PATRICK BILLINGSLEY
The University of Chicago

Weak Convergence of Measures:
Applications in Probability

SOCIETY FOR INDUSTRIAL AND APPLIED MATHEMATICS

PHILADELPHIA, PENNSYLVANIA 1971

10 9 8 7 6 5

ISBN 0-89871-176-2

Contents

Preface

The pages that follow are a record of ten lectures I gave during September of 1970 at a Regional Conference in the Mathematical Sciences at Iowa City, sponsored by the Conference Board of the Mathematical Sciences and the University of Iowa with support from the National Science Foundation.

The lectures cover a subset of the material in my book, *Convergence of Probability Measures*, proceeding in a direct path (at a pace brisker than that of the book) from the beginnings of the subject to its applications in limit theory for dependent random variables.

The present treatment departs from the book in many particulars. For example, the mapping theorems are here proved via Skorokhod's representation theorem, Prokhorov's theorem is proved by construction of a content, and the maximal inequality of § 6 is given a proof devised by Michael Wichura. Most important, the limit theorems at the end are proved under a new set of conditions which apply fairly broadly (not quite so broadly as the ones in my book) but at the same time make possible relatively simple proofs. There is a bibliography of recent papers.

Prerequisite for reading this account are the basic parts of metric-space topology and measure-theoretic probability, and one result of functional analysis, the Arzelà–Ascoli theorem.

Thanks are due those who attended the conference for sharp cross-examination and interesting commentary. Special thanks are due Robert V. Hogg, who organized the conference, and John D. Cryer, who took an excellent set of notes.

Chicago PATRICK BILLINGSLEY

May 1971

This page intentionally left blank

Weak Convergence of Measures: Applications in Probability

Patrick Billingsley

1. Introduction. Let Ω be the unit interval $[0, 1]$, let $\mathscr{B}$ consist of the Borel sets in $[0, 1]$, and let P denote Lebesgue measure on $\mathscr{B}$, so that $(\Omega, \mathscr{B}, \mathsf{P})$ is a probability space. Define

$$X_t(\omega) = 0$$

for $0 \leqq t \leqq 1$ and $\omega \in \Omega$, and define

$$Y_t(\omega) = \begin{cases} 0 & \text{if} \quad t \neq \omega, \\ 1 & \text{if} \quad t = \omega \end{cases}$$

for t and ω in the same ranges. Then $\mathsf{P}\{X_t = 0\} = \mathsf{P}\{Y_t = 0\} = 1$, so that the stochastic processes $\{X_t : 0 \leqq t \leqq 1\}$ and $\{Y_t : 0 \leqq t \leqq 1\}$ have the same finite-dimensional distributions, in the sense that

$$\mathsf{P}\{X_{t_1} \leqq x_1, \cdots, X_{t_k} \leqq x_k\} = \mathsf{P}\{Y_{t_1} \leqq x_1, \cdots, Y_{t_k} \leqq x_k\}$$

for all choices of the t_i and x_j. On the other hand,

$$\sup_{0 \leqq t \leqq 1} X_t(\omega) = 0, \qquad \sup_{0 \leqq t \leqq 1} Y_t(\omega) = 1$$

for all ω.

Thus the finite-dimensional distributions of a stochastic process by no means determine the distribution of every function of the process. In order to derive the distribution of a given function of a process one may require, beyond a specification of the finite-dimensional distributions, some kind of regularity condition that in effect involves all time points simultaneously (separability of the process is such a condition).

Something analogous arises in connection with limit theorems. Suppose $X_t(\omega) = 0$, as before, and suppose

$$X_t^{(n)}(\omega) = \begin{cases} nt & \text{if} \quad 0 \leqq t \leqq \dfrac{1}{n}, \\[2ex] 2 - nt & \text{if} \quad \dfrac{1}{n} \leqq t \leqq \dfrac{2}{n}, \\[2ex] 0 & \text{if} \quad \dfrac{2}{n} \leqq t \leqq 1. \end{cases}$$

In other words, for each ω, $X_t^{(n)}(\omega)$ is, as a function of t, the function with graph shown in Fig. 1.

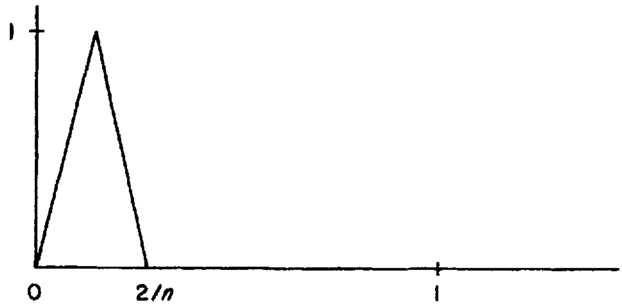

FIG. 1

For each positive t, $X_t^{(n)}(\omega) = 0$ if $t > 2/n$, and hence

$$\lim_{n\to\infty} \mathsf{P}\{X_{t_1}^{(n)} \leqq x_1, \cdots, X_{t_k}^{(n)} \leqq x_k\} = \mathsf{P}\{X_{t_1} \leqq x_1, \cdots, X_{t_k} \leqq x_k\}$$

for all choices of the t_i and x_j. In other words, the finite-dimensional distributions of the process $\{X_t^{(n)}: 0 \leqq t \leqq 1\}$ converge to those of $\{X_t: 0 \leqq t \leqq 1\}$. Nonetheless,

$$\sup_{0 \leqq t \leqq 1} X_t^{(n)}(\omega) = 1, \quad \sup_{0 \leqq t \leqq 1} X_t(\omega) = 0.$$

(This is the advanced-calculus example of a sequence pointwise but not uniformly convergent; that it involves no real randomness makes it simple but not irrelevant.)

Thus the convergence of the finite-dimensional distributions does not imply the convergence of the distribution of every function of the processes. In order to derive the limiting distribution of a given function of the processes one may require, beyond convergence of the finite-dimensional distributions, some convergence condition that in effect involves all time points simultaneously. The theory of weak convergence of probability measures on metric spaces provides such conditions. When applied in appropriate spaces of functions, the theory gives a powerful way of treating convergence problems that go beyond finite-dimensional cases, problems that arise in a natural way in applications.

2. Weak convergence. Let S be a metric space. We shall assume throughout all that follows, without further mention of the fact, that S is both *separable* and *complete*. The interior and closure of a set A we denote A° and A^- respectively, and its boundary $A^- - A^\circ$ we denote ∂A. The distance between points is $\rho(x, y)$, and the distance from x to a set A is

$$\rho(x, A) = \inf\{\rho(x, y) : y \in A\};$$

$\rho(x, A)$ is continuous in x. The δ-neighborhood of a set A is

$$A^\delta = \{x : \rho(x, A) < \delta\}. \tag{2.1}$$

The class of bounded, continuous real-valued functions on S we denote $C(S)$.

The open sets in S generate a σ-field denoted $\mathcal{S}$; the elements of $\mathcal{S}$ are called Borel sets. We work with Borel sets exclusively: each set introduced is assumed to be a Borel set, and each set constructed during the course of a proof can be shown, by an argument ordinarily omitted, to be a Borel set. Similarly, all functions are measurable with respect to $\mathcal{S}$.

We shall be concerned with nonnegative, completely additive set functions P on $\mathcal{S}$ satisfying $P(S) = 1$—that is, with probability measures. If P_n and P are probability measures (on $(S, \mathcal{S})$), we say P_n converges weakly to P, and write $P_n \Rightarrow P$, if

$$\lim_{n\to\infty} \int f\, dP_n = \int f\, dP \tag{2.2}$$

for all functions f in $C(S)$. (We omit the region of integration if it is the entire space.)

A set whose boundary satisfies $P(\partial A) = 0$ is called a P-continuity set. The definition of weak convergence is framed in terms of the convergence of integrals of functions. The following basic theorem characterizes it in terms of the convergence of measures of sets.

THEOREM 2.1. *These four conditions are equivalent*:

(a) $P_n \Rightarrow P$,

(b) $\limsup_n P_n(F) \leqq P(F)$ *for all closed* F,

(c) $\liminf_n P_n(G) \geqq P(G)$ *for all open* G,

(d) $\lim_n P_n(A) = P(A)$ *for all* P*-continuity sets* A.

Proof. We first show that (a) implies (b). Suppose (a) holds and suppose F is closed. Given a positive ε, choose a positive δ such that (see the definition (2.1))

$$P(F^\delta) < P(F) + \varepsilon,$$

which is possible because $F^\delta \downarrow F$ as $\delta \downarrow 0$, F being closed. Now let

$$\varphi(t) = \begin{cases} 1 & \text{if } t \leqq 0, \\ 1 - t & \text{if } 0 \leqq t \leqq 1, \\ 0 & \text{if } t \geqq 1, \end{cases}$$

and define

$$f(x) = \varphi\left(\frac{1}{\delta}\rho(x, F)\right).$$

Since f is nonnegative and assumes the value 1 on F,

$$P_n(F) = \int_F f\, dP_n \leqq \int f\, dP_n.$$

Since f vanishes outside F^δ and never exceeds 1,

$$\int f\,dP = \int_{F^\delta} f\,dP \leqq P(F^\delta).$$

Finally, since $P_n \Rightarrow P$ and f is an element of $C(S)$,

$$\lim_n \int f\,dP_n = \int f\,dP.$$

Therefore,

$$\limsup_n P_n(F) \leqq \lim_n \int f\,dP = \int f\,dP \leqq P(F^\delta) < P(F) + \varepsilon.$$

Since ε was arbitrary, (b) follows.

A simple complementation argument proves the equivalence of (b) and (c). Now (b) and (c) together imply (d) because they imply

$$\begin{aligned} P(A^-) &\geqq \limsup_n P_n(A^-) \geqq \limsup_n P_n(A), \\ &\geqq \liminf_n P_n(A) \geqq \liminf_n P_n(A^\circ) \geqq P(A^\circ), \end{aligned}$$

and, if $P(\partial A) = 0$, the extreme terms are equal so that there is equality throughout.

The proof will be complete if we show that (d) implies (a). So we assume (d) and prove that (2.2) holds for f in $C(S)$. Choose α and β so that $\alpha < f(x) < \beta$ for all x. Now there are only countably many γ for which $P\{x:f(x) = \gamma\} > 0$. Given ε, choose α_i so that $\alpha = \alpha_0 < \alpha_1 < \cdots < \alpha_k = \beta$, $\alpha_i - \alpha_{i-1} < \varepsilon$, and $P\{x:f(x) = \alpha_i\} = 0$. If y is a boundary point of

$$\{x:\alpha_{i-1} < f(x) \leqq \alpha_i\}, \tag{2.3}$$

then $f(y)$ is either α_{i-1} or α_i, so the sets (2.3) are P-continuity sets. Now

$$\sum_{i=1}^{k} \alpha_{i-1} P_n\{x:\alpha_{i-1} < f(x) \leqq \alpha_i\} \leqq \int f\,dP_n \leqq \sum_{i=1}^{k} \alpha_i P_n\{x:\alpha_{i-1} < f(x) \leqq \alpha_i\} \tag{2.4}$$

and

$$\sum_{i=1}^{k} \alpha_{i-1} P\{x:\alpha_{i-1} < f(x) \leqq \alpha_i\} \leqq \int f\,dP \leqq \sum_{i=1}^{k} \alpha_i P\{x:\alpha_{i-1} < f(x) \leqq \alpha_i\}. \tag{2.5}$$

Since the sets (2.3) are P-continuity sets, it follows from the condition (d) that the extreme terms in (2.4) converge respectively to the extreme terms of (2.5). The latter differ by at most ε since $\alpha_i - \alpha_{i-1} < \varepsilon$. Hence the limits superior and inferior of $\int f\,dP_n$ are within ε of $\int f\,dP$, and (2.2) follows because ε was arbitrary.

Suppose now that the probability measures P and Q satisfy

$$\int f\,dP = \int f\,dQ \tag{2.6}$$

for all f in $C(S)$. If $P_n \equiv Q$, then $P_n \Rightarrow Q$, and from Theorem 2.1 we conclude that $Q(F) \leqq P(F)$ for all closed F. Since the roles of Q and P may be interchanged in this argument, P and Q agree for closed sets and hence also for open sets. Let $\mathscr{G}$ consist of those sets A that are F_σ-sets (that is, $F_n \uparrow A$ for some sequence of closed sets) and are at the same time G_δ-sets (that is, $G_n \downarrow A$ for some sequence of open sets). A closed set F is of course an F_σ and is a G_δ because it is the limit of the open sets $F^{1/n}$. Thus $\mathscr{G}$ contains the closed sets and similarly the open sets, and it is not hard to show that $\mathscr{G}$ is a finitely-additive field, which of course generates the σ-field $\mathscr{S}$. If P and Q agree for open sets and closed sets, they certainly must agree for sets in $\mathscr{G}$ and hence for all sets in $\mathscr{S}$. *Thus* (2.6) *implies P and Q are identical.*

If $P_n \Rightarrow P$ and $P_n \Rightarrow Q$, then (2.6) must hold, so P and Q are the same. Thus there can be at most one weak limit.

It is possible to metrize the space of probability measures on S in such a way that it becomes a separable complete metric space and convergence in the metric is equivalent to weak convergence. We shall never use this fact; we shall instead continue to work directly with the notion of convergence: $P_n \Rightarrow P$.

THEOREM 2.2. *If each sequence $\{P_{n_i}\}$ of $\{P_n\}$ contains a further subsequence $\{P_{n_{i_m}}\}$ such that $P_{n_{i_m}} \Rightarrow P$ as $m \to \infty$, then $P_n \Rightarrow P$ as $n \to \infty$.*

Proof. If P_n does not converge weakly to P, then (2.2) fails for some f in $C(S)$, so that

$$\left| \int f \, dP_{n_i} - \int f \, dP \right| \geqq \varepsilon$$

for some positive ε and all P_{n_i} in some subsequence. But then no subsequence of $\{P_{n_i}\}$ can converge weakly to P.

The classical case of weak convergence concerns the real line R^1 with the ordinary metric and probability measures on the class $\mathscr{R}^1$ of Borel sets on the line. Such a probability measure P is completely determined by its distribution function F, defined by

$$F(x) = P(-\infty, x].$$

Suppose $\{P_n\}$ is a sequence of probability measures on $(R^1, \mathscr{R}^1)$ with distribution functions F_n.

THEOREM 2.3. *If $P_n \Rightarrow P$, then $F_n(x) \to F(x)$ at all continuity points x of F, and conversely.*

Proof. If x is a continuity point of F, then $(-\infty, x]$, whose boundary is $\{x\}$, is a P-continuity set, so that $F_n(x) = P_n(-\infty, x] \to F(x) = P(-\infty, x]$.

Suppose on the other hand there is convergence of the distributions functions at continuity points of the limit. Given ε, choose continuity points a and b of F so that $a < b$, $F(a) < \varepsilon$, and $F(b) > 1 - \varepsilon$. And given an f in $C(R^1)$, choose continuity points x_i of F such that $a = x_0 < x_1 < \cdots < x_k = b$ and $|f(x) - f(x_i)| < \varepsilon$ for $x_{i-1} \leqq x \leqq x_i$. Let S be the Riemann sum $\sum_{i=1}^k f(x_i)[F(x_i) - F(x_{i-1})]$ and let S_n be this sum with F_n replacing F. Let M be an upper bound for $|f(x)|$. By hypothesis, $S_n \to S$, and since

$$\left| \int f \, dP - S \right| \leqq \varepsilon + MF(a) + M(1 - F(b)) < (2M + 1)\varepsilon$$

and

$$\left|\int f\,dP_n - S_n\right| \leqq \varepsilon + MF_n(a) + M(1 - F_n(b)) \to \varepsilon + MF(a)$$

$$+ M(1 - F(b)) < (2M + 1)\varepsilon,$$

the limits superior and inferior of $\int f\,dP_n$ are within $(4M + 2)\varepsilon$ of $\int f\,dP$, which completes the proof since ε was arbitrary.

A probability measure on the class $\mathcal{R}^k$ of Borel sets in k-dimensional Euclidean space R^k is determined by its distribution function

$$F(x_1, \cdots, x_k) = P\{y = (y_1, \cdots, y_k): y_1 \leqq x_1, \cdots, y_k \leqq x_k\}.$$

The analogue to Theorem 2.3 holds and can be proved by a consideration of k-dimensional Riemann sums.

3. Random elements and convergence in distribution. Let S be a metric space as before, and let $(\Omega, \mathcal{B}, \mathsf{P})$ be a probability measure space. A mapping $X:\Omega \to S$ is called a random element of S if it is measurable in the sense that $\{\omega: X(\omega) \in A\} = X^{-1}A \in \mathcal{B}$ for each $A \in \mathcal{S}$. Special cases are random variables ($S = R^1$) and random vectors ($S = R^k$). The distribution of X is the measure $P = \mathsf{P}X^{-1}$ on $\mathcal{S}$:

$$P(A) = \mathsf{P}X^{-1}(A) = \mathsf{P}\{\omega: X(\omega) \in A\} = \mathsf{P}\{X \in A\}.$$

Suppose in addition to X we have a sequence of random elements X_n of S, defined on spaces $(\Omega_n, \mathcal{B}_n, \mathsf{P}_n)$, with distributions $P_n = \mathsf{P}_n X_n^{-1}$. If $P_n \Rightarrow P$ we say X_n converges in distribution to X and write $X_n \Rightarrow X$. Every result about weak convergence has an anologue about convergence in distribution, and vice versa, and to pass from the one to the other requires not a proof, but merely a translation. For example, by Theorem 2.1, $X_n \Rightarrow X$ if and only if $\limsup_n \mathsf{P}_n\{X_n \in F\} \leqq \mathsf{P}\{X \in F\}$ for all closed sets F.

Suppose X_n and Y_n are both random elements of S defined on Ω_n. We shall drop the subscript n from P_n.

THEOREM 3.1. *If $X_n \Rightarrow X$ and the distance $\rho(X_n, Y_n)$ converges to 0 in probability, then $Y_n \Rightarrow X$.*

Proof. Clearly $\mathsf{P}\{Y_n \in F\} \leqq \mathsf{P}\{X_n \in (F^\delta)^-\} + \mathsf{P}\{\rho(X_n, Y_n) \geqq \delta\}$. The second term here goes to 0 since $\rho(X_n, Y_n)$ converges in probability to 0, and since $X_n \Rightarrow X$, we have

$$\limsup_n \mathsf{P}\{Y_n \in F\} \leqq \mathsf{P}\{X \in (F^\delta)^-\}.$$

Since $(F^\delta)^- \downarrow F$ as $\delta \downarrow 0$ if F is closed, the result follows.

Let us use the term *Lebesgue interval* to refer to the probability measure space $(\Omega, \mathcal{B}, \mathsf{P})$, where Ω is the unit interval, $\mathcal{B}$ is the σ-field of Borel sets in Ω and P is Lebesgue measure on $\mathcal{B}$.

THEOREM 3.2. *For each probability measure P on $(S, \mathcal{S})$, there is a random element of S, defined on the Lebesgue interval, with distribution P.*

Proof. For each k, construct a decomposition $\mathscr{A}_k = \{A_{k1}, A_{k2}, \cdots\}$ of S into disjoint sets of diameter less than $1/k$, and arrange that $\mathscr{A}_{k+1}$ refines $\mathscr{A}_k$. And construct a decomposition $\mathscr{I}_k = \{I_{k1}, I_{k2}, \cdots\}$ of the unit interval into subintervals whose lengths satisfy $|I_{ku}| = P(A_{ku})$, and arrange that $\mathscr{I}_{k+1}$ refines $\mathscr{I}_k$. Finally, arrange the indexing so that $A_{ku} \supset A_{k+1,v}$ if and only if $I_{ku} \supset I_{k+1,v}$. The construction can be carried out inductively because, if $l_1, l_2, \cdots$ are nonnegative numbers adding to the length of an interval, that interval can be split into subintervals of lengths $l_1, l_2, \cdots$.

Let x_{ku} be some point in A_{ku}, and define a random element X_k by

$$X_k(\omega) = x_{ku} \quad \text{if} \quad \omega \in I_{ku}. \tag{3.1}$$

Since $\{X_k(\omega), X_{k+1}(\omega), \cdots\}$ is contained in some one element of $\mathscr{A}_k$, its diameter is at most $1/k$; thus $\{X_k(\omega)\}$ is a Cauchy sequence for each ω, the limit $X(\omega) = \lim_k X_k(\omega)$ exists, and this limit satisfies

$$\rho(X(\omega), X_k(\omega)) \leqq \frac{1}{k}. \tag{3.2}$$

If a prime denotes a sum or union extended over those u for which A_{ku} meets a given set F, then

$$\begin{aligned} \mathsf{P}\{X_k \in F\} &\leqq \mathsf{P}\{X_k \in \textstyle\bigcup' A_{ku}\} = \sum' \mathsf{P}\{X_k \in A_{ku}\} \\ &= \textstyle\sum' |I_{ku}| = \sum' P(A_{ku}) \leqq P((F^{1/k})^-). \end{aligned}$$

If F is closed, it follows that

$$\limsup_k \mathsf{P}\{X_k \in F\} \leqq P(F).$$

Thus the distribution of X_k converges to P, and hence, by (3.2) and Theorem 3.1 (with $Y_k = X$), X has distribution P.

We turn now to the Skorokhod representation theorem, an extension of Theorem 3.2. In addition to P, consider a sequence $\{P_n\}$ of probability measures on S.

THEOREM 3.3. *If $P_n \Rightarrow P$, then there exist on the Lebesgue interval random elements X_n and X which have respective distributions P_n and P and satisfy* $\lim_n X_n(\omega) = X(\omega)$ *for each ω.*

Proof. Construct the decompositions $\mathscr{A}_k$ of the preceding proof, but this time require that each A_{ku} be a P-continuity set. (Since $\partial\{y : \rho(x, y) < \delta\} \subset \{y : \rho(x, y) = \delta\}$, the spheres about x are P-continuity sets except for countably many radii, so S can be covered by countably many P-continuity sets of diameter less than $1/k$. The usual procedure for rendering the sets disjoint preserves P-continuity because $\partial(A \cap B) \subset (\partial A) \cup (\partial B)$.)

Consider the decompositions $\mathscr{I}_k$ as before, and, for each n, construct successively finer partitions $\mathscr{I}_k^{(n)} = \{I_{k1}^{(n)}, I_{k2}^{(n)}, \cdots\}$ with $|I_{ku}^{(n)}| = P_n(A_{ku})$. Inductively arrange the indexing so that (here $I < J$ for intervals means the right endpoint of I does not exceed the left endpoint of J) $I_{ku}^{(n)} < I_{kv}^{(n)}$ if and only if $I_{ku} < I_{kv}$. In other words, ensure that for each k the families $\mathscr{I}_k, \mathscr{I}_k^{(1)}, \mathscr{I}_k^{(2)}, \cdots$ are ordered similarly.

Define X_k by (3.1), as before, where $x_{ku} \in A_{ku}$, and define

$$X_k^{(n)}(\omega) = x_{ku} \quad \text{if} \quad \omega \in I_{ku}^{(n)}.$$

Again $X_k(\omega)$ converges to an $X(\omega)$ satisfying (3.2), and $X_k^{(n)}(\omega)$ converges ($k \to \infty$) to an $X^{(n)}(\omega)$ satisfying

$$\rho(X^{(n)}(\omega), X_k^{(n)}(\omega)) \leqq \frac{1}{k}. \tag{3.3}$$

And again X has distribution P and $X^{(n)}$ has distribution P_n.

Since $\sum_u [P(A_{ku}) - P_n(A_{ku})] = 0$, we have

$$\begin{aligned}\sum_u \big||I_{ku}| - |I_{ku}^{(n)}|\big| &= \sum_u |P(A_{ku}) - P_n(A_{ku})| \\ &= 2\sum_u{}'' [P(A_{ku}) - P_n(A_{ku})] = 2\sum_u [P(A_{ku}) - P_n(A_{ku})]^+,\end{aligned}$$

where the next-to-last sum extends over those u for which the summand is positive. Each summand goes to 0 as $n \to \infty$ because the A_{ku} are P-continuity sets, and it follows by dominated convergence that

$$\lim_{n\to\infty} \sum_u \big||I_{ku}| - |I_{ku}^{(n)}|\big| = 0. \tag{3.4}$$

Fix k and u_0, let α and α_n be the left endpoints of I_{ku_0} and $I_{ku_0}^{(n)}$ respectively, and let $\sum'$ indicate summation over the set of u for which $I_{ku} < I_{ku_0}$ (which is the same as the set for which $I_{ku}^{(n)} < I_{ku_0}^{(u)}$). Then (3.4) implies

$$\alpha = \sum_u{}' |I_{ku}| = \lim_{n\to\infty} \sum_u{}' |I_{ku}^{(n)}| = \lim_{n\to\infty} \alpha_n.$$

Similarly the right endpoint of $I_{ku}^{(n)}$ converges as $n \to \infty$ to the right endpoint of I_{ku}.

Hence, if ω is interior to I_{ku}, then ω lies in $I_{ku}^{(n)}$ for all sufficiently large n, so that, by (3.2) and (3.3),

$$\rho(X(\omega), X^{(n)}(\omega)) \leqq \frac{2}{k}. \tag{3.5}$$

Thus, if ω is not an endpoint of any I_{ku}, then, for each k, (3.5) holds for all sufficiently large n. In other words, $\lim_n X^{(n)}(\omega) = X(\omega)$ if ω is not in the set of endpoints of the I_{ku}. This last set, being countable, has Lebesgue measure 0, so that, if $X^{(n)}(\omega)$ is redefined as $X(\omega)$ on this set, $X^{(n)}$ still has distribution P_n, and there is now convergence for all ω. This proves the theorem (with $X^{(n)}$ for X_n).

The theorem can be restated: Consider random elements X and X_n of S; they may all be defined on different probability spaces.

COROLLARY 1. *If $X_n \Rightarrow X$, then there exist on the Lebesgue interval random elements Y_n and Y which have the distributions of X_n and X respectively and which satisfy* $\lim_n Y_n(\omega) = Y(\omega)$ *for all* ω.

If P is a measure on S, and if f is a mapping from S to another metric space S', measurable in the sense that $f^{-1}A \in \mathscr{S}$ if A is in the σ-field $\mathscr{S}'$ of Borel sets in S', then Pf^{-1} is the probability measure on $\mathscr{S}'$ defined by $Pf^{-1}(A) = P(f^{-1}A)$. Suppose, in addition, we have a sequence of measures P_n. Let D_f be the set of discontinuities of f and assume it lies in $\mathscr{S}$.

COROLLARY 2. *If $P_n \Rightarrow P$ and $P(D_f) = 0$, then $P_n f^{-1} \Rightarrow Pf^{-1}$.*

Proof. Consider the random elements of Theorem 3.3. Now $\lim_n X_n(\omega) = X(\omega)$ for each ω, and if f is continuous at $X(\omega)$, which by hypothesis holds except on an ω-set of Lebesgue measure 0, then

$$\lim_n f(X_n(\omega)) = f(X(\omega)). \tag{3.6}$$

Thus (3.6) holds for almost all ω, and, since $f(X_n)$ and $f(X)$ have respective distributions $P_n f^{-1}$ and Pf^{-1}, it follows that $P_n f^{-1} \Rightarrow Pf^{-1}$.

COROLLARY 3. *If $X_n \Rightarrow X$ and $\mathsf{P}\{X \in D_f\} = 0$, then $f(X_n) \Rightarrow f(X)$.*

This corollary is a direct translation of Corollary 2. It can also be deduced from Corollary 1.

COROLLARY 4. *If random variables X_n and X satisfy $X_n \Rightarrow X$, then*

$$\mathsf{E}\{|X|\} \leqq \limsup_n \mathsf{E}\{|X_n|\}. \tag{3.7}$$

COROLLARY 5. *If random variables X_n and X satisfy $X_n \Rightarrow X$, and if the X_n are uniformly integrable in the sense that*

$$\lim_{\alpha \to \infty} \sup_n \int_{\{|X_n| > \alpha\}} |X_n| \, d\mathsf{P} = 0, \tag{3.8}$$

then X is integrable and

$$\lim_n \mathsf{E}\{X_n\} = \mathsf{E}\{X\}.$$

To prove Corollary 4, consider the random variables Y_n and Y guaranteed by Corollary 1. Fatou's lemma implies $\mathsf{E}\{|Y|\} \leqq \limsup_n \mathsf{E}\{|Y_n|\}$, and (3.7) follows because Y_n and Y have the distributions of X_n and X. (The E in (3.7) denotes expected value with respect to whatever probability measure governs the random variable in question.) Corollary 5 similarly reduces to a standard fact of integration theory.

Theorem 3.3 can be used to give simple proofs of many results in statistics, for example, those connected with the δ-method.

4. Prokhorov's theorem. A family Π of probability measures on S is said to be relatively compact if each sequence $\{P_n\}$ of elements of Π contains some subsequence $\{P_{n_i}\}$ converging weakly to some probability measure P. The limit P is not required to lie in Π, but of course it must be a probability measure on S.

It is possible to metrize the space of probability measures on S (see the remarks preceding Theorem 2.2), and Π is relatively compact if and only if it has compact closure in this metric. It is not necessary to go into this matter, however, because the definition above makes good sense as it stands.

The following theorem, due to Prokhorov, is basic to the application of weak convergence in probability theory. The family Π is said to be *tight* if, for each positive ε, there is a compact set K_ε for which $P(K_\varepsilon) > 1 - \varepsilon$ for every P in Π.

THEOREM 4.1. *The family* Π *is relatively compact if and only if it is tight.*

Proof. Suppose that Π is tight. There is a sequence $\{K_u\}$ of compact sets such that $K_1 \subset K_2 \subset \cdots$ and $P(K_u) > 1 - 1/u$ for all u. Let $\mathscr{A}$ be a countable collection of open spheres forming a base for the topology of S, and let $\mathscr{H}$ consist of the finite unions of sets of the form $A^- \cap K_u$ with $u \geqq 1$ and A an element of $\mathscr{A}$. Then $\mathscr{H}$ is countable and is closed under the formation of finite unions, and each set in $\mathscr{H}$ is compact.

Given a sequence $\{P_n\}$ in Π, select by the diagonal procedure a subsequence $\{P_{n_i}\}$ along which limits

$$\alpha(H) = \lim_{i\to\infty} P_{n_i}(H) \tag{4.1}$$

exist for all H in $\mathscr{H}$. Suppose there exists a probability measure P such that

$$P(G) = \sup_{H \subset G} \alpha(H) \tag{4.2}$$

for all open sets G. Then $P_{n_i} \Rightarrow P$ as $i \to \infty$ because, if $H \subset G$, $\alpha(H) = \lim_i P_{n_i}(H) \leqq \liminf_i P_{n_i}(G)$, whence $P(G) \leqq \liminf_i P_{n_i}(G)$ follows via (4.2), proving weak convergence. Thus it suffices to produce a P satisfying (4.2).

Clearly $\alpha(H)$, defined by (4.1) for all H in $\mathscr{H}$, has these properties:

$$\alpha(H_1) \leqq \alpha(H_2) \quad \text{if} \quad H_1 \subset H_2; \tag{4.3}$$

$$\alpha(H_1 \cup H_2) = \alpha(H_1) + \alpha(H_2) \quad \text{if} \quad H_1 \cap H_2 = 0; \tag{4.4}$$

$$\alpha(H_1 \cup H_2) \leqq \alpha(H_1) + \alpha(H_2). \tag{4.5}$$

Define

$$\beta(G) = \sup_{H \subset G} \alpha(H) \tag{4.6}$$

for open sets G, and then define

$$\gamma(M) = \inf_{M \subset G} \beta(G)$$

for arbitrary subsets M of S. Clearly $\gamma(G) = \beta(G)$ for open G.

Now suppose we succeed in proving that γ is an outer measure and that each closed set is γ-measurable (measurable with respect to γ). Then all sets in $\mathscr{S}$ will be γ-measurable (recall the γ-measurable sets form a σ-field) and the restriction P of γ to $\mathscr{S}$ will be a measure satisfying $P(G) = \gamma(G) = \beta(G)$, so (4.2) will hold for open G as required, and P will be a probability measure because

$$1 \geqq P(S) = \beta(S) \geqq \sup_u \alpha(K_u) \geqq \sup_u \left(1 - \frac{1}{u}\right).$$

We first prove that β *is finitely subadditive* (on open sets): If $H \subset G_1 \cup G_2$ and $H \in \mathscr{H}$, define $F_1 = \{x \in H : \rho(x, G_1^c) \geqq \rho(x, G_2^c)\}$ and $F_2 = \{x \in H : \rho(x, G_2^c)$

$\geqq \rho(x, G_1^c)\}$ (see Fig. 2). If $x \in F_1$ and $x \notin G_1$, then $x \in G_2$, so that, since G_2^c is closed, $\rho(x, G_1^c) = 0 < \rho(x, G_2^c)$, a contradiction. Thus $F_1 \subset G_1$; similarly $F_2 \subset G_2$. Since F_1 is compact, being a closed subset of the compact set H, and F_1 is inside the open set G_1, it follows by the definition of $\mathscr{H}$ that $F_1 \subset H_1 \subset G_1$ for some H_1 in $\mathscr{H}$; similarly $F_2 \subset H_2 \subset G_2$ for some H_2 in $\mathscr{H}$. But then $\alpha(H) \leqq \alpha(H_1 \cup H_2) \leqq \alpha(H_1) + \alpha(H_2) \leqq \beta(G_1) + \beta(G_2)$ by (4.3), (4.5), and (4.6). Since we can vary H inside $G_1 \cup G_2$, $\beta(G_1 \cup G_2) \leqq \beta(G_1) + \beta(G_2)$ follows.

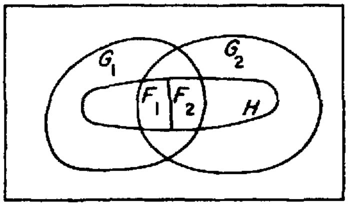

FIG. 2

Next, β *is countably subadditive* (on open sets): For, if $H \subset \bigcup_n G_n$, then, since H is compact, $H \subset \bigcup_{n \leqq n_0} G_n$ for some n_0, and therefore, by finite subadditivity, $\alpha(H) \leqq \beta(\bigcup_{n \leqq n_0} G_n) \leqq \sum_{n \leqq n_0} \beta(G_n) \leqq \sum_n \beta(G_n)$. Taking the supremum over H inside $\bigcup_n G_n$ gives $\beta(\bigcup_n G_n) \leqq \sum_n \beta(G_n)$.

And γ *is an outer measure*: Since γ is clearly monotone, we need only prove it countably subadditive. Given a positive ε and arbitrary subsets M_n of S, choose open sets G_n such that $M_n \subset G_n$ and $\beta(G_n) < \gamma(M_n) + \varepsilon/2^n$. Then, by the countable subadditivity of β, $\gamma(\bigcup_n M_n) \leqq \beta(\bigcup_n G_n) \leqq \sum_n \beta(G_n) < \sum_n \gamma(M_n) + \varepsilon$, whence, ε being arbitrary, we conclude $\gamma(\bigcup_n M_n) \leqq \sum_n \gamma(M_n)$.

It remains only to prove that *each closed set is γ-measurable.* We must show that, if F is closed and M arbitrary,

$$\gamma(M) \geqq \gamma(M \cap F) + \gamma(M \cap F^c) \tag{4.7}$$

(the reverse inequality follows by the subadditivity of γ). To prove (4.7) it suffices to prove

$$\beta(G) \geqq \gamma(G \cap F) + \gamma(G \cap F^c) \tag{4.8}$$

for open G, because then $G \supset M$ implies $\beta(G) \geqq \gamma(M \cap F) + \gamma(M \cap F^c)$ and taking the infimum over G gives (4.7).

To prove (4.8), choose, for given positive ε, an H_0 in $\mathscr{H}$ for which $H_0 \subset G \cap F^c$ and $\alpha(H_0) > \beta(G \cap F^c) - \varepsilon$. Now choose an H_1 in $\mathscr{H}$ for which $H_1 \subset G \cap H_0^c$ and $\alpha(H_1) > \beta(G \cap H_0^c) - \varepsilon$. Since H_0 and H_1 are disjoint and are contained in G (see Fig. 3), it follows by (4.4) that $\beta(G) \geqq \alpha(H_0 \cup H_1) = \alpha(H_0) + \alpha(H_1) > \beta(G \cap F^c) + \beta(G \cap H_0^c) - 2\varepsilon \geqq \gamma(G \cap F^c) + \gamma(G \cap F) - 2\varepsilon$. Since ε was arbitrary, this proves (4.8).

We turn to the converse problem of showing that a relatively compact Π must be tight. Consider a sequence $A_1, A_2, \cdots$ of open spheres of radius δ that cover S. For each ε, there exists an n such that, if $B_n = \bigcup_{i \leqq n} A_i$, $P(B_n) > 1 - \varepsilon$ for all P in Π, because otherwise for each n we have $P_n(B_n) \leqq 1 - \varepsilon$ for some P_n in Π,

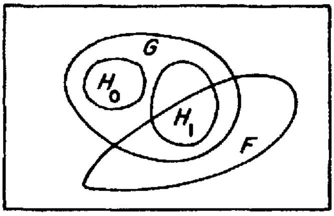

FIG. 3

and by relative compactness $P_{n_i} \Rightarrow P_0$ for some subsequence $\{P_{n_i}\}$ and probability measure P_0, which is impossible because $P_0(B_n) \leqq \liminf_i P_{n_i}(B_n) \leqq \liminf_i P_{n_i}(B_{n_i}) \leqq 1 - \varepsilon$ while $B_n \uparrow S$.

Thus for each positive ε and δ, there are finitely many spheres $A_1, \cdots, A_n$ of radius δ such that $P(\bigcup_{i \leqq n} A_i) > 1 - \varepsilon$ for all P in Π. Choose spheres $A_{k1}, \cdots, A_{kn_k}$ of radius $1/k$ such that $P(\bigcup_{i \leqq n_k} A_{ki}) > 1 - \varepsilon/2^k$. If K is the closure of the totally bounded set $\bigcap_{k \geqq 1} \bigcup_{i \leqq n_k} A_{ki}$, then K is compact and $P(K) > 1 - \varepsilon$ for all P in Π.

5. The space C. From here on we shall be concerned with the space C of continuous functions $x = x(t)$ on the closed unit interval, metrized by

$$\rho(x, y) = \sup_{0 \leqq t \leqq 1} |x(t) - y(t)|.$$

We denote by $\mathscr{C}$ the σ-field of Borel sets in C, and we shall be concerned with probability measures on $(C, \mathscr{C})$.

If $0 \leqq t_1 < t_2 < \cdots < t_k \leqq 1$, the mapping $\pi_{t_1 \cdots t_k}(x) = (x(t_1), \cdots, x(t_k))$ carries C continuously into R^k. Sets of the form $\pi_{t_1 \cdots t_k}^{-1} H$ with H an element of $\mathscr{R}^k$, a Borel set in R^k (k and $t_1, \cdots, t_k$ arbitrary), are called finite-dimensional sets, and the finite-dimensional sets form a finitely additive field. The closed sphere of radius r about x is the intersection of the finite-dimensional sets $\{y : |y(t) - x(t)| \leqq r\}$ with t ranging over the rationals; each open sphere is a countable union of closed spheres and each open set is a countable union of open spheres and hence lies in the σ-field generated by the finite-dimensional sets. Thus the finite-dimensional sets form a finitely additive field generating $\mathscr{C}$.

For a probability measure P on C, the various measures $P\pi_{t_1 \cdots t_k}^{-1}$ on the spaces R^k are called the finite-dimensional distributions of P. If two measures have the same finite-dimensional distributions, they agree for finite-dimensional sets, and hence, since these sets constitute a field generating $\mathscr{C}$, they are the same measure. Thus *the finite-dimensional distributions* $P\pi_{t_1 \cdots t_k}^{-1}$ *of* P *uniquely determine* P *itself.*

Suppose now that $x(t) \equiv 0$ and that x_n is the function given by

$$x_n(t) = \begin{cases} nt & \text{if } 0 \leqq t \leqq \frac{1}{n}, \\ 2 - nt & \text{if } \frac{1}{n} \leqq t \leqq \frac{2}{n}, \\ 0 & \text{if } \frac{2}{n} \leqq t \leqq 1; \end{cases}$$

let P be a unit mass at x (that is, $P(A)$ is 1 or 0 according as x lies in A or not) and let P_n be a unit mass at x_n. Now if $2/n$ is less than the smallest nonzero t_i, then x and x_n either both lie in $\pi_{t_1\cdots t_k}^{-1}H$ or else neither one does, so that $P\pi_{t_1\cdots t_k}^{-1}(H) = P_n\pi_{t_1\cdots t_k}^{-1}(H)$. Therefore there is weak convergence $P_n\pi_{t_1\cdots t_k}^{-1} \Rightarrow P\pi_{t_1\cdots t_k}^{-1}$ in R^k for each $t_1, \cdots, t_k$. On the other hand, the set $\{y : |y(t)| \leqq 1/2, 0 \leqq t \leqq 1\}$, the sphere of radius 1/2 about x, is a P-continuity set and $P_n(A) = 1$ does not converge to $P(A) = 0$. Thus P_n does not converge weakly to P.

This example shows that if there is convergence of the finite-dimensional distributions, that is, if

$$P_n\pi_{t_1\cdots t_k}^{-1} \Rightarrow P\pi_{t_1\cdots t_k}^{-1} \tag{5.1}$$

for all k and $t_1, \cdots, t_k$, it does not follow that there is weak convergence of P_n to P:

$$P_n \Rightarrow P. \tag{5.2}$$

(The converse proposition of course does hold because of Corollary 2 to Theorem 3.3.) Thus weak convergence in C involves considerations going beyond finite-dimensional ones, which is why it is useful (see the introduction).

On the other hand, (5.1) does imply (5.2) in the presence of relative compactness.

THEOREM 5.1. *If* (5.1) *holds for all k and $t_1, \cdots, t_k$, and if $\{P_n\}$ is relatively compact, then* (5.2) *holds.*

Proof. Since $\{P_n\}$ is relatively compact, each subsequence $\{P_{n_i}\}$ contains a further subsequence $\{P_{n_{i_m}}\}$ such that $P_{n_{i_m}} \Rightarrow Q$ as $m \to \infty$ for some probability measure Q on C. But then $P_{n_{i_m}}\pi_{t_1\cdots t_k}^{-1} \Rightarrow Q\pi_{t_1\cdots t_k}^{-1}$, so that, because of (5.1), $Q\pi_{t_1\cdots t_k}^{-1} = P\pi_{t_1\cdots t_k}^{-1}$. Thus P and Q have the same finite-dimensional distributions and, as observed above, this implies $P = Q$. Thus each subsequence of $\{P_n\}$ contains a further subsequence converging weakly to P, and (5.2) follows by Theorem 2.2.

Theorem 4.1 characterizes relative compactness by tightness. In order to apply Theorem 5.1 in concrete cases, we shall in turn characterize tightness by means of the Arzelà–Ascoli theorem.

For $x \in C$ and $\delta > 0$, the modulus of continuity is defined by

$$w_x(\delta) = \sup\{|x(s) - x(t)| : 0 \leqq s, t \leqq 1, |s - t| < \delta\}.$$

According to the Arzelà–Ascoli theorem, a set A in C has compact closure if and only if

$$\sup_{x\in C}|x(0)| < \infty \tag{5.3}$$

and

$$\lim_{\delta\to 0}\sup_{x\in C} w_x(\delta) = 0. \tag{5.4}$$

THEOREM 5.2. *A family Π of probability measures on C is tight (hence relatively compact) if and only if for each η there exists an a such that*

$$P\{x : |x(0)| > a\} < \eta, \qquad P \in \Pi, \tag{5.5}$$

and for each η and ε there exists a δ such that

$$P\{x: w_x(\delta) > \varepsilon\} < \eta, \qquad P \in \Pi. \tag{5.6}$$

Proof. Let A_a be the set in (5.5) and let $B_{\varepsilon,\delta}$ be the set in (5.6). If Π is tight, then, for each η there is a compact set K such that $P(K) > 1 - \eta$ for all P in Π. But, by the Arzelà–Ascoli theorem (see (5.3) and (5.4)), $K \subset A_a^c$ for large a and $K \subset B_{\varepsilon,\delta}^c$ for small δ, so (5.5) and (5.6) hold.

On the other hand, if (5.5) and (5.6) hold and η is given, choose a and δ_k so that $P(A_a) < \eta/2$ and $P(B_{k^{-1},\delta_k}) < \eta/2^{k+1}$ for all $k = 1, 2, \cdots$ and all P in Π. If K is the closure of $A_a^c \cap \bigcap_{k=1}^{\infty} B_{k^{-1},\delta_k}^c$, then $P(K) > 1 - \eta$ for all $P \in \Pi$, and, by the Arzelà–Ascoli theorem, K is compact.

Since an individual measure P forms a tight set (because it forms a relatively compact one), the inequalities in (5.5) and (5.6) hold for a single given P for large enough a and small enough δ. As sufficient conditions for tightness, therefore, we may relax (5.5) and (5.6) by allowing them to fail for finitely many P in Π, since these exceptional P may be provided for by increasing a and decreasing δ. Thus Theorem 5.1 has an alternate form in case Π is a sequence.

THEOREM 5.3. *A sequence $\{P_n\}$ of probability measures on C is tight (hence relatively compact) if and only if for each η there exist a and n_0 such that*

$$P_n\{x: |x(0)| > a\} < \eta, \qquad n \geqq n_0, \tag{5.7}$$

and for each η and ε there exist δ and n_0 such that

$$P_n\{x: w_x(\delta) > \varepsilon\} < \eta, \qquad n \geqq n_0. \tag{5.8}$$

The P_n we shall be concerned with arise as distributions of random elements of C, which we call random functions, constructed in the following way. Let $\xi_1, \xi_2, \cdots$ be a sequence of random variables defined on some probability measure space $(\Omega, \mathscr{B}, \mathrm{P})$, define $S_k = \xi_1 + \cdots + \xi_k$ (with $S_0 = 0$), and for $\omega \in \Omega$ and $0 \leqq t \leqq 1$ put

$$X_n(t, \omega) = \frac{1}{\sigma\sqrt{n}} S_{[nt]}(\omega) + \frac{nt - [nt]}{\sigma\sqrt{n}} \xi_{[nt]+1}(\omega), \tag{5.9}$$

where σ is a positive constant.

For each ω, $X_n(t, \omega)$ is, as a function of t, an element of C; at $t = k/n$ its value is $S_k(\omega)/\sigma\sqrt{n}$, and it varies linearly in the intervals between such points. This element of C we denote $X_n(\omega)$; this is the value at ω of a mapping from Ω to C, and we denote the mapping itself by X_n. On the other hand, if t is fixed, $\omega \to X_n(t, \omega)$ gives a mapping from Ω to R, and we denote this mapping by $X_n(t)$; it is, by (5.9) an ordinary random variable on Ω. If $A = \pi_{t_1 \cdots t_k}^{-1} H$ is a finite-dimensional set in $C (H \in \mathscr{R}^k)$, then

$$X_n^{-1}A = \{\omega: X_n(\omega) \in A\} = \{\omega: (X_n(t_1, \omega), \cdots, X_n(t_k, \omega)) \in H\}$$

lies in $\mathscr{B}$. Since $X_n^{-1}A \in \mathscr{B}$ for finite-dimensional sets A, and since the latter sets generate $\mathscr{C}$, it follows that $X_n^{-1}A \in \mathscr{B}$ for all $A \in \mathscr{C}$. In other words, $\omega \to X_n(\omega)$ is a measurable mapping and hence X_n is a random element of C.

We shall be interested in proving that the random functions X_n converge in distribution under appropriate conditions, and this will require establishing that their distributions are tight.

THEOREM 5.4. *If the sequence $\{\xi_n\}$ is stationary, and if for each ε there exists a λ, with $\lambda > 1$, and an n_0 such that*

$$\mathsf{P}\{\max_{i \leqq n}|S_i| \geqq \lambda\sigma\sqrt{n}\} \leqq \frac{\varepsilon}{\lambda^2}, \qquad n \geqq n_0, \tag{5.10}$$

then the distributions on C of the random functions (5.9) form a tight sequence.

Proof. Certainly the distributions P_n of the X_n satisfy (5.7) since $X_n(0, \omega) = 0$, and what we must verify is (5.8). For given ε and η, there exist, by the hypothesis (5.10) with $\varepsilon^2\eta/6^2$ in place of ε, a $\lambda > 1$ and a d_0 such that

$$\mathsf{P}\{\max_{i \leqq d}|S_i| \geqq \lambda\sigma\sqrt{d}\} \leqq \frac{\varepsilon^2\eta}{6^2\lambda^2} \tag{5.11}$$

if $d \geqq d_0$.

Define δ by

$$\delta = \frac{\varepsilon^2}{6^2\lambda^2}; \tag{5.12}$$

$\lambda > 1$ and there is no loss of generality in assuming $\varepsilon < 1$, so $0 < \delta < 1$. Given n, choose an integer d so that

$$2n\delta > d \geqq n\delta; \tag{5.13}$$

since δ is now fixed, there is such a d for all sufficiently large n, and moreover $d \geqq d_0$ for sufficiently large n, so that (5.11) is available. By (5.13) and (5.9),

$$P_n\{x : w_\delta(x) \geqq \varepsilon\} \leqq \mathsf{P}\{\max|S_j - S_i| \geqq \varepsilon\sigma\sqrt{n}\}, \tag{5.14}$$

where this maximum extends over i and j satisfying $0 \leqq i \leqq j \leqq n$, $|j - i| \leqq d$. If

$$M_{i,d} = \max_{i \leqq j \leqq i+d} |S_j - S_i|,$$

then the maximum in (5.14) is at most

$$3 \max_{k \leqq n/d} M_{kd,d},$$

and it follows by stationarity that

$$\begin{aligned} P_n\{x : w_\delta(x) \geqq \varepsilon\} &\leqq \mathsf{P}\left\{\max_{k \leqq n/d} M_{kd,d} \geqq \frac{1}{3}\varepsilon\sigma\sqrt{n}\right\} \\ &\leqq \frac{n}{d}\mathsf{P}\left\{\max_{i \leqq d}|S_i| \geqq \frac{1}{3}\varepsilon\sigma\sqrt{n}\right\}. \end{aligned} \tag{5.15}$$

By (5.13), $n/d < \delta^{-1}$, and by (5.13) and (5.12),

$$\frac{1}{3}\varepsilon\sigma\sqrt{n} \geqq \frac{1}{3}\varepsilon\sigma\sqrt{\frac{d}{2\delta}} \geqq \frac{1}{6}\frac{\varepsilon}{\sqrt{\delta}}\sigma\sqrt{d} = \lambda\sigma\sqrt{d},$$

so (5.11) and (5.15) imply

$$P_n\{x : w_\delta(x) \geqq \varepsilon\} \leqq \delta^{-1}\frac{\varepsilon^2\eta}{6^2\lambda^2} = \eta,$$

which completes the proof.

6. A maximal inequality. In order to apply Theorem 5.4, we need an effective way to bound the probability in (5.10). Let $\xi_1, \cdots, \xi_n$ be random variables (stationary or not), let $S_k = \xi_1 + \cdots + \xi_k$ ($S_0 = 0$), and put

$$M_n = \max_{k \leqq n} |S_k|. \tag{6.1}$$

What we need is an upper bound for $\mathsf{P}\{M_n \geqq \lambda\}$, and we shall derive one by an indirect approach.

Let

$$m_{ijk} = \min\{|S_j - S_i|, |S_k - S_j|\}, \tag{6.2}$$

and put

$$L_n = \max\{m_{ijk} : 0 \leqq i \leqq j \leqq k \leqq n\}. \tag{6.3}$$

Since $|S_k| \leqq |S_n - S_k| + |S_n|$ and $|S_k| \leqq |S_k| + |S_n|$, we have $|S_k| \leqq m_{0kn} + |S_n|$, and therefore

$$M_n \leqq L_n + |S_n|. \tag{6.4}$$

If $|S_n| = 0$, then obviously

$$|S_n| \leqq 2L_n + \max_{k \leqq n}|\xi_k|. \tag{6.5}$$

And this also holds if $|S_n| > 0$, because then there exists a k ($1 \leqq k \leqq n$) for which $|S_k| \geqq |S_n - S_k|$ and hence there exists a smallest such, so that $|S_{k-1}| < |S_n - S_{k-1}|$; then $|S_n - S_k| = m_{0kn} \leqq L_n$ and $|S_{k-1}| = m_{0,k-1,n} \leqq L_n$, and hence $|S_n| \leqq |S_{k-1}| + |\xi_k| + |S_n - S_k| \leqq 2L_n + |\xi_k|$, so (6.5) follows. Finally (6.4) and (6.5) combined give

$$M_n \leqq 3L_n + \max_{k \leqq n}|\xi_k|. \tag{6.6}$$

There are various ways of bounding the tails of the distributions of $|S_n|$ and $\max_{k \leqq n}|\xi_k|$, so (6.4) and (6.6) can be used to bound the tail of the distributions of M_n if we can bound $\mathsf{P}\{L_n \geqq \lambda\}$. We can derive such a bound by assuming bounds on the tails of the distributions of the quantities (6.2) of which L_n is the maximum.

THEOREM 6.1. *Suppose $u_1, \cdots, u_n$ are nonnegative numbers such that*

$$\mathsf{P}\{m_{ijk} \geqq \lambda\} \leqq \frac{1}{\lambda^4}\left(\sum_{i<l\leqq k} u_l\right)^2, \qquad 0 \leqq i \leqq j \leqq k \leqq n, \tag{6.7}$$

for $\lambda > 0$. Then

$$\mathsf{P}\{L_n \geqq \lambda\} \leqq \frac{K}{\lambda^4}\left(\sum_{0<l\leqq n} u_l\right)^2 \tag{6.8}$$

for $\lambda > 0$, where K is a universal constant.

The exponent 4 in (6.7) can be replaced by any $\alpha \geqq 0$ and the exponent 2 can be replaced by any $\beta > 1$, provided the corresponding replacements are made in (6.8); K then depends on α and β but on nothing else. The proof in this more general case differs from the proof below only notationally.

As a first illustration of Theorem 6.1, consider independent ξ_i with $\mathsf{E}\{\xi_i\} = 0$ and $\mathsf{E}\{\xi_i^2\} = \sigma^2$. Since the two quantities in the minimum in (6.2) are independent, it follows by Chebyshev's inequality that

$$\begin{aligned}\mathsf{P}\{m_{ijk} \geqq \lambda\} &= \mathsf{P}\{|S_j - S_i| \geqq \lambda\}\mathsf{P}\{|S_k - S_j| \geqq \lambda\}\\ &\leqq \frac{(j-i)\sigma^2}{\lambda^2}\frac{(k-j)\sigma^2}{\lambda^2} \leqq \frac{(k-i)^2\sigma^4}{\lambda^4},\end{aligned}$$

the last inequality being a case of $xy \leqq (x+y)^2$. Thus (6.7) holds with $u_l \equiv \sigma^2$.

The proof of Theorem 6.1 becomes simpler if we first generalize the result. Let T be a Borel subset of $[0, 1]$ and suppose $\eta = \{\eta_t : t \in T\}$ is a stochastic process with time running through T. We shall suppose that the paths of the process are right-continuous in the sense that if points s in T converge from the right to a point t in T, then $\eta_s \to \eta_t$ at all sample points (if T is finite, this imposes no restriction). Let

$$m_{rst} = \min\{|\eta_s - \eta_r|, |\eta_t - \eta_s|\}, \tag{6.9}$$

and define

$$L(\eta) = \sup\{m_{rst} : r \leqq s \leqq t, r, s, t \in T\}.$$

THEOREM 6.2. *Suppose μ is a finite measure on T such that*

$$\mathsf{P}\{m_{rst} \geqq \lambda\} \leqq \frac{1}{\lambda^4}\mu^2\{T \cap (r, t]\}, \qquad r \leqq s \leqq t, \tag{6.10}$$

for $\lambda > 0$ and r, s, t in T. Then

$$\mathsf{P}\{L(\eta) \geqq \lambda\} \leqq \frac{K}{\lambda^4}\mu^2(T) \tag{6.11}$$

for $\lambda > 0$, where K is a universal constant.

To deduce Theorem 6.1 from Theorem 6.2, we need only take $T = \{i/n : 0 \leqq i \leqq n\}$ and $\eta(i/n) = S_i, 0 \leqq i \leqq n$, and let μ have mass u_i at i/n, $1 \leqq i \leqq n$. We prove Theorem 6.2 by considering a succession of cases.

Case 1. Suppose first that $T = [0, 1]$ and that μ is Lebesgue measure. For positive $\theta_1, \theta_2, \cdots$, consider the event (we write $m(r, s, t)$ for m_{rst})

$$(6.12) \qquad \bigcap_{k=1}^{\infty} \bigcap_{i=1}^{2^k-1} \left\{ m\left(\frac{i-1}{2^k}, \frac{i}{2^k}, \frac{i+1}{2^k}\right) < \lambda\theta_k \right\}.$$

By (6.10), the complimentarity event has probability at most

$$(6.13) \qquad \sum_{k=1}^{\infty} \sum_{i=1}^{2^k-1} \frac{1}{(\lambda\theta_k)^4} \mu^2\left(\frac{i-1}{2^k}, \frac{i+1}{2^k}\right) \leqq \frac{4}{\lambda^4} \sum_{k=1}^{\infty} \frac{1}{\theta_k^4 2^k}.$$

We next show that, on the event (6.12), the inequality

$$(6.14) \qquad m\left(\frac{a}{2^k}, \frac{b}{2^k}, \frac{c}{2^k}\right) < 2 \sum_{j=1}^{k} \lambda\theta_j$$

holds for all integers k, a, b, c with $0 \leqq a \leqq b \leqq c \leqq 2^k$. If $k = 1$, then $a = 0$, $b = 1, c = 2$ is the only possibility, and (6.14) holds because the sample point is in (6.12). Suppose as induction hypothesis that (6.14) holds when k is replaced by $k - 1$. If, for example, $a = 2a'$ and $b = 2b'$ are even and $c = 2c' + 1$ is odd, then

$$m\left(\frac{a}{2^k}, \frac{b}{2^k}, \frac{c}{2^k}\right) \leqq M + m\left(\frac{2c'}{2^k}, \frac{2c'+1}{2^k}, \frac{2c'+2}{2^k}\right),$$

where M is the maximum of the two quantities

$$m\left(\frac{a'}{2^{k-1}}, \frac{b'}{2^{k-1}}, \frac{c'}{2^{k-1}}\right), \quad m\left(\frac{a'}{2^{k-1}}, \frac{b'}{2^{k-1}}, \frac{c'+1}{2^{k-1}}\right),$$

so that (6.14) in this case follows by the induction hypothesis and the assumption that the sample point is in (6.12). Clearly the other possibilities for the parities of a, b, and c can be handled the same way.

Thus (6.14) holds in general; since the process $\{\eta_t\}$ has right-continuous paths,

$$L(\eta) \leqq 2 \sum_{j=1}^{\infty} \lambda\theta_j$$

follows for sample points in (6.12). If we choose the θ_j so the right side of this inequality is λ at most, then $P\{L(\eta) \geqq \lambda\}$ has probability at most (6.13). If $\theta_j = C/2^{j/8}$ with C chosen to make the θ_j add to 1/2, we get

$$\mathsf{P}\{L(\eta) \geqq \lambda\} \leqq \frac{4}{\lambda^4} \sum_{k=1}^{\infty} \frac{C^{-4}}{2^{k/2}} = \frac{K}{\lambda^4},$$

which disposes of Case 1.

Case 2. Suppose $T = [0, 1]$ and μ is atomless—that is, $F(t) = \mu(0, t]$ is continuous. If F is strictly increasing and $F(1) = w$, define

$$\zeta(t) = w^{-1/2}\eta(F^{-1}(tw)), \qquad 0 \leqq t \leqq 1.$$

Then the process $\{\zeta(t)\}$ comes under Case 1 and the theorem holds for $\{\eta(t)\}$ because $L(\eta) = w^{1/2}L(\zeta)$. If $F(t)$ is continuous but not strictly increasing, consider first the measure having distribution function $F(t) + \varepsilon t$, and then let ε go to 0.

Case 3. Suppose that T is finite (which actually suffices for Theorem 6.1). There is no loss of generality in assuming T contains 0 and 1, so suppose T consists of the points

$$0 = t_0 < t_1 < \cdots < t_w = 1.$$

Let $\eta' = \{\eta'(t) : 0 \leqq t \leqq 1\}$ be a process defined by

$$\eta'(t) = \begin{cases} \eta(t_i) & \text{if} \quad t_i \leqq t < t_{i+1}, \\ \eta(1) & \text{if} \quad t = 1. \end{cases} \qquad 0 \leqq i < w,$$

If m'_{rst} denotes (6.9) for the process η', then m'_{rst} vanishes unless r, s, and t lie in different subintervals $[t_i, t_{i+1})$. Suppose

$$t_i \leqq r < t_{i+1} \leqq t_j \leqq s < t_{j+1} \leqq t_k \leqq t < t_{k+1}. \tag{6.15}$$

Then

$$m'_{rst} = m(t_i, t_j, t_k)$$

and hence, by the hypothesis of the theorem for the process η,

$$\mathsf{P}\{m'_{rst} \geqq \lambda\} \leqq \frac{1}{\lambda^4}\mu^2\{(t_i, t_k] \cap T\}. \tag{6.16}$$

Now let ν be the measure that corresponds to a uniform distribution of mass $\mu\{t_{l-1}\} + \mu\{t_l\}$ over the interval $[t_{l-1}, t_l]$, $1 \leqq l \leqq w$. Then

$$\mu\{(t_i, t_k] \cap T\} \leqq \nu[t_{i+1}, t_k] \leqq \nu(r, t],$$

so (6.16) implies

$$\mathsf{P}\{m'_{rst} \geqq \lambda\} \leqq \frac{1}{\lambda^4}\nu^2(r, t]. \tag{6.17}$$

Although (6.15) requires $t < 1$, (6.17) follows for $t = 1$ by a small modification of this argument.

Thus (6.17) holds for $0 \leqq r \leqq s \leqq t \leqq 1$, and Case 2 applies to the process η':

$$\mathsf{P}\{L(\eta') \geqq \lambda\} \leqq \frac{K}{\lambda^4}\nu^2(0, 1] \leqq \frac{K}{\lambda^4}(2\mu(T))^2.$$

Since $L(\eta') = L(\eta)$, if we replace the K that works in Cases 1 and 2 by $4K$, then the new K works in Cases 1, 2, and 3.

Case 4. For the general T and μ, consider finite sets

$$T_n : 0 \leqq t_{n0} < t_{n1} < \cdots < t_{nw_n} \leqq 1$$

that become dense in T, and let μ_n have mass $\mu\{(t_{n,i-1}, t_{n,i}] \cap T\}$ at the point t_{ni}. If $\eta^{(n)}$ is the process η with the time set cut back to T_n, then $L(\eta^{(n)}) \to L(\eta)$ by right-continuity of the paths. Since each $\eta^{(n)}$ is easily seen to come under Case 3, the general result follows by a passage to the limit.

7. Tightness. We can now combine the results of the preceding two sections to establish the tightness of the distributions of the random functions

$$X_n(t, \omega) = \frac{1}{\sigma\sqrt{n}} S_{[nt]}(\omega) + \frac{nt - [nt]}{\sigma\sqrt{n}} \xi_{[nt]+1}(\omega) \tag{7.1}$$

considered earlier.

Suppose that $\{\xi_1, \xi_2, \cdots\}$ is a stationary process with

$$\mathsf{E}\{\xi_k\} = 0, \quad \mathsf{E}\{\xi_k^2\} = \sigma_0^2, \tag{7.2}$$

and assume

$$J = 2 \sum_{k=1}^{\infty} |\mathsf{E}\{\xi_1 \xi_{1+k}\}| < \infty. \tag{7.3}$$

Put $S_k = \xi_1 + \cdots + \xi_k$.

If $r_k = \mathsf{E}\{\xi_1 \xi_{1+k}\}$, then

$$\mathsf{E}\{S_n^2\} = n\sigma_0^2 + 2 \sum_{k=1}^{n-1} (n - k) r_k,$$

and therefore

$$\frac{1}{n}\mathsf{E}\{S_n^2\} \leqq \sigma_0^2 + J. \tag{7.4}$$

If $R_k = \sum_{i=1}^k r_i$, then

$$\frac{1}{n}\mathsf{E}\{S_n^2\} = \sigma_0^2 + \frac{2}{n}\sum_{k=1}^{n-1} R_k,$$

and so

$$\frac{1}{n}\mathsf{E}\{S_n^2\} \to \sigma^2, \tag{7.5}$$

where

$$\sigma^2 = \sigma_0^2 + 2 \sum_{k=1}^{\infty} \mathsf{E}\{\xi_1 \xi_{1+k}\}. \tag{7.6}$$

We shall assume that (7.6) is positive, and we shall define the random function X_n by (7.1) with this value of σ.

We shall make one further assumption. Let $\mathscr{B}_1^k$ be the σ-field generated by the random variables $\xi_1, \cdots, \xi_k$, and let $\mathscr{B}_{k+1}^\infty$ be the σ-field generated by the random variables $\xi_{k+1}, \xi_{k+2}, \cdots$. We shall assume the existance of a finite B such that

$$\mathsf{P}(A_1 \cap A_2) \leqq B\mathsf{P}(A_1)\mathsf{P}(A_2), \qquad A_1 \in \mathscr{B}_1^k, \quad A_2 \in \mathscr{B}_{k+1}^\infty, \tag{7.7}$$

for all k, or, what is the same,

$$\begin{cases} \mathsf{P}(A_2|A_1) \leqq B\mathsf{P}(A_2), \\ \mathsf{P}(A_1|A_2) \leqq B\mathsf{P}(A_1), \end{cases} \qquad A_1 \in \mathscr{B}_1^k, \quad A_2 \in \mathscr{B}_{k+1}^{\infty}. \tag{7.8}$$

This is one way of requiring that the past and future do not unduly influence each other.

Of course (7.7) holds with $B = 1$ in the case of independent ξ_n. If $\{\eta_n\}$ is a stationary Markov chain whose stationary and transition probabilities satisfy

$$B = \sup_{ij} \frac{p_{ij}}{p_j} < \infty, \tag{7.9}$$

and if $\xi_n = \varphi(\eta_n)$, where φ is a numerical function on the state space, then (7.7) holds with (7.9) for B.

THEOREM 7.1. *Suppose that* $\{\xi_n\}$ *is stationary and satisfies* (7.2), *that the* J *in* (7.3) *is finite and the* σ *in* (7.6) *is positive, and that* (7.7) *holds; then the distributions of the random functions* (7.1) *are tight.*

Proof. By Theorem 5.4, it suffices to show that, for each ε there exists a $\lambda > 1$ such that

$$\mathsf{P}\{M_n \geqq \lambda\sigma\sqrt{n}\} \leqq \frac{\varepsilon}{\lambda^2} \tag{7.10}$$

for all sufficiently large n, where M_n is given by (6.1).

With the definition (6.2), we obtain by successively applying (7.7), stationarity, Chebyshev's inequality, and (7.4),

$$\begin{aligned} \mathsf{P}\{m_{ijk} \geqq \lambda\} &\leqq B\mathsf{P}\{|S_{j-i}| \geqq \lambda\}\mathsf{P}\{|S_{k-j}| \geqq \lambda\} \\ &\leqq B\frac{(j-i)(\sigma_0^2 + J)}{\lambda^2}\frac{(k-j)(\sigma_0^2 + J)}{\lambda^2} \leqq K_0\frac{(k-i)^2}{\lambda^4}, \end{aligned}$$

with $K_0 = B(\sigma_0^2 + J)^2$. By Theorem 6.1 with $u_l \equiv K_0^{1/2}$,

$$\mathsf{P}\{L_n \geqq \lambda\} \leqq \frac{K_0 K}{\lambda^4} n^2. \tag{7.11}$$

By stationarity,

$$\mathsf{P}\{\max_{k \leqq n} |\xi_k| \geqq \lambda\} \leqq n\mathsf{P}\{|\xi_1| \geqq \lambda\} \leqq \frac{n}{\lambda^2}\int_{\{|\xi_1| \geqq \lambda\}} \xi_1^2 \, d\mathsf{P}.$$

Those two inequalities and (6.6) give

$$\mathsf{P}\{M_n \geqq 4\lambda\} \leqq \frac{K_0 K}{\lambda^4} n^2 + \frac{n}{\lambda^2}\int_{\{|\xi_1| \geqq \lambda\}} \xi_1^2 \, d\mathsf{P}.$$

Replacing 4λ here by $\lambda\sigma\sqrt{n}$ leads to

$$\mathsf{P}\{M_n \geqq \lambda\sigma\sqrt{n}\} \leqq \frac{4^4 K_0 K}{\sigma^4 \lambda^4} + \frac{4^2}{\sigma^2\lambda^2}\int_{\{|\xi_1| \geqq \lambda\sigma/4\}} \xi_1^2 \, d\mathsf{P}.$$

For sufficiently large λ, the right side here is less than ε/λ^2, so that (7.10) holds for *all n*, which proves the theorem.

Suppose for the rest of this section, that the ξ_n are independent and identically distributed random variables satisfying (7.2) with $\sigma_0^2 > 0$. Then the hypotheses of Theorem 7.1 are satisfied (the J in (7.3) vanishes) and $\sigma = \sigma_0$. Thus the distributions of the X_n are tight. Furthermore, the central limit theorem holds in this case:

$$\frac{1}{\sigma\sqrt{n}} S_n \Rightarrow N, \tag{7.12}$$

where N is a normally distributed random variable with mean 0 and variance 1.

Fix $t, 0 < t \leqq 1$. If k_n is determined by $k_n n^{-1} \leqq t < (k_n + 1)n^{-1}$, then by (7.1) and stationarity,

$$\mathsf{P}\{|X_n(t) - \frac{1}{\sigma\sqrt{n}} S_{k_n}| \geqq \varepsilon\} = \mathsf{P}\{|\xi_1| \geqq \varepsilon\sigma\sqrt{n}\} \to 0. \tag{7.13}$$

Since $k_n/n \to t$, (7.12) implies $S_{k_n}/\sigma\sqrt{n} \Rightarrow \sqrt{t}N$, and it follows (Theorem 3.1) that

$$X_n(t) \Rightarrow \sqrt{t}N.$$

If $s < t$, determine j_n by $j_n n^{-1} \leqq s < (j_n + 1)n^{-1}$; if N_1 and N_2 are independent and each is normally distributed with mean 0 and variance 1, then

$$\begin{aligned} \mathsf{P}\left\{\frac{S_{j_n}}{\sigma\sqrt{n}} \leqq x, \frac{S_{k_n} - S_{j_n}}{\sigma\sqrt{n}} \leqq y\right\} &= \mathsf{P}\left\{\frac{S_{j_n}}{\sigma\sqrt{n}} \leqq x\right\} \mathsf{P}\left\{\frac{S_{k_n} - S_{j_n}}{\sigma\sqrt{n}} \leqq y\right\} \\ &\to \mathsf{P}\{\sqrt{s}N_1 \leqq x\}\mathsf{P}\{\sqrt{t - s}N_2 \leqq y\} \\ &= \mathsf{P}\{\sqrt{s}N_1 \leqq x, \sqrt{t - s}N_2 \leqq y\}. \end{aligned}$$

Thus

$$\left(\frac{S_{j_n}}{\sigma\sqrt{n}}, \frac{S_{k_n} - S_{j_n}}{\sigma\sqrt{n}}\right) \Rightarrow (\sqrt{s}N_1, \sqrt{t - s}N_2),$$

and it follows by Corollary 3 to Theorem 3.3 (consider the map $(x, y) \to (x, x + y)$ of R^2 into itself) that

$$\left(\frac{S_{j_n}}{\sigma\sqrt{n}}, \frac{S_{k_n}}{\sigma\sqrt{n}}\right) \Rightarrow (\sqrt{s}N_1, \sqrt{s}N_1 + \sqrt{t - s}N_2).$$

And now from (7.13) and the analogous relation for the point s, it follows (Theorem 3.1 again) that

$$(X_n(s), X_n(t)) \Rightarrow (\sqrt{s}N_1, \sqrt{s}N_1 + \sqrt{t - s}N_2).$$

To put it another way,

$$(X_n(s), X_n(t) - X_n(s)) \Rightarrow (\sqrt{s}N_1, \sqrt{t - s}N_2).$$

By a simple generalization of this argument, it follows that, if $t_1 < t_2 < \cdots < t_k$, then

$$(7.14) \qquad (X_n(t_1), X_n(t_2) - X_n(t_1), \cdots, X_n(t_k) - X_n(t_{k-1}))$$
$$\Rightarrow (\sqrt{t_1}N_1, \sqrt{t_2 - t_1}N_2, \cdots, \sqrt{t_k - t_{k-1}}N_k),$$

where $N_1, \cdots, N_k$ are independent, normally distributed random variables with mean 0 and variance 1. If $Y_i = \sqrt{t_i - t_{i-1}}N_i$ for $i \geqq 2$ and $Y_1 = \sqrt{t_1}N_1$, then (7.14) is the same as

$$(X_n(t_1), \cdots, X_n(t_k)) \Rightarrow (Y_1, Y_1 + Y_2, \cdots, Y_1 + \cdots + Y_k).$$

The distributions P_n of the X_n are tight and hence some subsequence of them converges (Prokhorov's theorem) to some probability measure P on C. Let W be a random function with distribution P (Theorem 3.2). Now $(W(t_1), W(t_2) - W(t_1), \cdots, W(t_k) - W(t_{k-1}))$ has the distribution of the limiting vector in (7.14). We have thus constructed a random function, a random element of C, whose increments $W(t) - W(s)$ are normally distributed with mean 0 and variance $t - s$—and the increments over nonoverlapping intervals are independent. Since this is a specification of the finite-dimensional distributions of the random function W, its distribution is unique. We call W Brownian motion.

Having constructed P and W from a subsequence, we return to a consideration of the whole sequence $\{P_n\}$. By (7.14), the finite-dimensional distributions of the P_n convergence to those of P; since $\{P_n\}$ is tight, Theorem 5.1 implies $P_n \Rightarrow P$. In other words,

$$(7.15) \qquad X_n \Rightarrow W$$

if X_n is defined by (7.1) and the ξ_n are independent and identically distributed with mean 0 and positive variance σ^2.

8. Limit theorems. We have shown that (7.15) holds in the independent case, and in this section we shall show that it holds more generally under the hypotheses of Theorem 7.1 together with a mixing condition. We start with a characterization of Brownian motion W.

THEOREM 8.1. *Suppose Y is a random element of C having independent increments and satisfying* $\mathsf{E}\{Y(t)\} = 0$ *and* $\mathsf{E}\{Y^2(t)\} = t$. *Then Y is distributed as W.*

Proof. We shall show that $Y(t + \delta) - Y(t)$ is normally distributed with mean 0 and variance δ, which will complete the identification of the finite-dimensional distributions of Y as those of W—which is enough.

Now for each n,

$$(8.1) \qquad Y(t + \delta) - Y(t) = \sum_{k=1}^{n} \left[Y\left(t + \frac{k}{n}\delta\right) - Y\left(t + \frac{k-1}{n}\delta\right) \right].$$

The summands here are independent and each one has mean 0 and variance δ/n. We shall show that Lyapounov's condition

$$\sum_{k=1}^{n} \mathsf{E}\left\{\left|Y\left(t+\frac{k}{n}\delta\right)-Y\left(t+\frac{k-1}{n}\delta\right)\right|^3\right\}\to 0 \tag{8.2}$$

holds (note that the total variance is δ), so the sum in (8.1) is asymptotically normal and hence $Y(t+\delta)-Y(t)$ *is* normal. Clearly (8.2) will follow if we prove that

$$\mathsf{E}\{|Y(s+h)-Y(s)|^3\} \leqq K_1 h^{3/2}, \tag{8.3}$$

where K_1 is independent of s and h.

To prove (8.3), fix s and h for the moment and define

$$\xi_i = Y\left(s+\frac{i}{n}h\right) - Y\left(s+\frac{i-1}{n}h\right), \qquad i=1,\cdots,n.$$

If $S_k = \xi_1 + \cdots + \xi_k$, then

$$Y(s+h)-Y(s) = S_n.$$

Because of the independence of the increments of Y, it follows by Chebyshev's inequality and the moment conditions that, in the notation (6.2),

$$\begin{aligned}\mathsf{P}\{m_{ijk}\geqq\lambda\} &= \mathsf{P}\left\{\left|Y\left(s+\frac{j}{n}h\right)-Y\left(s+\frac{i}{n}h\right)\right|\geqq\lambda\right\}\mathsf{P}\left\{\left|Y\left(s+\frac{k}{n}h\right)-Y\left(s+\frac{j}{n}h\right)\right|\geqq\lambda\right\}\\ &\leqq \frac{1}{\lambda^2}\left(\frac{j}{n}h-\frac{i}{n}h\right)\cdot\frac{1}{\lambda^2}\left(\frac{k}{n}h-\frac{j}{n}h\right)\leqq\frac{1}{\lambda^4}\frac{h^2}{n^2}(k-i)^2.\end{aligned}$$

Therefore, by Theorem 6.1,

$$\mathsf{P}\{L_n\geqq\lambda\}\leqq\frac{K}{\lambda^4}h^2,$$

with L_n defined by (6.3). By (6.6) and the fact that $|Y(s+h)-Y(s)| = |S_n| \leqq M_n$, we have

$$\begin{aligned}&\mathsf{P}\{|Y(s+h)-Y(s)|\geqq 4\lambda\}\\ &\qquad\leqq\frac{K}{\lambda^4}h^2+\mathsf{P}\left\{\max_{i\leqq n}\left|Y\left(s+\frac{i}{n}h\right)-Y\left(s+\frac{i-1}{n}h\right)\right|\geqq\lambda\right\}.\end{aligned}$$

Now $Y(t)$ is continuous in t, and hence the last term here goes to 0 as $n\to\infty$. Therefore (replace 4λ by λ),

$$\mathsf{P}\{|Y(s+h)-Y(s)|\geqq\lambda\}\leqq\frac{4^4K}{\lambda^4}h^2.$$

If $F(x)$ is the distribution function of $|Y(s+h)-Y(s)|^3$, then

$$1-F(x)\leqq\frac{4^4Kh^2}{x^{4/3}}.$$

Integration by parts gives

$$\int_a^b x\,dF(x) = a(1 - F(a)) + \int_a^b (1 - F(x))\,dx - b(1 - F(b)),$$

so that

$$\int_a^\infty x\,dF(x) = a(1 - F(a)) + \int_a^\infty (1 - F(x))\,dx. \tag{8.4}$$

Therefore,

$$\mathsf{E}\{|Y(s+h) - Y(s)|^3\} \leqq a + \int_a^\infty \frac{4^4Kh^2}{x^{4/3}}\,dx;$$

taking $a = h^{3/2}$ gives (8.3) with $K_1 = 1 + 3 \cdot 4^4 K$.

We next cast Theorem 8.1 in asymptotic form. We say a sequence $\{Y_n\}$ of random functions has asymptotically independent increments if the difference

$$\mathsf{P}\left[\bigcap_{i=1}^{k} \{Y_n(t_i) - Y_n(s_i) \leqq x_i\}\right] - \prod_{i=1}^{k} \mathsf{P}\{Y_n(t_i) - Y_n(s_i) \leqq x_i\}$$

goes to 0 for disjoint intervals $[s_i, t_i]$.

THEOREM 8.2. *Suppose that* $\{Y_n\}$ *has asymptotically independent increments, that* $\{Y_n^2(t): n = 1, 2, \cdots\}$ *is uniformly integrable for each* t, *and that* $\mathsf{E}\{Y_n(t)\} \to 0$ *and* $\mathsf{E}\{Y_n^2(t)\} \to t$ *as* $n \to \infty$ *for each* t; *suppose finally that the distributions of the* Y_n *are tight. Then* $Y_n \Rightarrow W$.

Proof. By Theorem 2.2, it is enough to show that if some subsequence of $\{Y_n\}$ converges in distribution to some Y, then Y is distributed as W. But the finite-dimensional distributions of such a Y are the limits of those of the subsequence, so Y has independent increments. Now $\{Y_n^2(t): n = 1, 2, \cdots\}$ is assumed uniformly integrable, and it follows that $\{Y_n(t): n = 1, 2, \cdots\}$ is also uniformly integrable; by Corollary 5 to Theorem 3.3, we can integrate to the limit along the subsequence: $\mathsf{E}\{Y(t)\} = \lim \mathsf{E}\{Y_n(t)\} = 0$ and $\mathsf{E}\{Y^2(t)\} = \lim \mathsf{E}\{Y_n^2(t)\} = t$. That Y is distributed as W now follows by the preceding theorem.

We turn finally to the problem of proving convergence in distribution for the random functions defined by

$$X_n(t) = \frac{1}{\sigma\sqrt{n}} S_{[nt]} + \frac{nt - [nt]}{\sigma\sqrt{n}} \xi_{[nt]+1}. \tag{8.5}$$

Here $S_k = \xi_1 + \cdots + \xi_k$, and the ξ_n satisfy the conditions of Theorem 7.1. That is, $\{\xi_n\}$ is stationary,

$$\mathsf{E}\{\xi_n\} = 0, \quad \mathsf{E}\{\xi_n^2\} = \sigma_0^2; \tag{8.6}$$

we have

$$\sum_{n=1}^{\infty} |E\{\xi_1 \xi_{1+n}\}| < \infty \tag{8.7}$$

and

(8.8) $$\sigma^2 = \sigma_0^2 + 2 \sum_{n=1}^{\infty} \mathsf{E}\{\xi_1 \xi_{1+n}\} > 0;$$

if $\mathscr{B}_1^n$ and $\mathscr{B}_n^\infty$ are respectively the σ-fields generated by the families $\{\xi_k : k \leqq n\}$ and $\{\xi_k : k \geqq n\}$, then

(8.9) $$\mathsf{P}(A_1 \cap A_2) \leqq B\mathsf{P}(A_1)\mathsf{P}(A_2), \quad A_1 \in \mathscr{B}_1^n, \quad A_2 \in \mathscr{B}_{n+1}^\infty,$$

where B is a finite constant independent of n. We shall also assume that

(8.10) $$|\mathsf{P}(A_1 \cap A_2) - \mathsf{P}(A_1)\mathsf{P}(A_2)| \leqq \alpha_n, \; A_1 \in \mathscr{B}_1^k, \quad A_2 \in \mathscr{B}_{k+n}^\infty,$$

for all k and n, where

(8.11) $$\lim_{n\to\infty} \alpha_n = 0.$$

This is a mixing condition which ensures that the distant future is virtually independent of the past.

THEOREM 8.3. *Suppose that* $\{\xi_1, \xi_2, \cdots\}$ *is stationary and satisfies* (8.6) *through* (8.9); *suppose also that* (8.10) *holds for some sequence satisfying* (8.11). *Then* X_n, *defined by* (8.5), *converges in distribution to* W.

Proof. We shall show that the X_n satisfy the hypotheses imposed on the Y_n in Theorem 8.2. We showed in Theorem 7.1 that the distributions of the X_n are tight; certainly $\mathsf{E}\{X_n(t)\} = 0$, and $\mathsf{E}\{X_n(t)\} \to t$ follows from (7.5).

From (8.10) it follows by induction that, if A_i is in the σ-field generated by $\{\xi_{a_i}, \cdots, \xi_{b_i}\}$, $i = 1, \cdots, k$, and if $a_i - b_{i-1} > l$, then

(8.12) $$|\mathsf{P}(A_1 \cap \cdots \cap A_k) - \mathsf{P}(A_1) \cdots \mathsf{P}(A_k)| \leqq k\alpha_l.$$

The increment $X_n(t_i) - X_n(s_i)$ is measurable with respect to the σ-field generated by $\xi_{[ns_i]}, \cdots, \xi_{[nt_i]+1}$ and $[ns_i] - [nt_{i-1}] - 1 \to \infty$ if $t_{i-1} < s_i$, so the asymptotic independence of the increments is a consequence of (8.11) and (8.12).

It remains only to prove the uniform integrability of $\{X_n^2(t) : n = 1, 2, \cdots\}$. By the definition (8.5) (recall $(x + y)^2 \leqq 2x^2 + 2y^2$),

$$X_n^2(t) \leqq \frac{2}{\sigma^2 n} S_{[nt]}^2 + \frac{2}{\sigma^2 n} \xi_{[nt]+1}^2.$$

The second term on the right is certainly integrable uniformly in n, and so it will suffice to prove $\{S_n^2/n : n = 1, 2, \cdots\}$ uniformly integrable. In the notation of § 6,

$$|S_n| \leqq M_n \leqq 3L_n + \max_{k \leqq n} |\xi_k|,$$

and hence

$$S_n^2 \leqq 18L_n^2 + 2 \max_{k \leqq n} \xi_k^2,$$

and so it suffices to prove separately the uniform integrability of L_n^2/n and of

$$m_n = \max_{k \leqq n} \xi_k^2/n. \tag{8.13}$$

In the course of proving Theorem 7.1 we established (7.11), which gives

$$\mathsf{P}\{L_n^2/n \geqq x\} \leqq \frac{K_0 K}{x^2}.$$

By (8.4) with the distribution function of L_n^2/n for F,

$$\int_{\{L_n^2/n \geqq a\}} \frac{1}{n} L_n^2 \, d\mathsf{P} \leqq a\frac{K_0 K}{a^2} + \int_a^\infty \frac{K_0 K}{x^2} dx = \frac{2K_0 K}{a},$$

which proves the uniform integrability of L_n^2/n.

To prove the uniform integrability of (8.13) we need a lemma.

LEMMA. *If ξ and η are nonnegative random variables and ξ is measurable $\mathscr{B}_1^k$ and η is measurable $\mathscr{B}_{k+1}^\infty$, then*

$$\mathsf{E}\{\xi\eta\} \leqq B\mathsf{E}\{\xi\}\mathsf{E}\{\eta\}. \tag{8.14}$$

Indeed, by the basic assumption (8.9), if $\xi = \sum_i \alpha_i I_{U_i}$ and $\eta = \sum_j \beta_j I_{V_j}$ are simple functions, then

$$\begin{aligned} \mathsf{E}\{\xi\eta\} &= \sum_{ij} \alpha_i\beta_j \mathsf{P}(U_i \cap V_j) \\ &\leqq B\sum_{ij} \alpha_i\beta_j \mathsf{P}(U_i)\mathsf{P}(V_j) = B\mathsf{E}\{\xi\}\mathsf{E}\{\eta\}, \end{aligned}$$

and (8.14) for the general case follows by a passage to the limit.

By an application of this lemma, (8.13) satisfies

$$\begin{aligned} \int_{\{m_n \geqq a\}} m_n \, d\mathsf{P} &\leqq \sum_{i,k=1}^n \int_{\{\xi_i^2 \geqq an\}} \frac{1}{n}\xi_k^2 \, d\mathsf{P} \\ &\leqq \sum_{i=1}^n \int_{\{\xi_i^2 \geqq an\}} \frac{1}{n}\xi_i^2 \, d\mathsf{P} + 2\sum_{1 \leqq i < k \leqq n} B\mathsf{P}\{\xi_i^2 \geqq an\}\mathsf{E}\left\{\frac{1}{n}\xi_k^2\right\}. \end{aligned}$$

By stationarity,

$$\int_{\{m_n \geqq a\}} m_n \, d\mathsf{P} \leqq \int_{\{\xi_1^2 \geqq a\}} \xi_1^2 \, d\mathsf{P} + \frac{B}{a}\mathsf{E}^2\{\xi_1^2\},$$

whence follows the uniform integrability of the m_n. This completes the proof of Theorem 8.3.

For a very simple application, consider a finite Markov chain $\{\eta_n\}$ with positive transition probabilities p_{ij} and stationary probabilities p_i, and suppose $\xi_n = \varphi(\eta_n)$, where φ is a real function on the state space. Then (8.9) holds with $B = \max_{ij} p_{ij}/p_j$. Moreover, the nth-order transition probability $p_{ij}^{(n)}$ converges to p_j at an exponential rate:

$$p_{ij}^{(n)} - p_j = O(\rho^n), \qquad 0 < \rho < 1.$$

Hence (8.10) holds with $\alpha_n = O(\rho^n)$.

If

$$\mathsf{E}\{\xi_n\} = \sum_i p_i\varphi(i) = 0, \tag{8.15}$$

then

$$\mathsf{E}\{\xi_1\xi_{n+1}\} = \sum_{ij} p_i p_{ij}^{(n)}\varphi(i)\varphi(j)$$
$$= \sum_{ij} (p_i p_{ij}^{(n)} - p_i p_j)\varphi(i)\varphi(j)$$

goes to zero exponentially, so that (8.7) holds. Thus Theorem 8.3 applies if (8.15) holds and if

$$\sigma^2 = \sum_i p_i\varphi^2(i) + 2\sum_{n=1}^{\infty}\sum_{ij} p_i p_{ij}^{(n)}\varphi(i)\varphi(j) \tag{8.16}$$

is positive.

Notice that Theorem 8.3 contains the central limit theorem for the process $\{\xi_n\}$ but does not presuppose it. If f maps C continuously into R^1, it follows by Corollary 3 to Theorem 3.3 that $f(X_n) \Rightarrow f(W)$. If $f(x) = x(1)$, then $f(X_n) = S_n/\sigma\sqrt{n}$ and $f(W)$ is normal with mean 0 and variance 1, so that

$$\frac{S_n}{\sigma\sqrt{n}} \Rightarrow W(1)$$

is the central limit theorem. On the other hand, the distribution of $f(W)$ is known for a variety of other functions f on C, and each of these leads to a limit theorem for $\{\xi_n\}$.

BIBLIOGRAPHY

These are papers which have appeared since the publication of my book mentioned in the preface. Papers listed without journal reference I have seen as preprints; copies can perhaps be obtained from the authors.

P. J. BICKEL, *Some contributions to the theory of order statistics*, Proc. Fifth Berkeley Symposium, 1 (1967), pp. 575–591.

———, *A distribution free version of the Smirnov two sample test in the p-variate case*, Ann. Math. Statist., 40 (1969), pp. 1–23.

P. J. BICKEL AND M. J. WICHURA, *Convergence criteria for multiparameter stochastic processes and some applications.*

PATRICK BILLINGSLEY, *Maxima of partial sums*, Probability and Information Theory, Lecture notes in mathematics #89, Springer-Verlag, Berlin-Heidelberg-New York, 1969, pp. 64–76.

A. A. BOROVKOV, *Convergence of weakly dependent processes to the Wiener process*, Theor. Probability Appl., 12 (1967), pp. 159–186.

———, *On convergence to diffusion processes*, Ibid., 12 (1967), pp. 405–431.

———, *On limit laws for service processes in multi-channel systems*, Siberian Math. J., 8 (1967), pp. 746–763.

B. M. BROWN, *Martingale central limit theorems*, Ann. Math. Statist., 42 (1971), pp. 59–66.

MARK BROWN, *Convergence in distribution of stochastic integrals.*

G. CHOQUET, *Sur les ensembles uniformément négligibles*, Sem. Choquet (Initiation à l'analyse), 9[e] année (1969/70), no. 6.

R. R. CORNWALL, *The approximation of perfect competition by a large, but finite, number of traders.*

ROY O. DAVIES, *A non-Prokhorov space.*

ALEJANDRO D. DE ACOSTA, *Existence and convergence of probability measures in Banach spaces*, Trans. Amer. Math. Soc., 152 (1970), pp. 273–298.

R. M. DUDLEY, *Convergence of Baire measures*, Studia Math., 27 (1966), pp. 251–268.

———, *Distances of probability measures and random variables*, Ann. Math. Statist., 39 (1968), pp. 1563–1572.

VÁCLAV FABIAN, *On uniform convergence of measures*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete, 15 (1970), pp. 139–143.

———, *On weak convergence of stochastic processes.*

D. H. FREMLIN, D. J. H. GARLING AND R. G. HAYDON, *On measures on topological spaces.*

DAVID A. FRIEDMAN, *Some invariance principles for functionals of a Markov chain*, Ann. Math. Statist., 38 (1967), pp. 1–7.

A. M. GARSIA, E. RODEMICH AND H. RUMSEY, JR., *A real variable lemma and the continuity of paths of some Gaussian processes*, Indiana Univ. Math. J., 20 (1970/71), pp. 565–578.

RICHARD HOLLY, *The motion of a large particle*, Trans. Amer. Math. Soc., 144 (1969), pp. 523–534.

DONALD L. IGLEHART, *Limit diffusion approximations for the many server queue and the repairman problem*, J. Appl. Probability, 2 (1965), pp. 429–441.

———, *Limit theorems for the multi-urn Ehrenfest model*, Ann. Math. Statist., 39 (1968), pp. 864–876.

———, *Diffusion approximations in collective risk theory*, J. Appl. Probability, 6 (1969), pp. 285–292.

———, *Diffusion approximations in applied probability*, Lectures in Applied Mathematics, vol. 12: Mathematics of the Decision Sciences, Part 2, 1967, pp. 235–254.

———, *Functional limit theorems for the queue GI/G/1 in light traffic*, Adv. Appl. Prob., to appear.

DONALD L. IGLEHART AND D. P. KENNEDY, *Weak convergence of the average of flag processes*, J. Appl. Prob., 7 (1970), pp. 747–753.

DONALD L. IGLEHART AND HOWARD M. TAYLOR, *Weak convergence of a sequence of quickest detection problems*, Ann. Math. Statist., 39 (1968), pp. 2149–2153.

DONALD L. IGLEHART AND WARD WHITT, *Multiple channel queues in heavy traffic, I*, Adv. Appl. Prob., 2 (1970), pp. 150–177.

———, *Multiple channel queues in heavy traffic, II: sequences, networks and batches*, Ibid., 2 (1970), pp. 355–369.

———, *The equivalence of functional central limit theorems for counting processes and associated partial sums*, Ann. Math. Statist., to appear.

D. P. KENNEDY, *A functional central limit theorem for k-dimensional renewal theory*, Ibid., 42 (1971), pp. 376–380.

———, *Weak convergence for the superposition and thinning of point processes*, Ibid., to appear.

J. KISYŃSKI, *On the generation of tight measures*, Studia Math., 30 (1968), pp. 141–151.

J. KUELBS, *The invariance principle for a lattice of random variables*, Ann. Math. Statist., 39 (1968), pp. 382–389.

THOMAS G. KURTZ, *Comparison of semi-Markov and Markov processes*, Ibid., 42 (1971), pp. 991–1002.

LUCIEN LECAM, *On the weak convergence of probability measures*, Ibid., 41 (1970), pp. 621–625.

THOMAS M. LIGGETT, *An invariance principle for conditioned sums of independent random variables*, J. Math. Mech., 18 (1968), pp. 559–570.

———, *Weak convergence of conditioned sums of independent random vectors*, Trans. Amer. Math. Soc., 152 (1970), pp. 195–213.

———, *On convergent diffusions: The densities and the conditioned process*, Indiana U. Math. J., 20 (1970), pp. 265–279.

———, *Convergence of sums of random variables conditioned on a future change of sign*, Ann. Math. Statist., 41 (1970), pp. 1978–1982.

R. M. LOYNES, *An invariance principle for reversed martingales*, Proc. Amer. Math. Soc., 25 (1970), pp. 56–64.

D. W. MÜLLER, *Verteilungs Invarianzprinzipien für das Gesetz der grossen Zahlen*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete, 10 (1968), pp. 173–192.

———, *Nonstandard proofs of invariance principles in probability theory*, Applications of Model Theory to Algebra, Analysis, and Probability, Holt, Reinhart and Winston, New York, 1969, pp. 186–194.

———, *On Glivenko-Cantelli convergence*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete, 16 (1970), pp. 195–210.

MICHAEL D. PERLMAN, *Characterizing measurability, distribution, and weak convergence of random variables in a Banach space by subsets of linear functionals*.

WALTER PHILIPP AND GEOFFREY R. WEBB, *An invariance principle for mixing sequences of random variables*.

GORDON PLEDGER AND R. J. SERFLING, *The waiting time of a vehicle in queue*.

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

———, *Applications of almost surely convergent constructions of weakly convergent processes*, Probability and Information Theory, Lectures notes in mathematics #89, Springer-Verlag, Berlin-Heidelberg-New York, 1969, pp. 187–200.

———, *Asymptotic results for rank statistics*, Nonparametric Techniques in Statistical Inference, M. L. Puri, ed., Cambridge University Press, 1970, pp. 21–37.

———, *Spacings revisited*.

R. PYKE AND D. ROOT, *An application of stopping times to obtain weak convergence*.

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

———, *Weak convergence and a Chernoff–Savage theorem for random sample sizes*, Ibid., 39 (1968), pp. 1675–1685.

———, *A note on Chernoff–Savage theorems*.

HERBERT ROBBINS AND DAVID SIEGMUND, *Boundary crossing probabilities for the Wiener process and sample sums*, Ann. Math. Statist., 41 (1970), pp. 1410–1429.

D. H. ROOT, *Construction of almost surely convergent stochastic processes*.

BENGT ROSÉN, *Asymptotic normality of sums of random elements with values in a real separable Hilbert space*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete, 13 (1969), pp. 221–255.

W. ROSENKRANTZ, *On rates of convergence for the invariance principle*, Trans. Amer. Math. Soc., 129 (1967), pp. 542–552.

I. R. SAVAGE AND J. SETHURAMAN, *Asymptotic distribution of the log likelihood ratio based on ranks in the two-sample problem*.

SIEGFRIED SCHACH, *Weak convergence results for a class of multivariate Markov processes*, Ann. Math. Statist., 42 (1971), pp. 451–465.

D. J. SCOTT, *An invariance principle for reversed martingales*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete, 20 (1971), pp. 9–27.

R. J. SERFLING, *Moment inequalities for maximum cumulative sum*, Ann. Math. Statist., 41 (1970), pp. 1227–1234.

GALEN R. SHORACK, *Asymptotic normality of linear combinations of functions of order statistics*, Ibid., 40 (1969), pp. 2041–2050.

———, *Asymptotic normality of linear combinations of order statistics*.

A. V. SKOROKHOD AND N. P. SLOBODENJUK, *Limit Theorems for Random Walks*, Naukova Dymka, Kiev, 1970. (In Russian.)

M. SREEHARI, *An invariance principle for random partial sums*, Sankhyā, 30 (1968), pp. 433–442.

S. M. STIGLER, *Linear functions of order statistics*.

MIRON STRAF, *A general Skorokhod space and its application to the weak convergence of stochastic processes with several parameters*.

FLEMMING TOPSØE, *A criterion for weak convergence of measures with an application to convergence of measures on D*[0, 1], Math. Scand., 25 (1969), pp. 97–104.

———, *Compactness of spaces of measures*, Studia Math., 36 (1970), pp. 195–212.

———, *Topology and Measure*. Lecture notes in mathematics #133, Springer-Verlag, Berlin-Heidelberg-New York, 1970.

———, *On the Glivenko–Cantelli theorem*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete, 14 (1970), pp. 239–250.

———, *Measure spaces connected by correspondences*.

A. Tortrat, *Calcul des Probabilités*, Masson, Paris, 1971.

Malati M. Vaidya, *On the convergence of measures in a reflexive, separable, Banach space*, Sankyā, 30 (1968), pp. 443–446.

B. B. Wells, Jr., *Weak compactness of measures*, Proc. Amer. Math. Soc., 20 (1969), pp. 124–130.

Ward Whitt, *Weak convergence theorems for queues in heavy traffic*.

———, *Multiple channel queues in heavy traffic, III: random server selection*, Adv. Appl. Prob., 2 (1970), pp. 370–375.

———, *Weak convergence theorems for priority queues: preemptive-resume discipline*, J. Appl. Prob., 8 (1971), pp. 74–94.

———, *Weak convergence of probability measures on the function space* $C[0, \infty)$, Ann. Math. Statist., 41 (1970), pp. 939–944.

———, *A guide to the application of limit theorems for sequences of stochastic processes*, Operations Res., 18 (1970), pp. 1207–1213.

———, *Stochastic Abelian and Tauberian theorems*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete, to appear.

———, *Weak convergence of first passage time processes*, J. Appl. Probability, to appear.

———, *Representation and convergence of point processes on the line*.

———, *Weak convergence involving a random time change*.

Michael Wichura, *On the weak convergence of non-Borel probabilities on a metric space*.

———, *Inequalities with applications to the weak convergence of random processes with multi-dimensional time parameters*, Ann. Math. Statist., 40 (1969), pp. 681–687.

———, *On the construction of almost uniformly convergent random variables with given weakly convergent image laws*, Ibid., 41 (1970), pp. 284–291.

———, *A note on the weak convergence of stochastic processes*.

———, *A note on the convergence of series of stochastic processes*.

Zheng Zeng-Tong, *Weak convergence of measures in metric spaces*, Acta Math. Sinica, 15 (1965), pp. 153–158.