Skorokhod, Dudley, Wichura). If the processes $\{\mathbb{Z}_n, n \geqslant 0\}$ take values in a metric space $(\mathcal{M}, m)$ and $\mathbb{Z}_n \Rightarrow \mathbb{Z}_0$, then there exists a probability space $(\Omega, \mathcal{A}, P)$ and processes $\{\mathbb{Z}_n^*, n \geqslant 0\}$ defined there such that $\mathbb{Z}_n \stackrel{\mathrm{d}}{=} \mathbb{Z}_n^*$ for all $n \geqslant 0$ and $m(\mathbb{Z}_n^*, \mathbb{Z}_0^*) \to 0$ a.s. as $n \to \infty$.

Skorokhod [59] gave the first version of this result in the case that $(\mathcal{M}, m)$ is complete and separable; Dudley [29] and Wichura [66] proved that the hypotheses of completeness and separability, respectively, could be dropped. See also Billingsley [6, p. 7]. Although the theorem does not tell how to construct the special almost surely convergent $\mathbb{Z}_n^*$ processes , it provides an extremely valuable conceptual tool. For example, in the case of the uniform empirical processes $\{\mathbf{U}_n\}$, the theorem yields the existence of probabilistically equivalent processes $\mathbf{U}_n^*, n \geqslant 1$, and a Brownian bridge process $\mathbf{U}^*$ all defined on a common probability space $(\Omega, \mathcal{A}, P)$ such that for each fixed $\omega \in \Omega$ the sequence of functions $\mathbf{U}_n^* = \mathbf{U}_n^*(\cdot, \omega)$ converge uniformly to the continuous function $\mathbf{U}^* = \mathbf{U}^*(\cdot, \omega)$ on [0, 1] as $n \to \infty$; that is,

$$||\mathbf{U}_n^* - \mathbf{U}^*|| \to 0 \text{ a.s.} \quad \text{as } n \to \infty. \quad (24)$$

The extreme usefulness of this point of view in dealing with weak convergence problems in statistics was recognized, explained, and advocated by Pyke [45,46] and has been effectively used to deal with a variety of problems involving two-sample rank statistics, linear combinations of order statistics (*see* L-STATISTICS), spacings*, minimum-distance estimators, and censored data* (the "product limit" estimator) by Pyke and Shorack [48], Pyke [46], Shorack [53–55], and Breslow and Crowley [10], to name only a few outstanding examples. See Pyke [45–47] for excellent expositions of this approach.

When the metric space $\mathcal{M}$ is the real line R^1, so that the $\mathbb{Z}_n$'s are just real-valued RVs, a very explicit construction of the $\mathbb{Z}_n^*$'s is possible using (11): If $F_n(z) \equiv P(\mathbb{Z}_n \leqslant z), n \geqslant 0$, then $F_n \stackrel{\mathrm{d}}{\to} F_0$ implies that $F_n^{-1}(u) \to F_0^{-1}(u)$ for almost all $u \in [0, 1]$ with respect to Lebesgue measure. Hence, if U is a uniform(0, 1) RV, then

$$\mathbb{Z}_n^* \equiv F_n^{-1}(U) \stackrel{\mathrm{d}}{=} \mathbb{Z}_n$$

by (11), and

$$\mathbb{Z}_n^* \equiv F_n^{-1}(U) \to F_0^{-1}(U) \equiv \mathbb{Z}_0^* \text{ a.s.}$$

as $n \to \infty$ [45].

The possibility of giving explicit, concrete constructions of the almost surely convergent versions $\mathbf{U}_n^*$ of the uniform empirical process $\mathbf{U}_n$ began to become apparent soon after the appearance of Skorokhod [60] (available in English translation in 1965), and Strassen [63] concerning the embedding of partial sum processes in Brownian motion*. This idea, together with a representation of uniform order statistics as ratios of partial sums of independent exponential RVs, was used by several authors, including Breiman [9], Brillinger [11], and Root [49], to give explicit constructions of the a.s. convergent processes $\{\mathbf{U}_n^*\}$ guaranteed to exist by Skorokhod's [59] theorem. The "closeness" or "rate" of this approximation was studied by Kiefer [37], Rosenkrantz [50], and Sawyer [52].

The method of Skorokhod embedding gave a relatively straightforward and clear construction of versions $\{\mathbf{U}_n^*\}$ converging almost surely. It quickly became apparent, however, that for some purposes the constructed versions $\{\mathbf{U}_n^*\}$ from this embedding method suffered from two inadequacies or deficiencies: The joint distributions of $\mathbf{U}_n^*$ and $\mathbf{U}_{n+1}^*$ (i.e., in n) were not correct; and rates of convergence for specific functions of the $\mathbf{U}_n$ process yielded by the construction were substantially less than those obtainable by other (direct, special) methods ($n^{-1/4}$ or less, rather than $n^{-1/2}$ or a little less). These difficulties were overcome by a Hungarian group of probabilists and statisticians in a remarkable series of papers published in 1975: see Csörgö and Révész [21] and Komlós et al. [40]. By combining the quantile or inverse probability integral transform (11), an ingenious dyadic approximation scheme, and careful analysis, the Hungarian construction* yields uniform empirical processes $\{\mathbf{U}_n^*\}$ which have the correct distributions jointly in n and which are (at least very nearly) as close as possible to a sequence $\{\mathbb{B}_n^*\}$ of Brownian bridge processes (each $B_n^* \stackrel{d}{=} \mathbf{U} =$ Brownian bridge) with the correct joint distributions in n dimensions:

Theorem 4. (Komlós et al. [40]). For the Hungarian construction $\{\mathbf{U}_n^*\}$ of the uniform empirical processes

$$||\mathbf{U}_n^* - \mathbb{B}_n^*|| = O(n^{-1/2}(\log n)^2) \text{ a.s. as } n \to \infty; \tag{25}$$

that is, there exists a positive constant $M < \infty$ such that

$$\limsup_{n\to\infty} \frac{n^{1/2}}{(\log n)^2} ||\mathbf{U}_n^* - \mathbb{B}_n^*|| \leqslant M \text{ a.s.} \tag{26}$$

Thus the supremum distance between the constructed uniform empirical processes $\mathbf{U}_n^*$ and the sequence of Brownian bridge processes $\mathbb{B}_n^*$ goes to zero only a little more slowly than $n^{-1/2}$ as $n \to \infty$. This fundamental strong approximation theorem has already proved itself to be of basic importance in a wide and growing range of problems, and has already been generalized and extended in several directions, some of which will be mentioned briefly. The monograph by Csörgö and Révész [23] contains an exposition and a variety of applications.

Other Limit Theorems for U_n

Weighted Metrics . The empirical process $\mathbf{U}_n$ is small near 0 and 1: $\mathbf{U}_n(0) = \mathbf{U}_n(1) = 0$ and $\text{var}[\mathbf{U}_n(t)] = t(1-t)$ for $0 \leqslant t \leqslant 1$. This has led to the introduction of various "weighted" metrics to account for and exploit the small values of $\mathbf{U}_n$ near 0 and 1. Supremum-type weighted metrics $||\cdot/q||$ (defined by

$$||(f-g)/q|| \equiv \sup_{0\leqslant t\leqslant 1} |(f(t)-g(t))/q(t)|$$

were first introduced by Chibisov [13], who gave conditions for $\mathbf{U}_n \Rightarrow \mathbf{U}$ with respect to $||\cdot/q||$: Essentially, q must satisfy

$$q^2(t)/[t(1-t)\log\log(\{t(1-t)\}^{-1})] \to \infty$$

as $t \to 0$ or 1 [e.g., $q(t) = [t(1-t)]^{1/2-\delta}$ with $0 < \delta < \frac{1}{2}$]. This convergence, which strengthens Donsker's theorem, was further investigated by Pyke and Shorack [48], O'Reilly [42], Shorack [56], and Shorack and Wellner [58], and successfully applied to statistical problems by Pyke and Shorack [48], Shorack [53,54], and subsequently by others. James [36] has established a corresponding weighted version of the Strassen-Finkelstein functional law of the iterated logarithm for $\mathbf{U}_n/b_n$.

The standardized or normalized empirical process $Z_n(t) = \mathbf{U}_n(t)/\sqrt{t(1-t)}, 0 < t < 1$, has also been thoroughly investigated, largely because it has the appealing feature of having $\text{var}[Z_n(t)] = 1$ for all $0 < t < 1$ and every $n \geqslant 1$. The limit theory of Z_n turns out to be closely linked to the Ornstein-Uhlenbeck process* and the classical work of Darling and Erdös [25] on normalized sums: see Jaeschke [35] and Eicker [31] for distributional limit theorems; and Csáki [19,20], Shorack [57], and Mason [41] for iterated logarithm-type results.

Oscillations of $\mathbf{U}_n$. The oscillation modulus $\omega_n(a)$ of $\mathbf{U}_n$, defined by

$$\omega_n(a) = \sup_{|t-s|\leqslant a} |\mathbf{U}_n(t) - \mathbf{U}_n(s)|$$
$$= n^{1/2} \sup_{|t-s|\leqslant a} |\Gamma_n(t) - \Gamma_n(s) - (t-s)|, \quad 0 < a \leqslant 1, \tag{27}$$

arises naturally in many statistical problems including tests for "bumps" of probability and density estimation*. Note that $\omega_n(1)$ is the classical Kuiper goodness-of-fit* statistic (see, e.g., Durbin [30, p. 33]). Cassels [12] established laws of the iterated logarithm for $\omega_n(a)$; Cressie [17,18] has investigated the limiting distribution of $\omega_n(a)$ for fixed a (which, by Donsker's theorem, is that of $\omega(a) \equiv \sup_{|t-s|\leqslant a} |\mathbf{U}(t) - \mathbf{U}(s)|$); and Stute [64] proved that if $a = a_n \to 0, a_n = n^{-\lambda}, 0 < \lambda < 1$, then

$$\lim_{n\to\infty} \frac{\omega_n(a_n)}{\{2a_n \log(1/a_n)\}^{1/2}} = 1 \text{ a.s.} \tag{28}$$

Stute [64] has exploited this result to obtain several interesting limit theorems for kernel estimates* of density functions. Shorack and Wellner [58] study related oscillation moduli and give weighted-metric convergence theorems related to those of Chibisov [13] and O'Reilly [42].

Quantile Processes . An important process closely related to the uniform empirical process $\mathbf{U}_n$ is the uniform quantile process* V_n defined on [0, 1] by

$$V_n(t) = n^{1/2}(\Gamma_n^{-1}(t) - t), \qquad 0 \leqslant t \leqslant 1, \tag{29}$$

where Γ_n^{-1} is the left-continuous inverse of Γ_n. Γ_n^{-1} and V_n are important for problems involving order statistics since $\Gamma_n^{-1}(i/n) = U_{n:i}$, the ith order statistic of the sample $U_1, \ldots, U_n$ of n i.i.d. uniform (0,1) RVs. There are many relationships between the processes $\mathbf{U}_n$ and V_n, such as the identity

$$V_n = -\mathbf{U}_n \circ \Gamma_n^{-1} + n^{1/2}(\Gamma_n \circ \Gamma_n^{-1} - I), \tag{30}$$

which shows that $V_n \Rightarrow V \equiv -\mathbf{U}$, because $\Gamma_n^{-1}(t) \to t$ uniformly in t a.s. and the second term in (30) has supremum norm equal to $n^{-1/2}$; a corresponding functional law of the iterated logarithm for V_n/b_n follows similarly.

For a sample from a general df F on R^1, the *quantile process* $\mathbf{Q}_n$ is defined by

$$\mathbf{Q}_n(t) = n^{1/2}(\mathbf{F}_n^{-1}(t) - F^{-1}(t)), \tag{31}$$

where $\mathbf{F}_n^{-1}$ denotes the left continuous inverse of the empirical df $\mathbf{F}_n$. By the inverse probability integral transform (11),

$$\mathbf{Q}_n(t) \overset{\mathrm{d}}{=} n^{1/2}(F^{-1} \circ \Gamma_n^{-1} - F^{-1}) = R_n \cdot V_n, \tag{32}$$

where the random difference quotient $R_n \equiv (F^{-1}(\Gamma_n^{-1}) - F^{-1})/(\Gamma_n^{-1} - I)$ can be shown to converge (under appropriate differentiability hypotheses on F^{-1}) to $dF^{-1}/dt = 1/(f \circ F^{-1})$. Thus, at least roughly,

$$\mathbf{Q}_n \Rightarrow \frac{1}{f \circ F^{-1}} V = \frac{-1}{f \circ F^{-1}} \mathbf{U}. \tag{33}$$

For precise formulations of this type of limit theorem, see Shorack [54] and Csörgö and Révész [22], who make use of strong approximation methods together with the deep theorems of Kiefer [38] concerning the process $\mathbf{D}_n = \mathbf{U}_n + V_n$.

MORE GENERAL SAMPLE SPACES AND INDEX SETS

Spurred by questions in many different areas of statistics, the theory of empirical processes has undergone rapid development. The basic theorems of Donsker and Strassen-Finkelstein in one dimension have been generalized to observations X with values in higher-dimensional Euclidean spaces R^k or more general sample spaces; to indexing by classes of sets or functions, and to observations which are dependent or nonidentically distributed. We focus on i.i.d. RVs in higher-dimensional spaces and indexing of these processes by sets and functions; dependent or nonidentically distributed RVs will be discussed in the following section.

A General "Donsker Theorem"

Now, as in the introduction, suppose that $X_1, X_2, \ldots, X_n$ are i.i.d. RVs with values in the measurable space $(\mathcal{X}, \mathcal{B})$, and consider the empirical measures $\mathbb{P}_n$ and empirical process G_n as processes "indexed" by sets C in some class of sets $\mathcal{C} \subset \mathcal{B}$. It turns out that the $\mathcal{C}$-empirical process $\{G_n(C) : C \in \mathcal{C}\}$ will converge weakly only if the class of sets $\mathcal{C}$ is not "too large." The most complete results to date are those of Dudley [70].

Theorem 5. (Dudley [70]). Under measurability and entropy* conditions (satisfied if $\mathcal{C}$ is not "too large"),

$$G_n \Rightarrow G_p \qquad \text{as } n \to \infty,$$

where G_P is a zero-mean Gaussian process indexed by sets $C \in \mathcal{C}$ with continuous sample functions and covariance

$$\begin{aligned} &\operatorname{cov}[G_P(A), G_P(B)] \\ &\quad = P(A \cap B) - P(A)P(B) \\ &\qquad\qquad\qquad \text{for all } A, B \in \mathcal{C}. \end{aligned}$$

This theorem generalizes and contains as special cases earlier results by Dudley [69], Bickel and Wichura [67], Neuhaus [77], and Straf [85] (all of which dealt with the case $\mathcal{X} = R^k$ and the class $\mathcal{C}$ of lower-left orthants, which yield the usual k-dimensional df $F(x)$ and empirical df $\mathbf{F}_n(x), x \in R^k$) as well as more recent results for convex sets due to Bolthausen [68]. Dudley's results have been used by Pollard [79] to treat chi-square goodness-of-fit* tests with data dependent cells.

If the empirical process G_n is considered as a process indexed by functions f in some class $\mathcal{F}, \{G_n(f) : f \in \mathcal{F}\}$, then a "Donsker theorem" will hold if the class $\mathcal{F}$ is not "too large." Roughly speaking, all the functions f in $\mathcal{F}$ must be sufficiently smooth and square integrable (with respect to P). Such a theorem under metric entropy conditions on the class $\mathcal{F}$ was first given by Strassen and Dudley [86] for the case when the sample space $\mathcal{X}$ is a compact metric space such as $[0, 1] \subset R^1$, or $[0, 1]^k \subset R^k$. In the case $\mathcal{X} = [0, 1]$, the weak convergence of G_n to $G = \{G(f) : f \in \mathcal{F}\}$ holds if the class $\mathcal{F}$ is any of the classes of Lipschitz functions

$$\begin{aligned} \mathcal{F}_\alpha = \{f : |f(x) - f(y)| \leqslant |x - y|^\alpha \\ \text{for all } x, y \in [0, 1]\} \end{aligned}$$

with $\alpha > \frac{1}{2}$; if $\alpha = \frac{1}{2}$, the convergence fails (there are "too many" functions in the class $\mathcal{F}_{1/2}$). Very recently similar (but more difficult) results have been given by Dudley [71] and Pollard [81] without the restriction to compact metric sample spaces $\mathcal{X}$.

Several applications of the properties of empirical processes indexed by functions to problems in statistics have been made: Giné [72] and Wellner [87] use such processes to study test statistics of interest for directional data*; Pollard [80] uses his Donsker theorem to give a central limit theorem for the cluster centers of a clustering method studied earlier in R^1 by Hartigan [73].

General Law of the Iterated Logarithm

In the same way that Dudley's weak convergence theorem in the preceding section generalizes Donsker's theorem, a law of the iterated logarithm for the $\mathcal{C}$-empirical process which generalizes the Strassen-Finkelstein theorem has been proved by Kuelbs and Dudley [76]. We introduce the sets of functions

$$\begin{aligned} \mathcal{H}^0 = \{h \in L^2(\mathcal{X}, \mathcal{B}, P) : \int h\,dP = 0 \\ \text{and } \int |h|^2 dP \leqslant 1\}, \\ \mathscr{G}^0_{\mathcal{C}} = \{g : \mathcal{C} \to R \text{ defined by } g(C) \\ = \int_C h\,dP, C \in \mathcal{C}; h \in \mathcal{H}^0\}; \end{aligned}$$

$\mathscr{G}^0_{\mathcal{C}}$ is the appropriate analog for the $\mathcal{C}$-empirical process of the set of functions $\mathcal{U}^0$ which arose in the Strassen-Finkelstein theorem.

Theorem 6. (Kuelbs and Dudley [76]). Under the same measurability and entropy conditions as required for weak convergence of the $\mathcal{C}$-empirical process (satisfied if $\mathcal{C}$ is not "too large"), with probability 1 every subsequence of

$$\left\{ \frac{G_n}{(2 \log \log n)^{1/2}} : n \geqslant 3 \right\}$$

restricted to $C \in \mathcal{C}$ has a uniformly convergent subsequence, and the set of limit functions is precisely $\mathscr{I}_{\mathcal{C}}^0$.

This theorem has consequences analogous to those of the Strassen-Finkelstein theorem, and generalizes earlier results for special sample spaces $\mathcal{X}$ and classes $\mathcal{C}$ due to Kiefer [75], Révész [82], Richter [84], and Wichura [88]; it contains the Strassen-Finkelstein theorem as a special case ($\mathcal{X} = [0,1], \mathcal{C} = \{[0,t] : 0 \leqslant t \leqslant 1\}$).

For the $\mathcal{F}$-empirical process (indexed by functions f in some collection $\mathcal{F}$), only partial results are available (see, e.g., Kaufman and Philipp [74]). However, if $\mathcal{F}$ is a class of functions satisfying the hypotheses of Dudley [17] or Pollard [81] sufficient for weak convergence, the following iterated logarithm law should hold: With probability 1 every subsequence of

$$\left\{\frac{G_n(f)}{(2\log\log n)^{1/2}} : n \geqslant 3, f \in \mathcal{F}\right\}$$

has a uniformly (in $f \in \mathcal{F}$) convergent subsequence, and the set of limit functions is

$$\mathscr{I}_{\mathcal{F}}^0 = \{g : \mathcal{F} \to R^1 \text{ defined by } g(f) = \int fh dP, f \in \mathcal{F}; h \in \mathcal{H}^0\}.$$

Almost Surely Convergent Versions; Strong Approximations

In higher-dimensional situations the Skorokhod-Dudley-Wichura theorem continues to guarantee the existence of almost surely convergent versions G_n^* of the empirical process G_n, and this again provides an extremely useful way to treat statistics representable as functions of $\mathbb{P}_n$ and G_n.

Concerning explicit strong approximations much less is known, the best results being those of Philipp and Pinzur [78] (for the case $\mathcal{X} = R^k$, general P, and $\mathscr{C} =$ the lower left orthants) and Révész [82,83] ($\mathcal{X} = [0,1]^k$, P uniform on $[0,1]^k$, and $\mathscr{C} =$ a class of sets with smooth boundaries).

DEPENDENT OR NONIDENTICALLY DISTRIBUTED OBSERVATIONS

In many cases of practical importance the observations are either nonidentically distributed, or dependent, or both. In comparison to the i.i.d. case treated in the preceding sections, present knowledge of the empirical measures $\mathbb{P}_n$ and corresponding empirical processes G_n is much less complete in these cases. A variety of results are available, however, for the most important case of $\mathcal{X} = R^k$ and $\mathscr{C} = \{(-\infty, x] : x \in R^k\}$, the lower-left orthants.

Independent, Nonidentically Distributed Observations

When the observations $X_1, \ldots, X_n$ have distributions $P_1, \ldots, P_n$ on $\mathcal{X}$, the natural empirical process to consider is

$$G_n = n^{1/2}(\mathbb{P}_n - \overline{P}_n),$$
$$\overline{P}_n = n^{-1}(P_1 + \cdots + P_n).$$

In the case $\mathcal{X} = R^1$ and $\mathscr{C} = \{(-\infty, x] : x \in R^1\}$, sufficient conditions for weak convergence of ("reduced versions" of) G_n have been given by Koul [103], Shorack [121,122], and Withers [128]. These authors also study the "weighted" or "regression" processes $\mathbf{W}_n = \sum_{i=1}^n c_{ni}(\delta_{X_i} - P_i)/(\sum_i^n c_{ni}^2)^{1/2}$, where the c_{ni}'s are appropriate (regression) constants (see also Hájek [102]); Shorack [122] gives convergence with respect to weighted metrics and convergence theorems for the related quantile processes; Withers [128] allows the observations to be dependent (strong mixing). Interesting inequalities for the limiting distributions of supremum functionals of the process are given by Sen et al. [120] and Rechtschaffen [115]; van Zuijlen [124,125] gives linear bounds and many useful inequalities.

In the case $\mathcal{X} = R^k$ and $\mathscr{C} =$ the lower-left orthants, conditions ensuring weak convergence of ("reduced" versions of) G_n have been given by Neuhaus [108] and Rüschendorf [119]. Many of the weak convergence theorems above are (naturally) formulated for triangular arrays of RVs with independent RVs in each row.

Although little is known about functional laws of the iterated logarithm analogous to the Strassen-Finkelstein theorem for independent nonidentically distributed observations, a recent inequality due to Bretagnolle [91] makes possible the following extension of the Chung-Smirnov law of the iterated logarithm in the case $\mathcal{X} = R^1, \mathscr{C} = \mathcal{O}_1$, and the observations form a *single* independent sequence. Let $\mathbf{F}_n(x) = \mathbb{P}_n(-\infty, x], \overline{F}_n(x) = \overline{P}_n(-\infty, x]$, and $G_n(x) = n^{1/2}(\mathbf{F}_n(x) - \overline{F}_n(x))$ for $x \in R^1$, so $||G_n|| = n^{1/2}||\mathbf{F}_n - \overline{F}_n|| = n^{1/2} \sup_x |\mathbf{F}_n(x) - \overline{F}_n(x)|$. Bretagnolle's [91] inequality says that the classical exponential bound of Dvoretzky et al. [97] for the i.i.d. case continues to hold (for arbitrary df's of the observations $F_1, \ldots, F_n$) if their absolute constant is increased by a factor of 4:

$$\Pr(||G_n|| \geqslant \lambda) = \Pr(n^{1/2}||\mathbf{F}_n - \overline{F}_n|| \geqslant \lambda) \leqslant 4C \exp(-2\lambda^2)$$

for all $n \geqslant 1$ and all $\lambda > 0$, where C is an absolute constant (weaker inequalities were given earlier by Singh [123] and Devroye [95]). A consequence of Bretagnolle's inequality is that

$$\limsup_{n\to\infty} \frac{||G_n||}{(2 \log \log n)^{1/2}} = \limsup_{n\to\infty} \frac{n^{1/2}||\mathbf{F}_n - \overline{F}_n||}{(2 \log \log n)^{1/2}} \leqslant \frac{1}{2} \text{ a.s.}$$

for independent observations $X_1, X_2, \ldots$ from a completely arbitrary sequence of df's $F_1, F_2, \ldots$

The results for G_n (and $\mathbf{W}_n$) sketched here have been applied by Koul [104] and Bickel [89] (regression problems), Shorack [121] (linear combinations of order statistics), Sen et al. [120] (strength of fiber bundles), and Gill [101] (censored survival data).

Dependent Observations

Billingsley [90, Sec. 22] proved two different weak convergence or Donsker theorems for the empirical process of a strictly stationary sequence of real-valued random variables with common continuous df F satisfying a weak or ϕ-mixing condition. Billingsley's results have subsequently been extended to other weaker (i.e., less restrictive) mixing conditions by Mehra and Rao [107] (who also consider the regression process $\mathbf{W}_n$ mentioned above and weighted metrics), Gastwirth and Rubin [99] (who introduced a new mixing condition intermediate between weak and strong mixing), and Withers [128]. Puri and Tran [110] provide linear in probability bounds, almost sure nearly linear bounds, and strengthened Glivenko-Cantelli theorems for $\mathbf{F}_n$ under a variety of mixing conditions.

When the (dependent) stationary sequence of observations has values in $\mathcal{X} = R^k$, Donsker theorems for the empirical process have been given by Rüschendorf [118] and Yoshihara [129]. The recent strong approximation results of Philipp and Pinzur [109] apply to strictly stationary R^k-valued observations with common continuous df satisfying a certain strong-mixing property. This strong approximation has, as corollaries, both Donsker (weak convergence) and Strassen-Finkelstein iterated logarithm theorems for the empirical processes of such variables.

An especially interesting Donsker theorem application for the empirical process of mixing variables is to robust location estimators under dependence by Gastwirth and Rubin [100].

Dependent and/or nonidentically distributed observations and the corresponding empirical processes also arise in studies of (a) problems involving finite populations [116,117]; (b) closely related problems concerning permutation tests* [89,127]; (c) residuals and "parameter-estimated empirical processes" [92,96,106]; (d) Fourier coefficients of an i.i.d. real-valued sample [98]; and (e) the spacings between the points of an i.i.d. sample [93,94,105,111,112,114,121]. An interesting variant on the latter set of problems is Kakutani's method of interval subdivision; see Van Zwet [126] and Pyke [113] for a discussion of Glivenko-Cantelli theorems; analogs of the Donsker theorem and the Strassen-Finkelstein theorem seem to be unknown.

MISCELLANEOUS TOPICS

This section briefly summarizes work concerning (a) censored survival data and the product limit estimator, (b) optimality properties of $\mathbb{P}_n$ as an estimator of P, and (c) large deviation theorems for empirical measures and processes.

Censored Survival Data; The Product Limit Estimator

In many important problems arising in medical or reliability settings, RVs $X_1, \ldots, X_n$ (i.i.d. with common df F) representing "survival times," cannot be observed. Instead, the statistician observes $(Z_1, \delta_1), \ldots, (Z_n, \delta_n)$, where Z_i is the smaller of the lifetime X_i and a censoring time $Y_i, Z_i = \min\{X_i, Y_i\}$, and δ_i equals 1 or 0 according as $Z_i = X_i$ or $Z_i = Y_i$. The statistician's goal is to estimate the df F of the survival times $\{X_i\}$, in spite of the *censoring*.

The nonparametric maximum likelihood estimator of F, the product-limit estimator (or Kaplan-Meier estimator*) $\hat{\mathbf{F}}_n$, was derived by Kaplan and Meier [139]:

$$1 - \hat{\mathbf{F}}_n(t) = \prod_{\{i: Z_{n:i} \leqslant t\}} (1 - 1/(n - i + 1))^{\delta_{n:i}},$$

where $Z_{n:1} \leqslant \cdots \leqslant Z_{n:n}$ and $\delta_{n:1}, \ldots, \delta_{n:n}$ denote the corresponding δ's. When there is no censoring, so $Z_i = X_i$ and $\delta_i = 1$ for all $i = 1, \ldots, n$, the product-limit estimator $\hat{\mathbf{F}}_n$ reduces to the usual empirical df $\mathbf{F}_n$.

Study of Donsker or weak convergence theorems for the corresponding empirical process

$$\hat{G}_n = n^{1/2}(\hat{\mathbf{F}}_n - F)$$

was initiated by Efron [134] under the assumption of i.i.d. censoring variables Y_i independent of the X_i (the random censorship model). Efron conjectured the weak convergence of $\hat{G}_n$, and used it in a study of two-sample statistics of interest for censored data. The weak convergence of $\hat{G}_n$ was first proved by Breslow and Crowley [132] under the assumption of i.i.d. censoring variables with common df G by use of a Skorokhod construction and long calculations. Gill [137], following Aalen [130,131], put the large-sample theory of $\hat{\mathbf{F}}_n$ and $\hat{G}_n$ in its natural setting by using the martingale* theory of counting processes together with a martingale (functional) central limit theorem due to Rebolledo [140] to give a simpler proof of the weak convergence under minimal assumptions on the independent censoring times $\{Y_i\}, \ldots, Y_n$. To state the theorem, let $C(t) = \int_0^t (1 - F)^{-2}(1 - G)^{-1}\, dF$ and set $K(t) = C(t)/(1 + C(t))$.

Theorem 7. (Breslow and Crowley [132]; Gill [137]).

$$\begin{aligned} \hat{G}_n &= n^{1/2}(\hat{\mathbf{F}}_n - F) \Rightarrow (1 - F) \cdot \mathbb{B} \circ C \\ &\stackrel{\mathrm{d}}{=} \left(\frac{1 - F}{1 - K}\right) \cdot \mathbf{U} \circ K \quad \text{as } n \to \infty \end{aligned}$$

where $\mathbb{B}$ denotes standard Brownian motion on $[0, \infty)$.

Gill [138] has given a refined and complete version of this theorem. Aalen [130,131] and Gill [137,138] have clarified the extremely important role which counting processes, and their associated martingales, play in the theory of empirical processes in the uncensored as well as the censored case.

Some preliminary iterated logarithm laws for G_n have been established by Földes and Rejtö [135,136]; iterated logarithm laws also follow from the strong approximations of $\hat{G}_n$ and other related processes provided by Burke et al. [133].

Optimality

Asymptotic minimax theorems demonstrating the asymptotic optimality of the empirical df $\mathbf{F}_n$ in a very large class of estimators of F and with respect to a large class of loss functions were first obtained by Dvoretzky et al. [144] in the i.i.d. case with $\mathcal{X} = R^1$, and by Kiefer and Wolfowitz [145] in the case $\mathcal{X} = R^k$; see also Levit [148]. An interesting representation theorem for the limiting distributions of regular estimates of a df F on [0, 1] has been established by Beran [141]. This asserts, roughly speaking, that the limiting process corresponding to any regular estimator of F has a representation as $\mathbf{U} \circ F + W$, where $\mathbf{U}$ is a Brownian bridge process and W is some process on [0, 1] independent of $\mathbf{U}$. Hence the empirical df $\mathbf{F}_n$ is an optimal estimator of F in this sense since $G_n = n^{1/2}(\mathbf{F}_n - F) \stackrel{\mathrm{d}}{=} \mathbf{U}_n \circ F \Rightarrow \mathbf{U} \circ F$ with $W = 0$ identically.

Motivated by questions in reliability, Kiefer and Wolfowitz [146] showed that the empirical df $\mathbf{F}_n$ remains asymptotically minimax for the problem of estimating a concave (or convex) df (even though $\mathbf{F}_n$ is itself not necessarily concave). Millar [149], using results of LeCam [147], put the earlier asymptotic minimax results in an elegant general setting and gave a geometric sufficient condition in order that the empirical df $\mathbf{F}_n$ be an asymptotically minimax estimator of F in a specified subset of df's. Millar's geometric criterion implies, in particular, that the empirical df is asymptotically minimax for estimating F in the classes of distributions with increasing or decreasing failure rates, or the class of distribution functions with decreasing densities on $[0, \infty)$; also, $\mathbf{F}_n$ is not asymptotically optimal as an estimator of a df symmetric at 0 (the symmetrized empirical df is optimal for this class). Wellner [15] established the asymptotical optimality of the product limit estimator in the case of randomly censored data.

There is a large literature concerning the power of various tests based on the empirical df and empirical processes; see Chibisov [142,143] on local alternatives, and Raghavachari [150] concerning the limiting distributions of Kolmogorov statistics under fixed alternatives.

Large Deviations*

Suppose that $X_1, \ldots, X_n$ are i.i.d. RVs with values in $\mathcal{X}$, common probability measure P on $\mathcal{X}$, and empirical measures $\mathbb{P}_n, n \geqslant 1$, as in the introduction. If Π is a collection of probability measures on $\mathcal{X}$ distant from P, then, by a Glivenko-Cantelli theorem,

$$\Pr(\mathbb{P}_n \in \Pi) \to 0 \quad \text{as } n \to \infty$$

since $\mathbb{P}_n \to P$ a.s. (in a variety of senses). In fact, this convergence to zero typically occurs exponentially fast as n increases, as demonstrated in problems concerning the Bahadur efficiency* of a variety of test statistics; see Groeneboom et al. [156], Bahadur and Zabell [154], and references therein. The constant appearing in the exponential rate is given by the Kullback-Liebler information of Π relative to P, $K(\Pi, P)$, defined by

$$K(\Pi, P) = \inf_{Q \in \Pi} K(Q, P),$$

$$K(Q, P) = \begin{cases} \int q \log q dP, & Q \ll P, q \equiv \dfrac{dQ}{dP}, \\ \infty, & \text{otherwise} \end{cases}$$

Theorem 8. (Groeneboom et al. [156]). If $\mathcal{X}$ is a Hausdorff space, Π is a collection of probability measures on $\mathcal{X}$ satisfying $K(\Pi^0, P) = K(\overline{\Pi}, P) = K(\Pi, P)$, where the interior Π^0 and closure $\overline{\Pi}$ of Π are taken relative to a certain topology τ, then

$$\Pr(\mathbb{P}_n \in \Pi) = \exp(-n[K(\Pi, P) + o(1)]) \quad \text{as } n \to \infty$$

$$[\text{i.e., } \lim_{n \to \infty} n^{-1} \log \Pr(\mathbb{P}_n \in \Pi) = K(\Pi, P)].$$

Groeneboom et al. [156] give several applications of this general theorem. In the special case of i.i.d. uniform (0,1) X's and

$$\Pi = \{P : \sup_t (F(t) - t) \geqslant \lambda > 0 \text{ with } F(t) = P(-\infty, t]\},$$

the number $K(\Pi, I)$ has been computed explicitly by Sethuraman [157], Abrahamson [152], Bahadur [153], and Siegmund [158]:

$$\begin{aligned} K(\Pi, I) &= \inf_{0 \leqslant t \leqslant 1-\lambda} \left\{ (\lambda + t) \log\left(\frac{\lambda + t}{t}\right) + (1 - \lambda - t) \log\left(\frac{1 - \lambda - t}{1 - t}\right) \right\} \\ &= (\theta_1 - \theta_2)\lambda + \theta_2 + \log(1 - \theta_2) \equiv g(\lambda), \end{aligned}$$

where $\theta_2 < 0 < \theta_1$ satisfy $\theta_1^{-1} + \theta_2^{-1} = \lambda^{-1}$ and $\theta_1 - \theta_2 = \log[(1 - \theta_2)/(1 - \theta_1)]$. The calculations of Siegmund [158] make the $o(1)$ term explicit in this case.

$$\Pr\left(\sup_{0 \leqslant t \leqslant 1} (\mathbf{F}_n(t) - t) > \lambda \right) \sim h(\lambda) \exp(-n g(\lambda)) \text{ as } n \to \infty,$$

where

$$h(\lambda) \equiv \left\{\lambda|\theta_2|^{-1}(1-\theta_2) \times \left[1+\left(\frac{|\theta_2|}{\theta_1}\right)^3\left(\frac{1-\theta_1}{1-\theta_2}\right)\right]\right\}^{-1/2}$$

Berk and Jones [155] have some related results.

REFERENCES

Introduction

1. Durbin, J. (1973). *Distribution Theory for Tests Based on the Sample Distribution Function*. Reg. Conf. Ser. Appl. Math. No. 9. SIAM, Philadelphia.
2. Gaenssler, P. and Stute, W. (1979). *Ann. Prob.*, **7**, 193–243.
3. Niederhausen, H. (1981). *Ann. Statist.*, **9**, 923–944.
4. Pyke, R. (1972). *Jeffrey-Williams Lectures, Canad. Math. Congr.*, Montreal, pp. 13–43.

The One-Dimensional Case

5. Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley, New York.
6. Billingsley, P. (1971). *Weak Convergence of Measures: Applications in Probability*. Reg. Conf. Ser. Appl. Math. No. 5. SIAM, Philadelphia.
7. Bolthausen, E. (1977). *Metrika*, **24**, 215–227.
8. Boos, D. D. (1981). *J. Amer. Statist. Ass.*, **76**, 663–670.
9. Breiman, L. (1968). *Probability*. Addison-Wesley, Reading, Mass.
10. Breslow, N. and Crowley, J. (1974). *Ann. Statist.*, **2**, 437–453.
11. Brillinger, D. R. (1969). *Bull. Amer. Math. Soc.*, **75**, 545–547.
12. Cassels, J. W. S. (1951). *Proc. Camb. Philos. Soc.*, **47**, 55–64.
13. Chibisov, D. M. (1964). *Select. Transl. Math. Statist. Prob.*, **6**, 147–156.
14. Chibisov, D. M. (1965). *Theor. Prob. Appl.*, **10**, 421–437.
15. Chung, K. L. (1949). *Trans. Amer. Math. Soc.*, **67**, 36–50.
16. Cramér, H. (1928). *Skand. Aktuarietidskr.*, **11**, 141–180.
17. Cressie, N. (1977). *J. Appl. Prob.*, **14**, 272–283.
18. Cressie, N. (1980). *Ann. Prob.*, **8**, 828–840.
19. Csáki, E. (1974). *Limit Theorems of Probability Theory*. Keszttheley, Hungary, pp. 47–58.
20. Csáki, E. (1977). *Zeit. Wahrscheinlichkeitsth. verw. Geb.*, **38**, 147–167.
21. Csörgö, M. and Révész, P. (1975). *Zeit. Wahrscheinlichkeitsth. verw. Geb.*, **31**, 261–269.
22. Csörgö, M. and Révész, P. (1978). *Ann. Statist.*, **6**, 882–894.
23. Csörgö, M. and Révész, P. (1981). *Strong Approximations in Probability and Statistics*. Academic Press, New York.
24. Darling, D. A. (1957). *Ann. Math. Statist.*, **28**, 823–838.
25. Darling, D. A. and Erdös, P. (1956). *Duke Math. J.*, **23**, 143–155.
26. Donsker, M. D. (1951). *Mem. Amer. Math. Soc.*, **6**, 1–12.
27. Donsker, M. D. (1952). *Ann. Math. Statist.*, **23**, 277–281.
28. Doob, J. L. (1949). *Ann. Math. Statist.*, **20**, 393–403.
29. Dudley, R. M. (1968). *Ann. Math. Statist.*, **39**, 1563–1572.
30. Durbin, J. (1973). *Distribution Theory for Tests Based on the Sample Distribution Function*. Reg. Conf. Ser. Appl. Math. No. 9. SIAM, Philadelphia.
31. Eicker, F. (1979). *Ann. Statist.*, **7**, 116–138.
32. Erdös, P. and Kac, M. (1946). *Bull. Amer. Math. Soc.*, **52**, 292–302.
33. Erdös, P. and Kac, M. (1947). *Bull. Amer. Math. Soc.*, **53**, 1011–1020.
34. Finkelstein, H. (1971). *Ann. Math. Statist.*, **42**, 607–615.
35. Jaeschke, D. (1979). *Ann. Statist.*, **7**, 108–115.
36. James, B. R. (1975). *Ann. Prob.*, **3**, 762–772.
37. Kiefer, J. (1969). *Zeit. Wahrscheinlichkeitsth. verw. Geb.*, **13**, 321–332.
38. Kiefer, J. (1970). In *Nonparametric Techniques in Statistical Inference*, M. L. Puri, ed. Cambridge University Press, Cambridge, England, pp. 299–319.
39. Kolmogorov, A. N. (1933). *Giorn. Ist. Ital. Attuari*, **4**, 83–91.
40. Komlós, J., Major, P., and Tusnády, G. (1975). *Zeit Wahrscheinlichkeitsth. verw. Geb.*, **32**, 111–131.
41. Mason, D. M. (1981). *Ann. Prob.*, **9**, 881–884.
42. O'Reilly, N. E. (1974). *Ann. Prob.*, **2**, 642–645.

43. Pollard, D. (1980). *Metrika*, **27**, 43–70.
44. Prohorov, Y. V. (1956). *Theor. Prob. Appl.*, **1**, 157–214.
45. Pyke, R. (1969). *Lect. Notes Math.*, **89**, 187–200.
46. Pyke, R. (1970). In *Nonparametric Techniques in Statistical Inference*, M. L. Puri, ed. Cambridge University Press, Cambridge, England, pp. 21–37.
47. Pyke, R. (1972). *Jeffrey-Williams Lectures*. *Canad. Math. Congr.*, Montreal, pp. 13–43.
48. Pyke, R. and Shorack, G. R. (1968). *Ann. Math. Statist.*, **39**, 755–771.
49. Root, D. H. (1969). *Ann. Math. Statist.*, **40**, 715–718.
50. Rosenkrantz, W. A. (1967). *Trans. Amer. Math. Soc.*, **129**, 542–552.
51. Sähler, W. (1968). *Metrika*, **13**, 149–169.
52. Sawyer, S. (1974). *Rocky Mountain J. Math.*, **4**, 579–596.
53. Shorack, G. R. (1972). *Ann. Math. Statist.*, **43**, 412–427.
54. Shorack, G. R. (1972). *Ann. Math. Statist.*, **43**, 1400–1411.
55. Shorack, G. R. (1974). *Ann. Statist.*, **2**, 661–675.
56. Shorack, G. R. (1979). *Stoch. Proc. Appl.*, **9**, 95–98.
57. Shorack, G. R. (1980). *Aust. J. Statist.*, **22**, 50–59.
58. Shorack, G. R. and Wellner, J. A. (1982). *Ann. Prob.*, **10**, 639–652.
59. Skorokhod, A. V. (1956). *Theor. Prob. Appl.*, **1**, 261–290.
60. Skorokhod, A. V. (1961). *Studies in the Theory of Random Processes*. Kiev University Press, Kiev. (English transl.: Addison-Wesley, Reading, Mass., 1965.)
61. Smirnov, N. V. (1939). *Math. Sb. N.S.*, **6**, 3–26.
62. Smirnov, N. V. (1944). *Uspekhi Mat. Nauk.*, **10**, 179–206.
63. Strassen, V. (1964). *Zeit. Wahrscheinlichkeitsth. verw. Geb.*, **3**, 211–226.
64. Stute, W. (1982). *Ann. Prob.*, **10**, 86–107.
65. von Mises, R. (1931). *Wahrscheinlichkeitsrechnung*. Franz Deuticke, Leipzig, Germany.
66. Wichura, M. J. (1970). *Ann. Math. Statist.*, **41**, 284–291.

More General Sample Spaces and Index Sets

67. Bickel, P. J. and Wichura, M. J. (1971). *Ann. Math. Statist.*, **42**, 1656–1670.
68. Bolthausen, E. (1978). *Zeit. Wahrscheinlichkeitsth. verw. Geb.*, **43**, 173–181.
69. Dudley, R. M. (1966). *Ill. J. Math.*, **10**, 109–126.
70. Dudley, R. M. (1978). *Ann. Prob.*, **6**, 899–929; correction: *Ann. Prob.*, **7**, 909–911.
71. Dudley, R. M. (1980). *Donsker classes of functions*. Preprint.
72. Giné, M. E. (1975). *Ann. Statist.*, **3**, 1243–1266.
73. Hartigan, J. A. (1978). *Ann. Statist.*, **6**, 117–131.
74. Kaufman, R. and Philipp, W. (1978). *Ann. Prob.*, **6**, 930–952.
75. Kiefer, J. (1961). *Pacific J. Math.*, **11**, 649–660.
76. Kuelbs, J. and Dudley, R. M. (1980). *Ann. Prob.*, **8**, 405–418.
77. Neuhaus, G. (1971). *Ann. Math. Statist.*, **42**, 1285–1295.
78. Philipp, W. and Pinzur, L. (1980). *Zeit. Wahrsheinlichkeitsth. verw. Geb.*, **54**, 1–13.
79. Pollard, D. (1979). *Zeit. Wahrscheinlichkeitsth. verw. Geb.*, **50**, 317–331.
80. Pollard, D. (1982). *Ann. Prob.*, **10**, 919–926.
81. Pollard, D. (1982). *J. Aust. Math. Soc. A*, **33**, 235–248.
82. Révész, P. (1976). *Ann. Prob.*, **5**, 729–743.
83. Révész, P. (1976). *Lect. Notes Math.*, **566**, 106–126.
84. Richter, H. (1974). *Manuscripta Math.*, **11**, 291–303.
85. Straf, M. L. (1971). *Proc. 6th Berkeley Symp. Math. Statist. Prob.*, Vol. 2. University of California Press, Berkeley, Calif., pp. 187–221.
86. Strassen, V. and Dudley, R. M. (1969). *Lect. Notes Math.*, **89**, 224–231.
87. Wellner, J. A. (1979). *Ann. Statist.*, **7**, 929–943.
88. Wichura, M. J. (1973). *Ann. Prob.*, **1**, 272–296.

Dependent or Nonidentically Distributed Observations

89. Bickel, P. J. (1973). *Ann. Math. Statist.*, **40**, 1–23.
90. Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley, New York.
91. Bretagnolle, J. (1980). Communication au *Colloque Int. du CNRS*, **307**, 39–44.

92. Burke, M. D., Csörgö, M., Csörgö, S., and Révész, P. (1979). *Ann. Prob.*, **7**, 790–810.
93. Cressie, N. (1979). *Biometrika*, **66**, 619–628.
94. del Pino, G. E. (1979). *Ann. Statist.*, **7**, 1058–1065.
95. Devroye, L. P. (1977). *J. Multivariate Anal.*, **7**, 594–597.
96. Durbin, J. (1973). *Ann. Statist.*, **1**, 279–290.
97. Dvoretzky, A., Kiefer, J., and Wolfowitz, J. (1956). *Ann. Math. Statist.*, **27**, 642–669.
98. Freedman, D. and Lane, D. (1980). *Ann. Statist.*, **8**, 1244–1251.
99. Gastwirth, J. L. and Rubin, H. (1975). *Ann. Statist.*, **3**, 809–824.
100. Gastwirth, J. L. and Rubin, H. (1975). *Ann. Statist.*, **3**, 1070–1100.
101. Gill, R. D. (1980). *Censoring and Stochastic Integrals*. Mathematische Centrum, Amsterdam, The Netherlands.
102. Hájek, J. (1965). *Bernoulli-Bayes-Laplace Seminar*. University of California Press, Berkeley, Calif., pp. 45–60.
103. Koul, H. L. (1970). *Ann. Math. Statist.*, **41**, 1768–1773.
104. Koul, H. L. (1971). *Ann. Math. Statist.*, **42**, 466–476.
105. Koziol, J. A. (1977). *J. R. Statist. Soc. B*, **39**, 333–336.
106. Loynes, R. M. (1980). *Ann. Statist.*, **8**, 285–298.
107. Mehra, K. and Rao, M. S. (1975). *Ann. Prob.*, **3**, 979–991.
108. Neuhaus, G. (1975). *Ann. Statist.*, **3**, 528–531.
109. Philipp, W. and Pinzur, L. (1980). *Zeit. Wahrscheinlichkeitsth. verw. Geb.*, **54**, 1–13.
110. Puri, M. L. and Tran, L. T. (1980). *J. Multivariate Anal.*, **10**, 405–425.
111. Pyke, R. (1965). *J. R. Statist. Soc. B*, **7**, 395–449.
112. Pyke, R. (1972). *Proc. 6th Berkeley Symp. Math. Statist. Prob.*, Vol. 1. University of California Press, Berkeley, Calif., pp. 417–427.
113. Pyke, R. (1980). *Ann Prob.*, **8**, 157–163.
114. Rao, J. S. and Sethuraman, J. (1975). *Ann. Statist.*, **3**, 299–313.
115. Rechtschaffen, R. (1975). *Ann. Statist.*, **3**, 787–792.
116. Rosen, B. (1964). *Ark. Math.*, **4**(28), 383–424.
117. Rosen, B. (1967). *Zeit. Wahrscheinlichkeitsth. verw. Geb.*, **7**, 103–115.
118. Rüschendorf, L. (1974). *J. Multivariate Anal.*, **4**, 469–478.
119. Rüschendorf, L. (1976). *Ann. Statist.*, **4**, 912–923.
120. Sen, P. K., Bhattacharya, B. B., and Suh, M. W. (1973). *Ann. Statist.*, **1**, 297–311.
121. Shorack, G. R. (1973). *Ann. Statist.*, **1**, 146–152.
122. Shorack, G. R. (1979). *Statist. Neerlandica*, **33**, 169–189.
123. Singh, R. S. (1975). *Ann. Prob.*, **3**, 371–374.
124. van Zuijlen, M. C. A. (1976). *Ann. Statist.*, **4**, 406–408.
125. van Zuijlen, M. C. A. (1978). *Ann. Prob.*, **6**, 250–266.
126. Van Zwet, W. R. (1978). *Ann. Prob.*, **6**, 133–137.
127. Wellner, J. A. (1979). *Ann. Statist.*, **7**, 929–943.
128. Withers, C. (1976). *Aust. J. Statist.*, **18**, 76–83.
129. Yoshihara, K. (1975). *Zeit. Wahrscheinlichkeitsth. verw. Geb.*, **33**, 133–138.

Miscellaneous Topics

Censored survival data; the product limit estimator

130. Aalen, O. O. (1978). *Zeit. Wahrscheinlichkeitsth. verw. Geb.*, **38**, 261–277.
131. Aalen, O. O. (1978). *Ann. Statist.*, **6**, 701–726.
132. Breslow, N. and Crowley, J. (1974). *Ann. Statist.*, **2**, 437–453.
133. Burke, M. D., Csörgö, S., and Horváth, L. (1981). *Zeit. Wahrscheinlichkeitsth. verw. Geb.*, **56**, 87–112.
134. Efron, B. (1967). *Proc. 5th Berkeley Symp. Math. Statist. Prob.*, Vol. 4. University of California Press, Berkeley, Calif., pp. 831–853.
135. Földes, A. and Rejtö, L. (1981). *Ann. Statist.*, **9**, 122–129.
136. Földes, A. and Rejtö, L. (1981). *Zeit. Wahrscheinlichkeitsth. verw. Geb.*, **56**, 75–86.
137. Gill, R. D. (1980). *Censoring and Stochastic Integrals*. Math. Center Tract No. 124. Mathematical Centre, Amsterdam, The Netherlands.
138. Gill, R. D. (1983). *Ann. Statist.*, **11**, 49–58.
139. Kaplan, E. L. and Meier, P. (1958). *J. Amer. Statist. Ass.*, **53**, 457–481.
140. Rebolledo, R. (1980). *Zeit. Wahrscheinlichkeitsth. verw. Geb.*, **51**, 269–286.

Optimality

141. Beran, R. (1977). *Ann. Statist.*, **5**, 400–404.
142. Chibisov, D. M. (1965). *Theor. Prob. Appl.*, **10**, 421–437.
143. Chibisov, D. M. (1969). *Sankhyā A*, **31**, 241–258.
144. Dvoretzky, A., Kiefer, J., and Wolfowitz, J. (1956). *Ann. Math. Statist.*, **27**, 642–669.
145. Kiefer, J. and Wolfowitz, J. (1959). *Ann. Math. Statist.*, **30**, 463–489.
146. Kiefer, J. and Wolfowitz, J. (1976). *Zeit. Wahrscheinlichkeitsth. verw. Geb.*, **34**, 73–85.
147. LeCam, L. (1972). *Proc. 6th Berkeley Symp. Math. Statist. Prob.*, Vol. 1. University of California Press, Berkeley, Calif., pp. 245–261.
148. Levit, B. Ya. (1978). *Theor. Prob. Appl.*, **23**, 371–377.
149. Millar, P. W. (1979). *Zeit. Wahrscheinlichkeitsth. verw. Geb.*, **48**, 233–252.
150. Raghavachari, M. (1973). *Ann. Statist.*, **1**, 67–73.
151. Wellner, J. A. (1982). *Ann. Statist.*, **10**, 595–602.

Large deviations

152. Abrahamson, I. G. (1967). *Ann. Math. Statist.*, **38**, 1475–1490.
153. Bahadur, R. R. (1971). *Some Limit Theorems in Statistics*. SIAM, Philadelphia.
154. Bahadur, R. R. and Zabell, S. L. (1979). *Ann. Prob.*, **7**, 587–621.
155. Berk, R. H. and Jones, D. H. (1979). *Zeit. Wahrscheinlichkeitsth. verw. Geb.*, **47**, 47–59.
156. Groeneboom, P., Oosterhoff, J., and Ruymgaart, F. H. (1979). *Ann. Prob.*, **7**, 553–586.
157. Sethuraman, J. (1964). *Ann. Math. Statist.*, **35**, 1304–1316.
158. Siegmund, D. O. (1982). *Ann. Prob.*, **10**, 581–588.

BIBLIOGRAPHY

Dudley, R. M. (1984). A Course on Empirical Processes. *Lecture Notes in Math.*, **1097**, 2–142. (École d'Été de Probabilités de St.-Flour, 1982.)

Gaenssler, P. (1983). *Empirical Processes*. Inst. Math. Statist., Lecture Notes-Monograph Series (S. S. Gupta, ed.), Vol. 3.

Pollard, D. (1984). *Convergence of Stochastic Processes*. Springer Verlag, New York.

Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.

See also CONVERGENCE OF SEQUENCES OF RANDOM VARIABLES; GLIVENKO–CANTELLI THEOREMS; HUNGARIAN CONSTRUCTIONS OF EMPIRICAL PROCESSES; LAW OF THE ITERATED LOGARITHM; and STOCHASTIC PROCESSES.

JON A. WELLNER