The Annals of Probability
1979, Vol. 7, No. 2, 193–243

SPECIAL INVITED PAPER

EMPIRICAL PROCESSES: A SURVEY OF RESULTS FOR INDEPENDENT AND IDENTICALLY DISTRIBUTED RANDOM VARIABLES

BY PETER GAENSSLER[1] AND WINFRIED STUTE[1]

Ruhr-University of Bochum, West Germany

We consider a sequence $\xi_1, \xi_2, \cdots$ of independent and identically distributed random observations. Let μ_n denote the empirical distribution for the sample $\{\xi_1, \cdots, \xi_n\}$. It is the aim of the present article to give a survey of various results in the theory of empirical distributions and empirical processes. Special emphasis is given to the developments of the last ten years.

0. Introduction and basic definitions. The idea of testing hypotheses on the basis of a distribution giving equal mass to each observation has been central to modern statistical inference since 1933, when Glivenko, Cantelli and Kolmogorov published their fundamental results on the convergence of the empirical distribution function. Pyke (1972) also made some comments on much earlier references. Since then a large literature has evolved, and it is the aim of the present article to survey various aspects of this theory. Because of the complexity of the subject, this review must necessarily be incomplete. For example, we do not mention any result for the two-sample case. Similarly, in an already long paper we have no place to review the rapidly increasing literature on empirical processes which are based on weakly dependent observations (such as mixing variables). Instead, to make the survey more accessible to nonexperts, we include a discussion of the main ideas of proofs.

The basic viewpoint of the paper is no doubt a probabilistic one. The reader who is interested in more statistical applications is referred to the monograph of Durbin (1973) and the literature cited there.

Throughout, let $\xi_1, \xi_2, \cdots$ be independent and identically distributed (i.i.d.) random variables on a probability space $(\Omega, \mathfrak{F}, \mathbb{P})$ with values in a sample space $(X, \mathfrak{B})$. By μ we shall always denote the probability distribution of ξ_1 on $\mathfrak{B}$, i.e.,

$$\mu(B) = \mathbb{P}(\{\xi_1 \in B\}) \qquad \text{for all } B \in \mathfrak{B}.$$

Usually $\xi_1, \xi_2, \cdots$ will be random vectors in the Euclidean space $\mathbb{R}^k$. In this case

Received May 24, 1976; revised February 17, 1978.

[1]Now at the University of Munich.

AMS 1970 *subject classifications.* Primary 60-02, 62D05; secondary 60B10, 60F05, 60F10, 60F15, 60K99, 62E15, 60E20.

Key words and phrases. Empirical distribution, empirical process, Glivenko-Cantelli convergence, LIL's for empirical processes, weighted empirical process, invariance principle, rates of convergence, random sample size, strong approximation.

$F \equiv F_{\xi_1}$ will always be the (right-continuous) distribution function (df) of ξ_1, i.e., $F(\mathbf{t}) = \mu((-\infty, \mathbf{t}])$, where $(-\infty, \mathbf{t}]$ is the extended interval with corner $\mathbf{t} \in \mathbb{R}^k$.

In testing hypotheses about the (unknown) distribution μ, statistical inferences are usually based on the independent observations $\xi_1(\omega), \xi_2(\omega), \cdots, \omega \in \Omega$. To get a statistical approximation to μ the following concept of empirical distribution has proved useful. For each $\omega \in \Omega$ and $C \in \mathfrak{B}$ let $\mu_n^\omega(C)$ denote the fraction of those $\xi_1(\omega), \cdots, \xi_n(\omega)$ which fall into the set C. The number $\mu_n^\omega(C)$ is called the *empirical measure* of C for the sample $\{\xi_1(\omega), \cdots, \xi_n(\omega)\}$. For notational convenience we write $\mu_n(C)$ instead of $\mu_n^\omega(C)$. If $\xi_1, \xi_2, \cdots$ are random vectors in an Euclidean space, F_n will always denote the df of μ_n, called the *empirical df.*

A much more illuminating expression for $\mu_n(C)$ may be obtained in terms of the indicator function 1_C of C, namely

$$\mu_n(C) = n^{-1}\Sigma_{i=1}^n 1_C(\xi_i).$$

From this we see that $\mu_n(C)$ is a properly normalized partial sum of independent Bernoulli-variables with mean $\mu(C)$. The strong law of large numbers (SLLN) therefore implies that $\mu_n(C) \to \mu(C)$ $\mathbb{P}$-almost surely ($\mathbb{P}$-a.s.). In Part 1 of this survey various aspects of the almost sure convergence of μ_n to μ are reviewed.

According to the classical central limit theorem (CLT), a weak convergence result should involve a different normalizing factor. For this, let

$$\beta_n(C) := n^{\frac{1}{2}}(\mu_n(C) - \mu(C)) = n^{-\frac{1}{2}}\Sigma_{i=1}^n(1_C(\xi_i) - \mu(C)).$$

Then, by the CLT, $\beta_n(C)$ is weakly convergent to a normal random variable with zero mean and variance $\mu(C)(1 - \mu(C))$. If C ranges over some parameter set $\mathcal{C} \subset \mathfrak{B}$, the collection of random variables $\{\beta_n(C) : C \in \mathcal{C}\}$ will be called the *empirical $\mathcal{C}$-process*. In Part 2 certain properties of the β_n-process are investigated, especially when $\xi_1, \xi_2, \cdots$ are random vectors in $\mathbb{R}^k$ and $\mathcal{C}$ is the class of all intervals $(-\infty, \mathbf{t}]$. If in addition ξ_1 is uniformly distributed over the unit cube $I^k = [0, 1]^k$, we use the symbol α_n. For example, if $k = 1$,

$$\alpha_n(t) = n^{\frac{1}{2}}(F_n(t) - t) \qquad \text{for} \quad 0 \leqslant t \leqslant 1.$$

We shall always refer to α_n as the *uniform empirical process*, or empirical process, on $I \equiv I^1$. Since for each uniform random variable ξ, the variate $F^{-1}(\xi)$ has distribution function F (where F^{-1} is the inverse function of F), this shows that β_n is a version of $\alpha_n \circ F$, i.e., both processes have the same finite dimensional distributions. From a probabilistic point of view the study of a general β_n-process may therefore be reduced to the study of the uniform empirical process. On the other hand, the process $\beta_n \circ F^{-1}$ is a version of α_n for continuous F's. The importance of this observation lies in the fact that with its help one may construct statistics which are distribution-free. In this survey we mainly consider the Kolmogorov and Smirnov statistics, defined by

$$D_n = \sup_{0 \leqslant \lambda \leqslant 1}|F_n(\lambda) - \lambda| \text{ and } D_n^+ = \sup_{0 \leqslant \lambda \leqslant 1}(F_n(\lambda) - \lambda).$$

It is known (cf. Simpson (1951)) that the corresponding quantities are no longer distribution-free in the case of multivariate observations.

In the second part we shall review some important distributional results for the α_n-process, both in the finite sample and limiting case. As usual, convergence in distribution is defined in the sense of weak convergence of measures. Recall that a sequence $\{\nu_j\}$ of (finite) Borel-measures on a metric space (S, d) is weakly convergent to ν_0 if and only if

$$\lim_{j\to\infty}\int f d\nu_j = \int f d\nu_0$$

for all real-valued, bounded, continuous functions f on S. The distribution of an S-valued random element will always be denoted by $\mathfrak{L}\{\cdot\}$. By definition, η_j converges in distribution to η_0 if $\mathfrak{L}\{\eta_j\} \to \mathfrak{L}\{\eta_0\}$ (weakly). The notation $\nu(f)$ is frequently used for the expectation of f w.r.t. the measure ν. The uniform empirical process is considered as a random element in the space $D = D[0, 1]$ of all right-continuous functions on I with left limits. In this context the limit process is tied down Brownian motion $B°(t) = B(t) - tB(1)$, $0 \leqslant t \leqslant 1$ (Brownian bridge). We always consider versions of B with continuous sample paths.

For complete separable metric spaces S, Skorokhod (1956) proved the following result.

THEOREM 0.1. *Suppose that $\{\eta_j\}$ is a sequence of S-valued random elements such that $\mathfrak{L}\{\eta_j\} \to \mathfrak{L}\{\eta_0\}$. Then there exist versions $\hat{\eta}_j$ of η_j* (i.e., *$\hat{\eta}_j$ and η_j have the same distribution*) *on an appropriate probability space $(\hat{\Omega}, \hat{\mathfrak{F}}, \hat{\mathbb{P}})$ such that $\hat{\eta}_j \to \hat{\eta}_0$ $\hat{\mathbb{P}}$-a.s.*

This was generalized to arbitrary separable metric spaces by Dudley (1968) and to arbitrary metric spaces, if η_0 takes values in a separable subset, by Wichura (1970).

For proving weak limit theorems such a result will often enable one to assume that $\{\eta_j\}$ is itself almost surely convergent. In Part 3 we shall consider the rate of this almost sure convergence for the special case of empirical processes. For further applications we refer the reader to Pyke (1969) and Billingsley (1971).

Before giving the detailed contents of the survey we have to establish some further notation. λ_k will always denote Lebesgue measure on $\mathbb{R}^k$. $[x]$ will be the greatest integer $n \leqslant x$, and $t \wedge s$ means the minimum of $t, s \in \mathbb{R}$. The sup-norm distance between two real functions f and g will be denoted by $\rho(f, g)$. Finally, the expectation w.r.t. $\mathbb{P}$ is written $\mathbb{E}(\cdot)$.

The three chapters of the survey are organized in the following way:

Part 1:

1.1. On Glivenko-Cantelli convergence.
1.2. On the speed of Glivenko-Cantelli convergence.
1.3. Functional LIL's for the empirical process.
1.4. Strong laws for the weighted empirical process.
1.5. On convergence of empirical probability measures.

Part 1

1.1. On Glivenko-Cantelli convergence. In statistical inference the standard procedures for testing hypotheses are usually based on a random sample $\{\xi_1, \xi_2, \cdots\}$ of i.i.d. observations. For such a sample the empirical probability measure $\mu_n(C)$ is defined by

$$\mu_n(C) = n^{-1}\Sigma_{i=1}^n 1_C(\xi_i),$$

i.e., $\mu_n(C)$ is the average number of points $\xi_1, \cdots, \xi_n$ falling into a subset C of the sample space X. To be precise, we will assume that there is a σ-field $\mathcal{B}$ on X, for which all ξ_i's are measurable with distribution μ on $\mathcal{B}$. Since the variables $1_C(\xi_1), 1_C(\xi_2), \cdots$ are again i.i.d. with common mean $\mu(C)$, the SLLN implies that for each $C \in \mathcal{B}$

$$\mu_n(C) \to \mu(C) \qquad \text{as} \quad n \to \infty,$$

except on a $\mathbb{P}$-null set (which may depend on C); $\mu_n(C)$ therefore provides a reasonable estimate of $\mu(C)$ as the sample size tends to infinity. For statistical inference based on this empirical measure it is important to construct strongly consistent tests against alternatives, i.e., as more observations are added a false hypothesis on μ is eventually rejected with probability one. A suitable statistic for constructing such tests is the so-called $\mathcal{C}$-discrepancy between μ_n and μ. This is defined for each determining subclass $\mathcal{C}$ of $\mathcal{B}$ by

$$D_n(\mathcal{C}, \mu) \equiv \sup_{C\in\mathcal{C}}|\mu_n(C) - \mu(C)|.$$

(Recall that a determining class has the property that if any two measures agree on that class then they must agree on the whole of $\mathcal{B}$.) For historical reasons we shall say that $\mathcal{C}$ is a Glivenko-Cantelli (GC)-class for the measure μ if and only if

$$D_n(\mathcal{C}, \mu) \to 0 \text{ as } n \to \infty \text{ on a set of probability one.} \tag{1.1.1}$$

Clearly, when such a result holds, strongly consistent tests can be constructed. In general it is too much to expect the GC-theorem to hold with $\mathcal{C} = \mathcal{B}$. Indeed, since μ_n^ω is concentrated on the countable set $\{\xi_1(\omega), \xi_2(\omega), \cdots\}$ for every $\omega \in \Omega$, the validity of (1.1.1) in this case would show that μ must be a discrete measure. On the other hand, it follows from a straightforward application of Scheffé's lemma that $\mathfrak{B}$ is a GC-class for every discrete μ. Hence to obtain GC-results for more general μ one must consider proper subclasses of $\mathfrak{B}$.

In most of the Glivenko-Cantelli theorems known to date the sample space X is a k-dimensional Euclidean space with Borel σ-field $\mathfrak{B}_k$. The classical result in this field was obtained for real-valued i.i.d. random variables and the class of all extended half intervals $(-\infty, \lambda]$ by Glivenko (1933) for continuous df $F = F_{\xi_1}$, and by Cantelli (1933) for general F's. In terms of the empirical df F_n, their result states that

$$(1.1.2) \qquad D_n^F \equiv \sup_{-\infty<\lambda<\infty}|F_n(\lambda) - F(\lambda)| \to 0 \quad \text{as} \quad n \to \infty \quad \text{almost surely.}$$

The analogous result for $k > 1$ has been well known for a long time. The first proof is apparently due to Wolfowitz (1960). See also Dehardt (1971), where the same result is derived by using compactness arguments in the space of all monotone functions on $\mathbb{R}^k$.

In 1953 Fortet and Mourier proved several theorems on the convergence of empirical measures in separable metric spaces. In particular they showed that for Euclidean spaces the class of all halfspaces is a GC-class for μ, whenever μ is absolutely continuous w.r.t. Lebesgue measure. Wolfowitz (1954, 1960) proved the same result for general μ's. In 1955 Blum considered the larger class of all lower layers $C \subset \mathbb{R}^k$, for which $\mathbf{s} \in C$ whenever $\mathbf{s} \leqslant \mathbf{t}$ for some $\mathbf{t} \in C$, together with all sets which are obtained by reversing, one at a time, the k (partial) inequalities in the definition of a lower layer. The corresponding GC-result then holds again when μ is absolutely continuous w.r.t. Lebesgue measure on $\mathbb{R}^k$. By Ranga Rao (1962), Glivenko–Cantelli convergence also holds in full generality for the intersections of at most m halfspaces, $m \in \mathbb{N}$ fixed. Finally, Sazonov (1963) showed by an example that in general GC-convergence fails to be true for the halfspaces in an infinite dimensional space.

Using compactness arguments in the space of all bounded, closed and convex subsets of $\mathbb{R}^k$ (endowed with the Hausdorff metric) Elker (1975) proved that the class of closed Euclidean balls is a GC-class for general μ's. Topsøe, Dudley and Hoffmann-Jørgensen (1976) showed by an example that Elker's result in this full generality is no longer true for infinite dimensional Banach spaces. The class $\mathcal{C} = \mathcal{C}_k$ of all convex Borel sets in $\mathbb{R}^k$ was first considered by Ahmad (1961) and Ranga Rao (1960). They proved (1.1.1) under the assumption that μ is absolutely continuous w.r.t. Lebesgue measure λ_k. The following simple example shows that $\mathcal{C}_k$ is not a GC-class for a general measure μ if $k > 1$. For this let μ_0 be the uniform distribution on the unit circle in the Euclidean plane. For each $\omega \in \Omega$ and $n \in \mathbb{N}$ let $C_n(\omega)$ denote the convex hull of the sample $\{\xi_1(\omega), \cdots, \xi_n(\omega)\}$. Since $\mu_n^\omega(C_n(\omega)) = 1$ and $\mu_0(C_n(\omega)) = 0$, we obtain $D_n(\mathcal{C}_2, \mu) = 1$, i.e., $\mathcal{C}_2$ is far from being a GC-class in this case.

Similar arguments (see Stute (1976b)) may be applied to show that the GC-theorem is also violated for $\mathcal{C}_k$ with $k \geqslant 2$, if for some $C \in \mathcal{C}_k$ the nonatomic part μ_c of the measure μ gives positive mass to the extreme points $e(C)$ of C. In other words, the following condition

(1.1.3) $$\mu_c(e(C)) = 0 \quad \text{for all} \quad C \in \mathcal{C}_k$$

is necessary for $\mathcal{C}_k$ to be a GC-class for μ. Furthermore (1.1.3) is also a sufficient condition if $k = 2$. To see that the sufficiency part is no longer valid for dimension $k = 3$, let μ be the uniform distribution on the envelope of some convex cone. Then (1.1.3) is obviously satisfied in this case, while on the other hand $D_n(\mathcal{C}_3, \mu) = 1$ by the same arguments as for μ_0 above.

To prove sufficiency one therefore requires more than (1.1.3). The following condition was shown to be suitable by Ranga Rao (1962):

(1.1.4) $$\mu_c(\partial C) = 0 \quad \text{for all} \quad C \in \mathcal{C}_k,$$

where ∂C denotes the boundary of C. It is easy to see (cf. Gaenssler and Stute (1976)) that (1.1.4) is in particular satisfied for each measure μ which is absolutely continuous w.r.t. Lebesgue measure. Stute (1976a) showed that $\mathcal{C}_k$ is also a GC-class for those μ's which are absolutely continuous w.r.t. an arbitrary product of k σ-finite measures on the real line.

By means of methods similar to those developed in Stute (1976a) a complete solution of the GC-problem in the $\mathcal{C}_k$-case was obtained by Elker (1975). See also Elker, Pollard and Stute (1977). For alternative approaches we refer the reader to Eddy and Hartigan (1977), and Topsøe (1977).

To state Elker's result, let $M_0 = \{\mathbf{x} \in \mathbb{R}^k : \mu(\{\mathbf{x}\}) > 0\}$ be the set of all μ-atoms. For $j = 1, 2, \cdots, k$ let M_j be the countable family of all j-flats L(j-dimensional linear manifolds in $\mathbb{R}^k$) such that μ still assigns positive mass to L after removing all those $L' \in M_i$, $i = 0, 1, \cdots, j - 1$, which are included in L. Furthermore, for $C \in \mathcal{C}_k$ and $L \in M_j$ let $\partial_L C$ denote the boundary of $C \cap L$ relative to L. Then $\mathcal{C}_k$ is a GC-class for μ if and only if

(1.1.5)

$$\mu\Big(\partial_L C - \bigcup_{i=0}^{j-1} \bigcup M_i\Big) = 0 \quad \text{for all} \quad C \in \mathcal{C}_k \quad \text{and every} \quad L \in M_j,\ 1 \leq j \leq k.$$

Clearly, when μ_c assigns zero mass to each hyperplane, then M_j is empty for all $1 \leq j < k$. In this case condition (1.1.5) is the same as (1.1.4).

In place of $\mathcal{C}_k$ Topsøe (1970) considered the larger class of all Borel sets B, for which $\partial B \subset \partial C$ for some $C \in \mathcal{C}_k$. As a main result he showed that in this case (1.1.4) is both necessary and sufficient for (1.1.1).

Vapnik and Chervonenkis (1971) obtained necessary and sufficient conditions for the GC-convergence using combinatorial arguments. See also Steele (1978). Below we will describe various methods of proof for the results stated so far. As shown by Dehardt (1971) the GC-convergence may be achieved by verifying the following Glivenko-Cantelli criterion (GCC):

(GCC): For every $\varepsilon > 0$ there exist a finite covering $\mathcal{C}(1), \cdots, \mathcal{C}(m)$ of $\mathcal{C}$ and μ-integrable functions $\underline{f}_j$ and $\bar{f}_j$ defined on $\mathbb{R}^k, j = 1, \cdots, m$, such that

$$\underline{f}_j \leq 1_C \leq \bar{f}_j \quad \text{for all} \quad C \in \mathcal{C}(j) \quad \text{and} \quad \mu\big(\bar{f}_j - \underline{f}_j\big) < \varepsilon, \qquad j = 1, \cdots, m.$$

In fact, it is easy to see that in this case

$$D_n(\mathcal{C}, \mu) \leqslant \sup_{j=1,\cdots,m} \max\{|\mu_n(f_j) - \mu(f_j)|, |\mu_n(\bar{f}_j) - \mu(\bar{f}_j)|\} + \varepsilon.$$

Since the first term on the right-hand side converges to zero $\mathbb{P}$-a.s. by the SLLN, we obtain $\limsup_{n\to\infty} D_n(\mathcal{C}, \mu) \leqslant \varepsilon$ on a set of probability one and therefore (1.1.1). Gaenssler (1974) considered a family $\{h_t : t \in T\}$ of μ-integrable functions on an abstract probability space $(X, \mathfrak{B}, \mu)$, where t ranges over some compact set T. Suppose that for some base $\mathfrak{T}_0$ of the topology on T the following condition holds: (1.1.6)

For every $t_0 \in T$ and each $\varepsilon > 0$ there exists a neighborhood $S \in \mathfrak{T}_0$ of t_0 such that $\mu(h_{t_0}) - \varepsilon < \mu(\inf_{s\in S} h_s) \leqslant \mu(\sup_{s\in S} h_s) < \mu(h_{t_0}) + \varepsilon$.

Then on a set of probability one

(1.1.7) $$\sup_{t\in T}|n^{-1}\Sigma_{i=1}^n h_t(\xi_i) - \mu(h_t)| \to 0 \quad \text{as} \quad n \to \infty,$$

where $\xi_1, \xi_2, \cdots$ are independent X-valued random elements with distribution μ. The method of proof is based on an argument which is similar to that employed for (GCC) above. Clearly, $\mathcal{C}$ is a GC-class for the measure μ if and only if (1.1.7) is satisfied for $T = \mathcal{C}$ and $h_t = 1_t$, $t \in T$. The problem now becomes one of defining a compact topology on $\mathcal{C}$ for which (1.1.6) holds. If $\mathcal{C}$ consists of all closed convex subsets of a fixed cube in $\mathbb{R}^k$, then it is compact in the Hausdorff topology by the Blaschke selection theorem (see Valentine (1964)). Condition (1.1.6) is in particular fulfilled (cf. Gaenssler (1974)) if μ satisfies (1.1.4). See also Krickeberg (1976), where the classical GC-result is derived by introducing a compact topology into the class of intervals of the form $(-\infty, u)$ or $(-\infty, u]$, $u \in \overline{\mathbb{R}}$.

There is also a more measure theoretic approach to the GC-theorems. In this setup one considers an arbitrary measurable space $(X, \mathfrak{B})$ with a fixed subfield $\mathfrak{B}_0$ of $\mathfrak{B}$ and a finite measure $u|\mathfrak{B}$. Then, by definition, a subfamily $\mathcal{C}$ of $\mathfrak{B}$ is a $(\mu, \mathfrak{B}_0)$-uniformity class if and only if

$$\lim_\alpha(\sup_{C\in\mathcal{C}}|\mu_\alpha(C) - \mu(C)|) = 0$$

for every net $(\mu_\alpha)_\alpha$ of (finite) measures on $\mathfrak{B}$ converging setwise to μ on $\mathfrak{B}_0$, i.e., $\lim_\alpha \mu_\alpha(B) = \mu(B)$ for each $B \in \mathfrak{B}_0$. The following characterization of $(\mu, \mathfrak{B}_0)$-uniformity classes may be found in Stute (1976a).

Criterion. In the above notation $\mathcal{C} \subset \mathfrak{B}$ is a $(\mu, \mathfrak{B}_0)$-uniformity class if and only if for every $\varepsilon > 0$ there exists a finite partition $\pi = \pi(\varepsilon)$ of X into $\mathfrak{B}_0$-sets such that for all $C \in \mathcal{C}$

$$\mu\left(\bigcup\ \{B \in \pi : B \cap C \neq \varnothing \neq B \setminus C\}\right) \leqslant \varepsilon.$$

From this criterion it is fairly easy to see that every $(\mu, \mathfrak{B}_0)$-uniformity class is even a $(\mu, \mathfrak{B}_0^*)$-uniformity class for the smallest subfield $\mathfrak{B}_0^*$, which contains the elements of $\pi(n^{-1})$, $n = 1, 2, \cdots$. The importance of this observation lies in the fact that $\mathfrak{B}_0^*$ is countable. In fact, using the ordinary SLLN one can find a set Ω_0 of probability one, such that $\mu_n^\omega(B) \to \mu(B)$ for each $\omega \in \Omega_0$ and all $B \in \mathfrak{B}_0^*$. In other words, μ_n^ω is setwise convergent to μ on $\mathfrak{B}_0^*$ for each $\omega \in \Omega_0$. By the very definition of the uniformity class property, we thus obtain for all such ω's

$$\lim_{n\to\infty}(\sup_{C\in\mathcal{C}}|\mu_n^\omega(C) - \mu(C)|) = 0,$$

i.e., $\mathcal{C}$ is a GC-class for the measure μ.

If, for example, $\mathcal{C}$ is the class of intervals on the real line and μ has no atoms, then an appropriate partition may be obtained by dividing $\mathbb{R}$ into finitely many intervals of μ-measure less than or equal to $\varepsilon/2$. For further applications see Stute (1976a), Gaenssler and Stute (1976), and Elker et al. (1979). Elker showed by an example that the uniformity class property is strictly stronger than GC-convergence.

In most of the results stated so far the sets C were subsets of an Euclidean space with a common geometrical structure, such as balls, halfspaces or intervals. In proving the GC-result this was strongly needed to verify one of the above criteria. For arbitrary sample spaces X such a geometrical argument will not be available, so that different criteria are required to obtain GC-results in this general setup. Since, by the Borel-Cantelli lemma, $\lim\sup_{n\to\infty} D_n(\mathcal{C}, \mu) \leqslant \varepsilon$ $\mathbb{P}$-a.s. if $\sum_{n\geqslant 1}\mathbb{P}(\{D_n(\mathcal{C}, \mu) > \varepsilon\}) < \infty$, we see that $\mathcal{C}$ is a GC-class for μ if the last series converges for all $\varepsilon > 0$. To obtain suitable upper bounds for the probabilities involved, Vapnik and Chervonenkis (1971) in their approach considered the two-sample $\mathcal{C}$-discrepancy

$$\bar{D}_n(\mathcal{C}, \mu) \equiv \sup_{C\in\mathcal{C}}|\mu_n(C) - \nu_n(C)|,$$

where ν_n is the empirical measure for the sample $\{\xi_{n+1}, \cdots, \xi_{2n}\}$. Assuming that both D_n and $\bar{D}_n$ are measurable functions they showed that for all large enough n:

$$\mathbb{P}(\{D_n(\mathcal{C}, \mu) > \varepsilon\}) \leqslant 2\mathbb{P}\big(\{\bar{D}_n(\mathcal{C}, \mu) > \varepsilon/2\}\big).$$

For estimating the right-hand side the following inequality was derived:

$$\text{(1.1.8)} \qquad \mathbb{P}\big(\{\bar{D}_n(\mathcal{C}, \mu) > \varepsilon/2\}\big) \leqslant 2\exp(-\varepsilon^2 n/8)\mathbb{E}(\Delta^{\mathcal{C}}(\xi_1, \cdots, \xi_{2n})),$$

where for each r-sample $\{x_1, \cdots, x_r\}$ in X, $\Delta^{\mathcal{C}}(x_1, \cdots, x_r)$ denotes the number of different subsamples of the form $\{x_1, \cdots, x_r\} \cap C$ induced by the sets in $\mathcal{C}$. The function

$$m^{\mathcal{C}}(r) = \max \Delta^{\mathcal{C}}(x_1, \cdots, x_r),$$

where the maximum is taken over all samples of size r, is called the growth function. Clearly, $m^{\mathcal{C}}(r) \leqslant 2^r$. As shown by Vapnik and Chervonenkis (1971) the growth function has the following remarkable property: it is either identically equal to 2^r or it is majorized by the power function $r^s + 1$, where s is the minimal value of r for which the equality $m^{\mathcal{C}}(r) = 2^r$ is violated. In the second case we therefore obtain from (1.1.8):

$$\mathbb{P}\big(\{\bar{D}_n(\mathcal{C}, \mu) > \varepsilon/2\}\big) \leqslant 2m^{\mathcal{C}}(2n)\exp(-\varepsilon^2 n/8) \leqslant 2\big[(2n)^s + 1\big]\exp(-\varepsilon^2 n/8).$$

Since the corresponding series converges for each $\varepsilon > 0$, we may conclude in this case that $\mathcal{C}$ is a GC-class for a general μ.

Vapnik and Chervonenkis (1971) computed the growth function for the class of all rays $(-\infty, t]$ on the real line, and for the class of all halfspaces in the k-dimensional Euclidean space. In the first case $m^{\mathcal{C}}(r) = r + 1$, so that $s = 2$. For the halfspaces one has (cf. Harding (1967))

$$m^{\mathcal{C}}(r) = 2\Sigma_{i=0}^{k}\binom{r-1}{i} \quad \text{if} \quad r > k+1$$
$$= 2^r \quad \text{if} \quad r \leqslant k+1$$

and therefore $s = k + 2$. For the class of all closed balls in $\mathbb{R}^k$, Dudley (1976) proved $s = k + 2$. Recently Topsøe (private communication) gave a much shorter proof which is based on Radon's theorem (cf. Valentine (1964), Theorem 1.26). Since for large $\mathcal{C}$'s the function $m^{\mathcal{C}}$ has an exponential growth it is sometimes more convenient to work with the logarithm of $\Delta^{\mathcal{C}}$ (w.r.t. base 2). Let

$$V(\mathcal{C}, \mu) = \liminf_{n\to\infty} n^{-1}\mathbb{E}(\log_2 \Delta^{\mathcal{C}}(\xi_1, \cdots, \xi_n)).$$

Clearly, $0 \leqslant V(\mathcal{C}, \mu) \leqslant 1$. Vapnik and Chervonenkis (1971) showed that $V(\mathcal{C}, \mu) = 0$ is a necessary and sufficient condition for $D_n(\mathcal{C}, \mu)$ to converge to zero in probability. A main result of Steele's (1978) work is that one actually has almost sure convergence.

Since $V(\mathcal{C}, \mu)$ is not at all easy to compute, Steele (1978) also introduced a new type of a $\Delta^{\mathcal{C}}$-function which is easier to handle and which is still effective in the study of Glivenko-Cantelli convergence. He also gave estimates for the corresponding $V(\mathcal{C}, \mu)$-term, which are closely related to the conditions (1.1.3) and (1.1.4) if $\mathcal{C} = \mathcal{C}_k$.

1.2. On the speed of Glivenko-Cantelli convergence. For a sequence of i.i.d. random variables with mean zero and variance $\sigma^2 < \infty$ the Hartman-Wintner law of the iterated logarithm (LIL) implies that for the sequence $\{S_n\}$ of partial sums one has

$$\limsup_{n\to\infty}(2n \log\log n)^{-\frac{1}{2}}|S_n| = \sigma \qquad \mathbb{P}\text{-a.s.}$$

Clearly, this is a remarkable improvement of the SLLN. On the other hand, it has been pointed out in the first section of this survey that the GC-theorem is a simple consequence of the SLLN, if $\mathcal{C}$ satisfies one of the criteria mentioned there. From this one might guess that a more refined technique would also yield rates of convergence, which are strongly related to the LIL, at least if $\mathcal{C}$ is not too large.

Let us first consider a single set $C \in \mathcal{C}$ with $0 < \mu(C) < 1$. By the LIL we obtain

(1.2.1)

$$\limsup_{n\to\infty}(2\log\log n)^{-\frac{1}{2}} n^{\frac{1}{2}} |\mu_n(C) - \mu(C)| = [\mu(C)(1-\mu(C))]^{\frac{1}{2}} \qquad \mathbb{P}\text{-a.s.}$$

Thus $D_n(\mathcal{C}, \mu)$ cannot tend to zero faster than $\mathcal{O}(n^{-\frac{1}{2}}(\log\log n)^{\frac{1}{2}})$. To simplify the notation let $c(\mu) = \sup_{C \in \mathcal{C}} [\mu(C)(1 - \mu(C))]^{\frac{1}{2}}$. If $\mathcal{C}$ is a finite system of sets, we may infer from (1.2.1) that

$$\text{(1.2.2)} \qquad \limsup_{n\to\infty}(2 \log\log n)^{-\frac{1}{2}} n^{\frac{1}{2}} D_n(\mathcal{C}, \mu) = c(\mu) \qquad \mathbb{P}\text{-a.s.}$$

Note that $c(\mu) \leqslant \frac{1}{2}$ and $c(\mu) = \frac{1}{2}$ if $\mu(C) = \frac{1}{2}$ for some $C \in \mathcal{C}$.

For infinite $\mathcal{C}$'s the uniform LIL (1.2.2) is not at all trivial. As in the classical GC-theorems the first result in this field was obtained for the class of all rays $(-\infty, \lambda]$ on the real line, i.e., for the maximum deviation D_n^F between the empirical df F_n and the theoretical df F (Smirnov (1944)). Actually more is known than (1.2.2) in this case. For example, Chung (1949) derived a simple series condition for a sequence of real numbers to be in the upper class of D_n^F. In Dvoretzky, Kiefer and Wolfowitz (1956) the uniform LIL is proved by means of a sharp exponential inequality for the tail probabilities of D_n^F (see (2.4.2) below). Csáki (1968) in his approach obtained (1.2.2) as a corollary to a more general LIL for a certain class of submartingales. For verifying his assumptions for the particular submartingale $\eta_n \equiv n \cdot \sup_{-\infty<\lambda<\infty}(F_n(\lambda) - F(\lambda))$, the exact expression for the distribution of η_n is needed (see Section 2.4 below).

In the multivariate case the corresponding LIL has been shown to hold by Kiefer (1961). Instead of (2.4.2) the proof is now based on the exponential inequality (2.4.3) below. The technique of Csáki (1968) breaks down in the multivariate case, since the exact distribution of η_n is not known (except for trivial cases) if $k > 1$. In the sequel we will give a brief survey on known results for various other classes of subsets of $\mathbb{R}^k$. For the uniform distribution on [0, 1] Cassels (1951) proved (1.2.2) for the class of all subintervals of [0, 1]. The extension to arbitrary dimension is due to Zaremba (1971) and Wichura (1973). Philipp (1971, 1973) considered the class of all rectangles with sides parallel to the coordinate planes, and the class of all ellipsoids. It is plausible that the uniform LIL is also true for the class of all halfspaces, but we know of no proof of this conjecture.

For the class $\mathcal{C}_k$ of all convex Borel subsets of $\mathbb{R}^k$ the situation is completely different. By the results of 1.1 $\mathcal{C}_k$ is in particular a GC-class for the uniform distribution on I^k. Philipp (1973) showed that in this case the uniform LIL is still valid in two dimensions. At the 1973 Oberwolfach meeting he conjectured that it is no longer true for $k \geqslant 3$. It turns out that the rate of convergence for $\mathcal{C}_k$ is a function of the dimension k. In fact, using a combinatorial argument, Schmidt (1975) showed that $D_n(\mathcal{C}_k, \mu) \geqslant c_k n^{-2/(k+1)}$ for every sequence of points in I^k, where c_k is a positive constant depending only on k. This implies that the LIL is violated for $\mathcal{C}_k$ if $k \geqslant 4$. By contrast, for a random sample from the uniform distribution on I^k, Stute (1977) showed that $\mathbb{P}$-almost surely

$$D_n(\mathcal{C}_k, \mu) = \mathcal{O}(n^{-2/(k+1)}(\log n)^{a_k}),$$

where $a_k = 2(k + 1)^{-1}$ if $k \geqslant 4$ and $a_3 = \frac{3}{2}$.

Combining these results we see that $D_n(\mathcal{C}_k, \mu)$ tends to zero as $n^{-2/(k+1)}$ up to a logarithmic factor, if $k \geqslant 3$. The same result holds if μ has a bounded Lebesgue density with compact support. Furthermore, it may be shown (see Stute (1975)) that without such a condition this rate of convergence may not hold.

Clearly, statements on rates of convergence need a technically more involved proof than the GC-results in the first section. Some methods of proof for the maximum deviation D_n^F between the empirical df F_n and the theoretical df F have already been mentioned. For arbitrary $\mathcal{C}$-discrepancies such arguments break down. Accordingly, different techniques are needed to obtain the upper bounds in the results stated so far. The proof is based on finding, for each $n \in \mathbb{N}$, a finite $\mathcal{C}$-approximating system $\mathcal{C}_n$ of a specified (small) cardinality. This means that for each $C \in \mathcal{C}$ one may obtain sets $C_n', C_n'' \in \mathcal{C}_n$ such that $C_n' \subset C \subset C_n''$ and $\mu(C_n'' \setminus C_n') \leqslant 2^{-n}$, say. Classical exponential inequalities may then be applied to get suitable upper bounds for the $\mathcal{C}_n$-discrepancy on a set of probability close to one. Together these will show that $D_n(\mathcal{C}, \mu)$ has the same asymptotic behaviour as $D_n(\mathcal{C}_n, \mu)$ on a set of probability one.

Under the assumption that $D_n(\mathcal{C}, \mu)$ is a measurable function for each $n \in \mathbb{N}$ the strong version of the GC-theorem has the following weak analogue:

(1.2.3) $$\lim_{n\to\infty} \mathbb{P}(\{D_n(\mathcal{C}, \mu) > \varepsilon\}) = 0 \text{ for each } \varepsilon > 0.$$

If $\mathcal{C}$ is a finite system of sets, it follows from known results (see Cramér (1938), Chernoff (1952)) that the convergence in (1.2.3) takes place at an exponential rate. More precisely, for each $\varepsilon > 0$ it is true that

(1.2.4) $$\lim_{n\to\infty} n^{-1} \log \mathbb{P}(\{D_n(\mathcal{C}, \mu) > \varepsilon\}) = \log p(\mathcal{C}, \varepsilon),$$

where $p(\mathcal{C}, \varepsilon)$ is a nonnegative constant strictly smaller than one (cf. Hoeffding (1963)). An argument similar to (GCC) of Section 1.1 now leads to the result that (1.2.4) is also valid for certain nonfinite $\mathcal{C}$'s. As shown by Sethuraman (1964) it is in particular satisfied for various classes of convex sets.

Note that (1.2.4) is a strengthening of the (strong) GC-result, since in this case $\Sigma_{n\geqslant 1}\mathbb{P}(\{D_n(\mathcal{C}, \mu) > \varepsilon\}) < \infty$ and therefore $\limsup_{n\to\infty} D_n(\mathcal{C}, \mu) \leqslant \varepsilon$ $\mathbb{P}$-a.s. for each $\varepsilon > 0$. See also Sen (1973b) and Csörgő (1974).

1.3. Functional LIL's for the empirical process. In 1964 Strassen obtained the following striking extension of the LIL for standard Brownian motion B on $\mathbb{R}_+$. For this let S_n be the continuous function on I defined by $S_n(t) = (2n \log\log n)^{-\frac{1}{2}} B(nt)$, $n > 2$. Then on a set of probability one the sequence of functions $\{S_n\}$ is relatively compact w.r.t. the topology of uniform convergence on I. The limit set consists of all functions f of the form $f(s) = \int_0^s g(x)\,dx$ with $\int_0^1 g^2(x)\,dx \leqslant 1$. Recall that for a sequence $M = \{z_n : n \in \mathbb{N}\}$ in a metric space (Z, d) a point $z \in Z$ is a limit point of M if and only if for each $\varepsilon > 0$ there exist arbitrarily large indices $n \in \mathbb{N}$ with $d(z_n, z) < \varepsilon$. Using the well-known Skorokhod embedding scheme (cf. Skorokhod (1965)) Strassen also obtained the corresponding LIL for the partial sum process of a sequence of i.i.d. random variables with finite variance.

In this section the analogue to Strassen's result is presented for the uniform empirical process, defined by

$$\alpha_n(t) = n^{\frac{1}{2}}(F_n(t) - t), \qquad 0 \leqslant t \leqslant 1.$$

By a technique similar to that of Strassen, Finkelstein (1971) proved the following functional LIL for the sequence of α_n's.

THEOREM 1.3.1. *On a set of probability one the sequence of functions* $\{(2 \log \log n)^{-\frac{1}{2}}\alpha_n\}$ *is relatively compact* w.r.t. *the topology of uniform convergence on I. The limit set L consists of all functions f of the form* $f(s) = \int_0^s g(x)\,dx$ *such that* $\int_0^1 g^2(x)\,dx \leqslant 1$ *and* $\int_0^1 g(x)\,dx = 0$.

In Kuelbs (1976) the same result was derived from a more general strong convergence theorem for the partial sums of i.i.d. random elements in the space $D[0, 1]$. He identified the limit set L as the unit ball of the reproducing kernel Hilbert space (RKHS) with reproducing kernel R given by the covariance structure of the Brownian bridge, namely

$$\begin{aligned} R(s, t) &= s(1 - t) \qquad \text{for } 0 \leqslant s \leqslant t \leqslant 1 \\ &= t(1 - s) \qquad \text{for } 0 \leqslant t \leqslant s \leqslant 1. \end{aligned} \tag{1.3.1}$$

According to Aronszajn (1950) a RKHS is a Hilbert space $(H, \langle\ \rangle_H)$ of (real- or complex-valued) functions on some set E with the following property: there exists a kernel function $R(s, t)$ of s and t in E such that

$$R(\cdot, t) \in H \qquad \text{for all} \quad t \in E$$

and

$$\langle f, R(\cdot, t)\rangle_H = f(t) \qquad \text{for all} \quad f \in H \text{ and } t \in E.$$

Below we will show that the limit set L in Finkelstein's theorem is equal to the unit ball of the RKHS for the kernel in (1.3.1).

For this let $Q(u, t)$ be defined by

$$\begin{aligned} Q(u, t) &= 1 - t \qquad \text{for } u \leqslant t \\ &= -t \qquad \text{for } u > t. \end{aligned}$$

Then we have

$$R(s, t) = \int_0^1 Q(u, t)Q(u, s)\,du \qquad \text{for all} \quad 0 \leqslant s, t \leqslant 1. \tag{1.3.2}$$

Let G be the closed subspace of $L^2(\lambda)$ (= square-Lebesgue-integrable functions on I) spanned by the set of functions $\{Q(\cdot, t) : 0 \leqslant t \leqslant 1\}$. Next let H denote the Hilbert space of all functions of the form

$$f(s) = \int_0^1 g(u)Q(u, s)\,du, \qquad 0 \leqslant s \leqslant 1, g \in G,$$

with inner product defined by

$$\langle f_1, f_2\rangle_H \equiv \int_0^1 g_1(u)g_2(u)\,du.$$

By (1.3.2), $R(\cdot, t) \in H$ for every $t \in I$. Furthermore,

$$\langle f, R(\cdot, t)\rangle_H = \int_0^1 g(u)Q(u, t)\,du = f(t).$$

Hence H is the RKHS with reproducing kernel R. Since

$$\int_0^1 Q(u, t)\, du = 0 \qquad \text{for every} \quad 0 \leqslant t \leqslant 1$$

it follows that

$$\int_0^1 g(u)\, du = 0 \qquad \text{for every} \quad g \in G. \tag{1.3.3}$$

Thus every $f \in H$ admits a representation

$$f(s) = \int_0^1 g(u) Q(u, s)\, du = \int_0^s g(u)\, du, \qquad g \in G.$$

To show that the limit set L is equal to the unit ball of H it therefore suffices to prove, in view of (1.3.3), that every $g \in L^2(\lambda)$ with zero mean is a member of G. Using standard approximation arguments one may assume w.l.o.g. that g is a step function on I. In this case, however, it is fairly easy to see that g is a linear combination of a finite number of $Q(\cdot, t)$-functions.

Let $\Omega_0 \subset \Omega$ denote the set of all ω's for which the assertion of Theorem 1.3.1 holds. Then Ω_0 belongs to the smallest subfield of $\mathfrak{F}$ for which each α_n is measurable. Consequently it suffices to prove Finkelstein's result for a particular version of $\{\alpha_n : n \in \mathbb{N}\}$. On the other hand, it has been pointed out before that by means of the Skorokhod embedding scheme the functional LIL for the partial sum process may be simply derived from the corresponding result for Brownian motion. Similarly, one could hope to obtain 1.3.1 from an embedding of $\{\alpha_n : n \in \mathbb{N}\}$ into a Gaussian process for which simpler estimates would yield the same result.

Müller (1970) introduced a two-parameter Gaussian process K with zero means and covariance structure given by

$$\mathrm{Cov}(K(n_1, t_1), K(n_2, t_2)) = \min(n_1, n_2)[\min(t_1, t_2) - t_1 t_2].$$

This process will be discussed further in 2.1 and 3.2. In particular, such a process can be obtained by summing up independent Brownian bridges B_i°:

$$K(n, t) \equiv \Sigma_{i=1}^n B_i^0(t), \qquad 0 \leqslant t \leqslant 1, n = 1, 2, \cdots .$$

Kiefer (1972a) proved that there is a process K of the above type and a version $\{\hat{\alpha}_n\}$ of $\{\alpha_n\}$ such that

$$\sup_{0 \leqslant t \leqslant 1} |\hat{\alpha}_n(t) - n^{-\frac{1}{2}} K(n, t)| = o(1) \quad \text{a.s.} \tag{1.3.4}$$

It therefore remains to prove that the assertions of Finkelstein's result are valid if α_n is replaced by $n^{-\frac{1}{2}}K(n, \cdot)$. Actually, this is shown to hold by Csörgő and Révész (1977) in their forthcoming monograph on strong approximations in probability and statistics.

Richter (1974) extended 1.3.1 to the case where $\xi_1, \xi_2, \cdots$ are i.i.d. random vectors with arbitrary df F on $\mathbb{R}^k$.

The various applications of the functional LIL are mainly based on the following simple lemma.

LEMMA. *If $M = \{z_n\}$ is a relatively compact sequence in a metric space (Z, d) with limit set L and if N is a continuous mapping of Z into some metric space (Z', d'), then $N(M)$ is relatively compact in (Z', d') with limit set $N(L)$.*

In particular, if Z is the space of all real-valued bounded functions on I under the sup-norm metric and Z' is the real line, we may apply the lemma to the function $N(f) = \|f\| = \sup_{0 \leqslant t \leqslant 1}|f(t)|$. Thus, by 1.3.1, the sequence $\{(2 \log \log n)^{-\frac{1}{2}} n^{\frac{1}{2}} D_n\}$ is relatively compact with limit set $N(L)$ on a set of probability one. We show that $N(L) = [0, \frac{1}{2}]$. First note that $\|f\| \leqslant \frac{1}{2}$ for every $f \in L$ and therefore $N(L) \subset [0, \frac{1}{2}]$. For the converse the following two properties of L are needed (see, for example, Richter (1974)): (a) $\alpha L \subset L$ for all $0 \leqslant \alpha \leqslant 1$ and (b) for every $\varepsilon > 0$ there exists $f \in L$ with $\|f\| > \frac{1}{2} - \varepsilon$.

If $\xi_1, \xi_2, \cdots$ have df F, the set $N(L)$ is equal to the interval $[0, c(F)]$, where $c(F) = \sup_{-\infty<\lambda<\infty}[F(\lambda)(1 - F(\lambda))]^{\frac{1}{2}}$. Hence on a set of probability one each point of this interval is the limit of some subsequence of $\{(2 \log \log n)^{-\frac{1}{2}} n^{\frac{1}{2}} D_n^F\}$. Clearly, this is an improvement of the Smirnov-Chung LIL (1.2.2) for D_n^F.

By definition, α_n is unchanged by permutations of the underlying sample $\{\xi_1, \cdots, \xi_n\}$. An empirical process which takes account of the order of occurrence of the ξ_i's is obtained by defining

$$Z_n^F(s, \mathbf{t}) = \frac{[ns]}{n^{\frac{1}{2}}}(F_{[ns]}(\mathbf{t}) - F(\mathbf{t})), \qquad 0 \leqslant s \leqslant 1, \mathbf{t} \in \mathbb{R}^k.$$

If F is concentrated on I^k we need only consider Z_n^F on the cube I^{k+1}. The corresponding functional LIL for Z_n^F was obtained by Wichura (1973).

THEOREM 1.3.2. *On a set of probability one the sequence of functions* $(2 \log \log n)^{-\frac{1}{2}} Z_n^F$ *is relatively compact* w.r.t. *the topology of uniform convergence on* I^{k+1}. *The limit set* L^* *consists of all functions of the form*

$$f(s, \mathbf{t}) = \int_0^s \int_{[0, \mathbf{t}]} g(v, \mathbf{w}) \mu(d\mathbf{w})\, dv$$

such that $f(s, \mathbf{1}) \equiv 0$ *and* $\int g^2 d(\mu \otimes \lambda_1) \leqslant 1$.

To show that the Finkelstein theorem is a consequence of 1.3.2 let $N(f)$ be defined, for each bounded function f on I^2, to be the function $t \to f(1, t)$ for $0 \leqslant t \leqslant 1$. Clearly, N is continuous w.r.t. the topologies of uniform convergence on I^2 and I. Since $\alpha_n = N(Z_n^F)$ (if $F(t) = t$ on I) the above lemma implies that with probability one $\{\alpha_n\}$ is relatively compact with limit set $N(L^*)$. We show that $L = N(L^*)$. For this note that $f(1, \cdot)$ has derivative $g_1 : w \to \int_0^1 g(v, w)\, dv$ if f has derivative g, and that $\int g_1^2\, d\lambda_1 \leqslant 1$ if $\int g^2\, d(\lambda_1 \otimes \lambda_1) \leqslant 1$. Hence $N(L^*) \subset L$. Conversely, each $f_1 \in L$ determines an element f of L^* by putting $f(s, t) = s f_1(t)$ (where the derivative of f is given by $g(v, w) = g_1(w)$ if g_1 is the derivative of f_1). Since $f_1 = N(f)$ this shows $L \subset N(L^*)$ and therefore $L = N(L^*)$.

Similarly, the following one-parameter version of the Smirnov-Chung LIL for the maximum deviation of F_n and F may be derived from 1.3.2 (cf. Wichura (1973), page 282).

COROLLARY 1.3.3. *On a set of probability one the sequence of functions* $s \to (2n \log \log n)^{-\frac{1}{2}}[ns] D_{[ns]}^F$ *is relatively compact* w.r.t. *the topology of uniform convergence on* I, *and its limit points coincide with the set of all nonnegative functions of the* *form*

$$f(s) = \int_0^s g(x)\,dx \quad \text{with} \quad \int_0^1 g^2\,d\lambda_1 \leqslant c^2(F).$$

1.4. Strong laws for the weighted empirical process. In this section various aspects of the strong law behaviour of the weighted empirical process are investigated. Let ψ denote a preassigned nonnegative (reasonable) function on I. The weighted empirical process is then defined by

$$\gamma_n(t) \equiv \gamma_n(t, F) = n^{\frac{1}{2}}\psi(t)(F_n(t) - F(t)), \qquad t \in \mathbb{R}.$$

Possible choices for ψ may be found in Section 2.2, and in 2.5, where the corresponding weak results are considered. For example, if F is concentrated on some bounded interval $[a, b]$, then $F_n(t)$ will usually yield a less satisfactory approximation of $F(t)$ for t's in a small neighbourhood of a or b. A weighted discrepancy might then be helpful for detecting certain properties of F over these portions of $[a, b]$.

By the usual arguments we may restrict our attention to the case when $\xi_1, \xi_2, \cdots$ are uniformly distributed on I. By Finkelstein's result 1.3.1 (and the lemma in 1.3) we obtain for each bounded ψ that on a set of probability one the sequence $\{(2 \log\log n)^{-\frac{1}{2}}\gamma_n\}$ is relatively compact in the topology of uniform convergence on I, with limit set $L_\psi \equiv \{\psi f : f \in L\}$. On the other hand, if ψ is unbounded on the interval $[\varepsilon, 1 - \varepsilon]$ for some $0 < \varepsilon < 1/2$, the Smirnov-Chung LIL implies that $\limsup_{n\to\infty} (2 \log\log n)^{-\frac{1}{2}}\|\gamma_n\| = \infty$ $\mathbb{P}$-a.s. Consequently, relative compactness may only be expected for those ψ's which are bounded on every interior interval of I. We therefore have to focus our attention on the behavior of ψ near 0 and 1. It follows from a result of Baxter (1955) that for

$$(1.4.1) \qquad \begin{aligned} \psi_0(t) &= [t(1 - t)]^{-\frac{1}{2}} && \text{for } 0 < t < 1 \\ &= 0 && \text{otherwise,} \end{aligned}$$

one gets $\limsup_{n\to\infty}(2 \log\log n)^{-\frac{1}{2}}\|\gamma_n\| = \infty$ $\mathbb{P}$-a.s. Hence we need only consider those ψ's which do not grow too fast at the endpoints of I.

Because $g(0) = g(1) = 0$ for all $g \in L_\psi$ it suffices to find conditions on ψ so that, on a set of probability close to one, $(2 \log\log n)^{-\frac{1}{2}}\gamma_n(t)$ is close to zero as soon as t is near 0 or 1 and n is large enough. For ψ satisfying certain regularity assumptions, the following condition was shown to be appropriate by James (1975):

$$(1.4.2) \qquad \int_0^1 \frac{\psi^2(t)}{\log\log t^{-1}(1 - t)^{-1}}\,dt < \infty.$$

Condition (1.4.2) is slightly weaker than square-integrability of ψ. It is easy to check that the above integral diverges for the weight function ψ_0 as defined in (1.4.1). As a main result James (1975) obtained the following

THEOREM 1.4.1. *Let ψ be a nonnegative function on I such that for some $0 < \delta \leqslant 1/2$, $t^{\frac{1}{2}}\psi(t)$ is monotone increasing on $(0, \delta]$, $(1-t)^{\frac{1}{2}}\psi(t)$ is nondecreasing on $[1-\delta, 1)$, and ψ is bounded on $[\delta, 1-\delta]$. Then, if* (1.4.2) *holds, the sequence $\{(2 \log\log n)^{-\frac{1}{2}}\gamma_n\}$ is relatively compact with limit set L_ψ on a set of probability one.*

Conversely, if the integral in (1.4.2) *diverges one gets*

$$\limsup_{n\to\infty}(2 \log\log n)^{-\frac{1}{2}}\|\gamma_n\| = \infty \qquad \mathbb{P}\text{-a.s.} \tag{1.4.3}$$

Finkelstein (1971, inequality (3)) proved that

$$|f(t)| \leqslant (t(1-t))^{\frac{1}{2}} \equiv q_0(t) \qquad \text{for all} \quad 0 \leqslant t \leqslant 1 \quad \text{and} \quad f \in L,$$

i.e., $\sup\{\|\psi f\| : f \in L\} \leqslant \|\psi q_0\|$. To prove equality consider the functions f_s defined, for each $0 < s < 1$, by $f_s(0) = 0$, $f_s(s) = q_0(s)$, $f_s(1) = 0$ and linear in between. Clearly, $f_s \in L$. Furthermore,

$$\begin{aligned}\sup\{\|\psi f\| : f \in L\} &\geqslant \sup\{\|\psi f_s\| : 0 < s < 1\}\\ &\geqslant \sup\{|\psi(s)|f_s(s) : 0 < s < 1\} = \|\psi q_0\|.\end{aligned}$$

An application of the last theorem therefore leads to the following

COROLLARY 1.4.2 (James). *Under the assumptions of* 1.4.1 (*including* (1.4.2)) *we have*

$$\limsup_{n\to\infty}(2 \log\log n)^{-\frac{1}{2}}\|\gamma_n\| = \|\psi q_0\| \ \mathbb{P}\text{-a.s.}$$

Since the weight function ψ_0 does not satisfy the condition (1.4.2), Theorem 1.4.1 provides another proof of (1.4.3) in this case. In spite of this, there could be some sequence $\{a_n\}$ of norming constants such that $\limsup_{n\to\infty} a_n\|\psi_0\alpha_n\|$ is finite and positive on a set of probability one; but Csáki (1974a) has shown that such sequences do not exist.

THEOREM 1.4.3. *For each sequence $\{a_n\}$ of norming constants we have* $\mathbb{P}$-a.s.

$$\limsup_{n\to\infty} a_n\|\psi_0\alpha_n\| = 0 \quad \text{or} \quad \infty$$

according as

$$\Sigma a_n^2/n \quad \textit{converges or diverges.}$$

If the series diverges the assertion is an immediate consequence of the Borel-Cantelli lemma. For, if $\Sigma c_n = \infty$, then $U_{1:n} := \min(\xi_1, \cdots, \xi_n) < c_n$ infinitely often with probability one. Since, for each $0 < t_0 < 1$,

$$a_n\|\psi_0\alpha_n\| \geqslant \frac{a_n n^{\frac{1}{2}}|F_n(t_0) - F_n(t_0 - 0)|}{2(t_0(1-t_0))^{\frac{1}{2}}},$$

we see that with $t_0 \equiv U_{1:n}$ and $c_n \doteq \varepsilon^2 a_n^2 n^{-1}$

$$a_n\|\psi_0\alpha_n\| \geqslant (2\varepsilon)^{-1} \quad \text{infinitely often with probability one.}$$

With $\varepsilon\downarrow 0$ this proves one part of 1.4.3. The converse half is more technical and must be omitted here.

We now consider the weight function ψ_ε which is obtained by truncating ψ_0 at $t = \varepsilon$ and $t = 1 - \varepsilon$, i.e.,

$$\begin{aligned}\psi_\varepsilon(t) &= (t(1-t))^{-\frac{1}{2}} && \text{if } \varepsilon \leqslant t \leqslant 1-\varepsilon \\ &= 0 && \text{otherwise, } 0 < \varepsilon < \tfrac{1}{2}.\end{aligned}$$

Since ψ_ε is bounded, the integrability condition (1.4.2) is trivially satisfied. Thus, by Corollary 1.4.2,

(1.4.4)

$$\limsup_{n\to\infty}(2\log\log n)^{-\frac{1}{2}} n^{\frac{1}{2}} \sup_{\varepsilon\leqslant t\leqslant 1-\varepsilon} \frac{|F_n(t)-t|}{(t(1-t))^{\frac{1}{2}}} = \|\psi_\varepsilon q_0\| = 1 \qquad \mathbb{P}\text{-a.s.}$$

It is now natural to ask what happens if ε is replaced by a sequence $\{\varepsilon_n\}$ tending to zero as $n \to \infty$. First note that as a consequence of Theorem 2 of Baxter (1955) one obtains

$$\text{(1.4.5)} \qquad \limsup_{n\to\infty} \frac{n^{\frac{1}{2}}|F_n(1/n) - 1/n)|}{((2/n)(1-1/n)\log\log n)^{\frac{1}{2}}} = \infty \qquad \mathbb{P}\text{-a.s.}$$

Thus, in order that the right-hand side of (1.4.4) (with ε replaced by ε_n) should be finite, one has to consider sequences $\{\varepsilon_n\}$ for which $n\varepsilon_n \to \infty$. Csörgő and Révész (1974) proved that for $\varepsilon_n = n^{-1}\log^4 n$ one gets

(1.4.6)

$$\limsup_{n\to\infty}(2\log\log n)^{-\frac{1}{2}} n^{\frac{1}{2}} \sup_{\varepsilon_n\leqslant t\leqslant 1-\varepsilon_n} \frac{|F_n(t)-t|}{(t(1-t))^{\frac{1}{2}}} = 2^{\frac{1}{2}} \qquad \mathbb{P}\text{-a.s.}$$

Notice that the constant on the right-hand side is now $2^{\frac{1}{2}}$ instead of 1. This is strongly related to a different LIL-type behaviour of the two-parameter Brownian motion $B = (B(s, t))$ (cf. Paranjape and Park (1973a)). Actually, the proof of (1.4.6) uses an embedding (cf. 3.2) of the empirical process into a two-parameter Gaussian process K of the form

$$K(s,t) = B(s,t) - tB(s,1), \qquad 0 \leqslant t \leqslant 1,\ 0 \leqslant s < \infty.$$

The factor $\log^4 n$ is determined by the error of this embedding, which is given by the estimate (3.2.3).

Csáki (1977) (see also Shorack (1977)) generalized (1.4.6) to sequences $\{\varepsilon_n\}$ satisfying

$$n\varepsilon_n(\log\log n)^{-1} \to \infty \quad \text{and} \quad (\log\log \varepsilon_n^{-1})(\log\log n)^{-1} \to c$$

for some $0 \leqslant c \leqslant 1$. In this case the right-hand side of (1.4.6) has to be replaced by $(c+1)^{\frac{1}{2}}$. This is closely connected with a pointwise LIL proved by Eicker (1970) and Kiefer (1972b):

$$\limsup_{n\to\infty}(2\log\log n)^{-\frac{1}{2}} n^{\frac{1}{2}}|F_n(\varepsilon_n) - \varepsilon_n| / (\varepsilon_n(1-\varepsilon_n))^{\frac{1}{2}} = 1 \qquad \mathbb{P}\text{-a.s.,}$$

whenever $n\varepsilon_n(\log\log n)^{-1} \to \infty$. Note that by (1.4.5) the LIL is violated for $\varepsilon_n = n^{-1}$. This is related to the fact that for triangular arrays of random variables no complete analogue of the Hartmen-Wintner LIL is available.

We will now investigate necessary and sufficient conditions for the Glivenko-Cantelli convergence of the weighted discrepancy between F_n and F, i.e., for

$$D_n^F(\psi) \equiv \sup_{t\in\mathbb{R}}\psi(t)|F_n(t) - F(t)|.$$

Again we restrict our attention to the case when $\xi_1, \xi_2, \cdots$ are uniformly distributed on I. If ψ is bounded on I then by the ordinary GC-theorem $D_n(\psi) \equiv D_n^F(\psi) \to 0$ $\mathbb{P}$-a.s. So, in order to obtain nontrivial results, we must consider unbounded ψ's. Let ψ be always such that it is bounded on $[\varepsilon, 1-\varepsilon]$ for some $\varepsilon > 0$, nonincreasing [nondecreasing] and continuous on $(0, \varepsilon)[(1-\varepsilon, 1)]$ with $\psi(0+0) = \psi(1-0) = \infty$. For such a ψ Wellner (1977a) showed that $D_n(\psi) \to 0$ $\mathbb{P}$-a.s. whenever $\int_0^1\psi(t)\,dt < \infty$. This is useful for proving strong laws of large numbers for linear functions of order statistics.

It is also true that

$$\limsup_{n\to\infty} D_n(\psi) = \infty \quad \mathbb{P}\text{-a.s.}$$

if $\int_0^1\psi(t)\,dt$ diverges. For this, it suffices to prove that for every $r \in \mathbb{N}$

$$\limsup_{n\to\infty} D_n(\psi) \geqslant r \quad \mathbb{P}\text{-a.s.} \tag{1.4.7}$$

By symmetry, we may assume w.l.o.g. that $\int_0^\varepsilon\psi(t)\,dt = \infty$. Then for all sufficiently large n, say $n \geqslant n_0$, the number $c_n = \max\{t \leqslant \varepsilon : \psi(t) = 2nr\}$ is well defined. Furthermore,

$$\Sigma_{n\geqslant n_0}c_n = (2r)^{-1}\Sigma_{n\geqslant n_0}c_n[2(n+1)r - 2nr] = \infty,$$

whence by virtue of the Borel-Cantelli lemma, $\xi_n \leqslant c$ (and therefore $U_{1:n} := \min(\xi_1, \cdots, \xi_n) \leqslant c_n$) infinitely often with probability one. Since ψ is continuous and nonincreasing on $(0, \varepsilon)$, this implies

$$\begin{aligned}\sup_{0\leqslant t\leqslant 1}\psi(t)|F_n(t) - t| &\geqslant 2^{-1}\psi(U_{1:n})|F_n(U_{1:n}) - F_n(U_{1:n} - 0)| \\ &\geqslant (2n)^{-1}\psi(c_n) = r\end{aligned}$$

infinitely often with probability one, whence (1.4.7).

The above results imply that for the weight function

$$\begin{aligned}\psi_1(t) &= t^{-1} \quad &\text{if } 0 < t \leqslant 1 \\ &= 0 \quad &\text{if } t = 0\end{aligned}$$

one gets $\limsup_{n\to\infty} D_n(\psi) = \infty$ $\mathbb{P}$-a.s. By a result of Shorack and Wellner (1977), $\limsup_{n\to\infty} a_n D_n(\psi_1)$ is either zero or infinite for each norming sequence $\{a_n\}$ with na_n^{-1} increasing. They showed that

$$\limsup_{n\to\infty} a_n D_n(\psi) = 0 \quad \text{or} \quad \infty \qquad \mathbb{P}\text{-a.s.}$$

according as

$$\Sigma a_n/n \quad \text{converges or diverges.}$$

Wellner (1978) also investigated the strong law behaviour of $\sup_{\varepsilon_n \leqslant t \leqslant 1}|F_n(t) - t|/t$ for various sequences $\{\varepsilon_n\}$ tending to zero.

1.5. On convergence of empirical probability measures. For an arbitrary separable metric space (S, d), the following property is sometimes useful to establish weak convergence on S (cf. Parthasarathy (1967), page 47, Theorem 6.6): there exists a countable class $\{f_j\}$ of bounded, continuous functions on S such that for each sequence $\nu_0, \nu_1, \nu_2 \cdots$ of finite Borel measures on S we have $\nu_n \to \nu_0$ weakly if and only if

$$\lim_{n\to\infty} \int f_j d\nu_n = \int f_j d\nu_0, \qquad j = 1, 2, \cdots.$$

We shall say that $\{f_j\}$ is a determining class for weak convergence on S. If $\xi_1, \xi_2, \cdots$ is a sequence of i.i.d. S-valued random elements with distribution μ, the existence of such a class may be applied to show that

$$\mu_n^{\omega} \to \mu \quad \text{weakly as} \quad n \to \infty \qquad \mathbb{P}\text{-a.s.} \tag{1.5.1}$$

In fact, by the SLLN we have $\mu_n(f_j) \to \mu(f_j)$ $\mathbb{P}$-almost surely for all $j = 1, 2, \cdots$, so that (1.5.1) is an easy consequence of the very definition of a determining class.

In this full generality (1.5.1) is apparently due to Varadarajan (1958). If S is the Euclidean space (1.5.1) follows immediately from the ordinary Glivenko-Cantelli theorem. Conversely, if μ has a continuous df F on $\mathbb{R}^k$, the Glivenko-Cantelli convergence may be derived from (1.5.1) by means of the Pólya-Cantelli theorem (cf. Pólya-Szegő (1972), page 81, example 127).

It is well known (cf. Prokhorov (1956)) that for complete separable metric spaces (S, d) weak convergence in the space of all finite Borel measures on S may be metrized by

$$r(\mu, \nu) := \inf\{\varepsilon > 0 : \mu(F) \leqslant \nu(F^{\varepsilon}) + \varepsilon \quad \text{for all closed} \quad F \subset S\}.$$

Here $F^{\varepsilon} := \{x \in S : d(x, F) < \varepsilon\}$ is the outer ε-parallel set of F. This result was generalized to arbitrary separable metric spaces by Dudley (1968). Because of (1.5.1) we therefore get $r(\mu_n, \mu) \to 0$ $\mathbb{P}$-almost surely. As to the speed of this convergence Zuker (1974) showed that for uniform samples on I^k

$$r(\mu_n, \mu) = \mathcal{O}\left(n^{-1/(k+1)}(\log n)^{\frac{1}{2}}\right) \qquad \mathbb{P}\text{-a.s.} \tag{1.5.2}$$

Dudley (1969) obtained similar estimates for the mean of $r(\mu_n, \mu)$. For $k = 1$, he also related $r(\mu_n, \mu)$ to the maximum deviation between F_n and F:

$$2^{-1}D_n^F \leqslant r(\mu_n, \mu) \leqslant 2D_n^F.$$

From this and the Smirnov-Chung LIL one may infer, using a tail event argument, that

$$\limsup_{n\to\infty}(2 \log\log n)^{-\frac{1}{2}} n^{\frac{1}{2}} r(\mu_n, \mu) = C \qquad \mathbb{P}\text{-a.s.}$$

for some constant $\frac{1}{4} \leqslant C \leqslant 1$. Hence (1.5.2) is not best possible in this case. Zuker (1974), Proposition IV. 4.4) proved that for $0 \leqslant \alpha < k$ there is a probability measure $\mu^{k,\alpha}$ on I^k and a $\delta > 0$ such that for all n, $\mathbb{P}\{\rho(\mu_n, \mu^{k,\alpha})n^{1/(2+\alpha)} > \delta\} >$

δ. It follows that $\limsup_{n\to\infty}\rho(\mu_n, \mu^{k,\alpha})n^{1/(2+\alpha)} \geqslant \delta$ $\mathbb{P}$-a.s. We also refer to Fortet and Mourier (1953), Dudley (1966b) and Zuker (1974) for the convergence of μ_n to μ in the so-called dual-bounded Lipschitz metric.

Part 2

2.1. Functional limit theorems (invariance principles). The study of invariance principles for the empirical process originally started with Doob's (1949) "Heuristic approach to the Kolmogorov-Smirnov theorems." Noticing that

$$\mathbb{E}(\alpha_n(t)) = 0, \qquad 0 \leqslant t \leqslant 1, \tag{2.1.1}$$

and

$$\mathbb{E}(\alpha_n(t)\alpha_n(s)) = t \wedge s - st, \qquad 0 \leqslant s, t \leqslant 1, \tag{2.1.2}$$

are the same as the corresponding moments of a tied-down Brownian motion $B^\circ(t) = B(t) - tB(1)$, $0 \leqslant t \leqslant 1$, and that, by the multivariate central limit theorem, the finite-dimensional distributions of α_n are asymptotically the same as those of B°, he assumed ("until a contradiction frustrates our devotion to heuristic reasoning") that in calculating asymptotic distributional results for the α_n-processes, one may simply replace the α_n's by B°. The argument for this reasoning was justified by calculation in the following now classical result by Donsker (1951, 1952):

For each function $H : D \to \mathbb{R}$, which is continuous in the uniform topology on D, $\mathfrak{L}\{H(\alpha_n)\} \to \mathfrak{L}\{H(B^\circ)\}$ as $n \to \infty$.

Note that there is a slight gap in the last statement, since $\mathfrak{L}\{H(\alpha_n)\}$ need not be well defined for a general continuous H. This comes from the fact (see Chibisov (1965)) that α_n is nonmeasurable for the Borel-σ-field of the sup-norm (ρ –) metric on D. In fact, as was pointed out in Dudley (1966a), what has been really proved by Donsker is that for all such H

$$\lim_{n\to\infty}\mathbb{P}^*(\{H(\alpha_n) \leqslant \lambda\}) = \mathbb{P}(\{H(B^\circ) \leqslant \lambda\}) \tag{2.1.3}$$

at all continuity points λ of the df of $H(B^\circ)$. Here

$$\mathbb{P}^*(B) := \inf\{\mathbb{P}(A) : B \subset A \in \mathfrak{F}\}, \qquad B \subset \Omega,$$

denotes the outer measure of $\mathbb{P}$. Note that since Brownian motion has continuous sample paths, the process B° takes values in a separable subspace of D. Hence the right-hand side of (2.1.3) is well defined. Anyway, letting

$$H(f) := \sup_{0\leqslant t\leqslant 1}|f(t)| \quad \text{and} \quad H(f) := \sup_{0\leqslant t\leqslant 1} f(t), \qquad f \in D,$$

respectively, no measurability questions will occur here. We therefore obtain for all $\lambda \in \mathbb{R}$

$$\lim_{n\to\infty}\mathbb{P}\left(\left\{n^{\frac{1}{2}}D_n \leqslant \lambda\right\}\right) = \mathbb{P}(\{\sup_{0\leqslant t\leqslant 1}|B^\circ(t)| \leqslant \lambda\})$$

and

$$\lim_{n\to\infty}\mathbb{P}\left(\left\{n^{\frac{1}{2}}D_n^+ \leqslant \lambda\right\}\right) = \mathbb{P}(\{\sup_{0\leqslant t\leqslant 1}B^\circ(t) \leqslant \lambda\}).$$

Though these results look very promising at first sight it is not an easy task to identify the limit distributions on the right-hand side. For a more detailed discussion of this see 2.2 below. Moreover, the (correct) statement of Donsker's result in this form is not quite satisfactory in so far as only real functions of α_n are considered. The so-called functional version, however, could not be formulated before 1956, when Prokhorov published his fundamental paper on convergence of random processes and limit theorems in probability theory. In this connection it is important to know that there is a metric d on D such that (D, d) is separable and complete (cf. Skorokhod (1956) and Billingsley (1968)) and such that the corresponding σ-field $\mathfrak{B}(d)$ of Borel sets in D equals the σ-field spanned by the projections $\pi_t : f \to f(t)$, $0 \leqslant t \leqslant 1, f \in D$. In particular this shows that α_n for each $n \in \mathbb{N}$ is a random element in (D, d) with $\mathfrak{L}\{\alpha_n\}$ completely determined by its finite-dimensional marginals. The functional version of Donsker's result is now concerned with the weak convergence of $\mathfrak{L}\{\alpha_n\}$ on $\mathfrak{B}(d)$ to some limit measure P_0. We already remarked that the finite-dimensional distributions of α_n were weakly convergent to the corresponding finite-dimensional distributions of $B°$. Thus, if $\mathfrak{L}\{\alpha_n\} \to P_0$ for some P_0, then by the uniqueness theorem necessarily $P_0 = \mathfrak{L}\{B°\}$. In order to achieve this convergence it remains to show that in the weak topology $\mathfrak{L}\{\alpha_n\}$, $n = 1, 2, \cdots$ is relatively sequentially compact (for in this case every limit point (and such exist) must be equal to $\mathfrak{L}\{B°\}$). It follows from Prokhorov's (1956) results that $\mathfrak{L}\{\alpha_n\}$, $n = 1, 2, \cdots$ is relatively sequentially compact if and only if $\{\mathfrak{L}\{\alpha_n\} : n \in \mathbb{N}\}$ is tight. Using an Arzela-Ascoli type theorem for the space D and noticing that $\alpha_n(0) = 0$ $\mathbb{P}$-a.s. for all $n \in \mathbb{N}$, by Theorem 15.2 in Billingsley (1968), it therefore remains to show that

$$(2.1.4) \qquad \lim_{\delta\to 0}\limsup_{n\to\infty}\mathbb{P}\big(\{\omega'_{\alpha_n}(\delta) \geqslant \varepsilon\}\big) = 0 \qquad \text{for all } \varepsilon > 0,$$

where $\omega'_f(\delta) := \inf_{\{t_i\}}\max_{0<i\leqslant r}\operatorname{osc}_f[t_{i-1}, t_i)$ and the infimum extends over the finite sets $\{t_i\}$ of points $0 = t_0 < t_1 < \cdots < t_r = 1$ satisfying $t_i - t_{i-1} > \delta$, $i = 1, \cdots, r$. In practice, however, it is much more convenient to work with the modulus $\omega''_f(\delta) := \sup \min\{|f(t) - f(t_1)|, |f(t_2) - f(t)|\}$, where the supremum extends over all $t - \delta \leqslant t_2 \leqslant t \leqslant t_1 \leqslant t + \delta$; this may always be done if the expected limit process has sample paths which are left continuous at 1 (cf. Billingsley (1968), Theorem 15.4 or Gaenssler and Stute (1977), Satz 8.5.6). For the ω''-criterion to be applicable in our case one needs easily verifiable conditions which imply (2.1.4) (with ω' replaced by ω'').

Criterion. Let η_n, $n = 1, 2, \cdots$ be an arbitrary sequence of random elements in D. Suppose that for some constants $a, K \geqslant 0$ and $b > 1$ and some continuous nondecreasing function $G : I \to \mathbb{R}$

$$(2.1.5) \qquad \mathbb{P}(\{|\eta_n(t) - \eta_n(t_1)| > \varepsilon, |\eta_n(t_2) - \eta_n(t)| > \varepsilon\}) \leqslant K\varepsilon^{-a}|G(t_2) - G(t_1)|^b$$

for all $0 \leqslant t_1 \leqslant t \leqslant t_2 \leqslant 1$, $\varepsilon > 0$ and $n \in \mathbb{N}$. Then

$$(2.1.6) \qquad \lim_{\delta\to 0}\limsup_{n\to\infty}\mathbb{P}\big(\{\omega''_{\eta_n}(\delta) \geqslant \varepsilon\}\big) = 0 \qquad \text{for all } \varepsilon > 0.$$

A proof may be found, e.g., in Billingsley (1968), or in Gaenssler and Stute (1977), the arguments being completely different. While in Billingsley's monograph several useful fluctuation inequalities are central to the proof, it is pointed out in the second textbook (Remark 8.5.15) that a condition like (2.1.6) (for fixed n without the lim sup) occurs when studying the realizability of stochastic processes in the space of all functions with right and left limits. In particular (2.1.6) is true under the assumptions of (2.1.5). By Chebyshev's inequality (2.1.5) is easily seen to hold (with $a = a_1 + a_2$), if for some $a_1, a_2 \geq 0$

$$(2.1.7) \qquad \mathbb{E}(|\eta_n(t) - \eta_n(t_1)|^{a_1}|\eta_n(t_2) - \eta_n(t)|^{a_2}) \leq K|G(t_2) - G(t_1)|^b.$$

For $G(t) = t$, $0 \leq t \leq 1$, the above criterion is due to Chentsov (1956). As far as the uniform empirical process $\eta_n = \alpha_n$ is concerned, it is easy to see that (2.1.5) holds with $a_1 = a_2 = 2$, $K = 6$, $b = 2$ and $G(t) = t$, $0 \leq t \leq 1$. Summarizing, we thus obtain the following invariance principle for the uniform empirical process.

THEOREM 2.1.1. *On* $(D, \mathfrak{B}(d))$, $\mathfrak{L}\{\alpha_n\} \to \mathfrak{L}\{B^\circ\}$ *as* $n \to \infty$.

One of the most important properties of weak convergence is that it is preserved under continuous mappings. In our case this means that $\mathfrak{L}\{H(\alpha_n)\} \to \mathfrak{L}\{H(B^\circ)\}$ for each d-continuous function on D. From the point of application, however, it would be more desirable to obtain an analogous result for the (larger) class of all ρ-continuous functions on D. Surprisingly this is an easy consequence of the continuous sample path property of B° together with the fact that every d-convergent sequence in D is ρ-convergent, if the limit belongs to $C = C([0, 1])$. A proof of it may either use a general continuous mapping theorem (cf. Billingsley (1968), Theorem 5.1) or may involve almost surely convergent constructions. To sketch the second proof, let $\hat{\alpha}_n$ and $\hat{B}^\circ$ be versions of α_n and B°, respectively (defined on some p-space $(\hat{\Omega}, \hat{\mathfrak{F}}, \hat{\mathbb{P}})$), such that $d(\hat{\alpha}_n, \hat{B}^\circ) \to 0$ (see Theorem 0.1 in the introduction) and therefore $\rho(\hat{\alpha}_n, \hat{B}^\circ) \to 0$ $\hat{\mathbb{P}}$-almost surely. Hence, by continuity, $H(\hat{\alpha}_n) \to H(\hat{B}^\circ)$ $\hat{\mathbb{P}}$-almost surely and therefore

$$\mathfrak{L}\{H(\alpha_n)\} = \mathfrak{L}\{H(\hat{\alpha}_n)\} \to \mathfrak{L}\{H(\hat{B}^\circ)\} = \mathfrak{L}\{H(B^\circ)\}.$$

(We always have to assume that H is $B(d)$-measurable). In summary we obtain

COROLLARY 2.1.2. *Let* $H : D \to \mathbb{R}$ *be* $\mathfrak{B}(d)$*-measurable and* ρ*-continuous at* $B^\circ(\omega)$ *for* $\mathbb{P}$*-almost all* $\omega \in \Omega$. *Then* $\mathfrak{L}\{H(\alpha_n)\} \to \mathfrak{L}\{H(B^\circ)\}$ *as* $n \to \infty$.

A list of possible choices for H may be found in Sahler (1968) or Durbin (1973). See also Anderson and Darling (1952). Theorem 2.1.1 may be equally applied to obtain limit results in the nonuniform case. Since $f \to f \circ F$ is ρ-continuous on D for each df F, and since $\alpha_n \circ F$ is a version of β_n, Theorem 2.1.1 implies that for F's concentrated on the unit interval, $\mathfrak{L}\{\beta_n\} \to \mathfrak{L}\{B^\circ \circ F\}$ as $n \to \infty$. The covariance structure of $B^\circ \circ F$ is given by

$$R^F(s, t) = F(t \wedge s) - F(t)F(s), \qquad 0 \leq t, s \leq 1.$$

Furthermore, $B^\circ \circ F$ is tied down to zero at 0 and 1 and has continuous sample paths if F is continuous.

Before moving to the multidimensional case we also mention another proof of tightness which is indicated in problem 9, page 296 of Breiman (1968). Using the fact that the ordered sample $(U_{1:n}, \cdots, U_{n:n})$ of n independent uniformly distributed random variables $\xi_1, \cdots, \xi_n$ has the same distribution as $(S_1/S_{n+1}, \cdots, S_n/S_{n+1})$, where $S_i = \sum_{j=1}^{i} y_j$, $i = 1, 2, \cdots$ are partial sums of independent variables y_j exponentially distributed with parameter 1, the crucial condition (2.1.6) (with $\eta_n = \alpha_n$) will follow from the ordinary invariance principle for the partial sum process pertaining to the sequence y_j, $j = 1, 2, \cdots$. The arguments are slightly more involved than those used on page 286 in Breiman's book to derive Theorem 13.16 there.

The same representation of the empirical process may also be applied to construct explicitly versions $\hat{\alpha}_n$ of α_n which converge almost surely (cf. 3.1 below). For a third (though classical) approach we refer to Parthasarathy (1967). We now turn to the multidimensional version of 2.1.1. For this let $\xi_1, \xi_2, \cdots$ be independent $\mathbb{R}^k$-valued random vectors with common df F on I^k. The corresponding empirical process β_n is now defined by

$$\beta_n(\mathbf{t}) := n^{\frac{1}{2}}(F_n(\mathbf{t}) - F(\mathbf{t})), \qquad \mathbf{t} \in I^k, n \in \mathbb{N},$$

where F_n is the empirical df pertaining to $\xi_1, \cdots, \xi_n$. As an appropriate space of sample functions we consider $D_k \equiv D(I^k)$ consisting of all real functions f on I^k such that for each $\mathbf{t}$ $\lim_{n\to\infty} f(\mathbf{t}_n)$ exists for all sequences $\mathbf{t}_n$, $n = 1, 2, \cdots$ approaching $\mathbf{t}$ in some quadrant with corner $\mathbf{t}$ and such that f is continuous from above. Then it was shown by Neuhaus (1971) and Straf (1971) that, as in the one-dimensional case, there exists a separable and complete metric d_k an D_k for which tightness may be described through the behaviour of a specified modulus of continuity. Furthermore the corresponding σ-field $\mathfrak{B}(d_k)$ of Borel sets equals the smallest σ-field of subsets of D_k for which all coordinate-mappings are measurable. This implies that β_n is a random element in D_k. In order to formulate a Chentsov-type inequality in the multidimensional case, for each interval $B = (\mathbf{s}, \mathbf{t}] = \Pi_{i=1}^{k}(s_i, t_i]$ and every f, one has to consider the increment $f(B)$ of f around B, namely

$$f(B) \equiv f((\mathbf{s}, \mathbf{t}])$$
$$\equiv \Sigma_{\varepsilon_1=0,1} \cdots \Sigma_{\varepsilon_k=0,1}(-1)^{k-\Sigma\varepsilon_i} f(s_1 + \varepsilon_1(t_1 - s_1), \cdots, s_k + \varepsilon_k(t_k - s_k)).$$

Disjoint intervals $(\mathbf{s}, \mathbf{t}]$ and $(\mathbf{s}', \mathbf{t}']$ are called neighbours if they abut and if for some $i \in \{1, \cdots, k\}$ they have the same ith-face $\Pi_{j\neq i}(s_j, t_j] = \Pi_{j\neq i}(s_j', t_j']$. Let μ be an arbitrary finite measure on I^k. In dimension k the analogue of (2.1.5) is now as follows (where η_n are arbitrary random elements in D_k):

$$\text{(2.1.8)} \qquad \mathbb{P}(\{|\eta_n(B)| > \varepsilon, |\eta_n(B')| > \varepsilon\}) \leq K\varepsilon^{-a}[\mu(B \cup B')]^b$$

for every pair of neighbouring intervals in I^k. Note that in dimension 1 the measure μ and the function G correspond to each other in so far as G may be viewed as the df of μ. Moreover the former continuity assumption on G now transforms to the condition that μ has continuous marginals. In particular, if $\eta_n = \beta_n$ and μ is the distribution of ξ_1, Bickel and Wichura (1971) proved the moment inequality

$$\mathbb{E}(|\beta_n(B)|^2|\beta_n(B')|^2) \leqslant 3[\mu(B \cup B')]^2,$$

from which (2.1.8) and therefore tightness follow immediately via Chebyshev's inequality. Inspection of the finite dimensional distributions now leads to the following result (Neuhaus (1971), Straf (1971), Bickel and Wichura (1971)):

THEOREM 2.1.3. *Let $\xi_1, \xi_2, \cdots$ be independent random vectors with a continuous* df *F on I^k. Then $\mathcal{L}\{\beta_n\} \to \mathcal{L}\{B^{\circ F}\}$ on $(D_k, \mathfrak{B}(d_k))$ as $n \to \infty$, where $B^{\circ F}$ is a centered Gaussian process with continuous sample paths tied down to zero at $1 \in I^k$ and the lower boundary L_k of I^k and such that*

$$\mathrm{Cov}(B^{\circ F}(\mathbf{t}), B^{\circ F}(\mathbf{s})) = F(\mathbf{t} \wedge \mathbf{s}) - F(\mathbf{t})F(\mathbf{s}), \qquad \mathbf{t}, \mathbf{s} \in I^k.$$

Here $\mathbf{t} \wedge \mathbf{s} = (\min(t_1, s_1), \cdots, \min(t_k, s_k))$ and $L_k = \{\mathbf{t} \in I^k : \min(t_1, \cdots, t_k) = 0\}$. Theorem 2.1.3 is also a consequence of Theorem 2 in Dudley (1966a) which states that $\mathcal{L}\{\beta_n\} \to_w \mathcal{L}\{B^{\circ F}\}$ as $n \to \infty$, and where, by definition, $\mathcal{L}\{\beta_n\} \to_w \mathcal{L}\{B^{\circ F}\}$ if and only if

$$\text{(2.1.9)} \qquad \lim_{n\to\infty} \overline{\int} f d\mathcal{L}\{\beta_n\} = \lim_{n\to\infty} \int_{-} f d\mathcal{L}\{\beta_n\} = \int f d\mathcal{L}\{B^{\circ F}\}$$

for all real-valued, bounded, ρ-continuous functions on D_k($\int^-$ and $\int_-$ denote upper and lower integrals, respectively). There should be made some further remarks at this point. First note that the metric d_k was not available for $k > 1$ before 1971, so that in order to obtain a well-defined distribution of β_n one had to find a suitable (large enough) σ-algebra on D_k making β_n measurable. For this Dudley (1966a) considered the smallest σ-field $\mathfrak{S}_k$ of subsets of D_k containing all open ρ-balls $S(f, \varepsilon) := \{g \in D_k : \rho(f, g) < \varepsilon\}$, $f \in D_k$, $\varepsilon > 0$. In his thesis Wichura (1968) proved that $\mathfrak{S}_k$ equals the smallest σ-field of subsets of D_k for which all coordinate-mappings are measurable. Thus $\mathfrak{S}_k = \mathfrak{B}(d_k)$, i.e., we need not distinguish between $\mathcal{L}\{\beta_n\}|\mathfrak{S}_k$ and $\mathcal{L}\{\beta_n\}|\mathfrak{B}(d_k)$. Secondly the continuous sample path property of $B^{\circ F}$ ensures that $\mathcal{L}\{B^{\circ F}\}$ is supported by a (ρ –) separable subspace of D_k and hence may be extended to the σ-field of all (ρ –) Borel sets on D_k. This means that the right-hand side of (2.1.9) is well defined. Moreover, as was shown by Wichura (1968), when checking w-convergence on $\mathfrak{S}_k(= \mathfrak{B}(d_k))$, in the presence of a tight limit measure one may confine oneself to those f's which are also $\mathfrak{S}_k$-measurable. For such an f, (2.1.9) may easily be verified by considering versions $\hat{\beta}_n$ and $\hat{B}^{\circ F}$ of β_n and $B^{\circ F}$ such that $d_k(\hat{\beta}_n, \hat{B}^{\circ F}) \to 0$ and therefore $\rho(\hat{\beta}_n, \hat{B}^{\circ F}) \to 0$ almost surely. In fact, by the dominated convergence theorem, for each real-valued, bounded, ρ-continuous and $\mathfrak{S}_k = \mathfrak{B}(d_k)$ – measurable f:

$$\lim_{n\to\infty} \int f d\mathcal{L}\{\beta_n\} = \lim_{n\to\infty} \int f d\mathcal{L}\{\hat{\beta}_n\} = \int f d\mathcal{L}\{\hat{B}^{\circ F}\} = \int f d\mathcal{L}\{B^{\circ F}\}.$$

In summary we see that Dudley's result is equivalent to 2.1.3.

Surprisingly much more is known than 2.1.3. For this let

$$Z_n^F(s, \mathbf{t}) = \frac{[ns]}{n^{1/2}}(F_{[ns]}(\mathbf{t}) - (\mathbf{t})), \quad (s, \mathbf{t}) \in I^{k+1}$$

be the sequential empirical process as defined in 1.3. The functional law of the iterated logarithm for Z_n^F is contained in 1.3.2. To obtain a weak limit result notice that Z_n^F is a random element in D_{k+1} for each $n \in \mathbb{N}$. Furthermore, by the independence of the ξ_i's, for each interval B in I^k

$$n^{-2}\mathbb{E}\big(|j^{1/2}\beta_j(B) - r^{1/2}\beta_r(B)|^2|m^{1/2}\beta_m(B) - j^{1/2}\beta_j(B)|^2\big)$$
$$\leqslant \left[\frac{j-r}{n}\right]\mu(B)\left[\frac{m-j}{n}\right]\mu(B)$$

for all $0 \leqslant r < j \leqslant m \leqslant n$, which yields the desired Chentsov-type inequality for all neighbouring intervals in I^{k+1} with equal first face (and μ replaced by $\lambda_1 \otimes \mu$). For the remaining intervals similar calculations as those needed for the β_n-process finally show that

$$\mathbb{E}\big(|Z_n^F(B)|^2|Z_n^F(B')|^2\big) \leqslant K[(\lambda_1 \otimes \mu)(B \cup B')]^2$$

for all $n \in \mathbb{N}$ and every pair of neighbouring intervals in I^{k+1}. Inspection of the finite dimensional distributions therefore leads to the following invariance principle for the sequential empirical process (Bickel and Wichura (1971)):

THEOREM 2.1.4. *Let $\xi_1, \xi_2, \cdots$ be independent random vectors with common continuous* df F *on I^k. Then $\mathfrak{L}\{Z_n^F\} \to \mathfrak{L}\{Z^F\}$ on $(D_{k+1}, \mathfrak{B}(d_{k+1}))$, where Z^F is a centered Gaussian process with continuous sample paths and such that for all $(s_i, \mathbf{t}_i) \in I^{k+1}$*

$$\mathrm{Cov}\big(Z^F(s_1, \mathbf{t}_1), Z^F(s_2, \mathbf{t}_2)\big) = \min(s_1, s_2)[F(\mathbf{t}_1 \wedge \mathbf{t}_2) - F(\mathbf{t}_1)F(\mathbf{t}_2)].$$

Hence Z^F is a Brownian bridge for fixed s and a Brownian motion for fixed $\mathbf{t}$. This is related to the fact that for fixed $\mathbf{t}$, $Z_n^F(\cdot, \mathbf{t})$ is the partial sum process for $1_{(-\infty, \mathbf{t}]}(\xi_i)$, $i = 1, \cdots$. Furthermore, since $Z_n^F(1, \mathbf{t}) = \beta_n(\mathbf{t})$, 2.1.3 is an immediate consequence of 2.1.4 by the continuous mapping theorem.

For $k = 1$ Müller (1970) considered the D-valued process $s \to Z_n^F(s, \cdot)$, $s \geqslant 0$, and proved weak convergence to a process $K : s \to K(s, \cdot) \in D$, for which $(s, t) \to K(s, t)$ is Gaussian with zero means and the same covariance structure as Z^F. The process $\bar{Z}_n^F$ defined as $\bar{Z}_n^F(s, \mathbf{t}) := n^{\frac{1}{2}}(F_{[n/s]}(\mathbf{t}) - F(\mathbf{t}))$, $\mathbf{t} \in I^k$, $0 < s \leqslant 1$, involves the whole sequence $\xi_1, \xi_2, \cdots$. By the Glivenko-Cantelli theorem, for every fixed n, as $s \to 0$, $\sup_{\mathbf{t} \in I^k}|\bar{Z}_n^F(s, \mathbf{t})| \to 0$ $\mathbb{P}$-a.s. Hence $\bar{Z}_n^F(s, \mathbf{t})$ can be extended to an element in D_{k+1} (for this $[x]$ is defined to be the least integer not smaller than x), which is equal to zero at the lower boundary of I^{k+1}. Since a Chentsov-type inequality is not available in this case the proof of tightness has to follow classical lines using appropriate fluctuation inequalities (cf. Billingsley (1968) or Shorack and Smythe (1976)). A different method which uses martingale arguments (involving Brown's (1971) inequality) has been applied by Neuhaus and Sen (1977) to obtain the following result.

THEOREM 2.1.5. *Under the assumptions of the last theorem,* $\mathfrak{L}\{\overline{Z}_n^F\} \to \mathfrak{L}\{\overline{Z}^F\}$ *on* $(D_{k+1}, \mathfrak{B}(d_{k+1}))$, *where* $\overline{Z}^F$ *equals the Gaussian process* Z^F *occurring in* 2.1.4.

If one looks at generalized Kolmogorov-Smirnov statistics $D_n(\mathcal{C}, \mu)$, the corresponding empirical ($\mathcal{C}$ −) process

$$\beta_n(C) := n^{\frac{1}{2}}(\mu_n(C) - \mu(C)), \qquad C \in \mathcal{C},$$

has to be considered as a (random) set function on $\mathcal{C}$ rather than a function of points. By the multivariate central limit theorem one obtains weak convergence of the finite dimensional distributions of β_n to those of a Gaussian process $B^\circ = B^{\circ\mu} = (B^{\circ\mu}(C))_{C \in \mathcal{C}}$ with zero means and covariance function

$$\operatorname{Cov}(B^\circ(C), B^\circ(C')) = \mu(C \cap C') - \mu(C)\mu(C'), \qquad C, C' \in \mathcal{C}.$$

However, before stating an invariance principle for the β_n-process in this general setup, one has to clarify in which sense weak convergence should be understood. One possible way is to look at β_n as a random function in the space $\mathfrak{B}(\mathcal{C})$ of all real-valued, bounded functions on $\mathcal{C}$ with sup-norm $\|f\| := \sup\{|f(C)| : C \in \mathcal{C}\}$. Since $(\mathfrak{B}(\mathcal{C}), \|\cdot\|)$ is nonseparable for each infinite system $\mathcal{C}$, this will cause severe measurability difficulties for β_n when endowing $\mathfrak{B}(\mathcal{C})$ with the Borel-σ-field of the sup-norm topology on $\mathfrak{B}(\mathcal{C})$. Therefore it will be more convenient to work with the smallest σ-field $\mathcal{S}_0$ containing all open $\|\cdot\|$-balls in $\mathfrak{B}(\mathcal{C})$ (cf. Dudley (1966a)). Furthermore, when studying weak convergence on nonseparable metric spaces (Dudley (1966a), Wichura (1968)), it turns out that the limit measure should be defined on the whole of all Borel sets rather than on $\mathcal{S}_0$. For example, if $\mathcal{C}$ is a compact metric space and $C(\mathcal{C})$ is the separable subspace of all continuous functions on $\mathcal{C}$, this may be easily achieved if the limit process has a version with continuous sample paths. On the other hand it is known (cf. Strassen and Dudley (1969)) from central limit theorems for $C(S)$-valued random elements (where S is a compact metric space and $C(S)$ denotes the space of all continuous functions on S) that a CLT may fail to hold if S in some sense is too "large" (which is measured in terms of the ε-entropy of S). With this in mind a corresponding limit theorem for the β_n-process might be only expected if $\mathcal{C}$ is sufficiently "poor." For example, if $\mathcal{C} = \mathcal{C}_k'$, the class of all convex closed subsets of I^k, then $\mathcal{C}$ is compact under the Hausdorff metric (by the well-known Blaschke selection theorem). Dudley (1974) obtained upper and lower bounds for the ε-entropy of $\mathcal{C}_k'$ which guarantee that for the uniform distribution on I^k the limit process B° has a version with continuous sample paths if $k = 1, 2$ but not if $k \geqslant 3$ (see Dudley (1973), Theorem 4.3). The corresponding invariance principle for the uniform empirical process $(\alpha_n(C))_{C \in \mathcal{C}_2'}$ has been obtained by de Hoyos (1972) (incorrectly) and by Bolthausen (1978):

(2.1.10) $$\mathfrak{L}\{\alpha_n(C) : C \in \mathcal{C}_2'\} \to_w \mathfrak{L}\{B^\circ(C) : C \in \mathcal{C}_2'\}$$

where again weak ($w -$) convergence has to be understood in the sense of Dudley (1966a) and Wichura (1968). The proof of tightness relies on a technique similar to that employed by Strassen and Dudley (1969) to prove a CLT for the case of Lipschitz-continuous independent summands in $C(S)$. In the present situation, instead of continuity Bolthausen's approach uses the fact that the summands $1_C(\xi_i) - \mu(C)$ are up to the Lipschitz-continuous $\mu(C)$ monotone in C. To make this work one has to approximate the elements C in $\mathcal{C}_2'$ from above by sets $\hat{C}_m$ belonging to a certain finite class $\hat{\mathcal{C}}_m$, such that C and $\hat{C}_m$ are within 2^{-m} w.r.t. to their Hausdorff-distance and card $(\hat{\mathcal{C}}_m)$ is sufficiently small (see Dudley (1974)). In Pyke (1975) a result similar to (2.1.10) has been stated for the class $\mathcal{C} = \mathcal{P}_m$ of all convex polygonal regions with at most $m(> k)$ vertices ($k \geqslant 1$ arbitrary).

Quite recently Dudley (1978) obtained a generalization of the D-space to more general parameter sets.

In Strassen and Dudley (1969) the empirical process B_n is also considered as defined on a set $\mathcal{H}$ of real functions rather than a class of sets, putting

$$\beta_n(f) := n^{\frac{1}{2}} \int f d(\mu_n - \mu), \qquad f \in \mathcal{H}.$$

The corresponding invariance principle then follows from the abovementioned CLT for $C(S)$-valued random elements, if $\mathcal{H} = S$ is compact in the sup-norm topology on I^k (see also Giné (1974) and Jain and Marcus (1975)).

2.2. Limiting distributions. As was already pointed out the invariance principle provides a powerful tool when determining limit distributions of test statistics which are functions of the empirical process. In practice the statistician has to choose this function according to the importance attached to a certain property of the underlying df F. Accordingly a large literature has evolved to cover the numerous cases of interest. Since a detailed study of this is beyond the scope of this survey the interested reader is referred, e.g., to the articles of Anderson and Darling (1952), Barton and Mallows (1965), Sahler (1968), Rényi (1973), and the monograph of Durbin (1973).

In this section we shall recall only the basic results for the Kolmogorov-Smirnov statistics with possible weights. For this let ψ be a preassigned (reasonable) nonnegative weight function on $I = [0, 1]$. Put $H(f) := \sup_{0 \leqslant t \leqslant 1} \psi(t) f(t), f \in D$. Then H is continuous in the uniform topology on D for all bounded ψ. We may therefore apply Corollary 2.1.2 to obtain $\mathcal{L}\{K_n^+(\psi)\} \to \mathcal{L}\{H(B^0)\}$ for all such ψ's, where

$$K_n^+(\psi) \equiv H(\alpha_n) = \sup_{0 \leqslant t \leqslant 1} \psi(t) \alpha_n(t).$$

Similarly for

$$K_n(\psi) \equiv \sup_{0 \leqslant t \leqslant 1} \psi(t) |\alpha_n(t)|.$$

In particular, if $\psi \equiv 1$, then $K_n^+(\psi) = n^{\frac{1}{2}} D_n^+$ and $K_n(\psi) = n^{\frac{1}{2}} D_n$, the normalized Smirnov and Kolmogorov statistics. Kolmogorov (1933) and Smirnov (1944) proved that

$$\lim_{n\to\infty}\mathbb{P}\Big(\big\{n^{\frac{1}{2}}D_n \leqslant \lambda\big\}\Big) = K_0(\lambda) \equiv \Sigma_{j=-\infty}^{\infty}(-1)^j\exp(-2j^2\lambda^2) \quad \text{if} \quad \lambda > 0$$
$$\equiv 0 \quad \text{otherwise;}$$

and

$$\lim_{n\to\infty}\mathbb{P}\Big(\big\{n^{\frac{1}{2}}D_n^+ \leqslant \lambda\big\}\Big) = K_0^+(\lambda) \equiv 1 - \exp(-2\lambda^2) \quad \text{if} \quad \lambda > 0$$
$$\equiv 0 \quad \text{otherwise.}$$

Doob (1949) showed that K_0 and K_0^+ occur as the distribution of $\sup_{0\leqslant t\leqslant 1}|B°(t)|$ and $\sup_{0<t\leqslant 1}B°(t)$, respectively. If $\psi(t) = (d + ct)^{-1}$, $d, d + c > 0$, then the distribution of $K_n^+(\psi)$ is related to the probability that the sample path of $B°$ crosses the line $y(t) = d + ct$. Doob (1949) proved the latter to be equal to $\exp(-2d(d + c))$. The corresponding distribution for the two-sided case has been considered, e.g., in Durbin (1973), page 22. If $\psi(t) = 1$ for $0 \leqslant a \leqslant t \leqslant b \leqslant 1$ and 0 otherwise, we arrive at a statistic which was proposed by Manija (1949) for detecting discrepancies over a central portion of I. Rényi (1953) derived limit results for various ψ's, such as $\psi_1(t) = t^{-1}$, $\psi_2(t) = (1 - t)^{-1}$, $\psi_3(t) = (t(1 - t))^{-1}$, $0 \leqslant a < t < b \leqslant 1$, which give additional mass near 0 and 1. That the boundedness of ψ is essential (though not necessary) in order to obtain a nondegenerate limit distribution may be seen from $\psi = \psi_1$ with $a = 0$. According to a result of Daniels (1945) (for a nice proof see Rényi (1973)),

$$\mathbb{P}\big(\big\{\sup_{0<t\leqslant 1}t^{-1}(F_n(t) - t) > \varepsilon\big\}\big) = (1 + \varepsilon)^{-1} \quad \text{for all} \quad \varepsilon \geqslant 0,$$

i.e.,

$$\lim_{n\to\infty}\mathbb{P}(\{K_n^+(\psi_1) > \varepsilon\}) = \lim_{n\to\infty}\big(1 + \varepsilon n^{-\frac{1}{2}}\big)^{-1} = 1 \quad \text{for all} \quad \varepsilon > 0.$$

The same holds true for the function $\psi(t) = (t(1 - t))^{-\frac{1}{2}}$, $0 \leqslant a < t < b \leqslant 1$, when $a = 0$ or $b = 1$. Remarkably enough it may be shown (see Theorem 2.5.2 below) that $a_n\sup_{0<t<1}(t(1 - t))^{-\frac{1}{2}}\alpha_n(t) - b_n$ has a nondegenerate limit for a suitable sequence of norming and centering constants a_n and b_n. The case $0 < a < b < 1$ has been investigated before by Anderson and Darling (1952).

In 2.5 below we shall establish sufficient conditions on arbitrary ψ's so that, e.g., $\mathfrak{L}\{K_n^+(\psi)\} \to \mathfrak{L}\{\sup_{0\leqslant t\leqslant 1}\psi(t)B°(t)\}$. It is clear that the limit distributions of $K_n^+(\psi)$ and $K_n(\psi)$ occur as the solution of the corresponding one- and two-sided boundary-value problem for the Brownian bridge $B°$. Equivalently, using Doob's (1949) transformation

$$B(t) = (1 + t)B°\left(\frac{t}{1 + t}\right), \quad t \geqslant 0 \quad \text{(in distribution);}$$

this is related to an associated boundary-value problem for Brownian motion B. For general ψ an attack on this was made in Durbin (1973) and in the literature cited there. Rates of convergence for the boundary-value problem may be found in 2.3 below. The case of a df F with possible points of discontinuity has been considered by Schmid (1958) and Carnal (1962). It turns out that the probabilities in question converge also in this case, but the limiting distributions are no longer independent of F. Actually, they only depend on the values of F at the discontinuity points but not on the form of F on the intervals in between.

If the underlying random variables $\xi_1, \xi_2, \cdots$ take their values in the k-dimensional Euclidean space $\mathbb{R}^k$ according to a df F similar limit results are available, at least for bounded ψ. In particular, for $\psi \equiv 1$ the weak convergence result for $K_n(\psi)$ has been obtained already in 1958 by Kiefer and Wolfowitz (without using the multidimensional invariance principle of course). It should be noted, however, that even for continuous F's the limit df depends on F and is still unknown, except for trivial cases.

2.3. Rates of convergence. Recall that for two p-measures μ and ν being defined on the σ-field of Borel sets of a separable metric space (S, d) their Prokhorov distance is defined by

$$r(\mu, \nu) := \inf\{\varepsilon > 0 : \mu(F) \leqslant \nu(F^\varepsilon) + \varepsilon \quad \text{for all closed} \quad F \subset S\}.$$

It is known that for each sequence $\eta_0, \eta_1, \eta_2, \cdots$ of S-valued random elements, $\mathfrak{L}\{\eta_n\} \to \mathfrak{L}\{\eta_0\}$ if and only if $r(\mathfrak{L}\{\eta_n\}, \mathfrak{L}\{\eta_0\}) \to 0$ as $n \to \infty$. In particular, if $S = D$, $\eta_0 = B^\circ$ and $\eta_n = \alpha_n$ for $n \geqslant 1$, then by 2.1.1 $r(\mathfrak{L}\{\alpha_n\}, \mathfrak{L}\{B^\circ\}) \to 0$. Concerning the rate of convergence, we mention the following results by Müller (1970) and Komlós-Major-Tusnády (1975). See also Dudley (1972).

THEOREM 2.3.1.

Müller: $\qquad r(\mathfrak{L}\{\alpha_n\}, \mathfrak{L}\{B^\circ\}) = \mathcal{O}(n^{-\frac{1}{4}}\log n).$

Komlós et al: $\qquad r(\mathfrak{L}\{\alpha_n\}, \mathfrak{L}\{B^\circ\}) = \mathcal{O}(n^{-\frac{1}{2}}\log n).$

While Müller's proof involves intricate methods of nonstandard analysis, the second result follows from a deep strong approximation theorem. In fact, by Theorem 3.1.1 below, one can find versions $\hat{\alpha}_n$ of α_n and $\hat{B}_n^\circ$ of B° (defined on an appropriate p-space $(\hat{\Omega}, \hat{\mathfrak{F}}, \hat{\mathbb{P}})$) such that for some constants $C, K, \lambda > 0$:

$$\hat{\mathbb{P}}\Big(\big\{\sup_{0 \leqslant t \leqslant 1}|\hat{\alpha}_n(t) - \hat{B}_n^\circ(t)| > n^{-\frac{1}{2}}(C \log n + x)\big\}\Big) \leqslant K \exp(-\lambda x) \qquad \text{for all} \quad x \in \mathbb{R}.$$

In particular for all sufficiently large R we thus obtain

$$\hat{\mathbb{P}}\Big(\big\{\sup_{0 \leqslant t \leqslant 1}|\hat{\alpha}_n(t) - \hat{B}_n^\circ(t)| \geqslant Rn^{-\frac{1}{2}}\log n\big\}\Big) = \mathcal{O}(n^{-1}). \tag{2.3.1}$$

Furthermore, for any closed $F \subset D$ and any $\varepsilon > 0$

$$\begin{aligned}\mathbb{P}(\{\alpha_n \in \mathfrak{F}\}) &= \hat{\mathbb{P}}(\{\hat{\alpha}_n \in F\}) \\ &\leqslant \hat{\mathbb{P}}(\{\hat{B}_n^\circ \in F^\varepsilon\}) + \hat{\mathbb{P}}(\{\sup_{0 \leqslant t \leqslant 1}|\hat{\alpha}_n(t) - \hat{B}_n^\circ(t)| \geqslant \varepsilon\}),\end{aligned}$$

so that by setting $\varepsilon = Rn^{-\frac{1}{2}}\log n$ the desired result is an immediate consequence of (2.3.1). In their 1974 paper Komlós et al. showed that the order $n^{-\frac{1}{2}}\log n$ cannot be improved. (2.3.1) may also be applied to obtain rates of convergence for a large class of continuous functionals of α_n. For this purpose let $H : D \to \mathbb{R}$ be a function which satisfies a Lipschitz condition of the following type:

$$|H(f) - H(g)| \leqslant L \sup_{0 \leqslant t \leqslant 1} |f(t) - g(t)|, \qquad f, g \in D.$$

Assume further that $H(B^\circ)$ has a density which is bounded by a constant $M > 0$. Then for each $\lambda \in \mathbb{R}$

$$\begin{aligned}
&\mathbb{P}(\{H(\alpha_n) \leqslant \lambda\}) - \mathbb{P}(\{H(B^\circ) \leqslant \lambda\}) \\
&\quad \leqslant \hat{\mathbb{P}}\Big(\Big\{H(\hat{\alpha}_n) \leqslant \lambda, \sup_{0 \leqslant t \leqslant 1} |\hat{\alpha}_n(t) - \hat{B}_n^\circ(t)| < Rn^{-\frac{1}{2}} \log n\Big\}\Big) \\
&\qquad + \mathcal{O}(n^{-1}) - \hat{\mathbb{P}}(\{H(\hat{B}_n^\circ) \leqslant \lambda\}) \\
&\quad \leqslant \hat{\mathbb{P}}\Big(\Big\{H(\hat{B}_n^\circ) \leqslant \lambda + LRn^{-\frac{1}{2}} \log n\Big\}\Big) + \mathcal{O}(n^{-1}) - \hat{\mathbb{P}}(\{H(\hat{B}_n^\circ) \leqslant \lambda\}) \\
&\quad \leqslant MLRn^{-\frac{1}{2}} \log n + \mathcal{O}(n^{-1}).
\end{aligned}$$

By symmetry

$$\mathbb{P}(\{H(\alpha_n) \leqslant \lambda\}) - \mathbb{P}(\{H(B^\circ) \leqslant \lambda\}) \geqslant -MLRn^{-\frac{1}{2}} \log n - \mathcal{O}(n^{-1}),$$

so that in summary

$$\sup_{-\infty < \lambda < +\infty} |\mathbb{P}(\{H(\alpha_n) \leqslant \lambda\}) - \mathbb{P}(\{H(B^\circ) \leqslant \lambda\})| = \mathcal{O}\big(n^{-\frac{1}{2}} \log n\big).$$

It is easy to see that the assumptions on H are particularly fulfilled for the functionals

$$H(f) := \sup_{0 \leqslant t \leqslant 1} f(t) \quad \text{and} \quad H(f) := \sup_{0 \leqslant t \leqslant 1} |f(t)|,$$

whence by the above considerations

$$\sup_{-\infty < \lambda < +\infty} \Big|\mathbb{P}\Big(\Big\{n^{\frac{1}{2}} D_n^{(+)} \leqslant \lambda\Big\}\Big) - K_0^{(+)}(\lambda)\Big| = \mathcal{O}\big(n^{-\frac{1}{2}} \log n\big).$$

It follows from the results of the next section that in these cases the exact order is $\mathcal{O}(n^{-\frac{1}{2}})$. Though the functional $H(f) = \int_0^1 f^2(t)\,dt$ does not satisfy the Lipschitz condition the strong approximation result may equally well be applied to obtain the same $\mathcal{O}(n^{-\frac{1}{2}} \log n)$ rate of convergence for the Cramér-von Mises statistic $H(\alpha_n)$ (Csörgő (1976)).

As a third application we consider the problem of estimating rates of convergence in the so-called boundary-value problem. For this let g_1 and g_2 denote two (reasonable) functions on I such that $g_1(t) < g_2(t)$ for all $t \in I$. Let

$$Q_n(a, b) \equiv \mathbb{P}(\{g_1(t) \leqslant \alpha_n(t) \leqslant g_2(t) \quad \text{for all} \quad a < t < b\}),$$

$0 \leqslant a < b \leqslant 1$, be the probability that the sample path of α_n lies between g_1 and g_2 on (a, b). As before it is then easy to see that

$$\begin{aligned}
Q_n(a, b) \leqslant \mathbb{P}(\{&g_1(t) - \varepsilon_n \leqslant B^\circ(t) \leqslant g_2(t) + \varepsilon_n \\
&\text{for all} \quad a < t < b\}) + \mathcal{O}(n^{-r})
\end{aligned}$$

and

$$Q_n(a, b) \geqslant \mathbb{P}(\{g_1(t) + \varepsilon_n \leqslant B^\circ(t) \leqslant g_2(t) - \varepsilon_n \quad \text{for all} \quad a < t < b\}) + \mathcal{O}(n^{-r})$$

(uniformly for all possible choices of g_1 and g_2), where $\varepsilon_n = Rn^{-\frac{1}{2}}\log n$ and $r = r(R) > 0$ becomes large with R.

If we put $g_1(t) = -g_2(t) \equiv -\lambda$, $\lambda > 0$, then the above inequalities yield the result

$$\sup_\lambda |\mathbb{P}(\{\sup_{a<t<b}|\alpha_n(t)| \leqslant \lambda\}) - \mathbb{P}(\{\sup_{a<t<b}|B^\circ(t)| \leqslant \lambda\}) = \mathcal{O}(n^{-\frac{1}{2}}\log n)$$

(because $\mathcal{L}\{\sup_{a<t<b}|B^\circ(t)|\}$ has a bounded density), and similarly for $\sup_{a<t<b}\alpha_n(t)$ by setting $g_1 \equiv -\infty$ and $g_2 \equiv \lambda$. If, e.g., $g_1(t) = -g_2(t) = -\lambda(1 - t)$ and $b < 1$ we obtain an estimate of the convergence rate for a Rényi-type statistic, namely

$$\sup_\lambda \left|\mathbb{P}\left(\left\{\sup_{a<t<b}\frac{|\alpha_n(t)|}{1-t} \leqslant \lambda\right\}\right) - \mathbb{P}(\{\sup_{a/(1-a)<t<b/(1-b)}|B(t)| \leqslant \lambda\})\right| = \mathcal{O}(n^{-\frac{1}{2}}\log n).$$

Notice that in the second term the time interval transformed into $(a/(1-a), b/(1-b))$ and B° has to be replaced by Brownian motion B. This follows from the fact that

$$B(t) = (1+t)B^\circ\left(\frac{t}{1+t}\right), \qquad t \geqslant 0, \quad \text{in distribution.}$$

The last results improve on estimates by Nikitin (1972) who, instead of the more powerful Komlós-Major-Tusnády inequality, used a certain Skorokhod-embedding procedure which is discussed in 3.1 below.

2.4. Exact distributions. The results of 2.2 and 2.3 imply that for estimating tail probabilities for certain test statistics of α_n, one may consider the corresponding limiting distributions instead, at least if the sample size n is large enough. For small n, however, one should work with the exact distributions themselves.

As in 2.2 we shall recall only the basic results. A comprehensive account of these results and related problems may be found in the monograph of Durbin (1973) and in the literature cited there. Again let $\xi_1, \xi_2, \cdots$ be independent uniformly distributed random variables. For the distribution of D_n^+ and D_n we need to compute

$$\begin{aligned}\mathbb{P}(\{D_n^+ \leqslant \lambda\}) &= \mathbb{P}(\{F_n(t) \leqslant \lambda + t \quad \text{for all} \quad 0 \leqslant t \leqslant 1\}) \\ &= 1 - \mathbb{P}(\{F_n \text{ crosses the line } \ y(t) = \lambda + t\})\end{aligned}$$

and

$$\begin{aligned}\mathbb{P}(\{D_n \leqslant \lambda\}) &= \mathbb{P}(\{-\lambda + t \leqslant F_n(t) \leqslant \lambda + t \quad \text{for all} \quad 0 \leqslant t \leqslant 1\}) \\ &= 1 - \mathbb{P}(\{F_n \text{ leaves the region } \ y(t) = \pm\lambda + t\}).\end{aligned}$$

Many different techniques have evolved to attack this problem, even in a more general setup. Let us first mention the basic result by Epanechnikov (1968) and Steck (1971) (for a new and elementary proof see Pitman (1972)). For this let $0 \leq U_{1:n} \leq \cdots \leq U_{n:n} \leq 1$ denote the ordered sample pertaining to $\xi_1, \cdots, \xi_n$.

THEOREM 2.4.1. *If* $0 \leq u_1 \leq u_2 \leq \cdots \leq u_n \leq 1$ *and* $0 \leq v_1 \leq v_2 \leq \cdots \leq v_n \leq 1$ *are given constants such that* $u_i < v_i$ *for all* $i = 1, \cdots, n$, *then*

$$\mathbb{P}(\{u_i \leq U_{i:n} \leq v_i, i = 1, \cdots, n\}) = n! \det\left[(v_i - u_j)_+^{j-i+1}/(j-i+1)!\right],$$

where $x_+ = \max(x, 0)$ *and the matrix element is taken to be zero if* $j < i - 1$.

As an immediate application of this result one obtains the probability that the empirical df lies between two other df's G_1 and G_2, say. Suppose that G_1 is left-continuous and G_2 is right-continuous on (0, 1). Then it follows from 2.4.1 that

$$P_n(G_1, G_2) \equiv \mathbb{P}(\{G_1(t) \leq F_n(t) \leq G_2(t) \quad \text{for all} \quad 0 \leq t \leq 1\})$$
$$= n! \det\left[(v_i - u_j)_+^{j-i+1}/(j-i+1)!\right],$$

where $v_i = G_1^{-1}((i-1)/n)$ and $u_j = G_2^{-1}(j/n)$, and where G_1^{-1} and G_2^{-1} denote the left- and right-continuous inverses of G_1 and G_2, respectively. In particular, by setting for each $0 < \lambda < 1$

$$G_1(t) = 0, 0 \leq t < 1 \qquad \text{and} \qquad G_2(t) = 0, t = 0$$
$$= 1, t = 1 \qquad\qquad = \min(1, t + \lambda), 0 < t \leq 1,$$

then $\mathbb{P}(\{D_n^+ \leq \lambda\}) = P_n(G_1, G_2)$, where in this case $v_i \equiv 1$ and $u_j = (j/n - \lambda)_+$. Similarly for D_n. The distribution of D_n^+ has been given in terms of an incomplete Abel sum for the first time by Smirnov (1944) and later, but independently, by Birnbaum and Tingey (1951). They show that

$$(2.4.1) \quad \mathbb{P}(\{D_n^+ \leq \lambda\}) = 1 - \lambda \Sigma_{i=0}^{[n(1-\lambda)]} \binom{n}{i} \left(1 - \lambda - \frac{i}{n}\right)^{n-i} \left(\lambda + \frac{i}{n}\right)^{i-1}, \qquad 0 \leq \lambda \leq 1.$$

For the proof note that $D_n^+ \leq \lambda$ if and only if

$$U_{i:n} \geq 1 - \lambda - \frac{n-i}{n} \qquad \text{for all} \quad i = n, n-1, \cdots, K,$$

where K is the integer uniquely determined by the relation $K + 1 \geq n\lambda + 1 > K$. Since the joint distribution of $U_{1:n}, \cdots, U_{n:n}$ has the density $n!$ on $0 \leq t_1 \leq t_2 \leq \cdots \leq t_n \leq 1$ and 0 otherwise,

$$\mathbb{P}(\{D_n^+ \leq \lambda\})$$
$$= n! \int_{1-\lambda}^{1} \int_{1-\lambda-(1/n)}^{x_n} \cdots \int_{1-\lambda-(n-K)/n}^{x_{K+1}} \int_0^{x_K} \cdots \int_0^{x_2} dx_1 dx_2 \cdots dx_{K-1} dx_K \cdots dx_n,$$

which after some calculation yields the right-hand side of (2.4.1). Dempster (1959) considered the problem of calculating the probability that F_n crosses a general line $y(t) = \lambda + ct$, $\lambda \geq 0$, $\lambda + c \geq 1$. See also Durbin (1973). Takács (1967, 1970) obtained (2.4.1) from a general result (the proof of which is based on the classical ballot lemma) on the fluctuations of sample functions of stochastic processes with interchangeable increments and step functions as its sample paths. We also refer to a discussion of Vincze (1970), who gave an approach similar to Takács (1967) but using a generalized version of the ballot lemma due to Tusnády and Sarkadi (cf. Vincze (1970)). Finally, Nef (1964) derived (2.4.1) for values $\lambda = r/n, r = 0, 1, \cdots, n$, from the distribution of $Z_n(r)$, where $Z_n(r)$ denotes the number of (horizontal) intersections of the graph $F_n(\cdot)$ with the line $y(t) = (r/n) + t$. Notice that $D_n^+ \geqslant r/n$ if and only if $Z_n(r) \geqslant 1$.

The distribution of D_n has been given in terms of Laplace transforms by Darling (1960). Using the fact that in distribution nF_n is a Poisson process with parameter n conditioned to be n in $t = 1$, Durbin (1968, 1973) derives a transition matrix H_λ with $\mathbb{P}(\{D_n \leqslant \lambda\})$ given in terms of the $[n\lambda + 1]$th diagonal element of the corresponding n-fold product H_λ^n. Massey (1950) obtained $\mathbb{P}(\{D_n \leqslant \lambda\})$ as a solution of a certain difference equation.

In order to derive the LIL (1.2.2) for D_n, Dvoretzky, Kiefer and Wolfowitz (1956) made substantial use of the following large deviation theorem, which in this form is best possible according to the fact that

$$\begin{aligned}\lim_{n\to\infty}\mathbb{P}\Big(\big\{n^{\frac{1}{2}}D_n > \lambda\big\}\Big) &= \mathbb{P}(\{\sup_{0\leqslant t\leqslant 1}|B^\circ(t)| > \lambda\}) \\ &= 2\exp(-2\lambda^2)(1 + o(1)) \qquad \text{as} \quad \lambda \to \infty.\end{aligned}$$

THEOREM 2.4.2. (Dvoretzky-Kiefer-Wolfowitz). *There is a universal constant c such that for all $\lambda > 0$*

$$\mathbb{P}\Big(\big\{n^{\frac{1}{2}}D_n > \lambda\big\}\Big) \leqslant c\exp(-2\lambda^2). \tag{2.4.2}$$

The proof uses the expression (2.4.1) for the df of D_n^+. Using martingale arguments Wellner (1977c) showed the left-hand side of (2.4.2) (with 4λ instead of λ) to be less than or equal to the first absolute moment of a truncated gamma $(n, 1)$-variable. From this he obtained an estimate similar to (2.4.2) but with $1/50$ instead of 2 in the exponent.

Much less is known in the case of multivariate random vectors. Kiefer (1961) showed that for each $k \geqslant 2$ and every $\varepsilon > 0$, there exists a universal constant $c = c(\varepsilon, k)$ such that for all $n \in \mathbb{N}$

$$\mathbb{P}\Big(\big\{n^{\frac{1}{2}}D_n > \lambda\big\}\Big) \leqslant c\exp(-(2-\varepsilon)\lambda^2), \qquad \lambda > 0. \tag{2.4.3}$$

Kiefer (1961) also gave an example which shows that (2.4.3) in this form is best possible. Moreover it is sharp enough to yield the LIL (1.2.2). If F is the uniform df on I^2, the limiting process B° of α_n is the standard Brownian sheet B tied down at $1 \in I^2$ and the lower boundary of I^2 (see 2.1):

$$B^\circ(t_1, t_2) = B(t_1, t_2) - t_1t_2B(1, 1), \qquad 0 \leqslant t_i \leqslant 1.$$

Goodman (1976) obtained a lower bound for the tail probability of

$\sup_{0 \leqslant t_i \leqslant 1} B^{\circ}(t_1, t_2) \equiv S'$:

$$\mathbb{P}(\{S' \geqslant \lambda\}) \geqslant (2\lambda^2 + 1)\exp(-2\lambda^2), \qquad \lambda \geqslant 0.$$

Note that in one dimension $\mathbb{P}(\{S' \geqslant \lambda\}) = \exp(-2\lambda^2)$. The proof of Goodman is interesting in so far as it uses a result of Kuelbs (1973), which relates k-parameter Brownian motion to a 1-parameter Brownian motion taking its values in the space of all continuous functions on I^{k-1}. See also Paranjape and Park (1973b) and Orey and Pruitt (1973).

The exact distributions of various other statistics based on the empirical df are discussed, e.g., in Durbin (1973). See also Csáki (1974b).

A completely different representation of exact distributions may be obtained from calculating the corresponding series expansion. The first result in this direction is due to Smirnov (1944), who showed that for all $0 < \lambda < \mathcal{O}(n^{\frac{1}{6}})$

$$\mathbb{P}\left(\left\{n^{\frac{1}{2}} D_n^+ \leqslant \lambda\right\}\right) = 1 - \exp(-2\lambda^2)\left[1 - \frac{2\lambda}{3n^{\frac{1}{2}}} + \mathcal{O}(n^{-1})\right].$$

In an unpublished paper Karplevskaia (1949) proved that

$$\mathbb{P}\left(\left\{n^{\frac{1}{2}} D_n^+ \leqslant \lambda\right\}\right) = 1 - \exp(-2\lambda^2)\left[1 - \frac{2\lambda}{3n^{\frac{1}{2}}} + \frac{2\lambda^2}{3n}\left(1 - \frac{2\lambda^2}{3}\right) + \mathcal{O}(n^{-\frac{3}{2}})\right],$$

and Chan Li-Tsian (1955) gave one more term of the expansion. We also refer to Lauwerier (1963) for a complete series expansion in terms of inverse powers of $n^{\frac{1}{2}}$. The asymptotic expansion of $\mathcal{L}\{n^{\frac{1}{2}} D_n\}$ is again due to Chan Li-Tsian (1956). For further details see Gnedenko, Koroluk and Skorokhod (1960). Finally, we mention the work of Sanov (1957), Hoeffding (1965), Hoadley (1967), Stone (1974) and Groeneboom (1976), who investigate the first order asymptotic behaviour of $\mathbb{P}(\{\mu_n \in A\})$, where A is a fixed class of probability measures on the sample space.

2.5. Weak convergence of the weighted empirical process. In this section we establish the weak analogue of the LIL for the weighted empirical process as considered in 1.4. For this let ψ always denote a positive continuous function on (0, 1) approaching ∞ at the endpoints of I. The ρ_ψ-metric is defined by

$$\rho_\psi(f, g) := \rho(\psi f, \psi g) = \sup_{0<t<1} \psi(t)|f(t) - g(t)|.$$

Since (D, ρ_ψ) fails to be a separable metric space, weak convergence w.r.t. the ρ_ψ-metric ($\nu_n \to_\psi \nu$) has to be defined again in the sense of Dudley (1966a). The convergence $\mathcal{L}\{\alpha_n\} \to_\psi \mathcal{L}\{B^{\circ}\}$ for certain of these ρ_ψ-metrics plays an important role in statistical applications when studying asymptotic distributions of statistics based on ordered observations, such as linear rank statistics and linear combinations of order statistics. See Pyke and Shorack (1968) and Shorack (1972).

To start with the discussion assume first that $\mathcal{L}\{\alpha_n\} \to_\psi \mathcal{L}\{B^{\circ}\}$ and let $H : D \to \mathbb{R}$ be defined by $H(f) := \lim \sup_{t \to 0} \psi(t) f(t) \wedge 1$. Since H is continuous in the ρ_ψ-metric $\mathcal{L}\{H(\alpha_n)\} \to \mathcal{L}\{H(B^{\circ})\}$ as $n \to \infty$. Assume furthermore that $\psi(t)t \to 0$ as $t \to 0$ (a condition which turns out to be necessary in the situation considered below). Then $H(\alpha_n) = 0$ $\mathbb{P}$-almost surely for all $n \in \mathbb{N}$ and therefore $H(B^\circ) = 0$ $\mathbb{P}$-almost surely. This means that for each $\varepsilon > 0$ the function $\varepsilon\psi^{-1}$ is in the upper class of B°. Since B° and B have the same local behaviour near zero $\varepsilon\psi^{-1}$ is also in the upper class of Brownian motion B, i.e.,

(2.5.1) $\lim_{s\downarrow 0}\mathbb{P}\big(\{\sup_{0<t\leqslant s}\varepsilon^{-1}\psi(t)B(t) > 1\}\big) = 0 \qquad \text{for all} \quad \varepsilon > 0.$

Using the classical Kolmogorov test (see Itô and Mc Kean (1965)), O'Reilly (1974) showed that for ψ^{-1} nondecreasing in a neighbourhood $(0, \gamma)$ of zero, (2.5.1) is equivalent to the condition

(2.5.2) $\int_0^\gamma t^{-1}\exp\big(-\varepsilon h^2(t)\big)\,dt < \infty \qquad \text{for all} \quad \varepsilon > 0,$

where $h(t) \equiv t^{-\frac{1}{2}}\psi^{-1}(t)$, $t > 0$.

THEOREM 2.5.1 (O'Reilly). *Let $q = \psi^{-1}$ be a continuous positive function on* $(0, 1)$ *bounded away from zero on* $[\gamma, 1-\gamma]$ *for some* $\gamma > 0$, *nondecreasing (nonincreasing) on* $(0, \gamma]$ $([1-\gamma, 1))$. *Then*

(2.5.3) $\int_0^1 t^{-1}\exp\big(-\varepsilon h_i^2(t)\big)\,dt < \infty \qquad \text{for all} \quad \varepsilon > 0, i = 1, 2$

is both necessary and sufficient for $\mathfrak{L}\{\alpha_n\} \to_\psi \mathfrak{L}\{B^\circ\}$, *where* $h_1(t) = t^{-\frac{1}{2}}\psi^{-1}(t)$ *and* $h_2(t) = t^{-\frac{1}{2}}\psi^{-1}(1-t)$.

Notice that for symmetry reasons the convergence for $i = 2$ is needed to handle the critical point $t = 1$. The same result has been obtained by Chibisov (1964) for the class of weight functions $q = \psi^{-1}$ which are regularly growing of order $1/2$ in neighbourhoods of 0 and 1.

To give also a short idea of the sufficiency part of the proof, one first selects versions $\hat{\alpha}_n$ and $\hat{B}^\circ$ of α_n and B° such that $\rho(\hat{\alpha}_n, \hat{B}^\circ) \to 0$ almost surely as $n \to \infty$. It is enough to prove that $\rho_\psi(\hat{\alpha}_n, \hat{B}^\circ) \to 0$ in probability. The boundedness condition on ψ now guarantees that $\sup_{\gamma\leqslant t\leqslant 1-\gamma}\psi(t)|\hat{\alpha}_n(t) - \hat{B}^\circ(t)| \to 0$ almost surely as $n \to \infty$. Secondly, to handle the critical point 0, say, it remains to show that α_n and hence $\hat{\alpha}_n$ have in the limit the same sets of upper class functions as standard Brownian motion B. As shown by Chibisov (1964), the last statement is valid if instead of α_n it is valid for the processes

$$P_n(t) := n^{\frac{1}{2}}\left(\frac{N_n(t)}{n} - t\right), \qquad 0 \leqslant t \leqslant 1,$$

where N_n is a Poisson process with parameter n. We shall delay a discussion of this until Section 3.1. Since P_n, for each $n \in \mathbb{N}$, is a process with independent increments, standard maximal inequalities may be applied to show that, if (2.5.3) holds, then for all $\varepsilon > 0$

(2.5.4) $\lim_{s\to 0}\limsup_{n\to\infty}\mathbb{P}\big(\{\psi(t)|P_n(t)| > \varepsilon \quad \text{for some} \quad 0 < t \leqslant s\}\big) = 0,$

which is the desired result.

Wellner (1977c) considered the class of weight functions ψ having finite second moment $\int_0^1\psi^2(t)\,dt$. Instead of the Poisson arguments his proof is based on the fact (cf. Puri and Sen (1971), page 42) that $W_n(t) := \alpha_n(t)/(1-t)$, $0 \leqslant t < 1$, forms a martingale in t for each fixed $n \in \mathbb{N}$. Consequently Doob's martingale inequality may be applied to the discrete time parameter martingale $W_n(k/m)$, $k = 1, \cdots, [m\theta]$, to obtain an estimate of the tail probability of $\sup_{1 \leqslant k \leqslant [m\theta]} \psi(k/m)|\alpha_n(k/m)|$. Letting m tend to infinity, he arrives at the following inequality:

$$\mathbb{P}(\{\sup_{0<t\leqslant\theta} \psi(t)|\alpha_n(t)| \geqslant 4\varepsilon\}) \leqslant \varepsilon^{-1}\mathbb{E}(|T_n| 1_{\{|T_n| \geqslant \varepsilon\}}), \tag{2.5.5}$$

where $T_n = n^{-\frac{1}{2}}\sum_{i=1}^n y_i$ is a specified sum of i.i.d. rv's y_i having zero expectations and $\mathrm{Var}(y_i) = \int_0^\theta \psi^2(t)\,dt$. By Chebychev's inequality, the right-hand side of (2.5.5) is less than or equal to $\varepsilon^{-2}\mathbb{E}(T_n^2) = \varepsilon^{-2}\int_0^\theta \psi^2(t)\,dt$, which can be made arbitrarily small by letting $\theta \to 0$. This is closely related to an earlier result by Pyke and Shorack (1968).

The corresponding result for the sequential empirical process Z_n^F is contained in Wellner (1975). In this case the condition $\int_0^1 \psi^2(t)\,dt < \infty$ has to be replaced by the condition $\int_0^1\int_0^1 \psi^2(t_1, t_2)\,dt_1\,dt_2 < \infty$, and the monotonicity of ψ^{-1} by the monotonicity of $\psi^{-1}(t_1, \cdot)$ and $\psi^{-1}(\cdot, t_2)$.

As in the functional LIL (cf. Corollary 1.4.2) the case $\psi(t) = \psi_0(t) = (t(1-t))^{-\frac{1}{2}}$ is again excluded from the above considerations. In particular it may be seen from Chibisov (1964) that $\sup_{0<t<1}\psi_0(t)|\alpha_n(t)| \to \infty$ in probability as $n \to \infty$. Surprisingly, as was shown by Jaeschke (1975), there exist nonnegative norming constants a_n and b_n such that the distribution of $a_n \sup_{0<t<1}\psi_0(t)|\alpha_n(t)| - b_n$ has a nondegenerate limit as $n \to \infty$.

THEOREM 2.5.2. (Jaeschke). *Let* $L(x) := \exp[-e^{-x}\pi^{-\frac{1}{2}}]$, $x \in \mathbb{R}$ *and*

$$a_n = (2\log\log n)^{\frac{1}{2}},\ b_n = 2\log\log n + \tfrac{1}{2}\log\log\log n \quad (\textit{if defined}).$$

Then

$$\lim_{n\to\infty} \mathfrak{L}\{a_n \sup_{0<t<1}\psi_0(t)\alpha_n(t) - b_n\} = L$$

and

$$\lim_{n\to\infty} \mathfrak{L}\{a_n \sup_{0<t<1}\psi_0(t)|\alpha_n(t)| - b_n\} = L^2.$$

Notice that the constants a_n and b_n as well as the df $L^{\frac{1}{2}}$ with L as in the above theorem are exactly the same as those given by Darling and Erdös (1956) in their now classical result on the weak convergence of the maxima of normalized partial sums pertaining to a sequence of i.i.d. random variables. In fact, using Breiman's representation of the ordered sample $U_{1:n} \leqslant \cdots \leqslant U_{n:n}$ (cf. the remark after 2.1.2) the proof of 2.5.2 proceeds by showing that for each $0 < t_0 < 1$ both

$$\mathfrak{L}\{a_n \sup_{0<t\leqslant t_0}\psi_0(t)\alpha_n(t) - b_n\} \quad \text{and} \quad \mathfrak{L}\{a_n \sup_{t_0<t<1}\psi_0(t)\alpha_n(t) - b_n\}$$

have the same limit as $\mathfrak{L}\{a_n \max_{1\leqslant k\leqslant n}[k^{-\frac{1}{2}}\sum_{i=1}^k(N(i) - N(i-1) - 1)] - b_n\}$, namely $L^{\frac{1}{2}}$, where $N = (N(t))_{t \in \mathbb{R}_+}$ is a Poisson process with parameter 1. The first result of 2.5.2 then follows from the fact that $a_n \sup_{0 \leqslant t \leqslant t_0} \psi_0(t)\alpha_n(t) - b_n$ and $a_n \sup_{t_0 < t < 1} \psi_0(t)\alpha_n(t) - b_n$ are asymptotically independent.

In particular Theorem 2.5.2 implies that in spite of 1.4.2 and 1.4.3 there is still a weak version of the LIL for the weight function $\psi_0(t) = (t(1 - t))^{-\frac{1}{2}}$.

COROLLARY 2.5.3.

$$\sup_{0<t<1} n^{\frac{1}{2}} \left| \frac{F_n(t) - t}{(2t(1 - t)\log \log n)^{\frac{1}{2}}} \right| \to 1 \quad \textit{in probability as} \quad n \to \infty.$$

We also mention a result of Hira Lal Koul (1970) in the case of constant weights which may depend on the time index n. For this let $(c_{in})_{1 \leqslant i \leqslant n}$ be an arbitrary triangular array of real numbers. Put $s_n^2 := \Sigma_{i=1}^n c_{in}^2$ and let

$$L_n(t) := s_n^{-1} \Sigma_{i=1}^n c_{in}(1_{[0, t]}(\xi_i) - F(t)), \qquad 0 \leqslant t \leqslant 1,$$

where $\xi_1, \xi_2, \ldots$ are independent with common df F on I. It is easy to see that as for the ordinary empirical process a Chentsov-type inequality (see (2.1.7)) is also available in this case (with $G = F$). So, in order to obtain an invariance principle for the processes L_n, one merely has to find conditions on the constants c_{in}, which ensure the convergence of the finite-dimensional marginals. By the Cramér-Wold device and by the Lindeberg-Lévy theorem for triangular arrays this will follow, if, for example,

(2.5.6) $$s_n^{-2} \max_{1 \leqslant i \leqslant n} c_{in}^2 \to 0 \quad \text{as} \quad n \to \infty.$$

In summary we obtain

THEOREM 2.5.4. *Suppose that F is a continuous* df *on I. Then, if* (2.5.6) *holds,*

$$\mathfrak{L}\{L_n\} \to \mathfrak{L}\{B^\circ \circ F\} \quad \textit{on} \quad (D, \mathfrak{B}(d)) \quad \textit{as} \quad n \to \infty.$$

2.6. Empirical processes with random sample size. In many applied probability models statistical inference has to be made on the basis of observations $\xi_1, \cdots, \xi_n$, where the sample size n is in itself a random variable. For example, n might be the number of observations obtained within a fixed period of time. In this section we are first concerned with the weak convergence of the process $\beta_{N_n}^F \equiv \beta_{N_n}$ where F is a df on I^k and N_n, $n \in \mathbb{N}$, is some sequence of positive integer-valued random variables such that $N_n \to \infty$ in probability as $n \to \infty$. Previous work on this field has been done for the sequence of partial sums of i.i.d. random variables, starting with Anscombe (1972), Rényi (1960), and Blum et al. (1963). There the sequence N_n, $n \in \mathbb{N}$, was assumed to satisfy the following condition:

(2.6.1) $$N_n/n \to \nu \quad \text{in probability for some positive rv } \nu.$$

For the empirical process the following is known.

THEOREM 2.6.1. *Suppose that* (2.6.1) *holds. Then for every continuous* df F *on* I^k:

Pyke (1968), *Fernandez* (1970): $\mathcal{L}\{\beta_{N_n}\} \to \mathcal{L}\{B^\circ \circ F\}$, *if* $k = 1$ *and* $\nu \equiv 1$.

Csörgő (1974): $\mathcal{L}\{\beta_{N_n}\} \to \mathcal{L}\{B^\circ \circ F\}$, *if* $k = 1$ *and* $\nu > 0$ *arbitrary*.

Sen (1973a): $\mathcal{L}\{\beta_{N_n}\} \to \mathcal{L}\{B^\circ \circ F\}$, *if* $k \geqslant 1$ *and* $\nu > 0$ *arbitrary*.

In particular we see from 2.6.1 that $D_{N_n} = N_n^{-\frac{1}{2}} \sup_{\mathbf{t}\in I^k}|\beta_{N_n}(\mathbf{t})| \to 0$ in probability as $n \to \infty$. Csörgő (1973) remarked that the strong version of the Glivenko-Cantelli theorem cannot be proved in the general setup of 2.6.1.

In many applied models, when statistical inference is based on fixed time period observations, the counting variable N_n may be assumed to be independent of the sequence $\xi_1, \xi_2, \cdots$. Much attention has been given to the case when N_n is a Poisson variate with parameter n. In this case Kac (1949) proposed to study the modified empirical df.

$$F_n^*(t) := n^{-1}\Sigma_{i=1}^{N_n} 1_{(-\infty, t]}(\xi_i), \quad t \in \mathbb{R} \tag{2.6.2}$$

(where $F_n^* \equiv 0$ if $N_n = 0$). Note that F_n^* may take on values larger than 1. The main reason which led to this definition is related to the fact that the process nF_n^* has independent increments. In fact, it is only a matter of calculation to show that nF_n^* is a Poisson process with intensity function $nF = nF_{\xi_1}$ (one should always remember this way of constructing a Poisson process). Let

$$\beta_n^*(t) := n^{\frac{1}{2}}(F_n^*(t) - F(t)), \qquad t \in \mathbb{R},$$

denote the corresponding (modified) empirical process, a normalized Poisson process. Since in practice every test statistic based on β_n^* will be distribution free for continuous F's, we may and do assume that $F(t) = t$, $0 \leqslant t \leqslant 1$. In this case $\alpha_n^* \equiv \beta_n^*$ (when restricted to the unit interval) is a random element in D. By the CLT and the convolution property of Poisson variates, the finite dimensional distributions of α_n^* are asymptotically the same as those of standard Brownian motion B. Furthermore, by the independence of increments

$$\mathbb{E}(|\alpha_n^*(t) - \alpha_n^*(t_1)|^2|\alpha_n^*(t_2) - \alpha_n^*(t)|^2) = (t - t_1)(t_2 - t) \tag{2.6.3}$$

for all $0 \leqslant t_1 \leqslant t \leqslant t_2 \leqslant 1$. Since (2.6.3) holds uniformly in n, tightness is now in force by the same arguments as in 2.1. We thus arrived at the following

THEOREM 2.6.2. *On* $(D, \mathfrak{B}(d))$, $\mathcal{L}\{\alpha_n^*\} \to \mathcal{L}\{B\}$ *as* $n \to \infty$.

Note that

$$\alpha_n^*(t) = (N_n/n)^{\frac{1}{2}}\alpha_{N_n}(t) + tn^{-\frac{1}{2}}(N_n - n), \qquad 0 \leqslant t \leqslant 1, \tag{2.6.4}$$

where $T_n = n^{-\frac{1}{2}}(N_n - n)$ is asymptotically $\mathfrak{N}(0, 1)$ by the ordinary central limit theorem and $\mathcal{L}\{\alpha_n\} \to \mathcal{L}\{B^\circ\}$ by the invariance principle 2.1.1. We may therefore obtain versions $\hat{\alpha}_n$ and $\hat{N}_n$ of α_n and N_n such that $\rho(\hat{\alpha}_n, \hat{B}^\circ) \to 0$ and $\hat{T}_n \equiv n^{-\frac{1}{2}}$ $(\hat{N}_n - n) \to \hat{T}$ almost surely for some Brownian bridge $\hat{B}^0$ and some $\mathfrak{N}(0, 1)$-variable $\hat{T}$. Moreover we may assume w.l.o.g. that $(\hat{\alpha}_n)_{n\in\mathbb{N}}$ and $(\hat{N}_n)_{n\in\mathbb{N}}$ are independent, and that $\hat{T}$ is independent of $\hat{B}^\circ$. Hence

$$\hat{\alpha}_n^*(t) = (\hat{N}_n/n)^{\frac{1}{2}}\hat{\alpha}_{\hat{N}_n}(t) + t\hat{T}_n, \qquad 0 \leqslant t \leqslant 1,$$

is a version of α_n^* with $\rho(\hat{\alpha}_n^*, \hat{B}) \to 0$ almost surely, where

$$\hat{B}(t) = \hat{B}^\circ(t) + t\hat{T}, \qquad 0 \leqslant t \leqslant 1,$$

is a version of standard Brownian motion. This provides a different proof of 2.6.2, which does not use the Poisson argument and, from another point of view, shows why the limiting process is B and not B°.

If one wants to express the essence of the last discussion, one could say: a (normalized) Poisson process is a modified empirical process. We now turn to a converse of this statement: the uniform empirical process is a (normalized) Poisson process conditioned to be zero at one. For this let $(N_n(t))_{t\in\mathbb{R}_+}$ denote a Poisson process with parameter n, and let $0 < s_1 < s_2 < \cdots$ be the associated sequence of renewal times. It is fairly easy to show (see, e.g., Sections 3.3 and 7.5 in Gaenssler and Stute (1977)), that $(s_1, \cdots, s_n)$, under the condition $\{s_n \leqslant 1, s_{n+1} > 1\}$, has the same distribution as the ordered sample $0 \leqslant U_{1:n} \leqslant \cdots \leqslant U_{n:n} \leqslant 1$ of n independent uniform random variables $\xi_1, \cdots, \xi_n$. In other words, if F_n is the empirical df of $\xi_1, \cdots, \xi_n$, then in distribution nF_n is the same as $(N_n(t))_{0\leqslant t\leqslant 1}$ under the condition $N_n(1) = n$. This representation of the sample df has been already used by Kolmogorov in his fundamental paper (1933). Transforming N_n into the normalized Poisson process

$$P_n(t) = n^{\frac{1}{2}}\left(\frac{N_n(t)}{n} - t\right), \qquad 0 \leqslant t \leqslant 1,$$

the same fact might be also expressed in terms of the empirical process: the uniform empirical process α_n is, in distribution, the same as a normalized Poisson process conditioned to be zero at one. Noticing that, in distribution, B° is the same as Brownian motion B under the condition $B(1) = 0$, the invariance principle 2.1.1 is therefore equivalent to the following conditional functional limit theorem for the normalized Poisson process:

(2.6.5) $$\mathcal{L}\{P_n | P_n(1) = 0\} \to \mathcal{L}\{B | B(1) = 0\} \quad \text{as} \quad n \to \infty.$$

The same ideas which led to the distributional representation of α_n may be used to construct explicitly a version of α_n in terms of $(N_n(t))_{t\in\mathbb{R}_+}$. We shall delay this until Section 3.1.

At the end of this section we still have to make some remarks concerning the rate of convergence in 2.6.2. As for the empirical process the best known results in this direction follow from strong approximation results. It may be shown (see Stute (1976c)) that (2.3.1) holds equally well for α_n^* (which B_n instead of B_n°), so that again $r(\mathcal{L}\{\alpha_n^*\}, \mathcal{L}\{B\}) = \mathcal{O}(n^{-\frac{1}{2}}\log n)$. The same result holds for the class of Lipschitz functionals as considered at the end of 2.3. In particular, if one defines

$$D_n^{*+} := \sup_{0\leqslant t\leqslant 1}[F_n^*(t) - t] \quad \text{and} \quad D_n^* := \sup_{0\leqslant t\leqslant 1}|F_n^*(t) - t|,$$

the so-called one- and two-sided Kac statistics, then for example

$$\sup_{-\infty<\lambda<+\infty}\left|\mathbb{P}\left(\left\{n^{\frac{1}{2}}D_n^* \leqslant \lambda\right\}\right) - \mathbb{P}\left(\left\{\sup_{0\leqslant t\leqslant 1}|B(t)| \leqslant \lambda\right\}\right| = \mathcal{O}\left(n^{-\frac{1}{2}}\log n\right).$$

See also Nikitin (1972). Identification of the limit distributions yields

$$\lim_{n\to\infty}\mathbb{P}\left(\left\{n^{\frac{1}{2}}D_n^{*+} \leqslant \lambda\right\}\right) = \left(\frac{2}{\pi}\right)^{\frac{1}{2}}\int_0^\lambda \exp(-u^2/2)\,du, \quad \lambda > 0$$
$$= 0 \qquad \text{otherwise,}$$

and

$$\lim_{n\to\infty}\mathbb{P}\left(\left\{n^{\frac{1}{2}}D_n^* \leqslant \lambda\right\}\right) = \frac{4}{\pi}\sum_{k=0}^{\infty}(-1)^k/(2k+1)\exp\left[-(2k+1)^2\pi^2/8\lambda^2\right], \quad \lambda > 0$$
$$= 0 \quad \text{otherwise.}$$

The exact distributions as well as the limiting distribution of D_n^{*+} have been studied by Allen and Beekman (1966, 1967). The limiting distribution of D_n^* was found by Kac (1949). Csörgő (1972) and Suzuki (1972) obtained various results for other statistics based on F_n^* both in the finite and in the limiting case.

Part 3

3.1. Strong approximation results for the empirical process. As before we denote by α_n the uniform empirical process on the unit interval $I = [0, 1]$. Since by the invariance principle 2.1.1 $\mathfrak{L}\{\alpha_n\} \to \mathfrak{L}\{B^\circ\}$ on $(D, \mathfrak{B}(d))$ we may apply Theorem 0.1 from the introduction to obtain versions $\hat{\alpha}_n$ of α_n and $\hat{B}^\circ$ of B° such that $d(\hat{\alpha}_n, \hat{B}^\circ) \to 0$ and therefore $\rho(\hat{\alpha}_n, \hat{B}^\circ) \to 0$ almost surely as $n \to \infty$. We have already seen that the existence of such a sequence could be well applied in the various situations of 2.5. On the other hand, since these versions result from a general existence theorem, they fail to have any of the characteristic properties of the α_n's, except for equality in distribution.

In this section our main emphasis will be to give an explicit construction of the $\hat{\alpha}_n$'s. The first results in this field are due to Breiman (1968), Pyke and Root (1968), Brillinger (1969) and Rosenkrantz (1969). More recent work has been done by Komlós, Major and Tusnády (1975), whose methods rely on a rather delicate dyadic approximation procedure. The earlier proofs were based on the well-known Skorokhod embedding scheme (cf. Skorokhod (1965)). Because this method is more illuminating and easier to describe (see also Sawyer (1974)) we shall present it in greater detail.

For each $n \in \mathbb{N}$, let $Y_1^n, Y_2^n, \ldots$ be a sequence of independent exponential random variables with mean $1/n$, i.e., $\mathbb{P}(\{Y_1^n > x\}) = \exp(-nx)$ for all $x > 0$. The Y_i^n's then define a Poisson process N_n with parameter n in the following way:

$$N_n(t) = k \quad \text{if} \quad \sum_{i=1}^k Y_i^n \leqslant t \quad \text{and} \quad \sum_{i=1}^{k+1} Y_i^n > t,$$
$$\text{and} \quad N_n(t) = 0 \quad \text{if} \quad t < Y_1^n, \qquad t \in \mathbb{R}_+.$$

The normalized process

$$P_n(t) = n^{\frac{1}{2}}(N_n(t)/n - t), \quad t \in \mathbb{R}_+$$

converges weakly to Brownian motion on $D([0, T])$ for every finite $T > 0$. The Skorokhod representation is an embedding of a random variable ζ into Brownian motion B. Assume that $\mathbb{E}(\zeta) = 0$ and $\mathbb{E}(\zeta^2) = 1$. Then, if the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is rich enough, one may construct a Brownian stopping time τ such that $\mathcal{L}\{\zeta\} = \mathcal{L}\{B(\tau)\}$ and $\mathbb{E}(\tau) = \mathbb{E}(\zeta^2) = 1$. Furthermore, since Brownian motion has independent increments and satisfies the strong Markov property, one may apply this to obtain a sequence $\tau_0^1 \equiv 0 < \tau_1^1 < \tau_2^1 < \cdots$ of Brownian stopping times such that

(3.1.1)

$$B(\tau_i^1) - B(\tau_{i-1}^1), i = 1, 2, \cdots \quad \text{are independent and distributed as} \quad \zeta_i^1 \equiv 1 - Y_i^1$$

and

(3.1.2) $\tau_i^1 - \tau_{i-1}^1, i = 1, 2, \cdots$ are independent and distributed as τ_1^1.

Let $\zeta_i^n \equiv n^{\frac{1}{2}}(1/n - Y_i^n), i = 1, 2, \cdots$. Because Y_i^n is distributed as $n^{-1}Y_i^1$, we obtain in distribution

$$\zeta_i^n = n^{-\frac{1}{2}}(1 - Y_i^1) = n^{-\frac{1}{2}}(B(\tau_i^1) - B(\tau_{i-1}^1)).$$

By independence, for each $n \in \mathbb{N}$ the sequence of partial sums $\Sigma_{i=1}^k \zeta_i^n$ has the same joint distribution as $n^{-\frac{1}{2}}B(\tau_k^1), k = 1, 2, \cdots$. In terms of the scale transformed Brownian motion $B_n(t) = n^{-\frac{1}{2}}B(nt), t \in \mathbb{R}_+$, this implies that $Y_1^n + \cdots + Y_k^n$ are jointly distributed as $s_k^n \equiv kn^{-1} - n^{-\frac{1}{2}}B_n(\tau_k^n), k = 1, 2, \cdots$, where $\tau_k^n = n^{-1}\tau_k^1$. Defining $\hat{N}_n$ and $\hat{P}_n$ in the same way as N_n and P_n but with renewal times s_k^n instead of $\Sigma_{i=1}^k Y_i^n$, we obtain versions $\hat{P}_n$ of P_n for which we will show that

(3.1.3) $\sup_{0 \leqslant t \leqslant T}|\hat{P}_n(t) - B_n(t)| \to 0$ in probability for all $T > 0$.

W.l.o.g. let $T = 1$. For given t choose $k = k(w)$ such that $s_k^n \leqslant t < s_{k+1}^n$. Then

$$|\hat{P}_n(t) - B_n(t)| \leqslant |\hat{P}_n(t) - \hat{P}_n(s_k^n)| + |\hat{P}_n(s_k^n) - B_n(t)|$$
$$\leqslant n^{\frac{1}{2}}(s_{k+1}^n - s_k^n) + |B_n(\tau_k^n) - B_n(s_k^n)| + |B_n(s_k^n) - B_n(t)|.$$

Because of the continuity of Brownian sample paths and since $k \leqslant 2n$ on a set A_n with $\mathbb{P}(A_n) \to 1$, it therefore suffices to show that in probability

$$n^{\frac{1}{2}} \sup_{0 \leqslant k \leqslant 2n}(s_{k+1}^n - s_k^n) \to 0 \quad \text{and} \quad \sup_{1 \leqslant k \leqslant 2n}|\tau_k^n - s_k^n| \to 0 \qquad \text{as} \quad n \to \infty.$$

The first convergence follows from elementary calculations. For the second half

$$\sup_{1 \leqslant k \leqslant 2n}|\tau_k^n - s_k^n| \leqslant \sup_{1 \leqslant k \leqslant 2n} n^{-1}|\tau_k^1 - k| + n^{-1} \sup_{1 \leqslant k \leqslant 2n}|B(\tau_k^1)|,$$

which converges to zero in probability by the SLLN. This completes the proof of (3.1.3). The analogous result for the uniform empirical process uses the fact that an ordered sample $0 \leqslant U_{1:n} \leqslant \cdots \leqslant U_{n:n} \leqslant 1$ of n independent uniform random variables has the same joint distribution as $\bar{s}_k^n \equiv s_k^n / s_{n+1}^n, k = 1, \cdots, n$ (see, e.g., Breiman (1968), page 285). As Rényi (1973) pointed out the importance of this lies in the fact that with its help the investigation of ordered samples can be reduced to the investigation of sums of independent variables. In our case we may apply it to obtain a representation of the uniform empirical process in terms of the normalized Poisson process $\hat{P}_n$. For this, let $s_0^n = 0$ and define $\hat{\alpha}_n$ by

$$\hat{\alpha}_n(t) \equiv n^{\frac{1}{2}}\left(\frac{k}{n} - t\right) \quad \text{if} \quad \bar{s}_k^n \leqslant t < \bar{s}_{k+1}^n, \qquad 0 \leqslant t \leqslant 1.$$

Then $\hat{\alpha}_n$ is a version of α_n. Furthermore,

$$\hat{\alpha}_n(t) = \hat{P}_n(ts_{n+1}^n) + n^{\frac{1}{2}}(s_{n+1}^n - 1)t$$

whence, by (3.1.3), it is easy to see that

$$\text{(3.1.4)} \qquad \sup_{0 \leqslant t \leqslant 1} |\hat{\alpha}_n(t) - B_n^\circ(t)| \to 0 \qquad \text{in probability as} \quad n \to \infty.$$

Here $B_n^\circ(t) = B_n(t) - tB_n(1)$ is the Brownian bridge associated with B_n. Using sharper estimates it is even possible to show that $\rho(\hat{\alpha}_n, B_n^\circ) \to 0$ $\mathbb{P}$-almost surely. Brillinger (1969) obtained a bound for the rate of convergence:

$$\rho(\hat{\alpha}_n, B_n^\circ) = \mathcal{O}(\varepsilon_n) \ \ \mathbb{P}\text{-a.s., where} \quad \varepsilon_n = n^{-\frac{1}{4}}(\log n)^{\frac{1}{2}}(\log\log n)^{\frac{1}{4}}.$$

Kiefer (1972a) showed that even $\limsup_{n\to\infty} \rho(\hat{\alpha}_n, B_n^\circ)\varepsilon_n^{-1} = C^*$ with probability one for some specified positive constant C^*. Hence for the Breiman-Brillinger-Pyke-Root approach the results in this form yield the precise rate of convergence. Pyke also pointed out (see Brillinger (1969)) that the processes B_n° are different for each n. In fact,

$$B_n^\circ(t) = B_n(t) - tB_n(1) = n^{-\frac{1}{2}}(B(nt) - tB(n)), \qquad 0 \leqslant t \leqslant 1.$$

Instead of the Skorokhod embedding the work of Komlós, Major and Tusnády (1975) is based on a rigorous elaboration of the so-called quantile transformation technique, which may be found for the first time in a paper of Bártfai (1970).

THEOREM 3.1.1.(Komlós et al.). *On an appropriate p-space* $(\Omega, \mathfrak{F}, \mathbb{P})$ *there exist versions* $\hat{\alpha}_n$ *of* α_n *and* $\hat{B}_n^\circ$ *of* B° *such that for some constants* $C, K, \lambda > 0$

$$\text{(3.1.5)} \quad \mathbb{P}\Big(\Big\{\sup_{0 \leqslant t \leqslant 1} |\hat{\alpha}_n(t) - \hat{B}_n^\circ(t)| > n^{-\frac{1}{2}}(C \log n + x)\Big\}\Big) \leqslant K \exp(-\lambda x) \quad \textit{for all } x \in \mathbb{R} \textit{ and } n \in \mathbb{N}.$$

COROLLARY 3.1.2.

$$\sup_{0 \leqslant t \leqslant 1} |\hat{\alpha}_n(t) - \hat{B}_n^\circ(t)| = \mathcal{O}\left(n^{-\frac{1}{2}} \log n\right) \qquad \mathbb{P}\text{-a.s.}$$

The case $k \geqslant 1$ was dealt with by Csörgő and Révész (1975). They showed that for the uniform empirical process on I^k, one may find Brownian bridges $\{\hat{B}_n^\circ(\mathbf{t})\colon \mathbf{t} \in I^k\}$ and versions $\hat{\alpha}_n$ of α_n such that $\mathbb{P}$-a.s.

$$\sup_{\mathbf{t} \in I^k} |\hat{\alpha}_n(\mathbf{t}) - \hat{B}_n^\circ(\mathbf{t})| = \mathcal{O}\left(n^{-\frac{1}{2(k+1)}}(\log n)^{\frac{3}{2}}\right).$$

The more general problem of strong approximation of the empirical process $\{\beta_n(C)\colon C \in \mathcal{C}\}$ has been studied by Révész (1976) for $k = 2$ and some specified class $\mathcal{C}$ of sets with smooth boundaries.

3.2. Strong approximation of the two-parameter empirical process. Although the results of the last section yield a satisfactory rate of approximation for each single n, the shortcoming of this representation is that $\{\hat{\alpha}_n : n \in \mathbb{N}\}$ does not have the same joint distribution as $\{\alpha_n : n \in \mathbb{N}\}$. Consequently no strong law type behavior (such as the Smirnov-Chung LIL (1.2.2)) may be derived from a corresponding result of B°, nor do we get rates of convergence for probabilities involving more than one α_n. To overcome these difficulties Müller (1970) and Kiefer (1972) proposed to consider the empirical process both as a function of n and t. In 2.1.4 we mentioned the weak convergence of the two-parameter empirical process Z_n^F to a centered Gaussian process Z^F with covariance structure given by

$$\operatorname{Cov}(Z^F(s_1, t_1), Z^F(s_2, t_2)) = \min(s_1, s_2)[t_1 \wedge t_2 - t_1 t_2]$$

(if $k = 1$ and $F(t) = t$). It is clear that a strong approximation result for the two-parameter empirical process should therefore involve appropriate versions of Z_n^F and Z^F. Instead of this we first consider the process $\{n^{\frac{1}{2}}\alpha_n(t): n \in \mathbb{N}, 0 \leqslant t \leqslant 1\}$as a process of two parameters.

Let K be a centered Gaussian process on $\mathbb{R}_+ \times I$ with the same covariance structure, i.e.,

$$\operatorname{Cov}(K(x_1, t_1), K(x_2, t_2)) = \min(x_1, x_2)[t_1 \wedge t_2 - t_1 t_2],$$

$x_i \geqslant 0$, $0 \leqslant t_i \leqslant 1$. Hence K is a Z^F-process when restricted to I in the first coordinate. Furthermore, $(s, t) \to n^{-\frac{1}{2}}K(ns, t)$, $0 \leqslant s, t \leqslant 1$, defines a Z^F-process for each $n \in \mathbb{N}$. In the sequel we shall refer to K as a Kiefer process of first order. Note that a Kiefer process (on $\mathbb{N} \times I$) occurs when summing up independent Brownian bridges B_i°

$$K(k, t) \equiv \Sigma_{i=1}^k B_i^\circ(t), \qquad k \in \mathbb{N}, 0 \leqslant t \leqslant 1.$$

Using an intricate multidimensional Skorokhod embedding scheme Kiefer (1972a) showed that for some version $\{\hat{\alpha}_n : n \in \mathbb{N}\}$ of $\{\alpha_n : n \in \mathbb{N}\}$ (i.e., the $\hat{\alpha}_n$'s have the same joint distribution as the α_n's) and some Kiefer process $\hat{K}$ of the above type

$$\text{(3.2.1)} \qquad n^{-\frac{1}{2}} \sup_{0 \leqslant t \leqslant 1} |n^{\frac{1}{2}}\hat{\alpha}_n(t) - \hat{K}(n, t)| = \mathcal{O}\left(n^{-\frac{1}{6}}(\log n)^{\frac{2}{3}}\right) \text{ a.s.}$$

Answering a question of Kiefer (1972a), Komlós, Major and Tusnády (1975) showed that in using a certain dyadic approximation scheme instead of the Skorokhod embedding the exponent $-1/6$ on the right-hand side of (3.2.1) may be improved to $-\frac{1}{2}$:

$$\text{(3.2.2)} \qquad \sup_{1 \leqslant k \leqslant n} \sup_{0 \leqslant t \leqslant 1} |k^{\frac{1}{2}}\hat{\alpha}_k(t) - \hat{K}(k, t)| = \mathcal{O}(\log^2 n) \text{ a.s.}$$

As for the one-parameter empirical process (see 3.1) the last result follows from a sharp exponential bound for the corresponding tail probabilities:

(3.2.3)

$$\mathbb{P}\left(\left\{\sup_{1 \leqslant k \leqslant n} \sup_{0 \leqslant t \leqslant 1} |k^{\frac{1}{2}}\hat{\alpha}_k(t) - \hat{K}(k, t)| > (C \log n + x)\log n\right\}\right) \leqslant L \exp(-\lambda x)$$

for all x and n, where C, L and λ are positive absolute constants.

The results (3.2.2) and (3.2.3) may be expressed in terms of the Z_n^F- and Z^F-processes as well. For this note that for each $n \in \mathbb{N}$

$$(s, t) \to n^{-\frac{1}{2}}[ns]^{\frac{1}{2}}\hat{\alpha}_{[ns]}(t)$$

and

$$(s, t) \to n^{-\frac{1}{2}}\hat{K}(ns, t), \qquad 0 \leqslant s, t \leqslant 1,$$

are versions of Z_n^F and Z^F, respectively. Thus, e.g., (3.2.2) leads to the result that there exist versions $\hat{Z}_n^F$ and $\hat{Z}_{(n)}^F$ of Z_n^F and Z^F such that

$$\sup_{1 \leqslant k \leqslant n}\sup_{0 \leqslant t \leqslant 1}|\hat{Z}_n^F\Big(\frac{k}{n}, t\Big) - \hat{Z}_{(n)}^F\Big(\frac{k}{n}, t\Big)| = \mathcal{O}\big(n^{-\frac{1}{2}}\log^2 n\big) \quad \text{a.s.}$$

The inequality (3.2.3) also yields rates of convergence, e.g., in the boundary-value problem for the Z_n^F-processes. See Müller (1970) and Kiefer (1972a). The case of independent uniform random vectors $\xi_1, \xi_2, \cdots$ has been considered by Csörgő and Révész (1975). They show that for some version $\{\hat{\alpha}_n : n \in \mathbb{N}\}$ of $\{\alpha_n : n \in \mathbb{N}\}$

$$\sup_{\mathbf{t} \in I^k}|n^{\frac{1}{2}}\hat{\alpha}_n(\mathbf{t}) - K(n, \mathbf{t})| = \mathcal{O}\big(n^{(k+1)/2(k+2)}\log^2 n\big) \quad \text{a.s.}$$

Here K denotes a Kiefer process (of kth order), i.e., a centered Gaussian process with covariance function

$$\operatorname{Cov}(K(x_1, \mathbf{t}_1), K(x_2, \mathbf{t}_2)) = \min(x_1, x_2)\big[\Pi_{i=1}^k \min(t_1^i, t_2^i) - \Pi_{i=1}^k t_1^i t_2^i\big]$$

where $\mathbf{t}_j = (t_j^1, \cdots, t_j^k) \in I^k$ and $0 \leqslant y_j < \infty, j = 1, 2$.

The results of the last two sections will be treated in more detail in a forthcoming monograph by Csörgő and Révész (1977). See also Tusnády (1977).

Acknowledgment. The authors are especially grateful to David Pollard for his careful reading of the manuscript. Thanks are also due to the referees for their valuable suggestions and to Professor Patrick Billingsley for his encouragement and advice. We would like to thank Professor Ronald Pyke for his continuing interest during the preparation of this work.

REFERENCES

[1] Ahmad, S. (1961). Sur le théorème de Glivenko-Cantelli. *C. R. Acad. Sci. Paris Sér. A-B* **252** 1413–1417.

[2] Allen, J. L. and Beekman, J. A. (1966). A statistical test involving a random number of random variables. *Ann. Math. Statist.* **37** 1305–1309.

[3] Allen, J. L. and Beekman, J. A. (1967). Distributions of an M. Kac statistic. *Ann. Math. Statist.* **38** 1919–1923.

[4] Anderson, T. W. and Darling, D. A. (1952). Asymptotic theory of certain 'goodness of fit' criteria based on stochastic processes. *Ann. Math. Statist.* **23** 193–212.

[5] Anscombe, F. J. (1952). Large-sample theory of sequential estimation. *Proc. Cambridge Philos. Soc.* **48** 600–607.

[6] Aronszajn, N. (1950). The theory of reproducing kernels. *Trans. Amer. Math. Soc.* **68** 337–404.

[7] Bártfai, P. (1970). Über die Entfernung der Irrfahrtswege. *Studia Sci. Math. Hungar.* **5** 41–49.

[8] Barton, D. E. and Mallows, C. L.(1965). Some aspects of the random sequence. *Ann. Math. Statist.* **36** 236–260.

[9] Baxter, G. (1955). An analogue of the law of the iterated logarithm. *Proc. Amer. Math. Soc.* **6**

177–181.

[10] Bickel, P. J. and Wichura, M. J. (1971). Convergence criteria for multiparameter stochastic processes and some applications. *Ann. Math. Statist.* **42** 1656–1670.

[11] Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

[12] Billingsley, P. (1971). *Weak Convergence of Measures: Applications in Probability*. Regional Conference Series in Appl. Math., No. 5. SIAM, Philadelphia.

[13] Billingsley, P. and Topsøe, F. (1967). Uniformity in weak convergence. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **7** 1–16.

[14] Birnbaum, Z. W. and Tingey, F. H. (1951). One-sided confidence contours for probability distribution functions. *Ann. Math. Statist.* **22** 592–596.

[15] Blum, J. R. (1955). On the convergence of empiric distribution functions. *Ann. Math. Statist.* **26** 527–529.

[16] Blum, J. R., Hanson, D. L. and Rosenblatt, J. I. (1963). On the central limit theorem for the sum of a random number of independent random variables. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **1** 389–393.

[17] Bolthausen, E. (1978). Weak convergence of an empirical process indexed by the closed convex subsets of I^2. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete.* **43** 173–181.

[18] Breiman, L. (1968). *Probability*. Addison-Wesley, Reading.

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

[20] Brown, B. M. (1971). Martingale central limit theorems. *Ann. Math. Statist.* **42** 59–66.

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

[22] Carnal, H. (1962). Sur les théorèmes de Kolmogorov et Smirnov dans le cas d'une distribution discontinue. *Comment. Math. Helv.* **37** 19–35.

[23] Cassels, J. W. S. (1951). An extension of the law of the iterated logarithm. *Proc. Cambridge Philos. Soc.* **47** 55–64.

[24] Chan Li Tsian (1955). The exact statistical distribution of N. V. Smirnov and its asymptotic expansion. *Shuxue Jinzhan.* **1** 775–790 (in Chinese). (A Russian translation is in the "Mathematika" (1960) collection **4** 121–134.)

[25] Chan Li Tsian (1956). The exact statistical distribution of A. N. Kolmogorov and its asymptotic expansion. *Acta Math. Sinica* **6** 55–79 (in Chinese). (A Russian translation is in the "Mathematika" (1960) collection **4** 135–159.)

[26] Chentsov, N. N. (1956). Weak convergence of stochastic processes whose trajectories have no discontinuities of the second kind and the "heuristic" approach to the Kolmogorov-Smirnov tests. *Theor. Probability Appl.* **1** 140–144.

[27] Chernoff, H. (1952). A measure of asymptotic efficiency for tests based on the sum of observations. *Ann. Math. Statist.* **23** 493–507.

[28] Chibisov, D. M. (1964). Some theorems on the limiting behavior of the empirical distribution function. *Sel. Transl. Math. Statist. Prob.* **6** 147–156.

[29] Chibisov, D. M. (1965). An investigation of the asymptotic power of tests of fit. *Theor. Probability Appl.* **10** 421–437.

[30] Chung, K. L. (1949). An estimate concerning the Kolmogorov limit distribution. *Trans. Amer. Math. Soc.* **67** 36–50.

[31] Cramér, H. (1938). Sur un nouveau théorème limite de la théorie des probabilités. *Actualités Sci. Indust.* No. 736.

[32] Csáki, E. (1968). An iterated logarithm law for semimartingales and its application to empirical distribution function. *Studia Sci. Math. Hungar.* **3** 287–292.

[33] Csáki, E. (1974a). Some notes on the law of the iterated logarithm for empirical distribution function. Colloquia Mathematica Societatis János Bolyai. *Limit Theorems of Probability Theory* 47–58. Keszthely.

[34] Csáki, E. (1974b). Studies on the empirical d.f. MTA III. *Oszt. Közl.* **23** 239–327 (in Hungarian).

[35] Csáki, E. (1977). The law of the iterated logarithm for normalized empirical distribution function. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **38** 147–167.

[36] Csörgő, M. (1972). Distribution results for distance functions based on the modified empirical distribution function of M. Kac. *Ann. Inst. Statist. Math.* **24** 101–110.

[37] Csörgő, M. (1973). Glivenko-Cantelli type theorems for distance functions based on the modified empirical distribution function of M. Kac and for the empirical process with random sample size in general. *Lecture Notes in Math.* **296** 149–164.

[38] Csörgő, M. (1974). Convergence rates of the Glivenko-Cantelli laws. Unpublished.

[39] Csörgő, M. and Révész, P. (1974). Some notes on the empirical distribution function and the quantile process. Colloquia Mathematica Societatis János Bolyai. *Limit Theorems of Probability Theory* 59–71. Keszthely.

[40] Csörgő, M. and Révész, P. (1975). A new method to prove Strassen type laws of invariance principle, II. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **31** 261–269.

[41] Csörgő, M. and Révész, P. (1977). Strong approximations in probability and statistics. Preliminary draft.

[42] Csörgő, S. (1974). On weak convergence of the empirical process with random sample size. *Acta Sci. Math. Szeged* **36** 17–25. (Correction: 375–376.)

[43] Csörgő, S. (1976). On an asymptotic expansion for the von Mises ω^2 statistic. *Acta Sci. Math. Szeged* **38** 45–67.

[44] Daniels, H. E. (1945). The statistical theory of the strength of bundles of threads. *Proc. Roy. Soc. Ser. A* **183** 405–435.

[45] Darling, D. A. (1960). On the theorems of Kolmogorov-Smirnov. *Theor. Probability Appl.* **5** 356–361.

[46] Darling, D. A. and Erdös, P. (1956). A limit theorem for the maximum of normalized sums of independent random variables. *Duke Math. J.* **23** 143–155.

[47] Dehardt, J. (1971). Generalizations of the Glivenko-Cantelli theorem. *Ann. Math. Statist.* **42** 2050–2055.

[48] Dempster, A. P. (1959). Generalized D_n^+ statistics. *Ann. Math. Statist.* **30** 593–597.

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

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

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

[52] Dudley, R. M. (1966a). Weak convergence of probabilities on nonseparable metric spaces and empirical measures on Euclidean spaces. *Illinois J. Math.* **10** 109–126.

[53] Dudley, R. M. (1966b). Convergence of Baire measures. *Studia Math.* **27** 251–268.

[54] Dudley, R. M. (1968). Distances of probability measures and random variables. *Ann. Math. Statist.* **39** 1563–1572.

[55] Dudley, R. M. (1969). The speed of mean Glivenko-Cantelli convergence. *Ann. Math. Statist.* **40** 40–50.

[56] Dudley, R. M. (1972). Speeds of metric probability convergence. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **22** 323–332.

[57] Dudley, R. M. (1973). Sample functions of the Gaussian process. *Ann. Probability* **1** 66–103.

[58] Dudley, R. M. (1974). Metric entropy of some classes of sets with differentiable boundaries. *J. Approximation Theory* **10** 227–236.

[59] Dudley, R. M. (1976). Balls in $\mathbb{R}^k$ do not cut all subsets of $k + 2$ points. Preprint.

[60] Dudley, R. M. (1978). Central limit theorems for empirical measures. *Ann. Probability* **6** 899–929.

[61] Durbin, J. (1968). The probability that the sample distribution function lies between two parallel straight lines. *Ann. Math. Statist.* **39** 398–411.

[62] Durbin, J. (1973). *Distribution Theory for Tests Based on the Sample Distribution Function.* Regional Conference Series in Appl. Math., No. 9. SIAM, Philadelphia.

[63] Dvoretzky, A., Kiefer, J. and Wolfowitz, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Statist.* **27** 642–669.

[64] Eddy, W. F. and Hartigan, J. A. (1977). Uniform convergence of the empirical distribution function over convex sets. *Ann. Statist.* **5** 370–374.

[65] Eicker, F. (1970). On the probability that a sample distribution function lies below a line segment. *Ann. Math. Statist.* **41** 2075–2092.

[66] EICKER, F. (1970). A loglog law for double sequences of random variables. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **16** 107–133.

[67] EICKER, F. (1976). The asymptotic distribution of the supremum of the standardized empirical distribution function. Preprint, Univ. of Dortmund.

[68] ELKER, J. (1975). Über ein gleichmässiges Gesetz der grossen Zahlen. Diploma Thesis, Ruhr Univ. of Bochum.

[69] ELKER, J., POLLARD, D. and STUTE, W. (1979). Glivenko-Cantelli theorems for classes of convex sets. *F. Appl. Probability*. To appear.

[70] EPANECHNIKOV, V. A. (1968). The significance level and power of the two-sided Kolmogorov test in the case of small samples. *Theor. Probability Appl.* **13** 686–690.

[71] FERNANDEZ, P. J. (1970). A weak convergence theorem for random sums of independent random variables. *Ann. Math. Statist.* **41** 710–712.

[72] FINKELSTEIN, H. (1971). The law of the iterated logarithm for empirical distributions. *Ann. Math. Statist.* **42** 607–615.

[73] FORTET, R. and MOURIER, E. (1953). Convergence de la répartition empirique vers la répartition théorique. *Ann. Sci. École Norm. Sup.* **70** 267–285.

[74] GAENSSLER, P. (1973). On convergence of sample distributions. *Proc. 39th Session ISI* **45** 427–432.

[75] GAENSSLER, P. (1974). Around the Glivenko-Cantelli theorem. Colloquia Mathematica Societatis János Bolyai. *Limit Theorems of Probability Theory* 93–103. Keszthely.

[76] GAENSSLER, P. and STUTE, W. (1976). On uniform convergence of measures with applications to uniform convergence of empirical distributions. *Lecture Notes in Math.* **566** 45–56.

[77] GAENSSLER, P. and STUTE, W. (1977). *Wahrscheinlichkeitstheorie*. Springer, Berlin.

[78] GINÉ, M. E. (1974). On the central limit theorem for sample continuous processes. *Ann. Probability* **2** 629–641.

[79] GLIVENKO, V. (1933). Sulla determinazione empirica della legge di probabilitá. *Giorn. Ist. Ital. Attuari* **4** 92–99.

[80] GNEDENKO, B. V., KOROLUK, V. S. and SKOROKHOD, A. V. (1960). Asymptotic expansions in probability theory. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* **2** 153–170. Univ. of California Press.

[81] GOODMAN, V. (1976). Distribution estimates for functionals of the two-parameter Wiener process. *Ann. Probability* **4** 977–982.

[82] GROENEBOOM, P. (1976). Large deviation theorems for empirical distribution functions. Preprint, Math. Centrum, Amsterdam.

[83] HARDING, E. F. (1967). The number of partitions of a set of N points in k dimensions induced by hyperplanes. *Proc. Edinburgh Math. Soc.* **15** 285–289.

[84] HOADLEY, A. B. (1967). On the probability of large deviations of functions of several empirical cdf's. *Ann. Math. Statist.* **38** 360–381.

[85] HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.

[86] HOEFFDING, W. (1965). On probabilities of large deviations. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **1** 203–219. Univ. of California Press.

[87] DE HOYOS, A. G. (1972). Continuity and convergence of some processes parametrized by the compact convex sets in $\mathbb{R}^s$. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **23** 153–162.

[88] ITÔ, K. and MCKEAN, H. P., JR. (1965). *Diffusion Processes and their Sample Paths*. Springer, Berlin.

[89] JAESCHKE, D. (1975). Über die Grenzverteilung des Maximums der normierten empirischen Verteilungsfunktion. Dissertation, Univ. of Dortmund.

[90] JAESCHKE, D. (1979). The asymptotic distribution of the supremum of the standardized empirical distribution function on subintervals. *Ann. Statist.* **7** 108–115.

[91] JAIN, N. C. and MARCUS, M. B. (1975). Central limit theorems for $C(S)$-valued random variables. *J. Functional Analysis* **19** 216–231.

[92] JAMES, B. R. (1975). A functional law of the iterated logarithm for weighted empirical distributions. *Ann. Probability* **3** 762–772.

[93] KAC, M. (1949). On deviations between theoretical and empirical distributions. *Proc. Nat. Acad. Sci.* **35** 252–257.

[94] KIEFER, J. (1961). On large deviations of the empiric D. F. of vector chance variables and a law of iterated logarithm. *Pacific J. Math.* **11** 649–660.

[95] KIEFER, J. (1970). Deviations between the sample quantile process and the sample df. In *Nonparametric Techniques in Statistical Inference* (M. L. Puri, ed.) 299–319. Cambridge Univ. Press.

[96] KIEFER, J. (1972a). Skorohod embedding of multivariate rv's, and the sample df. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **24** 1–35.

[97] KIEFER, J. (1972b). Iterated logarithm analogues for sample quantiles when $p_n \downarrow 0$. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* **1** 227–244. Univ. of California Press.

[98] KIEFER, J. and WOLFOWITZ, J. (1958). On the deviations of the empiric distribution function of vector chance variables. *Trans. Amer. Math. Soc.* **87** 173–186.

[99] KOLMOGOROV, A. N. (1933). Sulla determinazione empirica di una legge di distribuzione. *Giorn. Ist. Ital. Attuari* **4** 83–91.

[100] KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1974). Weak convergence and embedding. Colloquia Mathematica Societatis János Bolyai. *Limit Theorems of Probability Theory* 149–165. Keszthely.

[101] KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1975). An approximation of partial sums of independent rv's, and the sample df, I. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **32** 111–131.

[102] KOUL, H. L. (1970) Some convergence theorems for ranks and weighted empirical cumulatives. *Ann. Math. Statist.* **41** 1768–1773.

[103] KRICKEBERG, K. (1976). An alternative approach to Glivenko-Cantelli theorems. *Lecture Notes in Math.* **566** 57–67.

[104] KUELBS, J. (1973). The invariance principle for Banach space valued random variables. *J. Multivariate Analysis* **3** 161–172.

[105] KUELBS, J. (1976). A strong convergence theorem for Banach space valued random variables. *Ann. Probability* **4** 744–771.

[106] LAUWERIER, H. A. (1963). The asymptotic expansion of the statistical distribution of N. V. Smirnov. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **2** 61–68.

[107] MANIJA, G. M. (1949). Generalization of the criterion of A. N. Kolmogorov for an estimate for the law of distribution for empirical data. *Dokl. Akad. Nauk SSSR* **69** 495–497 (in Russian).

[108] MASSEY, F. J. (1950). A note on the estimation of a distribution function by confidence limits. *Ann. Math. Statist.* **21** 116–119.

[109] MÜLLER, D. W. (1970). On Glivenko-Cantelli convergence. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **16** 195–210.

[110] NEF, W. (1964). Über die Differenz zwischen theoretischer und empirischer Verteilungsfunktion. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **3** 154–162.

[111] NEUHAUS, G. (1971). On weak convergence of stochastic processes with multidimensional time parameter. *Ann. Math. Statist.* **42** 1285–1295.

[112] NEUHAUS, G. and SEN, P. K. (1977). Weak convergence of tail-sequence processes for sample distributions and averages. *Mitt. Math. Sem. Giessen* **123** 25–35.

[113] NIKITIN, JA. JU. (1972). On a boundary-value problem for an empirical process. *Soviet Math. Dokl.* **13** 1081–1084.

[114] O'REILLY, N. E. (1974). On the weak convergence of empirical processes in sup-norm metrics. *Ann. Probability* **2** 642–651.

[115] OREY, S. and PRUITT, W. E. (1973). Sample functions of the N-parameter Wiener process. *Ann. Probability* **1** 138–163.

[116] PARANJAPE, S. R. and PARK, C. (1973a). Laws of iterated logarithm of multiparameter Wiener processes. *J. Multivariate Analysis* **3** 132–136.

[117] PARANJAPE, S. R. and PARK, C. (1973b). Distribution of the supremum of the two parameter Yeh-Wiener process on the boundary. *J. Appl. Probability* **10** 875–880.

[118] PARTHASARATHY, K. R. (1967). *Probability Measures on Metric Spaces*. Academic Press, New York.

[119] PHILIPP, W. (1971). Mixing sequences of random variables and probabilistic number theory. *Mem. Amer. Math. Soc.* **114**.

[120] PHILIPP, W. (1973). Empirical distribution functions and uniform distribution mod 1. In *Diophantine Approximation and its Applications* (C. F. Osgood, ed.). Academic Press, New York.

[121] PITMAN, E. J. G. (1972). Simple proofs of Steck's determinantal expressions for probabilities in the Kolmogorov and Smirnov tests. *Bull. Austral. Math. Soc.* **7** 227–232.

[122] PÓLYA, G. and SZEGŐ, G. (1972). *Problems and Theorems in Analysis I.* Springer, Berlin.

[123] PROKHOROV, YU. V. (1956). Convergence of random processes and limit theorems in probability theory. *Theor. Probability Appl.* **1** 157–214.

[124] PURI, M. L. and SEN, P. K. (1971). *Nonparametric Methods in Multivariate Analysis.* Wiley, New York.

[125] PYKE, R. (1968). The weak convergence of the empirical process with random sample size. *Proc. Cambridge Philos. Soc.* **64** 155–160.

[126] PYKE, R. (1969). Applications of almost surely convergent constructions of weakly convergent processes. *Lecture Notes in Math.* **89** 187–200. Springer, Berlin.

[127] PYKE, R. (1972). Empirical processes. *Jefferey-Williams Lectures, Canad. Math. Cong., Montreal* 13–43.

[128] PYKE, R. (1975). Multidimensional empirical processes: some comments. In *Statistical Inference and Related Topics I* (M. L. Puri, ed.) 45–58. Academic Press, New York.

[129] PYKE, R. and ROOT, D. H. (1968). An application of stopping times to obtain weak convergence. Technical Report No. 16, Univ. of Washington.

[130] PYKE, R. and SHORACK, G. (1968). Weak convergence of a two sample empirical process and a new approach to Chernoff-Savage theorems. *Ann. Math. Statist.* **39** 755–771.

[131] RAO, R. R. (1960). Generalization of a theorem of Pólya and applications. *Ann. Math. Statist.* **31** 241 (abstract).

[132] RAO, R. R. (1962). Relations between weak and uniform convergence of measures with applications. *Ann. Math. Statist.* **33** 659–680.

[133] RÉNYI, A. (1953). On the theory of order statistics. *Acta Math. Acad. Sci. Hungar.* **4** 191–231.

[134] RÉNYI, A. (1960). On the central limit theorem for the sum of a random number of independent random variables. *Acta Math. Acad. Sci. Hungar.* **11** 97–102.

[135] RÉNYI, A. (1973). On a group of problems in the theory of ordered samples. *Sel. Transl. Math. Statist. Prob.* **13** 289–298.

[136] RÉVÉSZ, P. (1976). On strong approximation of the multidimensional empirical process. *Ann. Probability* **4** 729–743.

[137] RICHTER, H. (1974). Das Gesetz vom iterierten Logarithmus für empirische Verteilungsfunktionen im $\mathbb{R}^k$. *Manuscripta Math.* **11** 291–303.

[138] ROSENKRANTZ, W. (1969). A rate of convergence for the von Mises statistic. *Trans. Amer. Math. Soc.* **139** 329–337.

[139] SAHLER, W. (1968). A survey on distribution-free statistics based on distances between distribution functions. *Metrika* **13** 149–169.

[140] SANOV, I. N. (1957). On the probability of large deviations of random variables. (in Russian). *Mat. Sb. NS* **42** (84) 11–44. (English translation in *Sel. Transl. Math. Statist. Prob.* **1** (1961), 213–244.)

[141] SAWYER, S. (1974). The Skorokhod representation. *Rocky Mountain J. Math.* **4** 579–596.

[142] SAZONOV, V. V. (1963). On the Glivenko-Cantelli theorem. *Theor. Probability Appl.* **8** 282–285.

[143] SCHMID, P. (1958). On the Kolmogorov and Smirnov limit theorems for discontinuous distribution functions. *Ann. Math. Statist.* **29** 1011–1027.

[144] SCHMIDT, W. M. (1975). Irregularities of distribution IX. *Acta Arith.* **27** 385–396.

[145] SEN, P. K. (1973a). On weak convergence of empirical processes for random numbers of independent stochastic vectors. *Proc. Cambridge Philos. Soc.* **73** 139–144.

[146] SEN, P. K. (1973b). An almost sure invariance principle for multivariate Kolmogorov-Smirnov statistics. *Ann. Probability* **1** 488–496.

[147] SETHURAMAN, J. (1964). On the probability of large deviations of families of sample means. *Ann. Math. Statist.* **35** 1304–1316.

[148] SHORACK, G. (1972). Functions of order statistics. *Ann. Math. Statist.* **43** 412–427.

[149] SHORACK, G. (1977). Some laws of the iterated logarithm type results for the empirical process. To appear in *Ann. Probability*.

[150] SHORACK, G. and SMYTHE, R. T. (1976). Inequalities for $\max|S_k|/b_k$ where $k \in \mathbb{N}^r$. *Proc. Amer. Math. Soc.* **54** 331–336.

[151] SHORACK, G. and WELLNER, J. (1978). Linear bounds on the empirical distribution function. *Ann. Probability* **6** 349–353.

[152] SIMPSON, P. B. (1951). Note on the estimation of a bivariate distribution function. *Ann. Math. Statist.* **22** 476–478.

[153] SKOROKHOD, A. V. (1956). Limit theorems for stochastic processes. *Theor. Probability Appl.* **1** 261–290.

[154] SKOROKHOD, A. V. (1965) *Studies in the Theory of Random Processes*. Addison-Wesley, Reading.

[155] SMIRNOV, N. V. (1939). On the deviation of the empirical distribution function. *Math. Sb. NS* **6** 3–26.

[156] SMIRNOV, N. V. (1944). An approximation to the distribution laws of random quantities determined by empirical data. *Uspehi Mat. Nauk* **10** 179–206.

[157] STECK, G. P. (1971). Rectangle probabilities for uniform order statistics and the probability that the empirical distribution function lies between two distribution functions. *Ann. Math. Statist.* **42** 1–11.

[158] STEELE, M. (1978). Empirical discrepancies and subadditive processes. *Ann. Probability* **6** 118–127.

[159] STONE, M. (1974). Large deviations of empirical probability measures. *Ann. Statist.* **2** 362–366.

[160] STRAF, M. L. (1971). Weak convergence of stochastic processes with several parameters. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* **2** 187–221. Univ. of California Press.

[161] STRASSEN, V. (1964). An invariance principle for the law of the iterated logarithm. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **3** 211–226.

[162] STRASSEN, V. (1967). Almost sure behavior of sums of independent random variables and martingales. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **2** 315–343. Univ. of California Press.

[163] STRASSEN, V. and DUDLEY, R. M. (1969). The central limit theorem and ε-entropy. *Lecture Notes in Math.* **89** 224–231. Springer, Berlin.

[164] STUTE, W. (1975). On uniformity classes of functions with an application to the speed of Mean-Glivenko-Cantelli convergence. Preprint, Ruhr Univ. Bochum.

[165] STUTE, W. (1976a). On a generalization of the Glivenko-Cantelli theorem. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **35** 167–175.

[166] STUTE, W. (1976b). A necessary condition for the convergence of the isotrope discrepancy. *Lecture Notes in Math.* **566** 138–140. Springer, Berlin.

[167] STUTE, W. (1976c). Approximation results for the M. Kac empirical process. Preprint, Ruhr Univ. Bochum.

[168] STUTE, W. (1977). Convergence rates for the isotrope discrepancy. *Ann. Probability* **5** 707–723.

[169] SUZUKI, G. (1972). Distributions of Kac-statistics. *Ann. Inst. Statist. Math.* **24** 415–421.

[170] TAKÁCS, L. (1967). *Combinatorial Methods in the Theory of Stochastic Processes*. Wiley, New York.

[171] TAKÁCS, L. (1970). Combinatorial methods in the theory of order statistics. In *Nonparametric Techniques in Statistical Inference* (M. L. Puri, ed.) 359–384. Cambridge Univ. Press.

[172] TOPSØE, F. (1970). On the Glivenko-Cantelli theorem. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **14** 239–250.

[173] TOPSØE, F. (1977). Uniformity in convergence of measures. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **39** 1–30.

[174] TOPSØE, F., DUDLEY, R. M. and HOFFMANN-JØRGENSEN, J. (1976). Two examples concerning uniform convergence w.r.t. balls in Banach spaces. *Lecture Notes in Math.* **566** 141–146.

[175] TUSNÁDY, G. (1977). Strong invariance principles. In *Recent Developments in Statistics* (J. R. Barra et al., eds.) 289–300. North-Holland, Amsterdam.

[176] VALENTINE, F. A. (1964). *Convex Sets*. McGraw-Hill, New York.

[177] VAPNIK, V. N. and CHERVONENKIS, A. YA. (1971). On uniform convergence of the frequencies of events to their probabilities. *Theor. Probability Appl.* **16** 264–280.

[178] VARADARAJAN, V. S. (1958). On the convergence of probability distributions. *Sankhya* **19** 23–26.

[179] VINCZE, I. (1970). On Kolmogorov-Smirnov type distribution theorems. In *Nonparametric Techniques in Statistical Inference* (M. L. Puri, ed.) 385–401. Cambridge Univ. Press.

[180] WELLNER, J. A. (1975). Convergence of the sequential uniform empirical process with bounds for centered beta r.v.'s and a log-log law. Ph.D. dissertation, Univ. of Washington.

[181] WELLNER, J. A. (1977a). A Glivenko-Cantelli theorem and strong laws of large numbers for functions of order statistics. *Ann. Statist.* **5** 473–480.

[182] WELLNER, J. A. (1977b). A law of the iterated logarithm for functions of order statistics. *Ann. Statist.* **5** 481–494.

[183] WELLNER, J. A. (1977c). A martingale inequality for the empirical process. *Ann. Probability* **5** 303–308.

[184] WELLNER, J. A. (1978). Limit theorems for the ratio of the empirical distribution function to the true distribution function. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **45** 73–88.

[185] WICHURA, M. (1968). On the weak convergence of non-Borel probabilities on a metric space. Ph.D. dissertation, Columbia Univ.

[186] WICHURA, M. (1970). On the construction of almost uniformly convergent random variables with given weakly convergent image laws. *Ann. Math. Statist.* **41** 284–291.

[187] WICHURA, M. (1973). Some Strassen-type laws of the iterated logarithm for multiparameter stochastic processes with independent increments. *Ann. Probability* **1** 272–296.

[188] WOLFOWITZ, J. (1954). Generalizations of the theorem of Glivenko-Cantelli. *Ann. Math. Statist.* **25** 131–138.

[189] WOLFOWITZ, J. (1960). Convergence of the empiric distribution function on half-spaces. In *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*. Stanford Univ. Press.

[190] ZAREMBA, S. K. (1971). Sur la discrépance des suites aléatoires. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **20** 236–248.

[191] ZUKER, M. (1974). Speeds of convergence of random probability measures. Ph.D. dissertation, M.I.T.

MATHEMATICAL INSTITUTE
UNIVERSITY OF MUNICH
THERESIENSTRASSE 39
D-8000 MUNICH 2
WEST GERMANY