equations as something "which Cauchy never even wrote" is obviously a wrong translation, what was meant being something like "which Cauchy did not write here". I have kept a list of dozens of lesser errors, but at least this translator does not fall to the depths of rendering Abel's famous statement "Cauchy est 'fou'" by "Cauchy is a fool" [**3**, p. 25].

The presentation of the book is scandalously bad, especially for a publisher with a great tradition of excellence. Apparently it was reproduced from a "camera-ready text," as the saying goes. The result is ugly and hard to read. The paper is flimsy and tears easily.

The foregoing review does not do justice to the book. To those mathematicians who would like to know how classical analysis developed, I can only say, Read it!

REFERENCES

1. U. Bottazzini, *Il calcolo sublime: Storia dell'analisi matematica da Euler a Weierstrass*, Boringhieri, Torino, 1981.

2. M. Kline, *Mathematical thought from ancient to modern times*, Oxford Univ. Press, New York, 1972.

3. I. Grattan-Guinness, *The development of the foundations of mathematical analysis from Euler to Riemann*, MIT Press, Cambridge, Mass., 1970.

4. J. V. Grabiner, *The origins of Cauchy's rigorous calculus*, MIT Press, Cambridge, Mass., 1981.

5. A. P. Youschkevitch, *The concept of function up to the middle of the* 19*th century*, Arch. Hist. Exact Sci. **16** (1976/7), 37–85.

C. TRUESDELL

BULLETIN (New Series) OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 17, Number 1, July 1987

0273-0979/87 $1.00 + $.25 per page

Empirical processes with applications to statistics, by Galen R. Shorack and Jon A. Wellner, Wiley Series in Probability and Mathematical Statistics, John Wiley and Sons, New York, Chichester, Brisbane, Toronto, Singapore, 1986, xxvii + 938 pp., $59.95. ISBN 0-471-86725-X.

This is an impressive book, the result of a colossal undertaking by two people who have witnessed much of, and contributed to, the modern development of empirical processes and their applications to statistics.

In their preface, on the main objectives of their study, the authors write:

> The study of the empirical process and the empirical distribution function is one of the major continuing themes in the historical development of mathematical statistics. The applications are manifold, especially since many statistical procedures can be viewed as functionals on the empirical process and the behavior of such procedures can be inferred from that of the empirical process itself. We consider the empirical process per se, as well as applications of order statistics, rank tests, spacings, censored data, and so on.

This review, especially the first half of it, is aimed at the broad audience of the American Mathematical Society: hence the early history of the subject of the book is given much more space than a review addressed to experts would normally allow. This approach has almost inevitably led to a somewhat more historical presentation of the book under review than would have otherwise been warranted. It has certainly resulted in a lengthy essay on the subject. I hope very much that the expert readers as well as the authors of the book themselves will not feel left out on occasion as a result of my trying to make the main subject itself more accessible to a larger audience.

The mathematical roots of the empirical process reach back to the very foundations of modern probability theory [Kolmogorov (1933a)]. Let $(\Omega, \mathscr{A})$ be a measurable space, and let $P(\cdot)$ be a nonnegative, normed ($P(\Omega) = 1$), σ-additive set function on $\mathscr{A}$. Then $(\Omega, \mathscr{A}, P)$ is called a probability space. A real-valued function $X(\omega)$, $\omega \in \Omega$, is called a random variable (rv) if for every Borel set B of the real line the ω-set $\{\omega \in \Omega : X(\omega) \in B\}$ belongs to $\mathscr{A}$. Thus a rv $X(\omega)$ is a real-valued measurable function on $(\Omega, \mathscr{A})$. The distribution function F of a rv X is defined by

$$F(x) = F_X(x) = P\{\omega \in \Omega : X(\omega) \leqslant x\}, \qquad -\infty < x < \infty.$$

Let $X_1, X_2, \ldots$ be independent rv's with distribution function F (i.e., the joint distribution of any finite number of $X_1, X_2, \ldots$ is the product measure generated by the corresponding product of so many F). The random distribution function

$$\mathbf{F}_n(x) = n^{-1} \sum_{i=1}^{n} 1_{(-\infty, x]}(X_i), \qquad -\infty < x < \infty,$$

which assigns mass $1/n$ to each value of $X_i = X_i(\omega)$ (the data value of X_i), is called the *empirical distribution function* of $X_1, \ldots, X_n$ ($1_A(\cdot)$ denotes the indicator function of the set A). The Glivenko (1933)–Cantelli (1933) theorem states that even though F may be unknown, it can be uniformly (in x) estimated for almost all $\omega \in \Omega$ by $\mathbf{F}_n$. Namely we have

$$\sup_{-\infty < x < \infty} |\mathbf{F}_n(x) - F(x)| \underset{\text{a.s.}}{\rightarrow} 0 \quad \text{as } n \to \infty,$$

i.e.,

$$P\left\{\lim_{n \to \infty} \sup_{-\infty < x < \infty} |\mathbf{F}_n(x) - F(x)| = 0\right\} = 1.$$

The latter theorem is a much improved and quite sophisticated version of Bernoulli's law (1713) of large numbers for independent binomial experiments.

It has been rightly called "the existence theorem for statistics as a branch of applied mathematics" [Pitman (1979)] and also "the fundamental theorem of statistics" [Loéve (1955)]. Roughly speaking it guarantees that statistics can make sense. It implies that the unknown probabilistic structure of the sequence $\{X_n\}_{n=1}^{\infty}$ can almost surely (a.s.) be discovered from the data via $\mathbf{F}_n$. In another fundamental work in 1933, Kolmogorov (1933b) describes the asymptotic probabilistic fluctuation of $D_n = \sup_{-\infty < x < \infty} |\mathbf{F}_n(x) - F(x)|$ with F continuous, at the rate of $n^{1/2}$ as follows:

$$(1)\quad \lim_{n \to \infty} P\{n^{1/2} D_n \leqslant y\}$$
$$= \sum_{k=-\infty}^{\infty} (-1)^k \exp(-2k^2 y^2) \quad \text{if } y > 0, \quad \text{zero otherwise,}$$
$$= L(y),$$

and thus he launches the study of what we call the *empirical process* today, defined by

$$n^{1/2}[\mathbf{F}_n(x) - F(x)], \qquad -\infty < x < \infty.$$

Important further steps of Smirnov (1939a, 1939b, 1944) followed. Let $Y_1, Y_2, \ldots$ be another sequence of independent random variables with the same continuous distribution function F as that of $X_1, X_2, \ldots$. Let $\mathbf{F}_m^*$ be the empirical distribution function of $Y_1, \ldots, Y_m$, and consider $D_{m,n} = \sup_{-\infty < x < \infty} |\mathbf{F}_n(x) - \mathbf{F}_m^*(x)|$. Then, as $m \to \infty$, $n \to \infty$, we have [Smirnov (1939b)]

$$(2)\qquad P\{N^{1/2} D_{n,m} \leqslant y\} \to L(y),$$

where $N = mn/(m+n)$. In his 1944 paper Smirnov obtained the exact (fixed $n \geqslant 1$) distribution of $n^{1/2} \sup_{-\infty < x < \infty} (\mathbf{F}_n(x) - F(x))$ from which he also derived the corresponding asymptotic ($n \to \infty$) distribution $1 - \exp(-2y^2)$, $y > 0$, by direct calculations.

At this stage it is enlightening to quote, using our notation here, from Feller (1948):

> The original proofs [Kolmogorov (1933b) and Smirnov (1939b)] are very intricate and are based on completely different methods. Kolmogorov's proof is based on an auxiliary theorem of equal depth proved in a separate paper [Kolmogorov (1933c)]. An alternative proof of Kolmogorov's theorem is due to Smirnov (1939a). However, Smirnov derives both theorems as corollaries to much deeper (but less useful) results concerning the number of intersections of the graphs of $\mathbf{F}_n(x)$ and $F(x) \pm \varepsilon n^{-1/2}$ and of $\mathbf{F}_n(x)$ and $\mathbf{F}_m^*(x) \pm \varepsilon N^{-1/2}$, respectively. It is therefore not surprising that Smirnov's proofs require a powerful technique and many auxiliary considerations. It is the purpose of the present paper to present unified proofs of the theorems which are based on methods of great generality. The new proof is not simple but simpler than the original ones.

Though he described them as "only routine manipulations," Feller's unified proofs constitute an impressive tour de force on generating functions and their limiting form, the Laplace transforms. As we will now see, however, the route of further developments took a different turn.

The maximum discrepancy between two empirical distributions $D_{n,m}$ shares with D_n the property that its distribution does not depend on F if F is continuous. For this reason it serves in establishing statistical tests of the hypothesis that $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$ are random samples from the same population. The random variable $D_{n,n}$ also served in establishing connection with diffusion processes. Gnedenko and Koroljuk (1951) showed that the distribution of $D_{n,n}$ reduces to a random walk problem with a well-known solution. The limiting procedure leads from random walks to diffusion processes, and this way it is also easier to see that the distribution of $n^{1/2}D_{n,n}$ tends to the limit in (2). On the other hand, the intricate calculations of Kolmogorov (1933b) and Smirnov (1939a, 1939b), as well as those of Feller (1948), have given impetus to exciting work on the convergence of stochastic processes, invariance principles, and convergence of probability measures. Roughly speaking this is also what this book is all about, plus some more in terms of applications to statistics.

Invariance principles have evolved from two major sources: partial sum processes and empirical processes. In two papers Kolmogorov (1931, 1933c) considers independent rv's X_i with mean zero, variance $\mathsf{E}X_i^2 = \sigma_i^2$ and $\mathsf{E}|X_i|^3 \leqslant \varepsilon\sigma_i^2$. He shows that for $\varepsilon > 0$ small, the probability that the trajectory $G = \{\sum_{i \leqslant k}\sigma_i^2, \sum_{i \leqslant k} X_i;\ k = 1, 2, \dots\}$ as a graph in the plane lies between two smooth curves differs little from a number obtainable from the solution of the heat equation that vanishes on the two curves. Today we would call this an *invariance principle*. Though Kolmogorov's result is reproduced in a book by Khinchine, it somehow escaped further attention. The first big steps which have indeed initiated a new methodology for proving asymptotic laws in probability theory were taken by Erdős and Kac (1946) [cf. also Kac (1946)]. They established the asymptotic distributions of the rv's

$$n^{-1/2} \max_{1 \leqslant k \leqslant n} S_k, \qquad n^{-1/2} \max_{1 \leqslant k \leqslant n} |S_k|, \qquad n^{-2} \sum_{k=1}^{n} S_k^2, \qquad n^{-3/2} \sum_{k=1}^{n} |S_k|,$$

where $S_k = \sum_{i=1}^{k} X_i$, and the X_i $(i = 1, 2, \dots)$ are independent random variables with a common distribution function F, and have mean zero and variance 1 $(\mathsf{E}X_1 = 0,\ \mathsf{E}X_1^2 = 1)$, in two steps. First they calculated their limiting distribution in terms of assuming convenient forms for F, and then they showed that the above functionals of partial sums did not remember the initially taken forms of F in the limit $(n \to \infty)$. They called this method of proof the *invariance principle*.

When talking about the origins of the invariance principle, another landmark is the paper of Doob (1949). He considers independent rv's $\xi_1, \xi_2, \dots$ which are uniformly distributed on $(0, 1)$ and their empirical distribution function

$$\mathbf{G}_n(t) = \frac{1}{n} \sum_{i=1}^{n} 1_{[0,t]}(\xi_i), \qquad 0 \leqslant t \leqslant 1.$$

There is no loss of generality, for if the distribution function F of the independent rv's $X_1, X_2, \dots$ is continuous, then $F(X_1)$, $F(X_2), \dots$ are independent uniform $(0, 1)$ rv's. Since $n\mathbf{G}_n(0) = 0$ with probability 1 and $n\mathbf{G}_n(t) - n\mathbf{G}_n(s)$ is the number of successes in n independent trials with probability $t - s$ of success in each trial, the random variable $n\mathbf{G}_n(t) - n\mathbf{G}_n(s)$ has expectation $n(t - s)$ and variance $n(t - s)(1 - (t - s))$. Hence for the *uniform empirical process* $\mathbf{U}_n(t) = n^{1/2}(\mathbf{G}_n(t) - t)$, $0 \leqslant t \leqslant 1$, we have $\mathsf{E}\mathbf{U}_n(t) = 0$ and

$$\mathsf{E}(\mathbf{U}_n(t) - \mathbf{U}_n(s))^2 = (t - s)(1 - (t - s)), \qquad 0 \leqslant s \leqslant t \leqslant 1.$$

This is the same covariance structure as that of a Brownian bridge $B(t)$, $0 \leqslant t \leqslant 1$, i.e., a Gaussian process with $\mathsf{E}B(t) = 0$ and $\mathsf{E}B(t)B(s) = s(1-t)$, $0 \leqslant s \leqslant t \leqslant 1$. Also by the multivariate central limit theorem, the limit of the joint distribution of $\mathrm{U}_n(t_1), \mathrm{U}_n(t_2), \ldots, \mathrm{U}_n(t_k)$ $(0 \leqslant t_1 < t_2 < \cdots < t_k \leqslant 1)$ is the corresponding finite-dimensional distribution of a Brownian bridge, i.e., that of $B(t_1), B(t_2), \ldots, B(t_k)$. Arguing along these lines, Doob (1949) concludes: "We shall assume, until a contradiction frustrates our devotion to heuristic reasoning, that in calculating asymptotic $\mathrm{U}_n(t)$ process distributions when $n \to \infty$ we may simply replace the $\mathrm{U}_n(t)$ processes by the $B(t)$ process." Since, as we have indicated above, direct evaluation of the limit distribution of $\sup_{0 \leqslant t \leqslant 1} \mathrm{U}_n(t)$ (resp. $\sup_{0 \leqslant t \leqslant 1} |\mathrm{U}_n(t)|$) is rather complicated, while the evaluation of the distribution of $\sup_{0 \leqslant t \leqslant 1} B(t)$ (resp. $\sup_{0 \leqslant t \leqslant 1} |B(t)|$) is easier, this approach is appealing. Indeed, besides posing his invariance argument, Doob (1949) proceeded to evaluate the distribution of the latter functionals of $B(t)$, leaving the justification of his heuristic approach open.

Inspired by the papers of Erdős and Kac (1946), and Doob (1949), the first steps towards providing a unifying theory for solving these types of problems for partial sum and empirical processes were taken by Donsker (1951, 1952). Concerning partial sums $\{S_n\}$, $n \geqslant 0$ $(S_0 \equiv 0)$, of independent identically distributed rv's $X_1, X_2, \ldots$ with mean zero and variance 1, Donsker's idea (1951) was that from these partial sums one should construct a sequence of stochastic processes $\{S_n(t),\ 0 \leqslant t \leqslant 1\}$ on $C[0,1]$ as follows: $S_n(k/n) = S_k/n^{1/2}$, and $S_n(t)$ is the linear interpolation of the latter for $k/n < t < (k+1)/n$. Thus one can study the limiting behavior of $S_k/n^{1/2}$ via that of $S_n(t)$ on $C[0,1]$. Indeed, using a multivariate central limit theorem, one can immediately say that the distributions of $(S_n(t_1), S_n(t_2), \ldots, S_n(t_k))$ converge to that of $(W(t_1), W(t_2), \ldots, W(t_k))$ for any fixed sequence $0 \leqslant t_1 < t_2 < \cdots < t_k \leqslant 1$ as $n \to \infty$, where $\{W(t),\ 0 \leqslant t \leqslant 1\}$ is a standard Brownian motion (Wiener process), i.e., a Gaussian process with $\mathsf{E}W(t) = 0$ and $\mathsf{E}W(s)W(t) = s$, $0 \leqslant s \leqslant t \leqslant 1$. Moreover, the distributional properties of $\{S_n(t);\ 0 \leqslant t \leqslant 1\}$ should coincide with those of $\{W(t);\ 0 \leqslant t \leqslant 1\}$ as $n \to \infty$. One possible way of saying this precisely is via Donsker's *functional central limit theorem* (1951), which extends the results of Erdős and Kac (1946): $h(S_n(t)) \to_{\mathrm{d}} h(W(t))$ as $n \to \infty$ for all h: $C[0,1] \to \mathbf{R}^1$ that are supremum norm $\|\ \|$ continuous almost surely with respect to Wiener measure W. Here $\to_{\mathrm{d}}$ means convergence in distribution, i.e., convergence of the sequence of distribution functions generated by $h(S_n(t))$ to that of $h(W(t))$. Donsker (1952) was first again in proving a similar functional central limit theorem for the empirical process, which justifies and extends Doob's conjecture (1949). Now if, in general, $\{X_n(t)\}_{n=0}^{\infty}$ is a sequence of stochastic processes (random elements) on a function space M endowed with a metric ρ (e.g. $(C[0,1], \|\ \|)$ is a complete separable metric space), and as $n \to \infty$

$$(X_n(t_1), \ldots, X_n(t_k)) \to_{\mathrm{d}} (X_0(t_1), \ldots, X_0(t_k))$$

for any fixed sequence $t_1 < \cdots < t_k$, then the statement that $h(X_n(t)) \to_{\mathrm{d}} h(X_0(t))$ for all ρ-norm continuous real-valued functionals h is not necessarily true. A complete methodology for proving the latter and some more, assuming convergence in distribution of the finite-dimensional distributions, was worked out in the fundamental papers of Prohorov (1956) and Skorokhod (1956). An excellent summary and further developments of these notions and techniques can be found in the books of Billingsley (1968) and Parthasarathy (1967).

Nowadays we have, roughly speaking, three main, frequently interacting methods for proving invariance principles:

(i) Classical weak convergence methods;

(ii) Vapnik-Čhervonenkis type combinatorial methods for indexing class of sets or class of functions, or metric entropy notions for counts of sets needed to cover a class of functions;

(iii) Strong and weak approximations of stochastic processes based on various forms of the Skorokhod embedding scheme (like, e.g., Strassen's invariance principle), or on various forms of the Hungarian construction.

There are now books available on all these three methods and their applications. In addition to the already mentioned books of Billingsley (1968) and Parthasarathy (1967) on the first method, the books by Dudley (1984), Gaenssler (1983), and Pollard (1984) mainly deal with the second one. The books by M. Csörgő and Révész (1981), M. Csörgő (1983), and M. Csörgő, S. Csörgő, and Horváth (1986) are mostly concerned with the Hungarian construction and its applications.

The present book is the first one which studies all three of the above methods, their interplay, and the vastness of their applications in statistics. The authors' remarkable technique with inequalities which they carefully develop for empirical processes throughout the text is a strong unifying theme of the book. The number of topics they cover in 900 pages is so extensive that it compelled them to provide a very helpful one-page short table of contents in addition to their 18-page regular table of contents. In their *References* they cite more than 500 papers, plus books on related material. Two basic techniques are stressed: reduction to the case of uniform $(0, 1)$ random variables on the unit interval, and use of Skorokhod versions of weakly convergent processes. The Vapnik-Čhervonenkis type approach is introduced only in the final chapter; nevertheless it succeeds in constituting a good introduction to this area. The Hungarian construction is well contrasted with the authors' own approach throughout the text. As an illustration of their overall approach we again quote from their preface:

> Good inequalities are a key to strong theorems. In Appendix A we review many of the classic inequalities of probability theory. Great care has been taken in the development of inequalities for the empirical process throughout the text; these are regarded as highly interesting in their own right. Exponential bounds and maximal inequalities appear at several points.
>
> Because of strong parallels between the empirical process and the partial sum process, many results for partial sums are also included. Chapter 2 contains most of these.
>
> Our main concern is with the empirical process for iid rv's, though we also consider the weighted empirical process of independent rv's in some detail. We ignore the large literature on mixing rv's, and confine our presentation for k-dimensions and general spaces to an introduction in the final chapter.
>
> We emphasize the special Skorokhod construction of various processes, as opposed to classic weak convergence, wherever possible. We feel this makes for simpler and more intuitive proofs. The Hungarian construction is also considered.

I found Chapter 1, *Introduction and survey of results*, very informative, a good key to the whole book. There one immediately (p. 3) learns about the inverse, or quantile transformation:

> THEOREM 1. (*The inverse transformation*). Let $\xi \cong$ Uniform $(0,1)$. *For a fixed distribution function (df) F, define its left continuous inverse by* $F^{-1}(t) \equiv \inf\{x : F(x) \geqslant t\}$ *for* $0 < t < 1$. *Then the rv* $X \equiv F^{-1}(\xi)$ *has df* F; *that is,* $X \equiv F^{-1}(\xi) \cong F$;

and also that (p. 9):

> The Skorokhod-Wichura-Dudley theorem (Theorem 2.3.4) is basic to much of our approach. We now illustrate its simplest special case; in this case a simple constructive proof is possible. Let $F_1, F_2, \dots$ and F_0 denote df's such that $F_n \to_{\mathrm{d}} F_0$ as $n \to \infty$. Define rv's X_n^* by $X_n^* \equiv F_n^{-1}(\xi)$ for $n \geqslant 0$ where ξ is a fixed Uniform $(0,1)$ rv; then $X_n^* \cong F_n(\xi)$ by Theorem 1. Moreover, and this is the fundamental result: THEOREM 4. (*Elementary Skorokhod theorem*) $X_n^* \to_{\mathrm{a.s.}} X_0^*$ *as* $n \to \infty$.

The use of this kind of a theorem is immediately well illustrated on p. 10:

> EXERCISE 6. (*Mann-Wald theorem*). *Suppose* $X_n \to_{\mathrm{d}} X_0$ *as* $n \to \infty$ *and* Ψ *is continuous except on a measurable set* Δ *for which* $P(X_0 \in \Delta) = 0$. *Then* $\Psi(X_n) \to_{\mathrm{d}} \Psi(X_0)$ *as* $n \to \infty$. HINT. Let $X_n^* = F_n^{-1}(\xi)$ for all $n > 0$ where $X_n \cong F_n$. Show that $\Psi(X_n^*) \to_{\mathrm{a.s.}} \Psi(X_0^*)$.

This in turn leads to their presentation of the problem of weak convergence on p. 14 as follows:

> We now turn to a generalization of the Mann-Wald theorem. Letting $\to_{\mathrm{f.d.}}$ mean that the finite-dimensional distributions of the process on the left converge to those of the process on the right, it is a minor exercise to show that
>
> (3) $$\mathbf{U}_n \underset{\mathrm{f.d.}}{\to} \mathbf{U} \quad \text{as } n \to \infty$$
>
> for a Brownian bridge $\mathbf{U}$ (see §2.2 for the definition of $\mathbf{U}$). However, this mode of convergence is not strong enough to yield the Mann-Wald theorem: that is, it does not follow from (3) that $h(\mathbf{U}_n) \to_{\mathrm{d}} h(\mathbf{U})$ for $\|\ \|$-continuous functions h. The concept of weak convergence, $\Rightarrow$, was designed to fill this need (we leave the precise definition of $\Rightarrow$ until Chapter 2). In a landmark paper, Doob (1949) suggested heuristically that
>
> (4) $$\mathbf{U}_n \Rightarrow \mathbf{U}_n \quad \text{as } n \to \infty,$$
>
> in a sense that carried with it the implication that
>
> (5) $$h(\mathbf{U}_n) \underset{\mathrm{d}}{\to} h(\mathbf{U}) \quad \text{as } n \to \infty \text{ for all } h \text{ that are } \|\ \|\text{-continuous a.s. } \mathbf{U}.$$

A use, other than that in (1), of (4) is presented immediately after, as they write on their Chapter 5:

> Once (4) was established, it was trivial to show results such as
>
> (8) $$\int_0^1 \mathbf{U}_n^2(t)\,dt \underset{\mathrm{d}}{\to} \int_0^1 \mathbf{U}^2(t)\,dt \quad \text{as } n \to \infty;$$

just note that $h(f) \equiv \int_0^1 f^2(t)\,dt$ is $\|\ \|$-continuous. The trick is to determine the distribution of $h(\mathbf{U})$; the solution of this problem for the h in (8) leads to some particularly fruitful methodology. This is explored in the next few paragraphs (see Kac and Siegert, 1947).

Contrasting the different methodologies, on p. 16 of Chapter 1 they write on their Chapters 2, 3 and 12 as follows.

In many ways the concept of weak convergence $\Rightarrow$ is a rather inconvenient one to work with. Technical manipulations became easier to deal with after Skorokhod (1956) and Komlós, Major, and Tusnády (1975) introduced their constructions. Thus Skorokhod effectively constructed a triangular array $\{\xi_{ni},\ 1 \leqslant i \leqslant n,\ n \geqslant 1\}$ of row-independent Uniform $(0,1)$ rv's and a Brownian bridge $\mathbf{U}$, all on a common probability space, that satisfy

(18) $$\|\mathbf{U}_n - \mathbf{U}\| \underset{\text{a.s.}}{\rightarrow} 0 \quad \text{for a special construction;}$$

here $\mathbf{U}_n$ is the empirical process of $\xi_{n1},\ldots,\xi_{nn}$. Since it is trivial from (18) that $h(\text{Skorokhod's } \mathbf{U}_n) \rightarrow_{\text{a.s.}} h(\mathbf{U})$ for any $\|\ \|$-continuous functional h, and since $h(\text{Skorokhod's } \mathbf{U}_n) \cong h(\text{any } \mathbf{U}_n)$, one obtains immediately from (18) the result (5) that $h(\text{any } \mathbf{U}_n) \rightarrow_{\text{d}} h(\mathbf{U})$. So far, (18) has only provided an alternative proof of (5). In what ways is it really superior to (4)? First, it can be understood and taught more easily than (4). Secondly, it is often possible to show that $h(\text{Skorokhod's } \mathbf{U}_n)$, or even $h_n(\text{Skorokhod's } \mathbf{U}_n)$, $\rightarrow_{\text{a.s.}} h(\mathbf{U})$ and to thereby establish the necessary $\|\ \|$-continuity of h in a fashion difficult or impossible to discover from (4). (Examples will be seen in the chapters on linear combinations of order statistics and rank statistics.) Given that Skorokhod's construction is based on a triangular array we know absolutely nothing about the joint distribution of Skorokhod's $(\mathbf{U}_1, \mathbf{U}_2, \ldots)$. Thus his construction can be used to infer $\rightarrow_{\text{d}}$ or $\rightarrow_{\text{p}}$ of h (any $\mathbf{U}_n$), but it is helpless and worthless for showing $\rightarrow_{\text{a.s.}}$ The Hungarian construction (begun in Csörgő and Révész, 1975a and fundamentally strengthened by Komlós et al. 1975), improves Skorokhod's construction in that it only uses a single sequence of Uniform $(0,1)$ rv's and a Kiefer process $\mathbf{K}$ (see §2.2 for the definition of the Kiefer process) on a common probability space that satisfy

(19) $$\overline{\lim_{n\to\infty}}\ \frac{\sqrt{n}}{(\log n)^2}\|\mathbf{U}_n - \mathbf{B}_n\| < \infty \quad \text{a.s.}$$

for the Hungarian construction; here $\mathbf{U}_n$ is the empirical process of $\xi_1,\ldots,\xi_n$ and

(20) $$\mathbf{B}_n \equiv \mathbf{K}(n,\cdot)/\sqrt{n} \quad \text{is a Brownian bridge}$$

as in (2.2.11). Since $h(\mathbf{B}_n) \cong h(\mathbf{U})$, this construction also yields (5). It is also capable of yielding $\rightarrow_{\text{a.s.}}$ for the original sequence, though the subscript n on $\mathbf{B}_n$ may make the problem difficult. Its real value is in the rate it establishes.

We should note here that one of the most important works which is cited by name throughout the book, Skorokhod (1961), is missing from the *References*. Skorokhod's theorem (1961) states that for any rv X with first moment 0 and finite second moment, one can define a probability space with a Brownian motion (Wiener process) and a stopping time τ such that $W(\tau) \cong X$ and $\mathsf{E}\tau = \mathsf{E}X^2$ (cf. Proposition 3, p. 38 of the book under review). This is the basic building block of the Skorokhod embedding scheme, which has also led to the first strong invariance principle, namely that of Strassen (1964) for partial sums of rv's. This, and also related works of Breiman (1968) and Brillinger (1969) for the empirical process $\mathbf{U}_n$ have played important roles and are rightly emphasized throughout the book. The above mentioned Kiefer process **K** is a Gaussian process with mean 0 and covariance function

$$\mathsf{E}\mathbf{K}(s_1, t_1)\mathbf{K}(s_2, t_2) = (s_1 \wedge s_2)[(t_1 \wedge t_2) - t_1 t_2].$$

The cited Hungarian construction was inspired by Strassen (1964, 1967), Kiefer (1969), and especially by Kiefer (1972), where he proved the first strong invariance principle for the empirical process $\mathbf{U}_n$ in terms of what nowadays we call a Kiefer process. Indeed, Kiefer (1972) established the result in (19) above with $n^{1/6}/(\log n)^{2/3}$, instead of $n^{1/2}/(\log n)^2$, via generalizing Skorokhod's embedding scheme to vector-valued rv's. While Kiefer's work in general is given well-deserved attention throughout this book, Kiefer (1972) is missed in the *References*, and looks like it was also missed in the text. Nevertheless, the above quotations are given here to convey the author's well-balanced view and mastery of their subject as manifested throughout their book.

The first three chapters (150 pages) provide the mathematical probabilistic background, a setting for further developments in the remaining 750 pages. The whole book is written at a rather advanced mathematical level. There is an interdependence table to help readers find their way around. The advanced mathematical level should not, however, deter anyone from trying, for this book is well motivated throughout, filled with clear illustrations and tables, and it can be read at several levels. It provides a remarkably accessible summary of the asymptotic theory that is currently available for empirical processes and their applications to statistics. As to these applications, the book creates a unified theory and treatment for a vast array of topics in nonparametric statistics which otherwise can only be found in several books, not necessarily relating to each other. Thus we can, for example, learn about contiguity, convergence of empirical and rank processes under contiguous location, scale and regression alternatives, as well as empirical and rank processes of residuals in one setting. The book gives an excellent treatment of orthogonal decomposition of processes, and of various statistics. Martingale methods, censored data and the product-limit estimators, Poisson and exponential representations as well as exact distributions are highlighted in a sequence of chapters. In further impressive chapters we see the authors' approach to linear and nearly linear bounds on the empirical distribution function $\mathbf{G}_n$, to exponential inequalities, and then their treatment of $\|\cdot/q\|$-metric convergence of empirical processes. Recent developments in the Hungarian construction as well are highlighted here. The interplay of their approach with the Hungarian construction is also well illustrated when writing about laws of the iterated logarithm and oscillations of the empirical process. The *uniform quantile process*

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

the *uniform empirical difference process*

$$\mathbf{D}_n = \mathbf{U}_n + \mathbf{V}_n \quad \text{on } [0,1]$$

of Bahadur (1966) and Kiefer (1967, 1970) and their extensions, the *normalized uniform empirical process*

$$\mathbf{Z}_n(t) = \mathbf{U}_n(t)/(t(1-t))^{1/2}, \quad 0 < t < 1,$$

and the *standardized quantile process*

$$\mathbf{Q}_n(t) = f(F^{-1}(t))n^{1/2}(\mathbf{F}_n^{-1}(t) - F^{-1}(t)), \quad 0 < t < 1,$$

where $f = F'$ is assumed to be positive on the open support of F, have played an important role in developing the theory as well as the applications of empirical processes. Three chapters are devoted to these processes, which in turn lead to a unified treatment of L-statistics, rank statistics, and spacings in another three chapters. Further considerations result in the introduction of symmetry processes and their statistical applications, bootstrapping of the empirical process, convergence of U-statistic empirical processes, and reliability theory and econometric functions. There is also a chapter devoted to large deviations, discussing topics like Bahadur efficiency, the Kullback-Leibler information number, and the Sanov problem. The treatment of the empirical process of independent but not identically distributed rv's is highlighted by Bretagnolle's inequality and exponential bound (1980). Appendix A was already mentioned above. There is also an Appendix B on martingales and counting processes. A list of special symbols, an author index, and a subject index (17 pp. combined), greatly facilitate the process of getting familiar with the book.

Naturally, in a huge book like this there are bound to be misprints. Given its size, I have not noticed too many. Here I will only mention two of them, for these may actually be somewhat misleading if one sees the formulae in question the first time around. Namely, in formula (10) of p. 12, and also in (7) of p. 14, the factor $(-1)^{k+1}$ multiplying $\exp(-2k^2\lambda)$ is missing. The thus-corrected formula is equivalent to the distribution function $L(y)$ of (1) above.

In summary, this book is an important addition to the literature in probability and statistics in general, and to the theory and applications of empirical processes in particular. It will certainly be one of the basic references on asymptotic theory for empirical processes for some time to come. It is intended for graduate students and research workers in statistics and probability. The prerequisite is a standard graduate course in probability and some exposure to nonparametric statistics. A reasonable number of exercises are included. Frequently these are results from themes the authors have not pursued in their book.

As mentioned already, there are also other recent books available, emphasizing other methods for proving invariance principles. For someone who is not familiar with any, or some of these methods, and would like to learn about them in one setting, the present book is best, I believe, to start with. Should one end up liking any one of the methods in particular, then naturally, one should have a look at the other books too. On the other hand, those who have already seen them should definitely keep also this one in mind. At $59.95 it is a steal.

REFERENCES

R. R. Bahadur (1966), *A note on quantiles in large samples*, Ann. Math. Statist. **37**, 577–580.

J. Bernoulli (1713), *Ars Coniectandi*. I–II, III–IV, Oswald's Klassiker der Exacten Wissenschaften, No. 108, W. Engelmann, Leipzig, 1899.

P. Billingsley (1968), *Convergence of probability measures*, Wiley, New York.

L. Brieman (1968), *Probability*, Addison-Wesley, Reading, Mass.

J. Bretagnolle (1980), *Statistique de Kolmogorov-Smirnov pour un enchantillon nonequireparti*, Colloq. Internat. CNRS **307**, 39–44.

D. R. Brillinger (1969), *The asymptotic representation of the sample distribution function*, Bull. Amer. Math. Soc. **75**, 545–547.

F. P. Cantelli (1933), *Sulla determinazione empirica delle leggi di probabilità*, Giorn. Ist. Ital. Attuari **4**, 421–424.

M. Csörgő (1983), *Quantile processes with statistical applications*, Regional Conf. Ser. on Appl. Math. vol. 42, SIAM, Philadelphia, Pa.

M. Csörgő, S. Csörgő and L. Horváth (1986), *An asymptotic theory for empirical reliability and concentration processes*, Lecture Notes in Statistics, vol. 33, Springer-Verlag, Berlin and New York.

M. Csörgő and P. Révész (1975), *A new method to prove Strassen-type laws of invariance principle*. I, II, Z. Wahrsch. Verw. Gebiete **31**, 255–260; 261–269.

______ (1981), *Strong approximations in probability and statistics*, Akadémiai Kiadó, Budapest—Academic Press, New York, 1981.

M. Donsker (1951), *An invariance principle for certain probability limit theorems*, Mem. Amer. Math. Soc. **6**, 1–12.

______ (1952), *Justification and extension of Doob's heuristic approach to the Kolmogorov-Smirnov theorems*, Ann. Math. Statist. **23**, 277–281.

J. L. Doob (1949), *Heuristic approach to the Kolmogorov-Smirnov theorems*, Ann. Math. Statist. **20**, 393–403.

R. M. Dudley (1984), *A course on empirical processes*, Lecture Notes in Math., vol. 1097, Springer-Verlag, Berlin and New York.

P. Erdős and M. Kac (1946), *On certain limit theorems of the theory of probability*, Bull. Amer. Math. Soc. **52**, 292–302.

W. Feller (1948), *On the Kolmogorov-Smirnov limit theorems for empirical distributions*, Ann. Math. Statist. **19**, 177–189.

P. Gaenssler (1983), *Empirical processes*, IMS Lecture Notes—Monograph Series, vol. 3, Institute of Mathematical Statistics, Hayward, Calif.

V. Glivenko (1933), *Sulla determinazione empirica della legge di probabilità*, Giorn. Ist. Ital. Attuari **4**, 92–99.

B. V. Gnedenko and V. S. Korolyuk (1951), *On the maximum discrepancy between two empirical distributions*, Selected Transl. Math. Statist. Prob. **1** (1961), 13–16; original in Dokl. Akad. Nauk SSSR **80**, 525.

M. Kac (1946), *On the average of a certain Wiener functional and a related limit theorem in calculus of probability*, Trans. Amer. Math. Soc. **59**, 404–414.

M. Kac and A. J. F. Siegert (1947), *An explicit representation of a stationary Gaussian process*, Ann. Math. Statist. **18**, 438–442.

J. Kiefer (1967), *On Bahadur's representation of sample quantiles*, Ann. Math. Statist. **38**, 1323–1342.

______ (1969), *On the deviations in the Skorokhod-Strassen approximation scheme*, Z. Wahrsch. Verw. Gebiete **13**, 321–332.

______ (1970), *Deviations between the sample quantile process and the sample df*, Nonparametric Techniques in Statistical Inference (M. L. Puri, ed.), Cambridge Univ. Press, Cambridge.

______ (1972), *Skorohod embedding of multivariate RV's and the sample DF*, Z. Wahrsch. Verw. Gebiete **24**, 1–35.

A. N. Kolmogorov (1931), *Eine Verallgemeinerung des Laplace-Liapunovschen Stazes*, Izv. Akad. Nauk SSSR Ser. Fiz-Mat., 959–962.

______ (1933a), *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Springer-Verlag, Berlin and New York.

______ (1933b), *Sulla determinazione empirica di una legge di distribuzione*, Giorn. Ist. Ital.

Attuari 4, 83–91.
______ (1933c), *Über die Grenzwertsätze der Wahrscheinlichkeitsrechnung*, Izv. Akad. Nauk SSSR Ser. Fiz-Mat., 363–372.
J. Komlós, P. Major and G. Tusnády (1975; 1976), *An approximation of partial sums of independent rv's and the sample df.* I, II, Z. Wahrsch. Verw. Gebiete **32**, 111–131; **34**, 33–58.
M. Loève (1955), *Probability theory*, Van Nostrand, New York.
K. R. Parthasarathy (1967), *Probability measures on metric spaces*, Academic Press, New York.
E. Pitman (1979), *Some basic theory for statistical inference*, Chapman & Hall, London.
D. Pollard (1984), *Convergence of stochastic processes*, Springer-Verlag, Berlin and New York.
Yu. V. Prohorov (1956), *Convergence of random processes and limit theorems in probability theory*, Theory Probab. Appl. **1**, 157–214.
A. V. Skorokhod (1956), *Limit theorems for stochastic processes*, Theory Probab. Appl. **1**, 261–290.
______ (1961), *Studies in the theory of random processes*, Kiev Univ.; Addison-Wesley, Reading, Mass., 1965 (translation).
N. V. Smirnov (1939a), *Ob uklonenijah empiričeskoi krivoi raspredelenija*, Recueil Mathématique (Matematičeskii Sbornik) N.S. **6 (48)**, 3–26.
______ (1939b), *An estimate of divergence between empirical curves of a distribution in two independent samples*, Vestnik Moskov. Univ. **2**, 3–14. (Russian)
______ (1944), *Approximate laws of distribution of random variables from empirical data*, Uspekhi Mat. Nauk **10**, 179–206. (Russian)
V. Strassen (1964), *An invariance principle for the law of the iterated logarithm*, Z. Wahrsch. Verw. Gebiete **3**, 211–226.
______ (1967), *Almost sure behaviour of sums of independent random variables and martingales*, Proc. Fifth Berkeley Sympos. Math. Statist. and Probab., vol. 2, pp. 315–343, Univ. of California Press, Berkeley, Calif.

MIKLÓS CSÖRGŐ

BULLETIN (New Series) OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 17, Number 1, July 1987

0273-0979/87 $1.00 + $.25 per page

Enigmas of chance. An autobiography, by Mark Kac, Harper and Row, New York, 1985, xxvii + 163 pp., $18.95. ISBN 0-06-015433-0

This wonderfully lively and colorful autobiography tells the story of a man who as a teen-ager fell under the spell of mathematics, never gave it up, and grew to become a brilliant, creative mathematician.

Born in 1914, literally with the opening gunfire of World War I, in Krzemieniec, a town in czarist Russia, as a son of a middle-class Jewish family, Mark Kac was raised in an intellectual tradition. His father held a Ph.D. degree in philosophy from Leipzig and an advanced degree from the University of Moscow, and when needed earned an income by tutoring in mathematics, classical languages, and history.

In 1925, Mark was admitted to the Lycée of Krzemieniec, a school with a long tradition and ambitious standards. In 1930, at the age of sixteen, he achieved his first success in mathematics: he found a new derivation of Cardano's solution of cubic equations and showed it to his mathematics teacher, who sent it on to a journal. By a chain of circumstances the paper