Convergence of Probability Measures

Patrick Billingsley

Departments of Statistics and Mathematics
The University of Chicago

JOHN WILEY & SONS, New York • Chichester • Brisbane • Toronto

Library of Congress Catalog Card Number: 68-23922
SBN 471 07242 7
Printed in the United States of America

20 19 18 17 16 15 14 13

TO MY MOTHER

Preface

Asymptotic distribution theorems in probability and statistics have from the beginning depended on the classical theory of weak convergence of distribution functions in Euclidean space—convergence, that is, at continuity points of the limit function. The past several decades have seen the creation and extensive application of a more inclusive theory of weak convergence of probability measures on metric spaces. There are many asymptotic results that can be formulated within the classical theory but require for their proofs this more general theory, which thus does not merely study itself. This book is about weak-convergence methods in metric spaces, with applications sufficient to show their power and utility.

The Introduction motivates the definitions and indicates how the theory will yield solutions to problems arising outside it. Chapter 1 sets out the basic general theorems, which are then specialized in Chapter 2 to the space of continuous functions on the unit interval and in Chapter 3 to the space of functions with discontinuities of the first kind. The results of the first three chapters are used in Chapter 4 to derive a variety of limit theorems for dependent sequences of random variables.

Although standard measure-theoretic probability and metric-space topology are assumed, no general (nonmetric) topology is used, and the few results required from functional analysis are proved in the text or in an appendix.

Mastering the impulse to hoard the examples and applications till the last, thereby obliging the reader to persevere to the end, I have instead spread them evenly through the book to illustrate the theory as it emerges in stages.

Chicago, March 1968 *Patrick Billingsley*

Acknowledgements

My thanks go to Søren Johansen, Samuel Karlin, David Kendall, Ronald Pyke, and Flemming Topsøe, who read large parts of the manuscript; the book owes much to their detailed suggestions, and I am very grateful. I should also like to thank Mary Woolridge for her typing, cheerful, swift, and error-free.

The writing of this book was supported in part by the Statistics Branch, Office of Naval Research, and in part by Research Grant No. 8026 from the Division of Mathematical, Physical, and Engineering Sciences of the National Science Foundation.

Contents

CONVERGENCE OF PROBABILITY MEASURES

Introduction

The De Moivre–Laplace limit theorem states that, if

$$F_n(x) = \mathsf{P}\left\{\frac{S_n - np}{\sqrt{npq}} \le x\right\} \tag{1}$$

is the distribution function of the normalized number of successes in n Bernoulli trials, and if

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}u^2}\, du \tag{2}$$

is the unit normal distribution function, then

$$F_n(x) \to F(x) \tag{3}$$

for all x ($n \to \infty$, the probability p of success fixed).

We say of arbitrary distribution functions F_n and F on the line that F_n *converges weakly to* F, which we indicate by writing $F_n \Rightarrow F$, if (3) holds for all continuity points x of F. Thus the De Moivre–Laplace theorem asserts that (1) converges weakly to (2); since (2) is everywhere continuous, the proviso about continuity points is vacuous in this case. If F_n and F are defined by

$$F_n(x) = \begin{cases} 0 & \text{if} \quad x < \dfrac{1}{n} \\[2ex] 1 & \text{if} \quad x \ge \dfrac{1}{n} \end{cases} \tag{4}$$

and

$$F(x) = \begin{cases} 0 & \text{if} \quad x < 0 \\ 1 & \text{if} \quad x \ge 0, \end{cases} \tag{5}$$

then again $F_n \Rightarrow F$, and this time the proviso does come into play: (3) fails at $x = 0$.

For a better understanding of this notion of weak convergence, which underlies a large class of limit theorems in probability theory, consider the probability measures P_n and P generated by arbitrary distribution functions F_n and F. These probability measures, defined on the class of Borel subsets of the line, are uniquely determined by the requirements

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

Since F is continuous at x if and only if the set $\{x\}$ consisting of x alone has P-measure 0, $F_n \Rightarrow F$ means that the implication

$$P_n(-\infty, x] \to P(-\infty, x] \qquad \text{if} \quad P\{x\} = 0 \tag{6}$$

holds for each x.

Let ∂A denote the boundary of a subset A of the line; ∂A consists of those points that are limits of sequences of points in A and are also limits of sequences of points outside A. Since the boundary of $(-\infty, x]$ consists of the single point x, (6) is equivalent to

$$P_n(A) \to P(A) \qquad \text{if} \quad P(\partial A) = 0, \tag{7}$$

where we have written A for $(-\infty, x]$. The fact of the matter is that $F_n \Rightarrow F$ holds if and only if the implication (7) is true for *every* Borel set A—a result proved in Chapter 1.

Let us distinguish by the term *P-continuity set* those Borel sets A for which $P(\partial A) = 0$, and let us say that P_n *converges weakly to* P, and write $P_n \Rightarrow P$, if $P_n(A) \to P(A)$ for each P-continuity set—that is, if (7) holds. As just asserted, $P_n \Rightarrow P$ if and only if the corresponding distribution functions satisfy $F_n \Rightarrow F$.

This reformulation of the concept of weak convergence clarifies the reason why we allow (3) to fail if F has a jump at x. Without this exemption, (4) would not converge weakly to (5), but this example may appear artificial. If we turn our attention to probability measures P_n and P, however, we see that $P_n(A) \to P(A)$ may fail if $P(\partial A) > 0$ even in the De Moivre–Laplace theorem. The measures P_n and P generated by (1) and (2) satisfy

$$P_n(A) = \mathsf{P}\left\{\frac{S_n - np}{\sqrt{npq}} \in A\right\} \tag{8}$$

and

$$P(A) = \frac{1}{\sqrt{2\pi}} \int_A e^{-\frac{1}{2}u^2}\, du \tag{9}$$

for Borel sets A. Now if A consists of the countably many points

$$\frac{k - np}{\sqrt{npq}}, \qquad n = 1, 2, \ldots, \qquad k = 0, 1, \ldots, n,$$

then $P_n(A) = 1$ for all n and $P(A) = 0$, so that $P_n(A) \to P(A)$ is impossible. Since ∂A is the entire real line, this does not violate (7).

Although the concept of weak convergence of distribution functions is tied to the real line (or to Euclidean space, at any rate), the concept of weak convergence of probability measures can be formulated for the general metric space, which is the real reason for preferring the latter concept. Let S be an arbitrary metric space, let $\mathscr{S}$ be the class of Borel subsets of S ($\mathscr{S}$ is the σ-field generated by the open sets), and consider probability measures P_n and P defined on $\mathscr{S}$. Exactly as before, we define weak convergence $P_n \Rightarrow P$ by requiring the implication (7) to hold for all Borel sets A. In Chapter 1 we investigate the general theory of this concept of convergence and see what it reduces to in various particular metric spaces. We prove there, for example, that P_n converges weakly to P if and only if

$$\int_S f\,dP_n \to \int_S f\,dP \quad \text{if} \quad f \in C(S), \tag{10}$$

where $C(S)$ denotes the class of bounded, continuous real-valued functions on S. (In order to conform with general mathematical usage, we take (10) as the definition of weak convergence, so that (7) becomes a necessary and sufficient condition instead of a definition.)

Chapter 2 concerns weak convergence in the space $C = C[0, 1]$ with the uniform topology; C is the space of all continuous real functions on the closed unit interval $[0, 1]$, metrized by taking the distance between two functions $x = x(t)$ and $y = y(t)$ to be

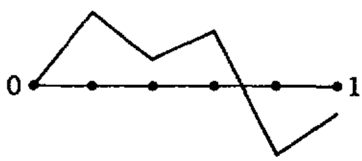
$$\rho(x, y) = \sup_{0 \le t \le 1} |x(t) - y(t)|. \tag{11}$$

An example of the sort of application made in Chapter 2 will show the utility of a general treatment of weak convergence—one that goes beyond the classical Euclidean case. Let $\xi_1, \xi_2, \ldots$ be a sequence of independent, identically distributed random variables defined on some probability space $(\Omega, \mathscr{B}, \mathsf{P})$. If the ξ_n have mean 0 and variance σ^2, then, according to the Lindeberg-Lévy central limit theorem, the distribution of the normalized sum

$$\frac{1}{\sigma\sqrt{n}} S_n = \frac{1}{\sigma\sqrt{n}} (\xi_1 + \cdots + \xi_n) \tag{12}$$

converges weakly, as n tends to infinity, to the normal distribution defined by (9).

We can formulate a refinement of the central limit theorem by proving weak convergence of the distributions of certain random functions constructed from the partial sums S_n. For each integer n and each sample point ω, construct on the unit interval the polygonal function that is linear on each of the subintervals $[(i-1)/n, i/n]$, $i = 1, 2, \ldots, n$, and has the value $S_i(\omega)/\sigma\sqrt{n}$ at the point i/n ($S_0(\omega) = 0$). In other words, construct the function $X_n(\omega)$ whose value at a point t of $[0, 1]$ is



(13)
$$X_n(t, \omega) = \frac{1}{\sigma\sqrt{n}} S_{i-1}(\omega) + \frac{t - (i-1)/n}{1/n} \frac{1}{\sigma\sqrt{n}} \xi_i(\omega) \qquad \text{if} \quad t \in \left[\frac{i-1}{n}, \frac{i}{n}\right].$$

For each ω, $X_n(\omega)$ is an element of the space C. Let P_n be the distribution of $X_n(\omega)$ in C, defined for Borel subsets A of C—Borel sets relative to the metric (11)—by

$$P_n(A) = \mathsf{P}\{\omega : X_n(\omega) \in A\}$$

(the definition is possible because the mapping $\omega \to X_n(\omega)$ turns out to be measurable in the right way). In Chapter 2 we prove

(14)
$$P_n \Rightarrow W,$$

where W is Wiener measure. We also prove the existence in C of Wiener measure, which describes the probability distribution of the path traced out by a particle under Brownian motion.

If $A = \{x : x(1) \leq \alpha\}$, then, since the value of the function $X_n(\omega)$ at $t = 1$ is $X_n(1, \omega) = S_n(\omega)/\sigma\sqrt{n}$,

$$P_n(A) = \mathsf{P}\left\{\omega : \frac{1}{\sigma\sqrt{n}} S_n(\omega) \leq \alpha\right\}.$$

It turns out that $W(\partial A) = 0$, so that (14) implies

$$\mathsf{P}\left\{\omega : \frac{1}{\sigma\sqrt{n}} S_n(\omega) \leq \alpha\right\} \to W\{x : x(1) \leq \alpha\}.$$

It also turns out that

$$W\{x : x(1) \leq \alpha\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\alpha} e^{-\frac{1}{2}u^2}\, du,$$

so that (14) does contain the Lindeberg–Lévy theorem.

The sum $\xi_1 + \cdots + \xi_n$ may be interpreted as the position at time n in a random walk. The central limit theorem asserts that this position (properly normalized) is, for n large, distributed approximately as the position at time $t = 1$ of a particle in Brownian motion. The relation (14) asserts that the entire path of the random walk during the first n steps is, for n large, distributed approximately as the path up to time $t = 1$ of a particle under Brownian motion.

To see in a concrete way that (14) contains information going beyond the central limit theorem, consider the set

$$A = \left\{x\colon \sup_{0\le t\le 1} x(t) \le \alpha\right\}.$$

Again it turns out that $W(\partial A) = 0$, so that (14) implies

$$(15)\qquad \lim_{n\to\infty} \mathsf{P}\left\{\omega\colon \frac{1}{\sigma\sqrt{n}} \max_{1\le k\le n} S_k(\omega) \le \alpha\right\} = \lim_{n\to\infty} P_n(A)$$

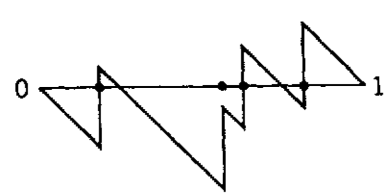
$$= W\left\{x\colon \sup_{0\le t\le 1} x(t) \le \alpha\right\}.$$

If we evaluate the right-most member of (15) (which we can do in a number of ways, for example, by computing the limit on the left for some specially selected sequence $\{\xi_n\}$ that makes the computation easy), then we have a limit theorem for the distribution of $\max_{k\le n} S_k$ under the hypotheses of the Lindeberg–Lévy theorem.

As a final example involving $X_n(\omega)$, take A to be the set of x in C for which the set $\{t\colon x(t) > 0\}$ has Lebesgue measure at most α (we assume $0 \le \alpha \le 1$). As before, $P_n(A) \to W(A)$. Since the Lebesgue measure of $\{t\colon X_n(t, \omega) > 0\}$ is essentially the fraction of the partial sums $S_1, S_2, \ldots, S_n$ that exceed 0, this argument leads to an arc sine law under the hypotheses of the Lindeberg–Lévy theorem. Chapter 2 contains the details of all these derivations.

We can in this way use the theory of weak convergence in C to obtain a whole class of limit theorems for functions of the partial sums $S_1, S_2, \ldots, S_n$. The fact that Wiener measure W is the weak limit of the distribution in C of the random function $X_n(\omega)$ can also be used to prove theorems about W, and W is interesting in its own right.

Chapter 3 specializes the theory of weak convergence to another space of functions on [0, 1]—the space $D = D[0, 1]$ of functions having discontinuities of at most the first kind. This is the natural space in which to analyze the behavior of empirical distribution functions, for example. Let $\xi_1, \xi_2, \ldots$ be independent random variables on $(\Omega, \mathscr{B}, \mathsf{P})$, each uniformly distributed on [0, 1]. Let $F_n(t, \omega)$ be the empirical distribution function of $\xi_1(\omega), \ldots, \xi_n(\omega)$; for $0 \le t \le 1$, $F_n(t, \omega)$ is the fraction of integers k, $1 \le k \le n$, for which $\xi_k(\omega) \le t$. Now let $Y_n(\omega)$ be the element of D whose value at t is



$$Y_n(t, \omega) = \sqrt{n}(F_n(t, \omega) - t).$$

If D is metrized in the right way, it becomes possible to speak of the distribution in D of the random function $Y_n(\omega)$ and to prove that this distribution converges weakly as n tends to infinity. Just as in the case of the random element of C defined by (13), we can then go on to derive the limiting distributions of

$$\sup_{0 \leq t \leq 1} \sqrt{n}\big(F_n(t, \omega) - t\big) = \sup_{0 \leq t \leq 1} Y_n(t, \omega)$$

and related quantities that arise in statistics.

Chapter 4 concerns weak convergence of the distributions of random functions derived from various dependent sequences of random variables. Many of the conclusions in Chapters 2, 3, and 4, although not requiring function-space concepts for their statement, could hardly have been derived without function-space methods.

Standard measure-theoretic probability and metric-space topology are used from the beginning. Although the point of view throughout is that of functional analysis (a function is a point in a space), nothing of functional analysis is assumed (beyond an initial willingness to view a function as a point in a space); all function-analytic results needed are proved in the text or else in Appendix I (which also gathers together for easy reference some results in metric-space topology).

Remarks. The main papers that lead to the development of this theory were Kolmogorov (1931), Erdös and Kac (1946 and 1947), Doob (1949), and Donsker (1951 and 1952). Prohorov (1953 and 1956) and Skorohod (1956) gave the theory its present form. Le Cam (1957) and Varadarajan (1958a and 1961a) have extended it to general topological spaces.

CHAPTER 1

Weak Convergence in Metric Spaces

1. MEASURES IN METRIC SPACES

Let S be a metric space. We shall study probability measures on the class $\mathscr{S}$ of Borel sets in S. Here $\mathscr{S}$ is the σ-field generated by the open sets—the smallest σ-field containing all the open sets—and a probability measure on $\mathscr{S}$ is a nonnegative, countably additive set function P with $P(S) = 1$.

If such probability measures P_n and P satisfy $\int_S f\,dP_n \to \int_S f\,dP$ for every bounded, continuous real function f on S, we say that P_n *converges weakly* to P and write $P_n \Rightarrow P$. Our aim in this chapter is to study this concept in detail; we begin with some properties of individual probability measures on $(S, \mathscr{S})$.

Although we must sometimes assume separability or completeness, most of the theorems in this chapter hold for an arbitrary metric space S. The spaces in our applications are usually separable and complete; since they rarely have further regularity properties, such as local compactness, we never impose further restrictions.

Measures and Integrals

THEOREM 1.1 *Every probability measure on $(S, \mathscr{S})$ is regular; that is, if $A \in \mathscr{S}$ and $\varepsilon > 0$, then there exist a closed set F and an open set G such that $F \subset A \subset G$ and $P(G - F) < \varepsilon$.*

Proof. Denote the metric on S by $\rho(x, y)$ and the distance from x to A by $\rho(x, A)$.† If A is closed, then we may take $F = A$ and $G = \{x: \rho(x, A) < \delta\}$

† For terminology and some theorems about metric spaces, see Appendix I.

for some δ, since the latter sets decrease to A as $\delta \downarrow 0$. Hence we need only show that the class $\mathscr{G}$ of Borel sets with the asserted property is a σ-field.† Given sets A_n in $\mathscr{G}$, choose closed sets F_n and open sets G_n such that $F_n \subset A_n \subset G_n$ and $P(G_n - F_n) < \varepsilon/2^{n+1}$. If $G = \bigcup_n G_n$, and if $F = \bigcup_{n \leq n_0} F_n$, with n_0 so chosen that $P(\bigcup_n F_n - F) < \varepsilon/2$, then $F \subset \bigcup_n A_n \subset G$ and $P(G - F) < \varepsilon$. Thus $\mathscr{G}$ is closed under the formation of countable unions; since $\mathscr{G}$ is obviously closed under complementation, the proof is complete.

Theorem 1.1 implies that P is determined by the values of $P(F)$ for closed sets F. Theorem 1.3 shows that P is determined by the values of $\int f\,dP$‡ for bounded, continuous real functions f defined on S. Denote by $C(S)$ the class of such functions f. It is shown on p. 222§ that each f in $C(S)$ is measurable $\mathscr{S}$. Everything depends on the following result, which shows how to approximate the indicator (or characteristic function) I_F of a closed set F by elements of $C(S)$.

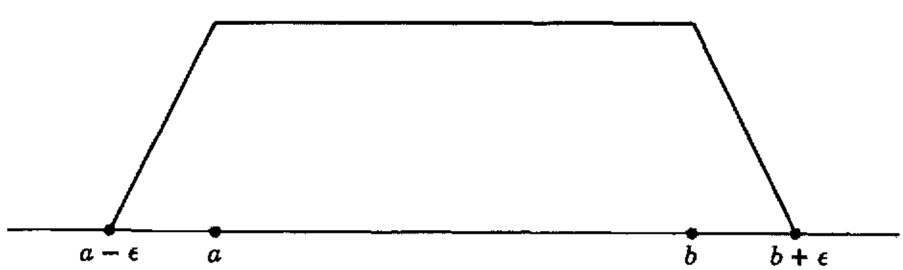
THEOREM 1.2 *If F is closed and ε positive, there is a function f in $C(S)$ such that $f(x) = 1$ if $x \in F$, $f(x) = 0$ if $\rho(x, F) \geq \varepsilon$, and $0 \leq f(x) \leq 1$ for all x. The function f may be taken to be uniformly continuous.*

Proof. Define a continuous function φ of a real variable by

$$\varphi(t) = \begin{cases} 1 & \text{if } t \leq 0, \\ 1 - t & \text{if } 0 \leq t \leq 1, \\ 0 & \text{if } 1 \leq t. \end{cases} \tag{1.1}$$

If

$$f(x) = \varphi\left(\frac{1}{\varepsilon}\,\rho(x, F)\right), \tag{1.2}$$



† We have defined the class $\mathscr{S}$ of Borel sets as the σ-field generated by the open sets, which is the same thing as the σ-field generated by the closed sets and is the one appropriate for the present theory. For related (mostly inappropriate) σ-fields, see Problem 6.

‡ When it is the entire space, we omit the region of integration.

§ Of Appendix II, a miscellany to which most measurability questions are relegated.

then f has the required properties—it is even uniformly continuous. The drawing graphs this f for $F = [a, b]$ on the line.

THEOREM 1.3 *Probability measures P and Q on $(S, \mathscr{S})$ coincide if*

$$\int f\,dP = \int f\,dQ \tag{1.3}$$

for each f in $C(S)$.

Proof. Suppose F is closed. Start with (1.1) and define, for each positive integer u,

$$\varphi_u(t) = \varphi(ut) \tag{1.4}$$

and

$$f_u(x) = \varphi_u(\rho(x, F)). \tag{1.5}$$

Then $\{f_u\}$ is a nonincreasing sequence of elements of $C(S)$ converging pointwise to I_F. By the bounded convergence theorem, $P(F) = \lim_u \int f_u\,dP$ and $Q(F) = \lim_u \int f_u\,dQ$, so that, if (1.3) holds for all f in $C(S)$, $P(F) = Q(F)$. Since P and Q agree for all closed sets, it follows by Theorem 1.1 that P and Q are identical.

Thus the values of $\int f\,dP$ for f in $C(S)$ completely determine the values of $P(A)$ for A in $\mathscr{S}$. This fact underlies the circle of ideas centering on the notion of weak convergence; although we have defined weak convergence by requiring the convergence of the integrals of functions in $C(S)$, in the next section we shall characterize it in terms of the convergence of the measures of certain sets.

Tightness

The following notion of tightness proves important both in the theory of weak convergence and in its applications. A probability measure P on $(S, \mathscr{S})$ is *tight* if for each positive ε there exists a compact (p. 217) set K such that $P(K) > 1 - \varepsilon$. Clearly, P is tight if and only if it has a σ-compact support.† By Theorem 1.1, P is tight if and only if $P(A)$ is, for each A in $\mathscr{S}$, the supremum of $P(K)$ over the compact subsets K of A.

In a space that is σ-compact, every probability measure is tight—which covers k-dimensional Euclidean space. The following result, which also covers the Euclidean case, is more useful.

† A support of a probability measure is a set A in $\mathscr{S}$ with $P(A) = 1$; a set is σ-compact if it can be represented as a countable union of compact sets. The characterization of a tight P as having a σ-compact support is inappropriate as a definition because it does not generalize in the right way to families of probability measures (see Section 6).

THEOREM 1.4 *If S is separable and complete, then each probability measure on $(S, \mathscr{S})$ is tight.*

Proof. Since S is separable, there is, for each n, a sequence $A_{n1}, A_{n2}, \ldots$ of open $1/n$-spheres covering S. Choose i_n so that $P(\bigcup_{i \le i_n} A_{ni}) > 1 - \varepsilon/2^n$. By the completeness hypothesis, the totally bounded set $\bigcap_{n \ge 1} \bigcup_{i \le i_n} A_{ni}$ has compact closure K (see p. 217). Since clearly $P(K) > 1 - \varepsilon$, the theorem follows.

Theorem 1.4 is false without the hypothesis of completeness; whether the hypothesis of separability can be suppressed is equivalent to the problem of measure. These matters are discussed in Appendix III.†

Remarks. Theorem 1.4 is due to Ulam (see Oxtoby and Ulam (1939)); LeCam (1957) introduced the term "tight."

PROBLEMS‡

1. Say that a function f separates sets A and B if $f(x) = 0$ for x in A, $f(x) = 1$ for x in B, and $0 \le f(x) \le 1$ for all x. If A and B are at positive distance, they can be separated by a uniformly continuous f [Theorem 1.2]. If A and B have disjoint closures but are at distance 0, they can be separated by a continuous f [$f(x) = \rho(x, A)/(\rho(x, A) + \rho(x, B))$] but not by a uniformly continuous f. There is no continuous f separating A and B if their closures meet; there is no f separating A and B if they meet.

2. Give examples of distinct topologies that give rise to the same class of Borel sets.

3. If S can be embedded as an open set in some complete metric space, then [Kelley (1955, p. 207)] it is topologically complete. Since a locally compact S is open in its completion, it is topologically complete. Hence Theorem 1.4 applies if S is separable and locally compact. Since such an S is σ-compact [being a union of open sets with compact closures and hence (p. 216) a countable such union], it also follows directly that each probability measure on it is tight; Euclidean space is an example.

4. Let S be a Hilbert space with a countably infinite orthonormal basis $x_1, x_2, \ldots$. Since S is separable and complete, Theorem 1.4 applies. However, no set with nonempty interior is compact [a nonempty interior must, for some x and ε, contain all the points $x + \varepsilon x_n$], so that S is neither locally compact nor [Baire's category theorem; Kelley (1955, p. 200)] σ-compact. If P assigns positive mass to each element of a countable, dense set, then P has no support locally compact in the relative topology.

5. Adapt Problem 4 to the general Banach space of countably infinite dimension [there exist points $x_1, x_2, \ldots$ with $\sup_n \|x_n\| < \infty$ and $\inf_{m \ne n} \|x_m - x_n\| > 0$; see Banach (1932, p. 83)]; $C[0, 1]$, important in probability, is such a space, which explains why a theory based on local compactness is of small utility in this subject. (See also Problem 5 in Section 3.)

---

† Although Theorem 1.4 as given suffices for all the applications in this book, it is natural to inquire after extensions. It is to questions of just this sort that Appendix III is devoted.

‡ Some problems involve concepts not required for an understanding of the text itself; there are no problems whose solutions are used later in the text. A simple assertion is understood to be prefaced by "show that." Square brackets contain hints or indications of solutions.

6. We have defined $\mathscr{S}$ as the σ-field generated by the open sets, which we can indicate by writing $\mathscr{S} = \sigma$(open sets). In the same way, define $\mathscr{S}_1 = \sigma$(closed G_δ sets) (a set is a G_δ if it is a countable intersection of open sets), define $\mathscr{S}_2 = \sigma(C(S))$ (the smallest σ-field with respect to which each function in $C(S)$ is measurable), and define $\mathscr{S}_3 = \sigma$(open spheres), $\mathscr{S}_4 = \sigma$(compact sets), and $\mathscr{S}_5 = \sigma$(compact G_δ sets). In a metric space each closed set is a G_δ. Use this fact and Theorem 1.2 to prove

$$\mathscr{S} = \mathscr{S}_1 = \mathscr{S}_2 \supset \mathscr{S}_3 \supset \mathscr{S}_4 = \mathscr{S}_5.$$

Show that $\mathscr{S} = \mathscr{S}_3$ if S is separable. Show that $\mathscr{S} = \mathscr{S}_5$ if S is σ-compact (which will be true if S is separable and locally compact). We may have $\mathscr{S}_2 \neq \mathscr{S}_3$ (even if S is locally compact): Take S uncountable and discrete. We may have $\mathscr{S}_3 \neq \mathscr{S}_4$ (even if S is separable and complete): Take S to be the Hilbert space in Problem 4. (The situation differs in the general topological space, where one must consider two classes of sets: The Borel sets are taken as the elements sometimes of $\mathscr{S}$ and sometimes of $\mathscr{S}_4$, and the Baire sets are taken as the elements sometimes of $\mathscr{S}_2$ and sometimes of $\mathscr{S}_5$—the terminology varies.)

7. In connection with tightness, this fact is interesting: Suppose P is defined on $(S, \mathscr{S})$, but suppose at the outset only that it is finitely additive. If, for each A in $\mathscr{S}$, $P(A) = \sup P(K)$ with K ranging over the compact subsets of A, then P is countably additive after all.

2. PROPERTIES OF WEAK CONVERGENCE

We have defined $P_n \Rightarrow P$ to mean that $\int f\, dP_n \to \int f\, dP$ for each f in the class $C(S)$ of bounded, continuous real functions on S. Note that, since the integrals $\int f\, dP$ completely determine P (Theorem 1.3), the sequence $\{P_n\}$ cannot converge weakly to two different limits at the same time. Note also that weak convergence depends only on the topology of S, not on the specific metric that generates it: Two metrics generating the same topology give rise to the same classes $\mathscr{S}$ and $C(S)$ and hence to the same notion of weak convergence.†

Portmanteau Theorem

The following theorem provides useful conditions equivalent to weak convergence; any one of these conditions could serve as the definition. A set A in $\mathscr{S}$ whose boundary ∂A satisfies $P(\partial A) = 0$ is called a *P-continuity set* (note that ∂A is closed and hence lies in $\mathscr{S}$).

THEOREM 2.1 *Let P_n, P be probability measures on $(S, \mathscr{S})$. These five conditions are equivalent:*

† If we topologize the space $Z(S)$ of all probability measures on $(S, \mathscr{S})$ by taking as the general basic neighborhood of P the set of Q such that $|\int f_i\, dP - \int f_i\, dQ| < \varepsilon$ for $i = 1, \ldots, k$, where ε is positive and the f_i lie in $C(S)$, then weak convergence is convergence in this topology. The topological structure of $Z(S)$, which will be of no direct concern to us, is discussed in Appendix III.

(i) $P_n \Rightarrow P$.
(ii) $\lim_n \int f\,dP_n = \int f\,dP$ *for all bounded, uniformly continuous real* f.
(iii) $\limsup_n P_n(F) \leq P(F)$ *for all closed* F.
(iv) $\liminf_n P_n(G) \geq P(G)$ *for all open* G.
(v) $\lim_n P_n(A) = P(A)$ *for all P-continuity sets* A.

A couple of examples will show the significance of these conditions. Let P be a unit mass at the point x ($P(A)$ is 1 or 0, according as x lies in A or not†), and let P_n be a unit mass at x_n. If $x_n \to x$, then $\int f\,dP_n = f(x_n) \to f(x) = \int f\,dP$ for all f in $C(S)$, so that $P_n \Rightarrow P$. If x_n does not converge to x, then, for some positive ε, we have $\rho(x_n, x) > \varepsilon$ for infinitely many n. If $f(y) = \varphi(\varepsilon^{-1}\rho(x, y))$ with φ defined by (1.1), then $f \in C(S)$, $f(x) = 1$, and $f(x_n) = 0$ for infinitely many n; hence P_n cannot converge weakly to P. Thus $P_n \Rightarrow P$ if and only if $x_n \to x$, which provides an example we shall often use. (Many putative weak-convergence theorems that are in fact not theorems can be disproved by specializing this example.) Since A is a P-continuity set if and only if $x \notin \partial A$, it is easy to check the equivalence of (i) and (v) in this case. If $x_n \to x$ but the x_n all differ from x, then there is strict inequality in (iii) for $F = \{x\}$ and strict inequality in (iv) for the complementary set $G = F^c$; moreover, if the x_n are all distinct and $A = \{x_2, x_4, \ldots\}$, then $P_n(A)$ does not converge to $P(A)$ or to anything else.

On the line with the ordinary metric, the DeMoivre–Laplace theorem also illustrates the conditions in the theorem. For a simpler example equally relevant, consider the measure P_n corresponding to a mass of $1/n$ at each of the points i/n, $i = 1, 2, \ldots, n$. Now P_n converges weakly to Lebesgue measure P confined to the unit interval, as follows from the fact that $\int f\,dP_n$ is an approximating sum to $\int f\,dP$ viewed as a Riemann integral. If A consists of the rationals, then $P_n(A) = 1$ does not converge to $P(A) = 0$; if G is an open set containing the rationals and having Legesgue measure near 0, then there is strict inequality in (iv).

We prove Theorem 2.1 by establishing the implications in the following diagram.

$$\begin{array}{ccccccc} & & & & & & \\ \text{(i)} & \to & \text{(ii)} & \to & \text{(iii)} & \leftrightarrow & \text{(iv)} \\ & & & & \updownarrow & & \\ & & & & \text{(v)} & & \end{array}$$

Of course, (i) $\to$ (ii) is trivial.

Proof of (ii) $\to$ (iii). Suppose (ii) holds and that F is closed. Suppose $\delta > 0$. For small enough ε, $G = \{x : \rho(x, F) < \varepsilon\}$ satisfies $P(G) < P(F) + \delta$,

† Each subset of S mentioned is assumed to lie in $\mathscr{S}$.

since the sets of this form decrease to F as $\varepsilon \downarrow 0$. If $f(x)$ is the function defined by (1.2), then f is uniformly continuous on S, $f(x) = 1$ on F, $f(x) = 0$ on the complement G^c of G, and $0 \leq f(x) \leq 1$ for all x. Since (ii) holds, we have $\lim_n \int f\,dP_n = \int f\,dP$, which, together with the relations

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

and

$$\int f\,dP = \int_G f\,dP \leq P(G) < P(F) + \delta,$$

implies

$$\limsup_n P_n(F) \leq \lim_n \int f\,dP_n = \int f\,dP < P(F) + \delta.$$

Since δ was arbitrary, (iii) follows.

Proof of (iii) $\to$ (i). Suppose that (iii) holds and that $f \in C(S)$. We shall first show that

$$\limsup_n \int f\,dP_n \leq \int f\,dP. \tag{2.1}$$

By transforming f linearly (with a positive coefficient for the first-degree term), we may reduce the problem to the case in which $0 < f(x) < 1$ for all x. For an integer k, temporarily fixed, let F_i be the closed set $F_i = \{x : i/k \leq f(x)\}$, $i = 0, 1, \ldots, k$. Since $0 < f(x) < 1$, we have

$$\sum_{i=1}^k \frac{i-1}{k} P\left\{x : \frac{i-1}{k} \leq f(x) < \frac{i}{k}\right\} \leq \int f\,dP < \sum_{i=1}^k \frac{i}{k} P\left\{x : \frac{i-1}{k} \leq f(x) < \frac{i}{k}\right\}.$$

The sum on the right is

$$\sum_{i=1}^k \frac{i}{k}\,[P(F_{i-1}) - P(F_i)] = \frac{1}{k} + \frac{1}{k}\sum_{i=1}^k P(F_i).$$

This and a similar transformation of the sum on the left yield

$$\frac{1}{k}\sum_{i=1}^k P(F_i) \leq \int f\,dP < \frac{1}{k} + \frac{1}{k}\sum_{i=1}^k P(F_i). \tag{2.2}$$

If (iii) holds, then $\limsup_n P_n(F_i) \leq P(F_i)$ for each i and hence (apply the right-hand inequality in (2.2) to P_n and the left-hand one to P)

$$\limsup_n \int f\,dP_n \leq \frac{1}{k} + \int f\,dP.$$

Letting $k \to \infty$, we obtain (2.1).

Applying (2.1) to $-f$ yields $\liminf_n \int f\,dP_n \geq \int f\,dP$, which, together with (2.1) itself, proves weak convergence.

The equivalence of (iii) and (iv) follows easily by complementation.

Proof of (iii) $\to$ *(v)*. Let A° denote the interior of A, and let A^- denote its closure. If (iii) holds, then so does (iv), and hence, for each A,

$$P(A^-) \geq \limsup_n P_n(A^-) \geq \limsup_n P_n(A) \geq \liminf_n P_n(A) \geq \liminf_n P_n(A^\circ) \geq P(A^\circ). \tag{2.3}$$

If $P(\partial A) = 0$, then the extreme terms equal $P(A)$ and $\lim_n P_n(A) = P(A)$ follows.

Proof of (v) $\to$ *(iii)*. Since $\partial\{x: \rho(x, F) \leq \delta\}$ is contained† in $\{x: \rho(x, F) = \delta\}$, these boundaries are disjoint for distinct δ, and hence at most countably many of them can have positive P-measure. Therefore, for some sequence of positive δ_k going to 0, the sets $F_k = \{x: \rho(x, F) \leq \delta_k\}$ are P-continuity sets. If (v) holds, then $\limsup_n P_n(F) \leq \lim_n P_n(F_k) = P(F_k)$ for each k; if F is closed, then $F_k \downarrow F$, so that (iii) follows. This completes the proof of Theorem 2.1.

Other Criteria

It is sometimes convenient to prove weak convergence by showing that $P_n(A) \to P(A)$ for some special class of sets A.

THEOREM 2.2 *Let $\mathscr{U}$ be a subclass of $\mathscr{S}$ such that* (i) *$\mathscr{U}$ is closed under the formation of finite intersections and* (ii) *each open set in S is a finite or countable union of elements of $\mathscr{U}$. If $P_n(A) \to P(A)$ for every A in $\mathscr{U}$, then $P_n \Rightarrow P$.*

Proof. If $A_1, \ldots, A_m$ lie in $\mathscr{U}$, then so do their intersections; hence, by the inclusion-exclusion formula,

$$\begin{aligned} P_n\left(\bigcup_{i=1}^m A_i\right) &= \Sigma_i P_n(A_i) - \Sigma_{ij} P_n(A_i A_j) + \Sigma_{ijk} P_n(A_i A_j A_k) - \cdots \\ &\to \Sigma_i P(A_i) - \Sigma_{ij} P(A_i A_j) + \Sigma_{ijk} P(A_i A_j A_k) - \cdots \\ &= P\left(\bigcup_{i=1}^m A_i\right). \end{aligned}$$

If G is open, then $G = \bigcup_i A_i$ for some sequence $\{A_i\}$ of elements of $\mathscr{U}$. Given ε, choose m so that $P(\bigcup_{i \leq m} A_i) > P(G) - \varepsilon$. By the relation just proved, $P(G) - \varepsilon < P(\bigcup_{i \leq m} A_i) = \lim_n P_n(\bigcup_{i \leq m} A_i) \leq \liminf_n P_n(G)$. Since ε was arbitrary, condition (iv) of the preceding theorem holds.

Let $S(x, \varepsilon)$ denote the (open) ε-sphere about x.

COROLLARY 1 *Let $\mathscr{U}$ be a class of sets such that* (i) *$\mathscr{U}$ is closed under the formation of finite intersections and* (ii) *for every x in S and every positive ε*

† The inclusion may be strict—in a discrete space, for example.

there is an A in $\mathscr{U}$ with $x \in A^\circ \subset A \subset S(x, \varepsilon)$. If S is separable and if $P_n(A) \to P(A)$ for every A in $\mathscr{U}$, then $P_n \Rightarrow P$.

Proof. Condition (ii)† implies that, for each point x of an open set G, $x \in A^\circ \subset A \subset G$ for some A in $\mathscr{U}$. Since S is separable, there exists (see p. 216) in $\mathscr{U}$ a finite or infinite sequence $\{A_i\}$ such that $G \subset \bigcup_i A_i^\circ$ and $A_i \subset G$, which implies $G = \bigcup_i A_i$. Thus $\mathscr{U}$ satisfies the hypotheses of Theorem 2.2.

COROLLARY 2 *Suppose that, for each finite intersection A of open spheres, we have $P_n(A) \to P(A)$, provided A is a P-continuity set. If S is separable, then $P_n \Rightarrow P$.*

Proof. The boundaries $\partial S(x, \varepsilon)$, being contained in the sets $\{y : \rho(x, y) = \varepsilon\}$, are disjoint (for fixed x) and hence have P-measure 0, with at most countably many exceptions. Since

$$\partial(A \cap B) \subset (\partial A) \cup (\partial B),$$

it follows that the hypotheses of Corollary 1 are satisfied by the class $\mathscr{U}$ of those P-continuity sets that are finite intersections of spheres, and the result follows.

Let us agree to call a subclass $\mathscr{V}$ of $\mathscr{S}$ a *convergence-determining class* if convergence $P_n(A) \to P(A)$ for all P-continuity sets A in $\mathscr{V}$ invariably entails the weak convergence of P_n to P. Corollary 2 becomes: In a separable space, the finite intersections of spheres constitute a convergence-determining class.

Let us further agree to call $\mathscr{V}$ a *determining class* if P and Q are identical whenever they agree on $\mathscr{V}$. The class of closed sets is a determining class and so is any field that generates $\mathscr{S}$. Although each convergence-determining class is clearly also a determining class, the following example shows that the converse fails. Let S be the half-open interval $[0, 1)$ with the ordinary metric; let $\mathscr{V}$ be the class of sets $[a, b)$ with $0 < a < b < 1$. Then $\mathscr{V}$ is a determining class but not (as may be seen by taking P_n $[P]$ a unit mass at $1 - 1/n$ $[0]$) a convergence-determining class. Although this one is artificial, we shall see that the applications abound with real examples of determining classes that are not convergence-determining classes.

We close this section with another condition for weak convergence. A sequence $\{x_n\}$ of real numbers converges to a limit x if and only if each subsequence $\{x_{n'}\}$ contains a further subsequence $\{x_{n''}\}$ that converges to x. (It is convenient to denote a sequence of integers by $\{n'\}$ rather than $\{n_k\}$ and a subsequence of $\{n'\}$ by $\{n''\}$ rather than $\{n_{k_i}\}$.) From this fact it is easy to deduce a weak-convergence analogue.

† This condition is slightly stronger than the requirement that the interiors of the elements of $\mathscr{U}$ form a base for the topology of S.

THEOREM 2.3 *We have $P_n \Rightarrow P$ if and only if each subsequence $\{P_{n'}\}$ contains a further subsequence $\{P_{n''}\}$ such that $P_{n''} \Rightarrow P$.*

We shall deal occasionally with weak convergence of P_t to P when t goes to infinity in a continuous manner. Of course, this is defined to mean that

$$\lim_{t\to\infty} \int f\,dP_t = \int f\,dP \tag{2.4}$$

for each f in $C(S)$. For f fixed, (2.4) holds if and only if

$$\lim_{n\to\infty} \int f\,dP_{t_n} = \int f\,dP \tag{2.5}$$

for each sequence $\{t_n\}$ going to infinity. Thus $P_t \Rightarrow P$ as $t \to \infty$ if and only if $P_{t_n} \Rightarrow P$ for each sequence $\{t_n\}$ going to infinity, and nothing really new is involved. We can also let t approach in a continuous manner some finite value t_0.

Remarks. Theorem 2.1 dates back at least to Alexandrov (1940–43). Theorem 2.2 is due to Kolmogorov and Prohorov (1954). For other accounts of the theory, see the books of Gikhman and Skorohod (1965), Hennequin and Tortrat (1965), and Parthasarathy (1967).

The Banach space $C(S)$ has for its adjoint $C^*(S)$ the space of finite signed measures on $\mathcal{S}$; the weak* topology, or the $C(S)$ topology of $C^*(S)$, relativized to the space $Z(S)$ of probability measures on $\mathcal{S}$, is the topology described in the first footnote in this section (hence the "weak" in our "weak convergence"); see Dunford and Schwartz (1958, pp. 262 and 419). Varadarajan (1958a and 1961a) investigates the topological structure of $Z(S)$; see also Appendix III.

For extensions of the theory to general topological spaces, see LeCam (1957), Varadarajan (1958a and 1961a), and Kallianpur (1961).

If the metric space S is not separable, the σ-field $\mathcal{S}_0$ generated by the spheres may be smaller than $\mathcal{S}$. Dudley (1966 and 1967) has a theory of weak convergence involving only sets in $\mathcal{S}_0$ and functions measurable $\mathcal{S}_0$.

If $P_n \Rightarrow P$, one can ask whether $P_n(A) \to P(A)$ holds uniformly on a given class of P-continuity sets; see Ranga Rao (1962) and Billingsley and Topsøe (1967).

PROBLEMS

1. If S is countable and discrete, then $P_n \Rightarrow P$ if and only if $P_n\{x\} \to P\{x\}$ for each one-point set $\{x\}$.

2. Let P_n and P be given by densities p and p_n with respect to a measure λ on $(S, \mathcal{S})$. If $p_n(x) \to p(x)$ except for x in a set of λ-measure 0, then $P_n \Rightarrow P$ [see Scheffé's theorem on p. 224]. Show by example that $p_n(x) \to p(x)$ may fail on a set of positive measure even though $P_n \Rightarrow P$.

3. Even though $P_n \Rightarrow P$, $\int f\,dP_n \to \int f\,dP$ may fail if f is bounded but not continuous or if f is continuous but not bounded (even if the integrals exist). Give examples. If S is compact, the second possibility does not exist. What if S is not compact but P has compact support?

4. The class of P-continuity sets (P fixed) form a field.

5. If $\mathscr{U}$ is a determining class, if $P_n(A) \to Q(A)$ for $A \in \mathscr{U}$, and if $P_n \Rightarrow P$, it does *not* follow that $P = Q$. [Define probability measures on the line by $P_n\{n^{-1}\} = P_n\{1 + n^{-1}\} = P\{0\} = P\{1\} = \frac{1}{2}$ and $Q\{0\} = 1$. Let B consist of the points 0, 1, n^{-1}, and $1 + n^{-1}$ ($n = 1, 2, \ldots$). Define $\mathscr{U}$ as the field of Borel sets A such that either (i) $A \cap B$ is finite and $0 \notin A$ or else (ii) $A^c \cap B$ is finite and $0 \notin A^c$. (This example is due to O. Björnsson.)]

6. Define what one should mean by determining classes and convergence-determining classes of elements of $C(S)$. Give an example of a determining class that is not a convergence-determining class. Show that a class uniformly dense in $C(S)$ is a convergence-determining class.

7. If f is bounded and upper semicontinuous (p. 218), then $P_n \Rightarrow P$ implies $\limsup_n \int f\, dP_n \leq \int f\, dP$.

8. Let $\{f_\theta\}$ be a family of real functions on S, equicontinuous at each x (for each x and ε, there exists a δ for which $\rho(x, y) < \delta$ implies, for all θ, $|f_\theta(x) - f_\theta(y)| < \varepsilon$). If $\{f_\theta\}$ is uniformly bounded and S is separable, then $P_n \Rightarrow P$ implies that $\int f_\theta\, dP_n \to \int f_\theta\, dP$ uniformly in θ. [First show that, for each ε, there exists a countable partition $\Delta_\varepsilon = \{D_{\varepsilon k}\}$ of S into P-continuity sets such that $|f_\theta(x) - f_\theta(y)| < \varepsilon$ for all θ if x and y lie in the same $D_{\varepsilon k}$. Approximate $\int f_\theta\, dP$ by $\Sigma_k f_\theta(x_k) P(D_{\varepsilon k})$ with x_k in $D_{\varepsilon k}$, and similarly for $\int f_\theta\, dP_n$, and apply Scheffé's theorem (p. 224).] (This result is due to Ranga Rao (1962); for extensions, see Billingsley and Topsøe (1967) and Topsøe (1967a and 1967b).)

3. SOME SPECIAL CASES

Euclidean Space

Let R^k denote k-dimensional Euclidean space, which we shall always take with the ordinary metric $\rho(x, y) = |x - y| = [\Sigma_{i=1}^k (x_i - y_i)^2]^{\frac{1}{2}}$. We denote by $\mathscr{R}^k$ the class of Borel sets; the elements of $\mathscr{R}^k$ we shall call *k-dimensional Borel sets, linear Borel sets* in case $k = 1$. By the notation $y \leq x$ [$y < x$], we mean $y_i \leq x_i$ [$y_i < x_i$] for all $i = 1, 2, \ldots, k$. (Note that $y < x$ is stronger than $y \leq x$ together with $y \neq x$.) An interval is a set

$$(a, b] = \{x : a < x \leq b\}. \tag{3.1}$$

Finally, let us denote by $e = (1, \ldots, 1)$ the vector whose coordinates are all 1.

The general probability measure P on $(R^k, \mathscr{R}^k)$ has a distribution function F, defined by

$$F(x) = P\{y : y \leq x\}, \qquad x \in R^k. \tag{3.2}$$

Let us relate weak convergence $P_n \Rightarrow P$ to the usual notion of convergence for the corresponding distribution functions F_n, F.

The function F is continuous at x if and only if for each positive ε there exists a positive δ such that $x - \delta e < y < x + \delta e$ implies $|F(x) - F(y)| < \varepsilon$. To say that F is continuous from above at x means that for each positive ε there exists a positive δ such that $x \leq y < x + \delta e$ implies $|F(x) - F(y)| < \varepsilon$.

From the definition (3.2) it follows that F is nondecreasing in each variable. Hence F is continuous from above at x if and only if $F(x)$ coincides with $\inf_{\delta>0} F(x + \delta e) = \inf_{\delta>0} P\{y:y \leq x + \delta e\}$. This infimum is just the P-measure of the intersection $\bigcap_{\delta>0} \{y:y \leq x + \delta e\} = \{y:y \leq x\}$. Therefore F is continuous from above at each x.

Since F is nondecreasing in each variable and is everywhere continuous from above, it is continuous at x if and only if it is continuous from below at that point, that is, if and only if for each positive ε there exists a positive δ such that $x - \delta e < y \leq x$ implies $|F(x) - F(y)| < \varepsilon$. Using the monotonicity once more, we see that this condition is in turn equivalent to $F(x) = \sup_{\delta>0} F(x - \delta e)$. The supremum being the P-measure of the union $\bigcup_{\delta>0} \{y:y \leq x - \delta e\} = \{y:y < x\}$, we see that F is continuous at x if and only if $F(x) = P\{y:y < x\}$. Since $\{y:y \leq x\} - \{y:y < x\}$ is exactly the boundary of $\{y:y \leq x\}$, F is continuous at x if and only if $\{y:y \leq x\}$ is a P-continuity set.

For distribution functions F_n and F, let us define $F_n \Rightarrow F$ to mean that there is convergence $F_n(x) \to F(x)$ at continuity points x of F. By what has just been proved, if $P_n \Rightarrow P$, then the corresponding distribution functions satisfy $F_n \Rightarrow F$. Now an interval $(a, b]$ is determined by the $2k$ $(k - 1)$-dimensional hyperplanes containing its faces; let $\mathscr{U}$ be the class of intervals for which all these hyperplanes have P-measure 0. Each vertex of an element of $\mathscr{U}$ is a continuity point of F, and $\mathscr{U}$ is closed under the formation of finite intersections. Since only countably many parallel hyperplanes can have positive P-measure, it follows by Corollary 1 to Theorem 2.2 that, if $P_n(A) \to P(A)$ for each A in $\mathscr{U}$, then $P_n \Rightarrow P$. Since $P(a, b]$ is a sum $\Sigma \pm F(x)$ with x ranging over the 2^k vertices of $(a, b]$ and similarly for $P_n(a, b]$, $F_n \Rightarrow F$ implies that $P_n(A) \to P(A)$ holds for each A in $\mathscr{U}$. Therefore $P_n \Rightarrow P$ and $F_n \Rightarrow F$ are equivalent.

Thus the notion of weak convergence reduces in R^k to the ordinary notion of the convergence of distribution functions. In other words, the sets $\{y:y \leq x\}$ form a convergence-determining class. The proof above also shows that the rectangles $(a, b]$ form a convergence-determining class.

The Circle

The same sort of result obviously holds if S is the unit circle in the complex plane: $P_n \Rightarrow P$ if and only if $P_n(A) \to P(A)$ for every arc A whose endpoints have P-measure 0. A sequence $\{x_1, x_2, \ldots\}$ of points of S (complex numbers of modulus 1) is said to be uniformly distributed if the allotment of points to each arc is proportional to its length, in the sense that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} I_A(x_j) = P(A), \tag{3.3}$$

where I_A is the indicator, or characteristic function, of A, and P is circular Lebesgue measure so normalized that $P(S) = 1$. If P_n denotes the nth empirical distribution of the sequence—the measure corresponding to a mass of $1/n$ at each of the points $x_1, x_2, \ldots, x_n$—this condition reduces to $P_n(A) \to P(A)$ for arcs A, so that the sequence is uniformly distributed if and only if $P_n \Rightarrow P$. Therefore, if every arc contains its proper quota of points, in the sense of (3.3), then so does every other Borel set whose boundary has Lebesgue measure 0. We prove in Section 7 a famous theorem of Weyl, according to which $\{x_1, x_2, \ldots\}$ is uniformly distributed if and only if $n^{-1} \sum_{j=1}^{n} (x_j)^u \to 0$ holds for every nonzero integer u.

The Space R^∞

The results for R^k carry over in all essential respects to the topological product of a countable sequence of copies of the line, that is, to the space R^∞ of sequences $x = (x_1, x_2, \ldots)$ of real numbers (see p. 218). The topology has as the basic neighborhoods of a point x the sets of the form

$$N_{k,\varepsilon}(x) = \{y : |x_i - y_i| < \varepsilon, \qquad i = 1, \ldots, k\} \tag{3.4}$$

with $\varepsilon > 0$ and $k = 1, 2, \ldots$. With this topology, R^∞ is a complete, separable metric space. (We shall always mean by R^∞ this topological product of a *countable* set of copies of the real line.)

Let π_k denote the natural projection from R^∞ to R^k, defined by $\pi_k(x) = (x_1, \ldots, x_k)$. A *finite-dimensional set*, or cylinder, is by definition a set of the form $\pi_k^{-1}H$ with $k \geq 1$ and $H \in \mathscr{R}^k$. Since each π_k is continuous and hence measurable (see p. 222), the finite-dimensional sets lie in the σ-field $\mathscr{R}^\infty$ of Borel sets in R^∞. Let $\mathscr{F}$ denote the class of finite-dimensional sets. Since each set (3.4) lies in $\mathscr{F}$ and since R^∞ is separable, $\mathscr{F}$ generates $\mathscr{R}^\infty$. Since $\mathscr{F}$ is a (finitely additive) field, it follows that $\mathscr{F}$ is a determining class.

For fixed k and x, the sets (3.4) for different values of ε have disjoint boundaries ($\varepsilon < \delta$ implies $N_{k,\varepsilon}^- \subset N_{k,\delta}^\circ$). Applying Corollary 1 of Theorem 2.2 to the class $\mathscr{U}$ of P-continuity sets in $\mathscr{F}$, we see therefore that $\mathscr{F}$ is even a convergence-determining class. Thus $P_n \Rightarrow P$ if and only if $P_n(A) \to P(A)$ holds for all finite-dimensional P-continuity sets A.

The Space C

In $C = C[0, 1]$, the space of continuous functions on $[0, 1]$ with the uniform metric $\rho(x, y) = \sup_t |x(t) - y(t)|$ (see p. 220), the situation differs markedly from that in R^∞. For points $t_1, \ldots, t_k$ in $[0, 1]$, let $\pi_{t_1 \cdots t_k}$ be the mapping that carries the point x of C to the point $(x(t_1), \ldots, x(t_k))$ of R^k. The finite-dimensional sets are now defined as sets of the form $\pi_{t_1 \cdots t_k}^{-1} H$ with $H \in \mathscr{R}^k$.

Since $\pi_{t_1\cdots t_k}$ is continuous, these sets lie in the class $\mathscr{C}$ of Borel sets in C. On the other hand, the closed sphere $\{y: \rho(x, y) \leq \varepsilon\}$ is the limit of the finite-dimensional sets $\{y: |x(i/n) - y(i/n)| \leq \varepsilon,\ i = 1, \ldots, n\}$; since C is separable, each open set is a countable union of open spheres and hence of closed spheres, so that the finite-dimensional sets generate $\mathscr{C}$. Since they form a field, the finite-dimensional sets are thus a determining class.

An example shows that the finite-dimensional sets do *not* form a convergence-determining class. Let P be a unit mass at 0 (the function that vanishes identically), and let P_n be a unit mass at the function x_n, where

$$x_n(t) = \begin{cases} nt & \text{if } 0 \leq t \leq \dfrac{1}{n}, \\[2ex] 2 - nt & \text{if } \dfrac{1}{n} \leq t \leq \dfrac{2}{n}, \\[2ex] 0 & \text{if } \dfrac{2}{n} \leq t \leq 1. \end{cases} \tag{3.5}$$

Since x_n does not converge to 0 uniformly (that is, in the topology of C), P_n cannot converge weakly to P. (For example, if $A = S(0, \frac{1}{2})$, then $P(\partial A) = 0$, while $P_n(A) = 0$ does not converge to $P(A) = 1$.) But $P_n(A) \to P(A)$ does hold for finite-dimensional P-continuity sets—in fact, if $A = \pi_{t_1\cdots t_k}^{-1} H$ and if $2/n$ is smaller than the least of the nonzero t_i, then $P_n(A) = P(A)$.

The finite-dimensional sets are thus a determining class in C but not a convergence-determining class. The difficulty, interest, and usefulness of weak convergence in C all spring from the fact that it involves considerations going beyond those of finite-dimensional sets.

Product Spaces

Let $S = S' \times S''$ be the product of metric spaces S' and S''. If S is separable (which requires that S' and S'' be separable), then the σ-fields $\mathscr{S}$, $\mathscr{S}'$, and $\mathscr{S}''$ of Borel sets in these spaces are related by $\mathscr{S} = \mathscr{S}' \times \mathscr{S}''$ (see p. 225).

The two marginal distributions of a probability measure P on $(S, \mathscr{S})$ are defined by $P'(A') = P(A' \times S'')$, $A' \in \mathscr{S}'$, and $P''(A'') = P(S' \times A'')$, $A'' \in \mathscr{S}''$.

THEOREM 3.1 *If S is separable, then a necessary and sufficient condition for $P_n \Rightarrow P$ is that $P_n(A' \times A'') \to P(A' \times A'')$ for each P'-continuity set A' and each P''-continuity set A'', where P' and P'' are the marginal distributions of P.*

Proof. Let ∂, ∂', and ∂'' denote the boundary operators in S, S', and S'', respectively. Since

$$\partial(A' \times A'') \subset \big((\partial' A') \times S''\big) \cup \big(S' \times (\partial'' A'')\big), \tag{3.6}$$

the condition is necessary.

To prove sufficiency, we apply Corollary 1 of Theorem 2.2 to the class $\mathscr{U}$ of sets $A' \times A''$ with A' a P'-continuity set and A'' a P''-continuity set. The class $\mathscr{U}$ is closed under the formation of finite intersections and, by hypothesis, $P_n(A) \to P(A)$ for A in $\mathscr{U}$.

Given (x', x'') in S and $\varepsilon > 0$, consider the sets

$$A_\delta = \{y' : \rho'(x', y') < \delta\} \times \{y'' : \rho''(x'', y'') < \delta\}.$$

For distinct δ, the sets $\partial'\{y' : \rho'(x', y') < \delta\}$ are disjoint and the sets $\partial''\{y'' : \rho''(x'', y'') < \delta\}$ are disjoint; therefore A_δ lies in $\mathscr{U}$ for some δ with $0 < \delta < \varepsilon$. If S is metrized by

$$\rho\big((x', x''), (y', y'')\big) = \max\{\rho'(x', y'), \rho''(x'', y'')\},$$

then A_δ is just the sphere with center (x', x'') and radius δ. Hence $\mathscr{U}$ satisfies the hypotheses of Corollary 1 of Theorem 2.2, as required.

The sufficiency of the condition in Theorem 3.1 implies that the measurable rectangles form a convergence-determining class (but says more: $P\big(\partial(A' \times A'')\big) = 0$ does not imply $P'(\partial' A') = P''(\partial'' A'') = 0\big)$.

For given probability measures P' and P'' on $(S', \mathscr{S}')$ and $(S'', \mathscr{S}'')$, the product measure $P' \times P''$ is a probability measure on $\mathscr{S}' \times \mathscr{S}''$ and hence, if S is separable, on $\mathscr{S}$. The following theorem, in which P_n' and P' are probability measures on $(S', \mathscr{S}')$ and P_n'' and P'' are probability measures on $(S'', \mathscr{S}'')$, is an immediate consequence of Theorem 3.1.

THEOREM 3.2 *If S is separable, then $P_n' \times P_n'' \Rightarrow P' \times P''$ if and only if $P_n' \Rightarrow P'$ and $P_n'' \Rightarrow P''$.*

PROBLEMS

1. Show directly that $F_n \Rightarrow F$ if and only if $F_n(x) \to F(x)$ for all x in a dense set and that $F_n \Rightarrow F$ if and only if $\limsup_n F_n(x) \le F(x)$ and $\liminf_n F_n(x - 0) \ge F(x - 0)$ for all x (here $F(x - 0) = \sup_{y<x} F(y)$).

2. If $k > 1$, the set of discontinuities of F, although having dense complement, need not be countable. A $(k - 1)$-dimensional hyperplane can contain at most countably many discontinuities if it is normal to none of the axes. [To see the problem, consider first the hyperplane $\{(x_1, x_2) : x_1 = -x_2\}$ in R^2.]

3. If $F_n \Rightarrow F$ and if F is continuous at each point of a closed set A, then

$$\sup_{x\in A} |F_n(x) - F(x)| \to 0.$$

4. The Lévy distance $\lambda(F, G)$ between two one-dimensional distribution functions is the infimum of those positive ε such that $F(x - \varepsilon) - \varepsilon \le G(x) \le F(x + \varepsilon) + \varepsilon$ for all x.

Interpret $\lambda(F, G)$ geometrically in terms of the graphs of F and G. Show that $F_n \Rightarrow F$ if and only if $\lambda(F_n, F) \to 0$; prove that the collection of one-dimensional distribution functions is a separable, complete metric space under λ.

5. Problem 5 in Section 1 adapts the problem preceding it to the general Banach space of countably infinite dimension. In C a simple direct analysis is possible: Work with the functions x_n defined by (3.5).

6. The uniform distribution on the unit square and the uniform distribution on its diagonal have identical marginal distributions. Relate to Theorem 3.2.

7. Extend the product-space theory at the end of the section by showing that, for a countable product of separable spaces, the finite-dimensional sets (appropriately defined) form a convergence-determining class (R^∞ is a special case).

4. CONVERGENCE IN DISTRIBUTION

The theory of weak convergence can be paraphrased as the theory of convergence in distribution. When stated in the terminology of the latter theory, which involves no new ideas, many results assume a compact and perspicuous form.

Random Elements

Let X be a mapping from a probability space $(\Omega, \mathscr{B}, \mathsf{P})$ into a metric space S. If X is measurable (in the sense that $X^{-1}\mathscr{S} \subset \mathscr{B}$; see p. 222), we call it a *random element*. We shall say X is defined *on* its domain Ω (or $(\Omega, \mathscr{B}, \mathsf{P})$) and *in* its range S and call it a random element *of* S. If $S = R^1$, we call X a random variable; if $S = R^k$, we call X a random vector; if $S = C$, we call X a random function.

Random variables and random vectors are familiar objects, and the Introduction contains (see formula (13) there) an example of a useful random function (although its measurability was not proved). A variety of random functions arise in a natural way in probability theory.

The *distribution*† of X is the probability measure $P = \mathsf{P}X^{-1}$ on $(S, \mathscr{S})$:

$$(4.1)\qquad P(A) = \mathsf{P}(X^{-1}A) = \mathsf{P}\{\omega: X(\omega) \in A\} = \mathsf{P}\{X \in A\}, \qquad A \in \mathscr{S}.$$

(We generally suppress the argument ω in this way.) In case $S = R^k$, we have also the associated *distribution function* of X, defined by

$$F(x) = P\{y: y \le x\} = \mathsf{P}\{X \le x\}, \qquad x \in R^k.$$

Note that P is a probability measure on a space of an arbitrary nature, whereas P is always defined on a metric space. For many questions, the distribution P contains all relevant information about the random element

† This has nothing to do with the distributions of Schwartz.

X. If h is a measurable function on S ($h^{-1}\mathscr{R}^1 \subset \mathscr{S}$), then, by the change-of-variable formula (see p. 223),

$$\int h(X)\, d\mathsf{P} = \int h\, dP, \tag{4.2}$$

in the sense that the two integrals exist or fail to exist together and have the same value if they do exist. In the usual expected-value notation, (4.2) becomes

$$\mathsf{E}\{h(X)\} = \int h\, dP. \tag{4.3}$$

Each probability measure on each metric space is the distribution of some random element on some probability space: Given P on $(S, \mathscr{S})$, if we take $(\Omega, \mathscr{B}, \mathsf{P}) = (S, \mathscr{S}, P)$, and if we take X to be the identity,

$$X(\omega) = \omega, \qquad \omega \in \Omega = S, \tag{4.4}$$

then X is a random element on Ω with values in S and has P as its distribution. Although the class of distributions thus coincides with the class of probability measures on metric spaces, we generally call a measure on a metric space a distribution only when it is indeed the distribution of some random element already under discussion.

Convergence in Distribution

We say a sequence $\{X_n\}$ of random elements *converges in distribution* to the random element X, and we write

$$X_n \xrightarrow{\mathscr{D}} X, \tag{4.5}$$

if the distributions P_n of the X_n converge weakly to the distribution P of X:

$$P_n \Rightarrow P. \tag{4.6}$$

Although this definition of course makes no sense unless the image space S (the range) and the topology on it are the same for all the random elements $X, X_1, X_2, \ldots$, the underlying probability spaces (the domains) may be all distinct. We ordinarily make no mention of these underlying spaces because their structures enter into the argument only by way of the distributions on S they induce. Thus we write $\mathsf{P}\{X_n \in A\}$ where we should write $\mathsf{P}_n\{X_n \in A\}$, and we write $\mathsf{E}\{f(X_n)\}$ where we should write $\int f(X_n)\, d\mathsf{P}_n$ or perhaps $\mathsf{E}_n\{f(X_n)\}$. Since $\int_S f(x)P(dx) = \int_\Omega f(X)\, d\mathsf{P}$ by the change-of-variable formula (4.3) and similarly for $\int f\, dP_n$, we have $X_n \xrightarrow{\mathscr{D}} X$ if and only if $\mathsf{E}\{f(X_n)\} \to \mathsf{E}\{f(X)\}$ for every $f \in C(S)$.

Theorem 2.1 asserts the equivalence of the following five statements. Call a set A in $\mathscr{S}$ an X-continuity set if $\mathsf{P}\{X \in \partial A\} = 0$.

(i) $X_n \xrightarrow{\mathscr{D}} X$.
(ii) $\lim_n \mathsf{E}\{f(X_n)\} = \mathsf{E}\{f(X)\}$ *for all bounded, uniformly continuous real* f.
(iii) $\limsup_n \mathsf{P}\{X_n \in F\} \leq \mathsf{P}\{X \in F\}$ *for all closed* F.
(iv) $\liminf_n \mathsf{P}\{X_n \in G\} \geq \mathsf{P}\{X \in G\}$ *for all open* G.
(v) $\lim_n \mathsf{P}\{X_n \in A\} = \mathsf{P}\{X \in A\}$ *for all X-continuity sets* A.

Each theorem about weak convergence can be similarly recast.

The following hybrid terminology is useful. If X_n are random elements of S, if P_n are the corresponding distributions, and if P is a probability measure on $(S, \mathscr{S})$, we say the X_n *converge in distribution* to P, and write

$$X_n \xrightarrow{\mathscr{D}} P, \tag{4.7}$$

in case $P_n \Rightarrow P$. There is the obvious corresponding version of Theorem 2.1.

It is a great convenience to be able to pass from one to another of the three equivalent concepts (4.5), (4.6), and (4.7), and we shall do so freely. This is largely a matter of expedient phraseology. For example, if random variables X_n have asymptotically a normal distribution with mean μ and variance σ^2, we shall express this fact by writing

$$X_n \xrightarrow{\mathscr{D}} N(\mu, \sigma^2). \tag{4.8}$$

It makes no difference whether we interpret $N(\mu, \sigma^2)$ as the measure on $(R^1, \mathscr{R}^1)$ with density $(2\pi\sigma^2)^{-\frac{1}{2}} e^{-(u-\mu)^2/2\sigma^2}$ and understand (4.8) in the sense of (4.7) or whether we interpret $N(\mu, \sigma^2)$ as a random variable with

$$\mathsf{P}\{N(\mu, \sigma^2) \in A\} = \frac{1}{\sqrt{2\pi}\,\sigma} \int_A e^{-(u-\mu)^2/2\sigma^2}\, du, \qquad A \in \mathscr{R}^1, \tag{4.9}$$

and understand (4.8) in the sense of (4.5). (There exist many such random variables $N(\mu, \sigma^2)$ on many probability spaces; the construction (4.4) gives one of them.) We shall restrict $N(\mu, \sigma^2)$ to this (dual) meaning, and we shall write N for $N(0, 1)$:

$$\mathsf{P}\{N \in A\} = \frac{1}{\sqrt{2\pi}} \int_A e^{-u^2/2}\, du, \qquad A \in \mathscr{R}^1. \tag{4.10}$$

Convergence in Probability

Many of the standard concepts and results for convergence in distribution of ordinary random variables generalize to random elements. If, for an element a of S,

$$\mathsf{P}\{\rho(X_n, a) \geq \varepsilon\} \to 0 \tag{4.11}$$

for each positive ε, we say X_n *converges in probability* to a and write

$$X_n \xrightarrow{P} a. \tag{4.12}$$

If a is conceived as a constant-valued random element, then, as is easily proved, $X_n \xrightarrow{P} a$ if and only if $X_n \xrightarrow{\mathscr{D}} a$. Alternatively, $X_n \xrightarrow{P} a$ if and only if the distribution of X_n converges weakly to the probability measure corresponding to a mass of 1 at the point a. The random elements X_n in (4.12) may, as usual, be defined on distinct probability spaces—only the range S need be common to them all.

If X_n and Y_n have a common domain, it makes sense to speak of the distance $\rho(X_n, Y_n)$—the function with value $\rho(X_n(\omega), Y_n(\omega))$ at ω. If S is separable, $\rho(X_n, Y_n)$ is a random variable (see p. 225). In the following theorem, we assume that, for each n, X_n and Y_n do have a common domain and that S is separable.

THEOREM 4.1 *If $X_n \xrightarrow{\mathscr{D}} X$ and $\rho(X_n, Y_n) \xrightarrow{P} 0$, then $Y_n \xrightarrow{\mathscr{D}} X$.*

Proof. If $F_\varepsilon = \{x: \rho(x, F) \leq \varepsilon\}$, then

$$\mathsf{P}\{Y_n \in F\} \leq \mathsf{P}\{\rho(X_n, Y_n) \geq \varepsilon\} + \mathsf{P}\{X_n \in F_\varepsilon\}.$$

Since F_ε is closed, the hypotheses imply

$$\limsup_n \mathsf{P}\{Y_n \in F\} \leq \limsup_n \mathsf{P}\{X_n \in F_\varepsilon\} \leq \mathsf{P}\{X \in F_\varepsilon\}.$$

If F is closed, then $F_\varepsilon \downarrow F$ as $\varepsilon \downarrow 0$ and the result follows by Theorem 2.1 (the version corresponding to convergence in distribution).

In the next theorem† we assume that, for each n, $Y_n, X_{1n}, X_{2n}, \ldots$ have a common domain and that S is separable.

THEOREM 4.2 *Suppose that, for each u, $X_{un} \xrightarrow{\mathscr{D}} X_u$ as $n \to \infty$ and that $X_u \xrightarrow{\mathscr{D}} X$ as $u \to \infty$. Suppose further that*

$$\lim_{u\to\infty} \limsup_{n\to\infty} \mathsf{P}\{\rho(X_{un}, Y_n) \geq \varepsilon\} = 0 \tag{4.13}$$

for each positive ε. Then $Y_n \xrightarrow{\mathscr{D}} X$ as $n \to \infty$.

Proof. Defining F_ε as before by $F_\varepsilon = \{x: \rho(x, F) \leq \varepsilon\}$, we have

$$\mathsf{P}\{Y_n \in F\} \leq \mathsf{P}\{X_{un} \in F_\varepsilon\} + \mathsf{P}\{\rho(X_{un}, Y_n) \geq \varepsilon\}.$$

By the hypothesis $X_{un} \xrightarrow{\mathscr{D}} X_u$ $(n \to \infty)$,

$$\limsup_{n\to\infty} \mathsf{P}\{Y_n \in F\} \leq \mathsf{P}\{X_u \in F_\varepsilon\} + \limsup_{n\to\infty} \mathsf{P}\{\rho(X_{un}, Y_n) \geq \varepsilon\}.$$

By (4.13) and the hypothesis $X_u \xrightarrow{\mathscr{D}} X$ $(u \to \infty)$,

$$\limsup_{n\to\infty} \mathsf{P}\{Y_n \in F\} \leq \mathsf{P}\{X \in F_\varepsilon\}.$$

The result follows as before.

---

† The remainder of this section is not central to the theory; after a cursory reading, it can be consulted as the need arises.

Suppose now that $X, X_1, X_2, \ldots$ all have a common domain and that S is separable. If

$$\mathsf{P}\{\rho(X_n, X) \geq \varepsilon\} \to 0$$

for every positive ε, we say that X_n converges in probability to X and write

$$X_n \xrightarrow{P} X. \tag{4.14}$$

Because of the assumptions here of separability and (more important) of a common domain for the X_n, this concept does not generalize (4.12).

THEOREM 4.3 *If* $X_n \xrightarrow{P} X$, *then*†

$$\mathsf{P}(\{X_n \in A\} + \{X \in A\}) \to 0 \tag{4.15}$$

for every X-continuity set A.

Proof. For each positive ε,

$$\mathsf{P}\{X_n \in A, X \notin A\} \leq \mathsf{P}\{\rho(X_n, X) \geq \varepsilon\} + \mathsf{P}\{\rho(X, A) < \varepsilon, X \notin A\}.$$

From this, the same inequality with A^c in place of A, and the assumption $X_n \xrightarrow{P} X$, we obtain

$$\limsup_{n\to\infty} \mathsf{P}(\{X_n \in A\} + \{X \in A\}) \\ \leq \mathsf{P}\{\rho(X, A) < \varepsilon, X \notin A\} + \mathsf{P}\{\rho(X, A^c) < \varepsilon, X \in A\}.$$

As $\varepsilon \to 0$, the right-hand member of this inequality tends to $\mathsf{P}\{X \in \partial A\} = 0$.

Since (4.15) implies $\mathsf{P}\{X_n \in A\} \to \mathsf{P}\{X \in A\}$, it follows from Theorem 4.3 that $X_n \xrightarrow{P} X$ implies $X_n \xrightarrow{\mathscr{D}} X$.

Product Spaces‡

Let X' and X'_n be random elements of S', and let X'' and X''_n be random elements of S''. In the rest of this section we assume that X' and X'' have the same domain, that X'_n and X''_n have the same domain for each n, and that S' and S'' are separable, so that (X', X'') and (X'_n, X''_n) are random elements of $S' \times S''$ (see p. 225). We seek conditions under which

$$(X'_n, X''_n) \xrightarrow{\mathscr{D}} (X', X''). \tag{4.16}$$

The random elements X' and X'' are by definition independent if $\mathsf{P}\{X' \in A', X'' \in A''\} = \mathsf{P}\{X' \in A'\}\,\mathsf{P}\{X'' \in A''\}$. If X' and X'' are independent and X'_n and X''_n are independent for each n, then, by Theorem 3.2, (4.16) is

† *Some notation:* E^c is the complement of E, $E_1 - E_2 = E_1 \cap E_2{}^c$ is the difference of E_1 and E_2, and $E_1 + E_2 = (E_1 - E_2) \cup (E_2 - E_1)$ is the symmetric difference of E_1 and E_2.
‡ The next-to-last footnote still applies.

equivalent to

$$X'_n \xrightarrow{\mathscr{D}} X', \qquad X''_n \xrightarrow{\mathscr{D}} X''. \tag{4.17}$$

Although (4.16) implies (4.17) even without independence, the converse fails—take (X', X'') distributed uniformly on the unit square and (X'_n, X''_n) distributed uniformly on its diagonal. We shall be concerned with cases in which X' and X'' are independent but X'_n and X''_n satisfy some condition weaker than independence. If they satisfy (4.16) with independent X' and X'', then it is natural to regard X'_n and X''_n as asymptotically independent.

By Theorem 3.1, (4.16) holds if and only if

$$\mathsf{P}\{X'_n \in A', X''_n \in A''\} \to \mathsf{P}\{X' \in A', X'' \in A''\} \tag{4.18}$$

for all X'-continuity sets A' and all X''-continuity sets A''.

THEOREM 4.4 *If $X'_n \xrightarrow{\mathscr{D}} X'$ and $X''_n \xrightarrow{P} a''$, then $(X'_n, X''_n) \xrightarrow{\mathscr{D}} (X', a'')$.*

Proof. We must verify (4.18) with X'' identically equal to a''. Suppose that A' is an X'-continuity set and that A'' is an X''-continuity set (that is, $a'' \notin \partial A''$). If $a'' \in A''$, then $\mathsf{P}\{X''_n \notin A''\} \to 0$, and (4.18) follows from $X'_n \xrightarrow{\mathscr{D}} X'$ and

$$\mathsf{P}\{X'_n \in A'\} - \mathsf{P}\{X''_n \notin A''\} \le \mathsf{P}\{X'_n \in A', X''_n \in A''\} \le \mathsf{P}\{X'_n \in A'\}.$$

If $a'' \notin A''$, then (4.18) follows from

$$\mathsf{P}\{X'_n \in A', X''_n \in A''\} \le \mathsf{P}\{X''_n \in A''\} \to 0.$$

In the next theorem we assume that X''_n converges in probability to a random element Y'' that is not necessarily constant, which requires that Y'' and all the (X'_n, X''_n) have the same domain $(\Omega, \mathscr{B}, \mathsf{P})$. Let $\mathscr{B}_0$ be a (finitely additive) field contained in $\mathscr{B}$, and denote by $\sigma(\mathscr{B}_0)$ the σ-field generated by $\mathscr{B}_0$.

THEOREM 4.5 *Suppose that X' and X'' are independent and that X'' has the same distribution as Y''. If $X''_n \xrightarrow{P} Y''$, if*

$$\mathsf{P}(\{X'_n \in A'\} \cap E) \to \mathsf{P}\{X' \in A'\}\, \mathsf{P}(E) \tag{4.19}$$

for each X'-continuity set A' and each set E in the field $\mathscr{B}_0$, and if each X'_n is measurable $\sigma(\mathscr{B}_0)$, then $(X'_n, X''_n) \xrightarrow{\mathscr{D}} (X', X'')$.

Note that (4.19) implies $X'_n \xrightarrow{\mathscr{D}} X'$ (take $E = \Omega$). Note also that the domain of (X', X'') need not be that common to Y'' and the (X'_n, X''_n). We may replace (X', X'') in the conclusion by (Y', Y'') if the domain of Y'' supports some random element Y' that is independent of Y'' and has the proper distribution; but it would be an unnecessary restriction to assume in general the existence of such a Y'.

Proof. Fix an X'-continuity set A' and an X''-continuity set A''. We are to prove (4.18), which, since X' and X'' are independent and X'' has the distribution of Y'', is the same thing as

$$\mathsf{P}\{X'_n \in A',\, X''_n \in A''\} \to \mathsf{P}\{X' \in A'\}\,\mathsf{P}\{Y'' \in A''\}. \tag{4.20}$$

Since $X''_n \xrightarrow{P} Y''$, it follows by Theorem 4.3 that (4.20) is in turn the same as

$$\mathsf{P}\{X'_n \in A',\, Y'' \in A''\} \to \mathsf{P}\{X' \in A'\}\,\mathsf{P}\{Y'' \in A''\}. \tag{4.21}$$

Write $E_n = \{X'_n \in A'\}$ and $\alpha = \mathsf{P}\{X' \in A'\}$, and let g denote the indicator of the set $\{Y'' \in A''\}$. Then (4.21) takes the form

$$\int_{E_n} g\, d\mathsf{P} \to \alpha \int g\, d\mathsf{P}. \tag{4.22}$$

Since each X'_n is measurable $\sigma(\mathscr{B}_0)$, each E_n lies in $\sigma(\mathscr{B}_0)$.

We shall prove that (4.22) holds if g is an *arbitrary* integrable function (measurable $\mathscr{B}$). If g is the indicator of a set in $\mathscr{B}_0$, then (4.22) holds because it is the same thing as (4.19). Clearly (4.22) then also holds if g is a simple function measurable $\mathscr{B}_0$. If g is integrable and measurable $\sigma(\mathscr{B}_0)$, then, for each positive ε, there is a simple function g_ε, measurable $\mathscr{B}_0$, with $\mathsf{E}\{|g - g_\varepsilon|\} < \varepsilon$; but then

$$\limsup_{n\to\infty} \left| \int_{E_n} g\, d\mathsf{P} - \alpha \int g\, d\mathsf{P} \right| \le (1 + |\alpha|)\, \mathsf{E}\{|g - g_\varepsilon|\},$$

so that (4.22) follows for all such g.

Finally, suppose that g is measurable $\mathscr{B}$ and integrable but not necessarily measurable $\sigma(\mathscr{B}_0)$. By the properties of conditional expected values† we still have, since $E_n \in \sigma(\mathscr{B}_0)$,

$$\int_{E_n} g\, d\mathsf{P} = \int_{E_n} \mathsf{E}\{g \,\|\, \sigma(\mathscr{B}_0)\}\, d\mathsf{P} \to \alpha \int \mathsf{E}\{g \,\|\, \sigma(\mathscr{B}_0)\}\, d\mathsf{P} = \alpha \int g\, d\mathsf{P}.$$

Thus (4.22) does hold for all integrable g, as required.

Remarks. The ideas in the proof of Theorem 4.5 come from Rényi (1958).

PROBLEMS

1. Prove for random variables that, if $X_n \xrightarrow{\mathscr{D}} X$ and $Y_n \xrightarrow{P} 0$, then $X_n + Y_n \xrightarrow{\mathscr{D}} X$ and $X_n Y_n \xrightarrow{P} 0$.

2. Prove for random vectors X_n, Y_n, X and random variables Z_n that (a) if $X_n \xrightarrow{\mathscr{D}} X$, $|X_n - Y_n| \le Z_n\,|X_n|$, and $Z_n \xrightarrow{P} 0$, then $Y_n \xrightarrow{\mathscr{D}} X$ and that (b) if $X_n \xrightarrow{\mathscr{D}} X$, $|X_n - Y_n| \le Z_n\,|Y_n|$, and $Z_n \xrightarrow{P} 0$, then $Y_n \xrightarrow{\mathscr{D}} X$. [Reduce (b) to (a) via the fact that $|x - y| \le \varepsilon\,|y|$ with $\varepsilon < \frac{1}{2}$ implies $|x - y| \le 2\varepsilon\,|x|$.]

---

† See Doob (1953) or Billingsley (1965). The central results in this book do not require conditional probabilities and expected values.

3. Three fair coins are tossed independently. Let E_{ij} be the event that coins i and j show the same face, let $X' = X'_n = X''_n = Y''$ be the indicator of E_{13}, let X'' be the indicator of E_{12}, and let $\mathscr{B}_0 = \{E_{12}, E_{23}\}$. The conclusion of Theorem 4.5 fails, although its hypotheses are satisfied except for the requirement that $\mathscr{B}_0$ be a field.

4. Let $X_1, X_2, \ldots$ be independent and have a common distribution P on S. Let $P_{n,\omega}$ be the empirical measure for $X_1(\omega), \ldots, X_n(\omega)$; $P_{n,\omega}(A)$ is the fraction of k, $1 \le k \le n$, for which $X_k(\omega) \in A$:

$$P_{n,\omega}(A) = \frac{1}{n} \sum_{k=1}^{n} I_A(X_k(\omega)).$$

Show that, if S is separable, then $P_{n,\omega} \Rightarrow P$ with probability 1. [Use the strong law of large numbers for Bernoulli trials and Corollary 1 to Theorem 2.2.] (This result is due to Varadarajan (1958b); see Ranga Rao (1962) for extensions.)

5. Show that random variables X_n and X satisfy $X_n \xrightarrow{\mathscr{D}} X$ if and only if

$$\mathsf{E}\{F(X_n)\} \to \mathsf{E}\{F(X)\} \tag{4.22}$$

for every continuous distribution function F [turn (1.1) around]. If Y has the distribution function F and is independent of X and all the X_n (take them all to have the same domain), then (4.22) is the same as $\mathsf{P}\{Y \le X_n\} \to \mathsf{P}\{Y \le X\}$.

6. For a probability measure P on $(S, \mathscr{S})$, (4.4) shows how to construct on a probability space $(\Omega, \mathscr{B}, \mathsf{P})$ a random element X with distribution P. If S is separable and complete, we can take P as Lebesgue measure on the Borel sets $\mathscr{B}$ of the unit interval Ω. [Let $\mathscr{A}_k = \{A_{ku}\}$ be a decomposition of S into P-continuity sets of diameter less than $1/k$ and let $\mathscr{I}_k = \{I_{ku}\}$ be a decomposition of Ω into subintervals with lengths $\mathsf{P}(I_{ku}) = P(A_{ku})$; arrange that $\mathscr{A}_{k+1}$ refines $\mathscr{A}_k$ and $\mathscr{I}_{k+1}$ refines $\mathscr{I}_k$. For $\omega \in I_{ku}$, give $X_k(\omega)$ some value in A_{ku}, show that $\{X_k(\omega)\}$ is fundamental for each ω, and use Theorem 4.1 to show that $X(\omega) = \lim_k X_k(\omega)$ has distribution P.] (Skorohod (1956) has the stronger result that, if $P_n \Rightarrow P$, random elements X_n and X with these distributions can be constructed on the unit interval in such a way that $X_n(\omega) \to X(\omega)$ for each ω.)

7. Show that $(X'_n, X''_n) \xrightarrow{\mathscr{D}} (X', X'')$ if X' and X'' are independent, $X''_n \xrightarrow{P} Y''$, where Y'' has the distribution of X'', and (4.19) holds for each E in the σ-field generated by Y''.

5. WEAK CONVERGENCE AND MAPPINGS

Continuous Mappings

If h is a measurable mapping of S into another metric space S' (with metric ρ' and σ-field $\mathscr{S}'$ of Borel sets), then each probability measure P on $(S, \mathscr{S})$ induces on $(S', \mathscr{S}')$ a unique probability measure Ph^{-1}, defined by $Ph^{-1}(A) = P(h^{-1}A)$ for $A \in \mathscr{S}'$. We need conditions under which $P_n \Rightarrow P$ implies $P_nh^{-1} \Rightarrow Ph^{-1}$. One such condition is that h be continuous, since then $f(h(x))$ is bounded and continuous on S whenever $f(y)$ is bounded and continuous on S', so that $P_n \Rightarrow P$ implies $\int f(h(x))P_n(dx) \to \int f(h(x))P(dx)$, a relation which, upon transformation of the integrals (see p. 223), becomes

$$\int f(y)P_nh^{-1}(dy) \to \int f(y)Ph^{-1}(dy).$$

For example, the natural projection π_k from R^∞ to R^k is continuous, so that $P_n \Rightarrow P$ implies $P_n\pi_k^{-1} \Rightarrow P\pi_k^{-1}$ for each k. Let us show that, conversely, if $P_n\pi_k^{-1} \Rightarrow P\pi_k^{-1}$ for each k, then $P_n \Rightarrow P$. From the continuity of π_k it follows easily that $\partial\pi_k^{-1}H \subset \pi_k^{-1}\,\partial H$ for $H \subset R^k$. Using special properties of π_k we shall prove that there is inclusion in the other direction. If $x \in \pi_k^{-1}\,\partial H$, so that $\pi_k x \in \partial H$, then there are points $\alpha^{(u)}$ in H and points $\beta^{(u)}$ in H^c such that $\alpha^{(u)} \to \pi_k x$ and $\beta^{(u)} \to \pi_k x$ $(u \to \infty)$. Since the points $(\alpha_1^{(u)}, \ldots, \alpha_k^{(u)}, x_{k+1}, \ldots)$ lie in $\pi_k^{-1}H$ and converge to x, and since the points $(\beta_1^{(u)}, \ldots, \beta_k^{(u)}, x_{k+1}, \ldots)$ lie in $(\pi_k^{-1}H)^c$ and also converge to x, $x \in \partial(\pi_k^{-1}H)$. Thus $\partial\pi_k^{-1}H = \pi_k^{-1}\,\partial H$. If $P_n\pi_k^{-1} \Rightarrow P\pi_k^{-1}$, then $P_n(A) \to P(A)$ for sets $A = \pi_k^{-1}H$ with $H \in \mathscr{R}^k$ and $P(\pi_k^{-1}\,\partial H) = 0$. Since $P(\pi_k^{-1}\,\partial H) = 0$ is equivalent to $P(\partial\pi_k^{-1}H) = 0$, $P_n(A) \to P(A)$ holds for all finite-dimensional P-continuity sets, and hence, since the finite-dimensional sets form a convergence-determining class, $P_n \Rightarrow P$.

We call the $P\pi_k^{-1}$ the *finite-dimensional distributions* or *measures* corresponding to P. We have shown that probability measures on $(R^\infty, \mathscr{R}^\infty)$ converge weakly if and only if all the corresponding finite-dimensional distributions converge weakly.

The finite-dimensional distributions of a probability measure P on $(C, \mathscr{C})$ we define as the various measures $P\pi_{t_1\cdots t_k}^{-1}$, where the $\pi_{t_1\cdots t_k}$ are the projections defined in Section 3. Since these projections are continuous, the weak convergence of probability measures on $(C, \mathscr{C})$ implies the weak convergence of the corresponding finite-dimensional distributions. But the converse fails because, as was shown by counterexample, the class of finite-dimensional sets is not convergence determining. Indeed, if P $[P_n]$ is a unit mass at 0 [the function (3.5)], then P_n does not converge weakly to P, even though $P_n\pi_{t_1\cdots t_k}^{-1} \Rightarrow P\pi_{t_1\cdots t_k}^{-1}$ holds for all sets $(t_1, \ldots, t_k)$. On the other hand, since the finite-dimensional sets form a determining class, a probability measure on $(C, \mathscr{C})$ is uniquely determined by its finite-dimensional distributions.

Main Theorem

We have seen that $P_n \Rightarrow P$ implies $P_nh^{-1} \Rightarrow Ph^{-1}$ if h is a continuous mapping of S into S', but we can weaken the continuity assumption. Assume only that h is measurable and let D_h be the set of discontinuities of h. Then $D_h \in \mathscr{S}$ (even if h is not measurable; see p. 225).

THEOREM 5.1 *If $P_n \Rightarrow P$ and $P(D_h) = 0$, then $P_nh^{-1} \Rightarrow Ph^{-1}$.*

Proof. We shall show that, if F is a closed subset of S', then

$$\limsup_{n\to\infty} P_nh^{-1}(F) \le Ph^{-1}(F).$$

Since $P_n \Rightarrow P$, we have

$$\limsup_n P_n(h^{-1}F) \leq \limsup_n P_n\big((h^{-1}F)^-\big) \leq P\big((h^{-1}F)^-\big).$$

Hence it suffices to prove $P\big((h^{-1}F)^-\big) = P(h^{-1}F)$, which follows from the assumption $P(D_h) = 0$ and the fact that $(h^{-1}F)^- \subset D_h \cup (h^{-1}F)$.

There are two immediate corollaries. For a random element X of S, $h(X)$ is a random element of S' (we still assume h measurable).

COROLLARY 1 *If* $X_n \xrightarrow{\mathscr{D}} X$ *and* $\mathsf{P}\{X \in D_h\} = 0$, *then* $h(X_n) \xrightarrow{\mathscr{D}} h(X)$.

COROLLARY 2 *If* $X_n \xrightarrow{P} a$ *and if* h *is continuous at* a, *then* $h(X_n) \xrightarrow{P} h(a)$.

For example, for ordinary random variables, $(X_n, Y_n) \xrightarrow{\mathscr{D}} (X, Y)$ implies $X_n + Y_n \xrightarrow{\mathscr{D}} X + Y$. Facts of this sort, used constantly in probability and statistics, all depend ultimately on Theorem 5.1.

Particular interest attaches to Theorem 5.1 when S' is the line, in which case h is an ordinary real, measurable function.

THEOREM 5.2 (i) *If* $P_n \Rightarrow P$, *then* $P_nh^{-1} \Rightarrow Ph^{-1}$ *for every real, measurable function* h *for which* $P(D_h) = 0$.

(ii) *If* $P_nh^{-1} \Rightarrow Ph^{-1}$ *for all bounded, continuous real* h, *then* $P_n \Rightarrow P$.

(iii) *If* $P_n \Rightarrow P$ *and* h *is a real, bounded, measurable function with* $P(D_h) = 0$, *then* $\int h\, dP_n \to \int h\, dP$.

Proof. Part (i) follows from Theorem 5.1. If $P_nh^{-1} \Rightarrow Ph^{-1}$, then, by change of variable, $\int f(h(x))P_n(dx) \to \int f(h(x))P(dx)$ for each f in $C(R^1)$. If h is bounded by M, then, taking

$$f(t) = \begin{cases} -M & \text{if } t \leq -M, \\ t & \text{if } -M \leq t \leq M, \\ M & \text{if } M \leq t, \end{cases}$$

we see that $\int h\, dP_n \to \int h\, dP$. Thus part (iii) is a consequence of part (i) and part (ii) follows by the definition of weak convergence. (We can strengthen (ii) by requiring h in the hypothesis to be uniformly continuous.)

In applications we are generally interested in establishing weak convergence in various metric spaces in order to be able, by applying part (i) of this theorem, to prove the weak convergence of measures induced on the line by various real functions h.

Integration to the Limit

In part (iii) of Theorem 5.2, the restriction that h be bounded can be relaxed. It is simplest to discuss this problem in terms of random variables X_n and X having distributions P_nh^{-1} and Ph^{-1}.

THEOREM 5.3 *If $X_n \xrightarrow{\mathscr{D}} X$, then* $\mathsf{E}\{|X|\} \le \liminf_n \mathsf{E}\{|X_n|\}$.

Proof. Take

$$h(x) = \begin{cases} |x| & \text{for } |x| \le \alpha \\ 0 & \text{for } |x| > \alpha \end{cases}$$

in part (iii) of Theorem 5.2; if $\mathsf{P}\{|X| = \alpha\} = 0$, then

$$\int_{\{|X| \le \alpha\}} |X| \, d\mathsf{P} = \lim_{n \to \infty} \int_{\{|X_n| \le \alpha\}} |X_n| \, d\mathsf{P} \le \liminf_{n \to \infty} \mathsf{E}\{|X_n|\}.$$

The result follows if we let α tend to infinity through values satisfying $\mathsf{P}\{|X| = \alpha\} = 0$.

The variables X_n are said to be *uniformly integrable* if

$$\lim_{\alpha \to \infty} \sup_n \int_{\{|X_n| \ge \alpha\}} |X_n| \, d\mathsf{P} = 0.$$

If the X_n are uniformly integrable, then

$$\sup_n \mathsf{E}\{|X_n|\} < \infty. \tag{5.1}$$

On the other hand, if

$$\sup_n \mathsf{E}\{|X_n|^{1+\varepsilon}\} < \infty$$

for some positive ε, then the X_n are uniformly integrable because in this case

$$\int_{\{|X_n| \ge \alpha\}} |X_n| \, d\mathsf{P} \le \frac{1}{\alpha^{\varepsilon}} \mathsf{E}\{|X_n|^{1+\varepsilon}\}.$$

The X_n are also uniformly integrable if there is a random variable Y such that $\mathsf{E}\{|Y|\} < \infty$ and

$$\mathsf{P}\{|X_n| \ge \alpha\} \le \mathsf{P}\{|Y| \ge \alpha\}, \qquad n \ge 1, \qquad \alpha > 0, \tag{5.2}$$

because in this case (see (3) on p. 223)

$$\int_{\{|X_n| \ge \alpha\}} |X_n| \, d\mathsf{P} \le \int_{\{|Y| \ge \alpha\}} |Y| \, d\mathsf{P}.$$

THEOREM 5.4 *Suppose $X_n \xrightarrow{\mathscr{D}} X$. If the X_n are uniformly integrable, then*

$$\mathsf{E}\{X_n\} \to \mathsf{E}\{X\}; \tag{5.3}$$

if X and the X_n are nonnegative and integrable, then (5.3) *implies that the X_n are uniformly integrable.*

Proof. If the X_n are uniformly integrable, the integrability of X follows from (5.1) and Theorem 5.3. Take

$$h_\alpha(x) = \begin{cases} x & \text{if } |x| < \alpha, \\ 0 & \text{if } |x| \ge \alpha, \end{cases}$$

in part (iii) of Theorem 5.2. Since $X_n \xrightarrow{\mathscr{D}} X$,

$$\mathsf{E}\{h_\alpha(X_n)\} \to \mathsf{E}\{h_\alpha(X)\} \tag{5.4}$$

if $\mathsf{P}\{|X| = \alpha\} = 0$. But

$$\mathsf{E}\{X_n\} - \mathsf{E}\{h_\alpha(X_n)\} = \int_{\{|X_n| \geq \alpha\}} X_n \, d\mathsf{P} \tag{5.5}$$

and

$$\mathsf{E}\{X\} - \mathsf{E}\{h_\alpha(X)\} = \int_{\{|X| \geq \alpha\}} X \, d\mathsf{P}. \tag{5.6}$$

Since these three relations imply

$$\limsup_{n\to\infty} |\mathsf{E}\{X_n\} - \mathsf{E}\{X\}| \leq \sup_n \int_{\{|X_n| \geq \alpha\}} |X_n| \, d\mathsf{P} + \int_{\{|X| \geq \alpha\}} |X| \, d\mathsf{P},$$

(5.3) does follow from uniform integrability.

On the other hand, if X and the X_n are integrable and (5.3) holds, it follows by (5.4), (5.5), and (5.6) that

$$\int_{\{|X_n| \geq \alpha\}} X_n \, d\mathsf{P} \to \int_{\{|X| \geq \alpha\}} X \, d\mathsf{P} \tag{5.7}$$

if $\mathsf{P}\{|X| = \alpha\} = 0$. Since X is integrable, we can, given ε, choose α so the right side of (5.7) is less than ε. But then the left side is less than ε for all n exceeding some n_0. Since the X_n are nonnegative and individually integrable, uniform integrability follows.

Since $X_n \xrightarrow{\mathscr{D}} X$ implies $X_n{}^r \xrightarrow{\mathscr{D}} X^r$, Theorems 5.3 and 5.4 immediately extend to moments higher than the first.

In none of this need X and the X_n (and the Y in (5.2)) have a common domain. If they do, and if $X_n(\omega) \to X(\omega)$ for almost all ω, or if $X_n \xrightarrow{P} X$, then $X_n \xrightarrow{\mathscr{D}} X$. Thus Theorem 5.3 contains Fatou's lemma and Theorem 5.4 contains the standard theorems about integration to the limit (the case involving (5.2) contains Lebesgue's dominated convergence theorem).

An Extension of Theorem 5.1†

Let h_n and h be measurable mappings from S to S'. We may ask whether $P_n \Rightarrow P$ implies $P_n h_n^{-1} \Rightarrow Ph^{-1}$ when h_n converges to h in some sense. Let E be the set of x such that $h_n x_n \to hx$ fails to hold for some sequence $\{x_n\}$ approaching x. (If h_n is identically equal to h, then $E = D_h$.) It is not hard to show that $x \in E^c$ if and only if for every ε there exist a k and a δ such that $i \geq k$ and $\rho(x, y) < \delta$ together imply $\rho'(hx, h_i y) < \varepsilon$. Assume that E lies in $\mathscr{S}$ (which is shown on p. 226 to hold automatically if S' is separable).

† This result is not used in the sequel.

THEOREM 5.5 *If $P_n \Rightarrow P$ and $P(E) = 0$, then $P_n h_n^{-1} \Rightarrow Ph^{-1}$.*

Proof. We shall show that $P(h^{-1}G) \leq \liminf_n P_n(h_n^{-1}G)$ for open subsets G of S'. From the above characterization of the points of E^c, it follows that, if $x \in E^c$ and if hx lies in the open set G, then there exist k and δ such that $h_i y \in G$ if $i \geq k$ and $\rho(x, y) < \delta$, so that x is interior to $T_k = \bigcap_{i \geq k} h_i^{-1}G$. Thus $h^{-1}G \subset E \cup \bigcup_k T_k^\circ$. Since $P(E) = 0$, we have $P(h^{-1}G) \leq P(\bigcup_k T_k^\circ)$; since $T_k^\circ \subset T_{k+1}^\circ$, we have, for given $\varepsilon > 0$, $P(h^{-1}G) < P(T_k^\circ) + \varepsilon$ for sufficiently large k. From $P_n \Rightarrow P$ it follows that $P(T_k^\circ) \leq \liminf_n P_n(T_k^\circ)$; since $T_k^\circ \subset h_n^{-1}G$ for large n, we have $P(T_k^\circ) \leq \liminf_n P_n(h_n^{-1}G)$. Therefore $P(h^{-1}G) \leq \liminf_n P_n(h_n^{-1}G) + \varepsilon$, which, since ε was arbitrary, completes the proof.

If $h_n = h$, this result reduces to Theorem 5.1. If h is everywhere continuous and h_n converges to h uniformly on compact sets, then E is empty, so that the hypothesis of the theorem is satisfied. More generally, $P(E) = 0$ if $P(D_h) = 0$ and if there is uniform convergence on compact sets. On the other hand, if $D_h = E = 0$, then h_n converges to h uniformly on compact sets.

Remarks. In the Euclidean case, Theorem 5.1 is sometimes called the Mann-Wald theorem (see Mann and Wald (1943) and Chernoff (1956)); Corollary 2 for rational functions h is Slutsky's theorem (see Slutsky (1925)). Theorem 5.5, which generalizes a result of Prohorov (1956), is due to H. Rubin; see Anderson (1963). For extensions of Theorems 5.1 and 5.5 and converses to them, see Topsøe (1967a and 1967b).

PROBLEMS

1. If a Borel function f has a derivative at a (this being the only assumption on f) and $X_n \xrightarrow{P} a$, then $f(X_n) = f(a) + f'(a)(X_n - a) + (X_n - a)Z_n$ with $Z_n \xrightarrow{P} 0$. Extend to higher derivatives.

2. Suppose f has a derivative at 0. If $X_n Y_n \xrightarrow{\mathscr{D}} Y$ and $Y_n \xrightarrow{P} 0$, then

$$X_n(f(Y_n) - f(0)) \xrightarrow{\mathscr{D}} f'(0)Y.$$

3. Suppose f_n and f have continuous derivatives with $f_n'(x) \to f'(x)$ uniformly in x. If $X_n Y_n \xrightarrow{\mathscr{D}} Y$ and $Y_n \xrightarrow{P} 0$, then $X_n(f_n(Y_n) - f_n(0)) \xrightarrow{\mathscr{D}} f'(0)Y$.

4. Suppose $X_n \xrightarrow{\mathscr{D}} X$. Give examples in which the X_n are integrable but X is not, and conversely. Show that $\mathsf{E}\{X_n\} \to \mathsf{E}\{X\}$ does not in general imply uniform integrability of the X_n.

5. Show that X_n are uniformly integrable if and only if $\sup_n \mathsf{E}\{|X_n|\} < \infty$ and for each ε there exists a δ such that, for all n, $\mathsf{P}(E) < \delta$ implies $\int_E |X_n| \, d\mathsf{P} < \varepsilon$.

6. Let P be Lebesgue measure on the unit interval and P_n correspond to a mass of $1/n$ at some point chosen from $((i - 1)/n, i/n)$, $i = 1, \ldots, n$. Show that $P_n \Rightarrow P$ and deduce from Theorem 5.2 that a bounded function continuous almost everywhere is Riemann integrable.

6. PROHOROV'S THEOREM

Relative Compactness

Let Π be a family of probability measures on $(S, \mathscr{S})$. We call Π *relatively compact* if *every sequence of elements of* Π *contains a weakly convergent subsequence;* that is, if for every sequence $\{P_n\}$ in Π there exist a subsequence $\{P_{n'}\}$ and a probability measure Q (defined on $(S, \mathscr{S})$, but not necessarily an element of Π) such that $P_{n'} \Rightarrow Q$.† Even though $P_{n'} \Rightarrow Q$ makes no sense if $Q(S) < 1$, it is to be emphasized that we do require that $Q(S) = 1$—we disallow any escape of mass, as discussed below.

We need to be able to decide whether or not a given family Π is relatively compact. For example, suppose we know of probability measures P_n and P on $(C, \mathscr{C})$ that the finite-dimensional distributions of P_n converge weakly to those of P. We have seen in Section 5 that P_n need not converge weakly to P. Suppose, however, that we also know that $\{P_n\}$ is relatively compact. Then each subsequence $\{P_{n'}\}$ contains a further subsequence $\{P_{n''}\}$ converging weakly to some limit Q. The finite-dimensional distributions of Q must be the weak limits of those of $\{P_{n''}\}$ and hence must coincide with the finite-dimensional distributions of P (for each $(t_1, \ldots, t_k)$, $P_{n''}\pi^{-1}_{t_1\cdots t_k} \Rightarrow Q\pi^{-1}_{t_1\cdots t_k}$ and $P_n\pi^{-1}_{t_1\cdots t_k} \Rightarrow P\pi^{-1}_{t_1\cdots t_k}$, so $Q\pi^{-1}_{t_1\cdots t_k} = P\pi^{-1}_{t_1\cdots t_k}$). But then, since a probability measure on C is completely determined by its finite-dimensional distributions (the finite-dimensional sets form a determining class), Q and P must themselves coincide. Thus each subsequence $\{P_{n'}\}$ contains a further subsequence converging weakly to P, and it follows by Theorem 2.3 that the entire sequence $\{P_n\}$ converges weakly to P. Note that, if $\{P_n\}$ does converge weakly to P, then it is relatively compact, so that relative compactness is not an excessive requirement.

Suppose we know that $\{P_n\}$ is relatively compact and that, for each $(t_1, \ldots, t_k)$, $P_n\pi^{-1}_{t_1\cdots t_k}$ converges weakly to some probability measure $\mu_{t_1\cdots t_k}$ on $(R^k, \mathscr{R}^k)$—the point being that we do not assume at the outset that the $\mu_{t_1\cdots t_k}$ are the finite-dimensional distributions of a probability measure on $(C, \mathscr{C})$. It still follows that each subsequence $\{P_{n'}\}$ contains a further subsequence $\{P_{n''}\}$ converging weakly to some limit. Since this limit must have the $\mu_{t_1\cdots t_k}$ as its finite-dimensional distributions, it is unique. Therefore $\{P_n\}$ converges weakly to *some* P.

These ideas provide a powerful technique for proving weak convergence in C and in other function spaces. One proves first that the finite-dimensional

---

† We take this as a *definition;* it is really a *sequential* compactness notion in $Z(S)$ as defined in the footnote on p. 11. See Appendix III for further topological information.

distributions converge weakly and then that the sequence in question is relatively compact. To use this method we need an effective criterion for relative compactness.

Consider R^1 first. Let Π be a family of probability measures on $(R^1, \mathscr{R}^1)$. Given a sequence $\{P_n\}$ from Π, we can apply to the sequence $\{F_n\}$ of corresponding distribution functions the classical Helly selection theorem (see p. 227). There exist a subsequence $\{F_{n'}\}$ and some function F such that

$$F_{n'}(x) \to F(x) \tag{6.1}$$

holds for all its continuity points x. The function F may be taken to be right-continuous, in which case there exists on $(R^1, \mathscr{R}^1)$ a finite measure μ such that

$$\mu(a, b] = F(b) - F(a). \tag{6.2}$$

Now it may happen that $\mu(R^1) < 1$. For instance, if P_n corresponds to a unit mass at the point n, then F is identically 0, no matter what subsequence $\{F_{n'}\}$ is selected, so that $\mu(R^1) = 0$. If P_n is the uniform distribution over $[-n, n]$, then $F(x) \equiv \frac{1}{2}$ is the only possibility—again $\mu(R^1) = 0$. In these examples, mass is "escaping to infinity" in an obvious and intuitive sense.

Suppose, however, that the measure μ determined by F via (6.2) satisfies $\mu(R^1) = 1$. Then μ is a probability measure and, since (6.1) holds at continuity points of F, we have (see Section 3) $P_{n'} \Rightarrow \mu$. Thus Π will be relatively compact if we somehow ensure that each of these measures μ we encounter satisfies $\mu(R^1) = 1$. Now $\mu(R^1) = 1$ if for every positive ε there exist a and b such that $\mu(a, b] \geq 1 - \varepsilon$. Suppose that for every positive ε there exist a and b such that

$$P_n(a, b] > 1 - \varepsilon, \qquad n = 1, 2, \ldots. \tag{6.3}$$

Since (6.3) persists if a is decreased and b increased, we may take a and b to be continuity points of F, in which case (6.3) and (6.1) together imply $\mu(a, b] \geq 1 - \varepsilon$.

It follows that a family Π of probability measures on $(R^1, \mathscr{R}^1)$ is relatively compact if for each positive ε there exist a and b such that $P(a, b] > 1 - \varepsilon$ for all P in Π, a condition which has the effect of preventing the escape of mass alluded to above. On the other hand, if this condition fails, then there exists some positive ε such that, no matter what a and b are, $P(a, b] \leq 1 - \varepsilon$ for some P in Π. Choose in Π a P_n with $P_n(-n, n] \leq 1 - \varepsilon$. If some subsequence $\{P_{n'}\}$ were to converge weakly to a probability measure Q, we would have, for each x,

$$Q(-x, x) \leq \liminf_{n'} P_{n'}(-x, x) \leq \liminf_{n'} P_{n'}(-n', n'] \leq 1 - \varepsilon,$$

which is impossible. Thus the condition is both necessary and sufficient.

Since an interval $(a, b]$ has compact closure, and since each compact set can be enclosed in such an interval, the condition can be recast: a family Π of probability measures on $(R^1, \mathscr{R}^1)$ is relatively compact if and only if for each positive ε there exists a compact set K such that $P(K) > 1 - \varepsilon$ for all P in Π. Now the condition has meaning in an arbitrary metric space.

A family Π of probability measures on the general metric space S is said to be *tight* if for every positive ε there exists a compact set K such that $P(K) > 1 - \varepsilon$ for all P in Π. If Π consists of a single measure alone, this definition reduces to the one introduced in Section 1. We have just seen in the case R^1 that tightness is necessary and sufficient for relative compactness. The following theorems, due to Prohorov, extend the sufficiency to arbitrary metric spaces and the necessity to separable, complete spaces.

THEOREM 6.1 *If Π is tight, then it is relatively compact.*

THEOREM 6.2 *Suppose S is separable and complete. If Π is relatively compact, then it is tight.*

We shall refer to these two theorems conjointly as Prohorov's theorem, calling Theorem 6.1 the direct half and Theorem 6.2 the converse half.

The Direct Theorem

We turn first to the proof of the direct half, which is the more useful in the applications. We shall prove the result successively for R^k, for R^∞, for S σ-compact (a countable union of compact sets), and, finally, for the general S. Each of the last three cases is handled by reducing it to the one preceding.

The case R^k. Here the proof is practically the same as the one already given for R^1. If $\{P_n\}$ is a sequence in Π, Helly's theorem (p. 227) implies that the sequence $\{F_n\}$ of corresponding distribution functions contains a subsequence $\{F_{n'}\}$ such that

$$F_{n'}(x) \to F(x) \tag{6.4}$$

at continuity points of F, where F is a function continuous from above. There exists on $(R^k, \mathscr{R}^k)$ a measure μ such that $\mu(a, b]$ is F differenced around the vertices of the k-dimensional rectangle $(a, b]$. Now $P_{n'} \Rightarrow \mu$ will follow if we prove $\mu(R^k) = 1$.

Given ε, choose a compact set K in R^k such that $P_{n'}(K) > 1 - \varepsilon$ for all n', which is possible because Π is tight. Now choose a and b such that $K \subset (a, b]$ and such that all 2^k vertices of $(a, b]$ are continuity points of F (this is possible because only countably many parallel $(k - 1)$-dimensional hyperplanes can have positive μ-measure). Since $P_{n'}(a, b]$ is $F_{n'}$ differenced around the vertices of $(a, b]$, (6.4) implies $P_{n'}(a, b] \to \mu(a, b]$. From $P_{n'}(a, b] \geq$

$P_{n'}(K) > 1 - \varepsilon$, it follows that $\mu(a, b] \geq 1 - \varepsilon$. Since ε was arbitrary, $\mu(R^k) = 1$. Thus Π is relatively compact.

For the case R^∞ we shall need the following lemma.

LEMMA 1 *If Π is a tight family on $(S, \mathscr{S})$ and if h is a continuous mapping from S to S', then $\{Ph^{-1}: P \in \Pi\}$ is a tight family on $(S', \mathscr{S}')$.*

Proof. Given ε, choose in S a compact set K such that $P(K) > 1 - \varepsilon$ for all P in Π. If $K' = hK$, then K' is compact (see p. 218) and $h^{-1}K' \supset K$, so that $Ph^{-1}(K') > 1 - \varepsilon$ for all P in Π.

The case R^∞. If Π is a tight family on $(R^\infty, \mathscr{R}^\infty)$, then, by Lemma 1, $\{P\pi_k^{-1}: P \in \Pi\}$ is, for each k, a tight family on $(R^k, \mathscr{R}^k)$. By the direct half of Prohorov's theorem for R^k, just proved via Helly's theorem, we can select from a given sequence $\{P_n\}$ in Π a subsequence $\{P_{n'}\}$ such that $P_{n'}\pi_k^{-1}$ converges weakly to a probability measure μ_k on $(R^k, \mathscr{R}^k)$. By the diagonal method (as on p. 219), we can choose the sequence $\{P_{n'}\}$ so that $P_{n'}\pi_k^{-1} \Rightarrow \mu_k$ holds for all k simultaneously.

Since the measures μ_k obviously satisfy the consistency conditions of Kolmogorov's existence theorem (see p. 228), there is a probability measure Q on $(R^\infty, \mathscr{R}^\infty)$ such that $Q\pi_k^{-1} = \mu_k$ for all k. (It was shown in Section 3 that the σ-field $\mathscr{R}^\infty$ of Borel sets in R^∞ coincides with the σ-field generated by the finite-dimensional sets, which is the σ-field that intervenes in Kolmogorov's existence theorem.) But then $P_{n'}\pi_k^{-1} \Rightarrow Q\pi_k^{-1}$ for each k, so that the finite-dimensional distributions of $P_{n'}$ converge to those of Q, which, as noted at the beginning of Section 5, implies $P_{n'} \Rightarrow Q$. Tightness implies relative compactness in R^∞.

For the σ-compact and the general cases, we shall need two further lemmas. Suppose that S_0 is a Borel subset of S:

$$S_0 \in \mathscr{S}. \tag{6.5}$$

Now S_0 is a metric space in its own right in the relative topology; let $\mathscr{S}_0$ denote the class of Borel sets for S_0. From (6.5) it follows (see p. 224) that

$$\mathscr{S}_0 = \{A : A \subset S_0, A \in \mathscr{S}\}; \tag{6.6}$$

in particular,

$$\mathscr{S}_0 \subset \mathscr{S}. \tag{6.7}$$

If P is a probability measure on $(S, \mathscr{S})$ with $P(S_0) = 1$, let P^r (r for restriction) be the probability measure on $(S_0, \mathscr{S}_0)$ got by restriction from $\mathscr{S}$ to $\mathscr{S}_0$ (see (6.7)). If P is a probability measure on $(S_0, \mathscr{S}_0)$, let P^e (e for extension) be the probability measure on $(S, \mathscr{S})$ with $P^e(A) = P(A \cap S_0)$ for $A \in \mathscr{S}$ (see (6.6)). Note that $P^e(S_0) = 1$.

If P is a probability measure on $(S, \mathscr{S})$ with $P(S_0) = 1$, then

$$(P^r)^e = P; \tag{6.8}$$

if P is a probability measure on $(S_0, \mathscr{S}_0)$, then

$$(P^e)^r = P. \tag{6.9}$$

If, in Lemma 1 and in Theorem 5.1, we take h to be the identity mapping from S_0 into S, we obtain the following lemma.

LEMMA 2 *If Π is a tight family on $(S_0, \mathscr{S}_0)$, then $\Pi^e = \{P^e : P \in \Pi\}$ is a tight family on $(S, \mathscr{S})$. If $P_n \Rightarrow P$ in $(S_0, \mathscr{S}_0)$, then $P_n{}^e \Rightarrow P^e$ in $(S, \mathscr{S})$.*

LEMMA 3 *If $P_n \Rightarrow P$ in $(S, \mathscr{S})$ and $P_n(S_0) = P(S_0) = 1$, then $P_n{}^r \Rightarrow P^r$ in $(S_0, \mathscr{S}_0)$.*

Proof. The general open set in S_0 is $G_0 = G \cap S_0$ with G open in S. Since $P_n{}^r(G_0) = P_n(G)$ and $P^r(G_0) = P(G)$, $\liminf_n P_n(G) \geq P(G)$ implies

$$\liminf_n P_n{}^r(G_0) \geq P^r(G_0).$$

The σ-compact case. If S is σ-compact, then it is separable and hence can be embedded homeomorphically into R^∞ (see p. 219). Since S is σ-compact, so is its image under the homeomorphism; in particular, this image is a Borel subset of R^∞. Thus S is homeomorphic to a Borel subset of R^∞. By Theorem 5.1, weak convergence persists under homeomorphism, and hence so does relative compactness of Π. Since compactness of sets persists under homeomorphism, so does tightness of Π. Therefore we can replace S by its homeomorphic image.

We may thus assume that S is a Borel subset of R^∞. If Π is tight in $(S, \mathscr{S})$, then, by Lemma 2 applied to R^∞ and its subset S, Π^e is tight in $(R^\infty, \mathscr{R}^\infty)$. Since, as already proved, Theorem 6.1 holds in this larger space, Π^e is relatively compact, so that, for each sequence $\{P_n\}$ in Π, the corresponding sequence $\{P_n{}^e\}$ contains a subsequence $\{P^e_{n'}\}$ converging weakly in the sense of $(R^\infty, \mathscr{R}^\infty)$ to some Q. By the tightness of Π itself, there is for each ε a compact subset K of S such that $P^e_{n'}(K) = P_{n'}(K) > 1 - \varepsilon$ for all n', so that $Q(S) \geq Q(K) \geq \limsup_{n'} P^e_{n'}(K) \geq 1 - \varepsilon$. Thus S supports Q as well as all the $P^e_{n'}$, and hence, by Lemma 3 and (6.9), $P_{n'}$ converges weakly in the sense of $(S, \mathscr{S})$ to Q^r. Tightness implies relative compactness if S is σ-compact.†

The general case. Whatever S is, if $S_0 = \bigcup_i K_i$, where K_i is a compact set in S such that $P(K_i) > 1 - 1/i$ for all P in Π, then S_0 supports each element of

† A separable S is homeomorphic to a subset of R^∞. The stronger assumption that S is σ-compact ensures that its image lies in $\mathscr{R}^\infty$, which is convenient because Lemmas 2 and 3 become awkward without the assumption $S_0 \in \mathscr{S}$.

Π and $\Pi^r = \{P^r : P \in \Pi\}$ is a tight family in $(S_0, \mathscr{S}_0)$. By the case just treated, Π^r is relatively compact. For each sequence $\{P_n\}$ in Π, therefore, the corresponding sequence $\{P_n{}^r\}$ contains a subsequence $\{P^r_{n'}\}$ converging weakly in the sense of $(S_0, \mathscr{S}_0)$ to some Q. By Lemma 2 and (6.8), $P_{n'}$ converges weakly in the sense of $(S, \mathscr{S})$ to Q^e. Thus Π is relatively compact, which proves Theorem 6.1 in full generality.

The Converse

We turn now to Theorem 6.2, the converse half of Prohorov's theorem. Note that it reduces to Theorem 1.4 in case Π consists of a single measure; the proof is a refinement of the earlier one.

Suppose that for each positive ε and δ there exists a finite collection $A_1, \ldots, A_n$ of δ-spheres such that $P(\bigcup_{i \le n} A_i) > 1 - \varepsilon$ for all P in Π. The following argument shows that Π is then tight. Given ε, choose, for each integer k, finitely many $1/k$-spheres $A_{k1}, \ldots, A_{kn_k}$ such that $P(\bigcup_{i \le n_k} A_{ki}) > 1 - \varepsilon/2^k$ for all P in Π. If K is the closure of the totally bounded set $\bigcap_{k \ge 1} \bigcup_{i \le n_k} A_{ki}$, then $P(K) > 1 - \varepsilon$ and, since S is assumed complete, K is compact (see p. 217).

We prove Theorem 6.2 by showing that, if the condition stated in the preceding paragraph fails, then Π is not relatively compact. Suppose then there exist positive ε and δ such that every finite collection $A_1, \ldots, A_n$ of δ-spheres satisfies $P(\bigcup_{i \le n} A_i) \le 1 - \varepsilon$ for some P in Π. Since S is assumed separable, it is the union of a sequence $A_1, A_2, \ldots$ of open spheres of radius δ. Let $B_n = \bigcup_{i \le n} A_i$ and choose P_n in Π so that $P_n(B_n) \le 1 - \varepsilon$. Suppose some subsequence $\{P_{n'}\}$ converged weakly to some limit Q. Since B_m is open, we would have $P(B_m) \le \liminf_{n'} P_{n'}(B_m)$ for each fixed m. But then, since $B_m \subset B_{n'}$ for large n', $P(B_m) \le \liminf_{n'} P_{n'}(B_{n'}) \le 1 - \varepsilon$ would follow; since B_m increases to S, this is impossible, which completes the proof.

For strengthened forms of Theorem 6.2, see Appendix III.

If X_n are random elements of S, we say $\{X_n\}$ is tight when $\{P_n\}$ is tight, where P_n is the distribution of X_n. If S is R^∞ or C, we identify the finite-dimensional distributions of X_n with those of P_n. The argument at the beginning of this section may be interpreted as asserting of random elements X_n and X of C that, if the finite-dimensional distributions of X_n converge weakly to those of X and $\{X_n\}$ is tight, then $X_n \xrightarrow{\mathscr{D}} X$.

Remarks. In his original proof of Theorem 6.1, Prohorov (1956) assumed S to be separable and complete; the present extension and its proof are due to Varadarajan (1958a and 1961a). See also LeCam (1957).

PROBLEMS

1. If Π is tight, then its elements have a common σ-compact support. The converse fails (unless, for example, Π consists of a single measure).

2. Let Π consist of the unit masses for points in A. Show directly from the definition that Π is relatively compact if and only if A^- is compact in S.

3. A sequence of probability measures on the line is tight if and only if the corresponding distribution functions satisfy $\lim_{x\to\infty} F_n(x) = 1$ and $\lim_{x\to-\infty} F_n(x) = 0$ uniformly in n. A class of normal distributions is tight if and only if the means and variances are bounded.

4. Show that, if F is a right-continuous, nondecreasing real function with $0 \le F(x) \le 1$, then there exist distribution functions F_n such that $F_n(x) \to F(x)$ for each continuity point x.

5. If $\{|X_n|^\delta\}$ is uniformly integrable (see p. 32) for some $\delta > 0$, then $\{X_n\}$ is tight.

6. Probability measures on a product $S' \times S''$ are tight if and only if the two sets of marginal distributions are tight in S' and S''.

7. Suppose S is separable and locally compact. Since S can be given a metric under which it is complete [Problem 3 of Section 1], Theorem 6.2 applies. From the fact that each compact set in S can be enclosed in the interior of a larger compact set, it follows further that $P_n \Rightarrow P$ if and only if $\int f\,dP_n \to \int f\,dP$ for every (bounded) continuous f with compact support (a result which cannot hold in C, for example, since there a continuous function with compact support vanishes identically; see Problem 5 in Section 3).

8. In connection with Lemma 3, note that, if Π is a tight family on $(S, \mathscr{S})$ and $P(S_0) = 1$ for all $P \in \Pi$, then it does not follow that $\Pi^r = \{P^r : P \in \Pi\}$ is a tight family on $(S_0, \mathscr{S}_0)$. [Take $S = [0, 1]$ and $S_0 = (0, 1)$ and consider point masses.] This can happen even if Π consists of a single measure. [See Remark 2 following Theorem 1 in Appendix III.]

7. FIRST APPLICATIONS

In this section we examine from the point of view of the preceding theory some standard probability results, results used frequently in the chapters that follow.

Smooth Functions

In proving Theorems 1.2 and 1.3 and in proving the implication (ii) $\to$ (iii) in Theorem 2.1, we used arguments involving the function (1.1). These arguments still go through if, in place of (1.1), we use any uniformly continuous φ such that $\varphi(t) = 1$ for $t \le 0$, $0 \le \varphi(t) \le 1$ for $0 \le t \le 1$, and $\varphi(t) = 0$ for $t \ge 1$. It is possible to construct such functions φ with smoothness properties stronger than uniform continuity. Outside $[0, 1]$, define φ as before; for $0 \le t \le 1$, define

$$\varphi(t) = \alpha^{-1} \int_t^1 e^{-1/s(1-s)}\,ds, \tag{7.1}$$

where

$$\alpha = \int_0^1 e^{-1/s(1-s)}\,ds.$$

Then, for each integer v, φ has, over the whole line, a bounded, continuous derivative of order v.

Let P_n and P be probability measures on $(R^1, \mathscr{R}^1)$.

THEOREM 7.1 *If $\int f\,dP_n \to \int f\,dP$ for every bounded, continuous function f having bounded, continuous derivatives of each order, then $P_n \Rightarrow P$.*

Proof. Consider the distribution functions F_n and F corresponding to P_n and P. If $\varphi_u(t) = \varphi(ut)$, with φ defined by (7.1), then, for each u,

$$\limsup_{n\to\infty} F_n(x) \le \lim_{n\to\infty} \int \varphi_u(y-x)P_n(dy) = \int \varphi_u(y-x)P(dy) \le F\left(x + \frac{1}{u}\right).$$

Using $\varphi_u(y - x + 1/u)$ in place of $\varphi_u(y - x)$, we see that $\liminf_n F_n(x) \ge F(x - 1/u)$ for each u. Hence $F_n(x) \to F(x)$ if F is continuous at x, and $P_n \Rightarrow P$ follows.

The Central Limit Theorem

Theorem 7.1 can be used to derive the classical central limit theorems for triangular arrays. For each n, let

$$\xi_{n1}, \ldots, \xi_{nk_n} \tag{7.2}$$

be independent random variables with mean 0 and finite variance σ^2_{nk}. (The probability space on which the variables (7.2) are defined may vary with n.) Let $S_n = \xi_{n1} + \cdots + \xi_{nk_n}$ and suppose its variance $s_n{}^2 = \sigma^2_{n1} + \cdots + \sigma^2_{nk_n}$ is positive. Recall that N is a random variable normally distributed with mean 0 and variance 1. We first prove Lindeberg's theorem:

THEOREM 7.2 *If*

$$\frac{1}{s_n{}^2} \sum_{k=1}^{k_n} \int_{\{|\xi_{nk}| \ge \varepsilon s_n\}} \xi^2_{nk}\, d\mathsf{P} \to 0 \tag{7.3}$$

$(n \to \infty)$ for each positive ε, then $S_n/s_n \xrightarrow{\mathscr{D}} N$.

Proof. It suffices to show that

$$\mathsf{E}\left\{f\left(\frac{S_n}{s_n}\right)\right\} \to \mathsf{E}\{f(N)\} \tag{7.4}$$

for every bounded, continuous f having bounded, continuous derivatives of each order. Fix such a function f and define

$$g(h) = \sup_x |f(x+h) - f(x) - f'(x)h - \tfrac{1}{2}f''(x)h^2| \tag{7.5}$$

(g is Borel measurable). By the mean value theorems of the second and third orders, there is a constant K, depending on f alone, such that

$$g(h) \le K \min\{h^2, |h|^3\}. \tag{7.6}$$

We have $g(h) \le Kh^2$ and $g(h) \le K\,|h|^3$; the first inequality is good for h large, the second for h near 0. Since f is fixed throughout the rest of the argument, so is K. From the definition (7.5) it follows that

$$(7.7)\quad |[f(x+h_1)-f(x+h_2)]-[f'(x)(h_1-h_2)+\tfrac{1}{2}f''(x)(h_1^2-h_2^2)]| \le g(h_1)+g(h_2).$$

If the ξ_{nk} were all normally distributed, $\mathsf{E}\{f(S_n/s_n)\}$ would coincide with $\mathsf{E}\{f(N)\}$. If we successively replace the ξ_{nk} by normal variables η_{nk} with mean 0 and variances σ_{nk}^2, we get a sequence

$$\mathsf{E}\{f(s_n^{-1}(\xi_{n1}+\cdots+\xi_{nk_n}))\},$$
$$\mathsf{E}\{f(s_n^{-1}(\xi_{n1}+\cdots+\xi_{nk_n-1}+\eta_{nk_n}))\},$$
$$\cdots\cdots\cdots\cdots\cdots\cdots$$
$$\mathsf{E}\{f(s_n^{-1}(\xi_{n1}+\eta_{n2}+\cdots+\eta_{nk_n}))\},$$
$$\mathsf{E}\{f(s_n^{-1}(\eta_{n1}+\cdots+\eta_{nk_n}))\}.$$

The first member of the sequence is $\mathsf{E}\{f(S_n/s_n)\}$ and the last one is $\mathsf{E}\{f(N)\}$. The idea now is to show that each member is so close to the next (for n large) that even the first and last members are close.

Since (7.4) involves the joint distribution of the variables (7.2) but not any properties of the probability space they are defined on, we may, by passing to a new space (say $(R^{2k_n}, \mathscr{R}^{2k_n})$), introduce random variables $\eta_{n1}, \ldots, \eta_{nk_n}$ such that η_{nk} is normally distributed with mean 0 and variance σ_{nk}^2 and such that the $2k_n$ random variables

$$(7.8)\qquad \xi_{n1}, \ldots, \xi_{nk_n}, \eta_{n1}, \ldots, \eta_{nk_n}$$

are independent. If

$$\zeta_{nk} = \sum_{1\le i<k} \xi_{ni} + \sum_{k<i\le k_n} \eta_{ni}, \qquad 1 \le k \le k_n,$$

then, since $\zeta_{nk_n} + \xi_{nk_n} = S_n$ and since $\zeta_{n1} + \eta_{n1}$ has the distribution of s_nN,

$$\left|\mathsf{E}\left\{f\left(\frac{S_n}{s_n}\right)\right\} - \mathsf{E}\{f(N)\}\right| \le \sum_{k=1}^{k_n} \left|\mathsf{E}\left\{f\left(\frac{\zeta_{nk}+\xi_{nk}}{s_n}\right) - f\left(\frac{\zeta_{nk}+\eta_{nk}}{s_n}\right)\right\}\right|.$$

Because of the independence of the sequence (7.8), the three variables ζ_{nk}, ξ_{nk}, and η_{nk} are independent for each value of k and therefore

$$\mathsf{E}\left\{f'\left(\frac{\zeta_{nk}}{s_n}\right)(\xi_{nk}-\eta_{nk})\right\} = \mathsf{E}\left\{f''\left(\frac{\zeta_{nk}}{s_n}\right)(\xi_{nk}^2-\eta_{nk}^2)\right\} = 0.$$

From (7.7) it follows that

$$\left|\mathsf{E}\left\{f\left(\frac{S_n}{s_n}\right)\right\} - \mathsf{E}\{f(N)\}\right| \le \sum_{k=1}^{k_n} \mathsf{E}\left\{g\left(\frac{\xi_{nk}}{s_n}\right) + g\left(\frac{\eta_{nk}}{s_n}\right)\right\}.$$

The proof will therefore be complete if we show that

$$\sum_{k=1}^{k_n} \mathsf{E}\left\{g\left(\frac{\xi_{nk}}{s_n}\right)\right\} \to 0 \tag{7.9}$$

and

$$\sum_{k=1}^{k_n} \mathsf{E}\left\{g\left(\frac{\eta_{nk}}{s_n}\right)\right\} \to 0. \tag{7.10}$$

Given $\varepsilon > 0$, split the expected value in (7.9) into an integral over $\{|\xi_{nk}| \le \varepsilon s_n\}$ and an integral over the complementary set. Use (7.6) to bound the integrand by $K\,|\xi_{nk}/s_n|^3$ on the first set and by $K\,|\xi_{nk}/s_n|^2$ on the second:

$$\mathsf{E}\left\{g\left(\frac{\xi_{nk}}{s_n}\right)\right\} \le K\varepsilon\frac{\sigma_{nk}^2}{s_n^{\,2}} + K\frac{1}{s_n^{\,2}}\int_{\{|\xi_{nk}|\ge\varepsilon s_n\}} \xi_{nk}^2\,d\mathsf{P}.$$

Thus

$$\sum_{k=1}^{k_n} \mathsf{E}\left\{g\left(\frac{\xi_{nk}}{s_n}\right)\right\} \le K\varepsilon + K\frac{1}{s_n^{\,2}}\sum_{k=1}^{k_n}\int_{\{|\xi_{nk}|\ge\varepsilon s_n\}} \xi_{nk}^2\,d\mathsf{P}, \tag{7.11}$$

and (7.9) follows from (7.3).

Since (7.11) also holds with η_{nk} in place of ξ_{nk}, to prove (7.10) we need only show that

$$\frac{1}{s_n^{\,2}}\sum_{k=1}^{k_n}\int_{\{|\eta_{nk}|\ge\varepsilon s_n\}} \eta_{nk}^2\,d\mathsf{P} \tag{7.12}$$

tends to 0 for each positive ε. But (7.12) is at most

$$\frac{1}{s_n^{\,2}}\sum_{k=1}^{k_n}\frac{1}{\varepsilon s_n}\,\mathsf{E}\{|\eta_{nk}|^3\} = \frac{1}{\varepsilon s_n^{\,3}}\sum_{k=1}^{k_n}\sigma_{nk}^3\,\mathsf{E}\{|N|^3\}.$$

Since

$$\frac{\sigma_{nk}^2}{s_n^{\,2}} \le \varepsilon^2 + \frac{1}{s_n^{\,2}}\int_{\{|\xi_{nk}|\ge\varepsilon s_n\}} \xi_{nk}^2\,d\mathsf{P},$$

(7.3) implies $\max_{k\le k_n}\sigma_{nk}/s_n \to 0$, which in turn implies $\sum_{k=1}^{k_n}\sigma_{nk}^3/s_n^{\,3} \to 0$. Thus (7.12) does tend to 0, which completes the proof.

If the ξ_{nk} have moments of order $2+\delta$, then the sum in (7.3) is at most $\varepsilon^{-\delta}s_n^{-2-\delta}\sum_{k=1}^{k_n}\mathsf{E}\{|\xi_{nk}|^{2+\delta}\}$, which proves Lyapounov's theorem:

THEOREM 7.3 *If, for some positive* δ,

$$\frac{1}{s_n^{2+\delta}}\sum_{k=1}^{k_n}\mathsf{E}\{|\xi_{nk}|^{2+\delta}\} \to 0,$$

then $S_n/s_n \xrightarrow{\mathscr{D}} N$.

Finally, from Theorem 7.2 we can deduce the Lindeberg–Lévy theorem:

THEOREM 7.4 *If $\xi_1, \xi_2, \ldots$ are independent and identically distributed with mean 0 and finite variance $\sigma^2 > 0$, then*

$$\frac{1}{\sigma\sqrt{n}} \sum_{k=1}^{n} \xi_k \xrightarrow{\mathscr{D}} N.$$

To prove this last result, take $k_n = n$ and $\xi_{ni} = \xi_i$; the sum in (7.3) is at most ξ_1^2/σ^2 integrated over $\{|\xi_1| \geq \varepsilon\sigma\sqrt{n}\}$.

Characteristic Functions

If P is a probability measure on $(R^k, \mathscr{R}^k)$, its characteristic function p is defined as

$$p(t) = \int e^{it\cdot x} P(dx), \qquad t \in R^k,$$

where $t \cdot x = \sum_{u=1}^{k} t_u x_u$ denotes inner product. Let Q be a second probability measure on $(R^k, \mathscr{R}^k)$, and let q be its characteristic function. We shall prove the uniqueness theorem:

THEOREM 7.5 *If $p(t) = q(t)$ for all t, then $P = Q$.*

For each t, $e^{it\cdot x}$ is, as a function of x, an element of $C(R^k)$ (or its real and imaginary parts are). Thus in the case R^k the uniqueness theorem refines Theorem 1.3, which asserts that P is determined by the values of $\int f\,dP$ for $f \in C(S)$. We shall prove uniqueness not by deriving an explicit inversion formula but by using the Weierstrass approximation theorem.

We know from Section 3 that the rectangles $(a, b]$ form a determining class; since such a rectangle is an increasing union of closed rectangles, it suffices to check that P and Q agree for sets

$$A = [a_1, b_1] \times \cdots \times [a_k, b_k].$$

By the argument for Theorem 1.3, this will follow if, for each integral u,

$$\int f\,dP = \int f\,dQ \tag{7.13}$$

holds when f has the form $f(x_1, \ldots, x_k) = f_1(x_1) \cdots f_k(x_k)$ with $f_j(s) = \varphi_u(\rho(s, [a_j, b_j]))$, where ρ denotes linear distance and φ_u is defined by (1.4).

Fix u. Given ε, choose r so large that each f_j vanishes outside the interval $[-r, r]$ and, at the same time, so large that, if I_r denotes the cube $\{x: |x_m| \leq r,\ m = 1, \ldots, k\}$, then $P(I_r^c) < \varepsilon$ and $Q(I_r^c) < \varepsilon$. Since $f_j(-r) = f_j(r)$, $f_j(s)$ can, by Weierstrass's theorem,† be uniformly

† For example, see Titchmarsh (1939) or, for Stone's generalization, Simmons (1963).

approximated in $[-r, r]$ by a finite trigonometric sum $\Sigma_l \alpha_l e^{il\pi s/r}$ of period $2r$. Multiplying together these sums for the different values of j, we see that f can be uniformly approximated in I_r by a finite trigonometric sum

$$g(x) = \Sigma_l \gamma_l e^{it^{(l)}\cdot x} \tag{7.14}$$

of period $2r$ in each variable. Choose (7.14) so that $|f(x) - g(x)| < \varepsilon$ for x in I_r. Since f is everywhere bounded by 1, g is bounded by $1 + \varepsilon$ in I_r and hence, by periodicity, in all of R^k. Thus $|f - g|$ is bounded by ε in I_r and by $2 + \varepsilon$ everywhere. Assuming $0 < \varepsilon < 1$, we have

$$\int |f - g| \, dP \leq \varepsilon + (2 + \varepsilon)P(I_r^c) < 4\varepsilon.$$

Similarly, $\int |f - g| \, dQ < 4\varepsilon$, so that

$$\left| \int f \, dP - \int f \, dQ \right| < \left| \int g \, dP - \int g \, dQ \right| + 8\varepsilon.$$

But $\int g \, dP = \int g \, dQ$ is an immediate consequence of $p = q$ and (7.14); since ε was arbitrary, (7.13) follows.

Suppose now that P_n and P are probability measures on $(R^k, \mathscr{R}^k)$ with characteristic functions p_n and p. We shall prove the continuity theorem:

THEOREM 7.6 *A necessary and sufficient condition for* $P_n \Rightarrow P$ *is that* $p_n(t) \to p(t)$ *for each* t.

The necessity follows from the fact that $e^{it\cdot x}$ is bounded and continuous in x for each t. We shall prove the sufficiency via two intermediate propositions.

If the $p_n(t)$ *converge pointwise to some limit and if* $\{P_n\}$ *is tight, then* $P_n \Rightarrow P$ *for some* P. Let $g(t) = \lim_n p_n(t)$; we make no assumptions whatever about g. If $\{P_n\}$ is tight, then, by Prohorov's theorem, it is relatively compact, so that each subsequence $\{P_{n'}\}$ contains some further subsequence $\{P_{n''}\}$ converging weakly to some limit whose characteristic function must then be $\lim_{n''} p_{n''}(t) = g(t)$. By the uniqueness theorem, there is thus only one possible such limit, which is the P we seek (see Theorem 2.3).

Notice that this proposition fails without the assumption of tightness [example: $k = 1$ and P_n uniform over $(-n, n)$]. Notice also that the proof just given exactly parallels the argument at the beginning of Section 6 involving tightness and finite-dimensional distributions for probability measures on $(C, \mathscr{C})$.

Now let us show that, *if the limit function* $g(t) = \lim_n p_n(t)$ *is continuous at* $t = 0$, *then* $\{P_n\}$ *is tight* (and hence weakly convergent to some limit). Clearly, $\{P_n\}$ is tight if each of the k corresponding sequences of one-dimensional marginal distributions is tight. The characteristic functions $p_n(s, 0, \ldots, 0)$ of the marginal distributions for the first coordinate converge pointwise to a limit $g(s, 0, \ldots, 0)$ continuous at $s = 0$, and similarly for the other coordinates. Thus each sequence of marginal distributions satisfies the hypotheses and it suffices to treat the case $k = 1$. In this case, by Fubini's theorem,

$$\frac{1}{u}\int_{-u}^{u}(1 - p_n(t))\,dt = \int\left[\frac{1}{u}\int_{-u}^{u}(1 - e^{itx})\,dt\right] P_n(dx)$$

$$= 2\int\left(1 - \frac{\sin ux}{ux}\right) P_n(dx) \tag{7.15}$$

$$\geq 2\int_{|x| \geq 2/u}\left(1 - \frac{1}{|ux|}\right) P_n(dx) \geq P_n\left\{x : |x| \geq \frac{2}{u}\right\}$$

(in particular, the first integral is real). Since g is continuous at the origin, there is, for each positive ε, a u such that $u^{-1}\int_{-u}^{u}|1 - g(t)|\,dt < \varepsilon$. By the bounded convergence theorem, for n beyond some n_0 we have

$$u^{-1}\int_{u}^{-u}|1 - p_n(t)|\,dt < 2\varepsilon$$

and hence, by (7.15), $P_n\{x : |x| \geq a\} < 2\varepsilon$ with $a = 2/u$. By increasing a if necessary, we can ensure that this inequality holds for the finitely many n preceding n_0: $\{P_n\}$ is indeed tight.

If g is the characteristic function of a probability measure, it is automatically continuous; the continuity theorem follows.

It is instructive to compare the following pairs of statements, of which all are true except 6^0. Here P_n and P are probability measures on $(R^k, \mathscr{R}^k)$ or on $(C, \mathscr{C})$, as appropriate; in the former case, p_n and p are their characteristic functions and g is a function on R^k; in the latter case, $P_n\pi_{t_1\cdots t_i}^{-1}$ and $P\pi_{t_1\cdots t_i}^{-1}$ are their finite-dimensional distributions and $\mu_{t_1\cdots t_i}^{-1}$ are probability measures on $(R^i, \mathscr{R}^i)$.

Space R^k	*Space C*
1. The function $p(t)$ determines P.	1^0. The measures $P\pi_{t_1\ldots t_i}^{-1}$ determine P.
2. If $p_n(t)$ converges to some $g(t)$ for each t and if $\{P_n\}$ is tight, then P_n converges weakly to some P.	2^0. If $P_n\pi_{t_1\ldots t_i}^{-1}$ converges weakly to some $\mu_{t_1\ldots t_i}$ for each $(t_1 \cdots t_i)$ and if $\{P_n\}$ is tight, then P_n converges weakly to some P.
3. Statement 2 fails without tightness.	3^0. Statement 2^0 fails without tightness.
4. If $p_n(t)$ converges to some $g(t)$ for each t and if g is continuous at 0, then $\{P_n\}$ is tight (and hence, by 2, weakly convergent).	
5. If $p_n(t) \to p(t)$ for each t and if $\{P_n\}$ is tight, then $P_n \Rightarrow P$.	5^0. If $P_n\pi_{t_1\cdots t_i}^{-1} \Rightarrow P\pi_{t_1\cdots t_i}^{-1}$ for each $(t_1 \cdots t_i)$ and if $\{P_n\}$ is tight, then $P_n \Rightarrow P$.
6. If $p_n(t) \to p(t)$ for each t, then $P_n \Rightarrow P$.	6^0. [If $P_n\pi_{t_1\cdots t_i}^{-1} \Rightarrow P\pi_{t_1\cdots t_i}^{-1}$ for each $(t_1 \cdots t_i)$, then $P_n \Rightarrow P$.]

Statement 1 is the uniqueness theorem for characteristic functions; 1^0 restates the fact that the finite-dimensional sets in C form a determining class. The proofs of 1 and 1^0 have nothing particular in common.

Statements 2 and 2^0 follow, respectively, from 1 and 1^0 by exactly parallel arguments.

Counterexamples substantiate 3 and 3^0.

Statement 4 has no analogue 4^0.

Now 5 and 5^0 follow, respectively, from 2 and 2^0, again by exactly parallel arguments; but the assumption of tightness in 5 is superfluous because of 4 and the fact that p must be continuous. Suppressing the tightness assumption leads from 5 to 6 and from 5^0 to 6^0; although 6 is true (because of 4 and the continuity of p), 6^0 is false (because there is no 4^0).

Not only is 6^0 false as it stands (as an implication applying to all P and all $\{P_n\}$), but *there is no single P for which it is true* (as an implication applying to all $\{P_n\}$): Define $h_n: C \to C$ by $h_n(x) = x + x_n$ with x_n defined by (3.5) and, given P, put $P_n = Ph_n^{-1}$. For some sufficiently small δ, the open set $A = \{x: x(t) < x(0) + \frac{1}{2},\ t \leq \delta\}$ satisfies $P(A) > 0$. If $A_n = A - x_n$, then $\limsup_n A_n$ is empty and hence

$$\liminf_n P_n(A) \leq \limsup_n P_n(A) = \limsup_n P(A_n) = 0.$$

By Theorem 2.1, P_n cannot converge weakly to P—but the finite-dimensional distributions do converge.

The argument just given also shows that there can *be* no 4^0. In this connection, notice that 3^0 follows from the fact that 6^0 is false; since 6 is true, 3 requires a special counterexample (the limit function g must be discontinuous at 0).

If there were a 4^0, most of Chapter 2 would be superfluous (and if there were a 4^0 for the space D, most of Chapter 3 would be superfluous). About 6^0: Since all the finite-dimensional distributions of P and all the P_n determine P and all the P_n themselves, conditions for $P_n \Rightarrow P$ can in principle be given in terms of finite-dimensional distributions alone.† But we must in effect deal with all sets $(t_1, \ldots, t_l)$ simultaneously as $n \to \infty$—it will not do to fix an arbitrary $(t_1, \ldots, t_l)$ and consider what happens as $n \to \infty$. And this is where tightness comes in.

The Cramér-Wold Device

By means of the following simple device due to Cramér and Wold, problems involving random vectors in R^k can often be reduced to problems involving only ordinary random variables in R^1. Suppose that k-dimensional random

---

† See Bartozyński (1961).

vectors $X_n = (X_{n1}, \ldots, X_{nk})$ and $X = (X_1, \ldots, X_k)$ satisfy

$$\sum_{j=1}^{k} t_j X_{nj} \xrightarrow{\mathscr{D}} \sum_{j=1}^{k} t_j X_j$$

for each point $t = (t_1, \ldots, t_k)$ of R^k. Then the characteristic functions $p_n(s) = \mathsf{E}\{\exp(is \sum_{j=1}^{k} t_j X_{nj})\}$ of these one-dimensional random variables converge to $p(s) = \mathsf{E}\{\exp(is \sum_{j=1}^{k} t_j X_j)\}$ for each real s. Taking $s = 1$, we see that

$$\mathsf{E}\{e^{it\cdot X_n}\} \to \mathsf{E}\{e^{it\cdot X}\}.$$

Since t was arbitrary, $X_n \xrightarrow{\mathscr{D}} X$ follows by the continuity theorem for characteristic functions.

THEOREM 7.7 *In R^k, X_n converges in distribution to X if and only if each linear combination of the components of X_n converges in distribution to the corresponding linear combination of the components of X.*

In the terminology of Section 2, the half-spaces in R^k form a convergence-determining class, a fact which apparently cannot be proved without ideas from Fourier analysis.

Local and Integral Limit Theorems

If P_n and P are probability measures on R^k with densities p_n and p with respect to Lebesgue measure and if

$$p_n(x) \to p(x) \tag{7.16}$$

for all x outside some set of Lebesgue measure 0, then, by Scheffé's theorem (p. 224),

$$P_n \Rightarrow P. \tag{7.17}$$

The relation (7.16), called a local limit theorem, implies the relation (7.17), called an integral limit theorem.

There is an analogous result in case P has a density but the mass for P_n is confined to a lattice in R^k. Let $\delta(n) = (\delta_1(n), \ldots, \delta_k(n))$ be a point of R^k with positive coordinates, let $\alpha(n) = (\alpha_1(n), \ldots, \alpha_k(n))$ be an arbitrary point in R^k, and denote by L_n the lattice consisting of all those points of R^k having the form

$$(u_1\delta_1(n) - \alpha_1(n), \ldots, u_k\delta_k(n) - \alpha_k(n)),$$

where $u_1, \ldots, u_k$ range independently over $0, \pm 1, \pm 2, \ldots$. If x is a point in L_n, then the interval $(x - \delta(n), x]$ (see (3.1) for the notation) is a cell of volume $v_n = \delta_1(n) \cdots \delta_k(n)$, and R^k is the disjoint union of these countably many cells.

Suppose now that P_n is a probability measure in R^k with $P_n(L_n) = 1$, and, for $x \in L_n$, let $p_n(x)$ be the mass (possibly 0) at that point. Let P have density p with respect to Lebesgue measure.

THEOREM 7.8 *Suppose that*

$$\max\{\delta_1(n), \ldots, \delta_k(n)\} \to 0. \tag{7.18}$$

Suppose further that, if $\{x_n\}$ is any sequence and x any point in R^k, and if x_n lies in L_n and varies with n in such a way that

$$x_n \to x, \tag{7.19}$$

then

$$\frac{p_n(x_n)}{v_n} \to p(x). \tag{7.20}$$

Then $P_n \Rightarrow P$.

Proof. Define a probability density q_n on R^k by setting $q_n(y) = p_n(x)/v_n$ if y lies in the cell $(x - \delta(n), x]$ determined by the point x of the lattice L_n. Since (7.19) implies (7.20), we have

$$q_n(y) \to p(y) \tag{7.21}$$

for each y. Let X_n have density q_n, and define Y_n on the same probability space by setting $Y_n = x$ if $X_n \in (x - \delta(n), x]$ with $x \in L_n$. What we are to prove is $Y_n \xrightarrow{\mathscr{D}} P$. Since $|X_n - Y_n| \leq |\delta(n)|$, this will follow by (7.18) and Theorem 4.1 if we prove $X_n \xrightarrow{\mathscr{D}} P$. But, in view of (7.21), this is a consequence of Scheffé's theorem.

Weak Convergence on the Circle and Torus

Let S be the unit circle in the complex plane and let $e_u(x) = x^u$ be the circular functions, $u = 0, \pm 1, \pm 2, \ldots$. The Fourier series of a measure P on S is defined by $p(u) = \int e_u(x)\, P(dx)$ for integral u. Since P is determined by the values of $\int f\, dP$ for f in $C(S)$ and since each f in $C(S)$ can, by Weierstrass's theorem, be uniformly approximated by linear combinations of the circular functions, p determines P. By the same reasoning as for the line, this uniqueness theorem implies a continuity theorem: $P_n \Rightarrow P$ if and only if the corresponding Fourier series satisfy $p_n(u) \to p(u)$ for each integer u. Since the circle is compact, there is this time no question that $\{P_n\}$ is tight; hence $\{P_n\}$ converges weakly to some limit if $\lim_n p_n(u)$ exists for each u. If P is normalized circular Lebesgue measure, then $p(u)$ vanishes for $u \neq 0$ (and, of course, is 1 for $u = 0$). Weyl's criterion follows immediately: $(x_1, x_2, \ldots)$ is uniformly distributed on the circle if and only if $n^{-1} \sum_{j=1}^{n} (x_j)^u \to 0$ for each nonzero u. In particular the powers x^j are uniformly distributed if x is not a root of unity.

The circle can be rolled out onto the interval $[0, 1)$ by means of the correspondence $e^{2\pi ix} \leftrightarrow x$. The Fourier series of a probability measure P on $[0, 1)$ is defined by $p(u) = \int e^{2\pi iux} P(dx)$ for integral u; of course, the uniqueness and continuity theorems for the circle can be restated for the interval $[0, 1)$. Although weak convergence here must refer to the topology that $[0, 1)$ inherits from the circle and not to the relative topology of the line (for example, $1 - 1/n \to 0$), we can clearly ignore this distinction if the limit measure assigns measure 0 to the point 0. Thus, if P_n, P are probability measures on R^1 with support $[0, 1)$ and if $P\{0\} = 0$, then $P_n \Rightarrow P$ if and only if

$$\int e^{2\pi iux} P_n(dx) \to \int e^{2\pi iux} P(dx)$$

for each integer u. A sequence $(x_1, x_2, \ldots)$ of real numbers is said to be uniformly distributed modulo 1 if the empirical distributions of the sequence $(\{x_1\}, \{x_2\}, \ldots)$ of fractional parts converge weakly to the uniform distribution on the unit interval. Weyl's criterion becomes: $(x_1, x_2, \ldots)$ is uniformly distributed modulo 1 if and only if $n^{-1} \sum_{j=1}^{n} e^{2\pi iux_j} \to 0$ for each nonzero integer u. This criterion is satisfied if $x_j = j\xi$, where ξ is irrational.

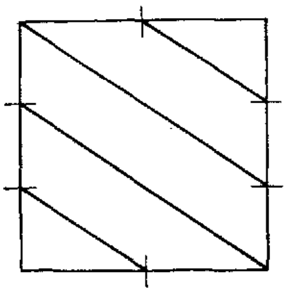
By the k-dimensional Weierstrass theorem, similar results hold for the k-dimensional torus—the product of k circles. (They even carry over to the general compact group, with the characters in the role of the circular functions.) Taking the case $k = 2$ and laying the torus out into a square, we see that, if P_n and P are probability measures in R^2 supported by the square $\{(x, y): 0 \le x, y < 1\}$, and if the lower and left-hand edges of the square have P-measure 0, then $P_n \Rightarrow P$ if and only if

$$\int e^{2\pi i(ux+vy)}\, dP_n \to \int e^{2\pi i(ux+vy)}\, dP$$

for all pairs of integers u and v.

We say a sequence $((x_1, y_1), (x_2, y_2), \ldots)$ in the plane is uniformly distributed modulo 1 if the empirical distributions, obtained by reducing all coordinates modulo 1, converge weakly to the uniform distribution on the square; this holds if and only if $n^{-1} \sum_{j=1}^{n} e^{2\pi i(ux_j+vy_j)} \to 0$ whenever u and v do not both vanish. This condition is satisfied if $(x_j, y_j) = (j\xi, j\eta)$, where ξ, η, and 1 are linearly independent (there do not exist integers u and v, not both 0, such that $u\xi + v\eta$ is integral). To take another case, suppose ξ and η satisfy exactly one constraint: $u\xi + v\eta$ is integral if and only if $u = 2w$ and $v = 3w$ for some integer w. Then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} e^{2\pi i(uj\xi+vj\eta)} = \begin{cases} 1 & \text{if } \dfrac{u}{2} = \dfrac{v}{3} \text{ is integral,} \\ 0 & \text{otherwise.} \end{cases}$$



By checking the Fourier coefficients, one sees that the empirical distributions converge weakly to a distribution uniform on the four slanting lines in the diagram.

The Cramér–Wold result has an obvious analogue here. For example, the planar sequence $((x_1, y_1), (x_2, y_2), \ldots)$ is uniformly distributed modulo 1 if and only if for all integral u and v, not both 0, the linear sequence $(ux_1 + vy_1, ux_2 + vy_2, \ldots)$ is uniformly distributed modulo 1.

Remarks. Theorem 7.2 is here proved by Lévy's version of Lindeberg's method; see Lévy (1925, pp. 246–249). See Feller (1966) for other approaches to the central limit theorem and for characteristic functions. There exists for linear spaces a theory of characteristic functions which ties in closely with weak convergence; see Prohorov (1960) and Gross (1963) and the references there.

See Cramér and Wold (1936) for their device. See Hardy and Wright (1960) for uniform distribution modulo 1, in particular for the last example here.

PROBLEMS

1. Replace the integrand in (7.1) by a $(2k)$th-degree polynomial so chosen that φ has bounded, continuous derivatives up to order k ($k = 3$ suffices for the proof of Lindeberg's theorem).

2. Use the Cramér–Wold device to derive a k-dimensional version of the Lindeberg-Lévy theorem.

3. A measurable function $h(x)$ on the line has a mean value if the limit

$$M\{h\} = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} h(x)\, dx$$

exists; it has a distribution if the probability measures (here λ is Lebesgue measure)

$$P_T(A) = \frac{1}{2T} \lambda\{x: |x| \le T,\, h(x) \in A\}, \qquad A \in \mathscr{R}^1,$$

converge weakly as $T \to \infty$. If h is almost periodic, then it is bounded and, for each t, e^{itx} is almost periodic (see Bohr (1932)); from the first proposition used in deriving the continuity theorem for characteristic functions, show that h has a distribution.

4. Consider on the line probability measures having moments of all orders. The measure P is said to be determined by its moments if $\int x^u\, P(dx) = \int x^u\, Q(dx)$, $u = 1, 2, \ldots$ implies $P = Q$. (See Feller (1966, pp. 224 and 487) for an example of a P not determined by its moments and for a criterion for P to be determined by its moments.) If $\sup_n |\int x^u\, P_n(dx)| < \infty$ for each u, and if $P_n \Rightarrow P$, then the moments of P_n converge to those of P [Theorem 5.4]. If the moments of P_n converge to those of P and if P is determined by its moments, then $P_n \Rightarrow P$. [First show $\{P_n\}$ is tight (Problem 5 in Section 6) and then imitate the proof given above for the continuity theorem for characteristic functions.] Use the Cramér–Wold technique to extend this result to R^k [consider product moments of all orders].

5. If k varies with n in such a way that $(k - np)/\sqrt{npq} \to x$, then

$$\binom{n}{k} p^k q^{n-k} npq \to \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

(see Feller (1957, p. 169)). Use Theorem 7.8 to deduce the De Moivre–Laplace theorem. Apply the same technique to the hypergeometric distribution (see Problem 2 on p. 17 and Problem 10 on p. 180 of Feller (1957)) and to the frequency count for a set of multinomial trials (leave out the count for one cell, so that the limiting distribution will have a density).

6. Use Problem 8 in Section 2 to prove that, if probability measures in R^k converge weakly, then the corresponding characteristic functions converge uniformly on bounded sets.

7. Generalize Theorem 7.8: Replace Lebesgue measure on R^k by a general measure on a separable S and consider decompositions $\{C_{ni}\}$ of S with $\lim_n \max_i \operatorname{diam} C_{ni} = 0$.

CHAPTER 2

The Space C

8. WEAK CONVERGENCE AND TIGHTNESS IN C

The Introduction explains some of the merits of proving weak convergence in the space $C = C[0, 1]$ of continuous functions on $[0, 1]$, where we give C the uniform topology by defining the distance between two points x and y (two functions x and y of $t \in [0, 1]$) as

$$\rho(x, y) = \sup_t |x(t) - y(t)|.$$

Weak Convergence

Although weak convergence in C does not in general follow from weak convergence of the finite-dimensional distributions, we saw at the beginning of Section 6 that it does in the presence of relative compactness. Since C is separable and complete (see p. 220) it follows by Prohorov's theorem (Theorems 6.1 and 6.2) that relative compactness of a family of probability measures on $(C, \mathscr{C})$ is equivalent to tightness of the family. Thus we have the following result.

THEOREM 8.1 *Let P_n, P be probability measures on $(C, \mathscr{C})$. If the finite-dimensional distributions of P_n converge weakly to those of P, and if $\{P_n\}$ is tight, then $P_n \Rightarrow P$.*

If we are to use this theorem to prove weak convergence in C, we must see just what tightness here means.

Tightness

The modulus of continuity of an element x of C is defined by

$$w_x(\delta) = w(x, \delta) = \sup_{|s-t|<\delta} |x(s) - x(t)|, \qquad 0 < \delta \leq 1. \tag{8.1}$$

Let $\{P_n\}$ be a sequence of probability measures on $(C, \mathscr{C})$.

THEOREM 8.2 *The sequence $\{P_n\}$ is tight if and only if these two conditions hold:*

(i) *For each positive η, there exists an a such that*

$$P_n\{x : |x(0)| > a\} \leq \eta, \qquad n \geq 1. \tag{8.2}$$

(ii) *For each positive ε and η, there exist a δ, with $0 < \delta < 1$, and an integer n_0 such that*

$$P_n\{x : w_x(\delta) \geq \varepsilon\} \leq \eta, \qquad n \geq n_0. \tag{8.3}$$

Condition (i) stipulates that $\{P_n \pi_0^{-1}\}$ be tight. In connection with (8.3), note that $w_x(\delta)$ is continuous in x and hence measurable.

Proof. Suppose $\{P_n\}$ is tight. Given ε and η, choose a compact set K such that $P_n(K) > 1 - \eta$ for all n. By the Arzelà–Ascoli theorem (p. 221), we have $K \subset \{x : |x(0)| \leq a\}$ for large enough a and $K \subset \{x : w_x(\delta) < \varepsilon\}$ for small enough δ, so that (i) and (ii) follow (with $n_0 = 1$ in (ii)). This proves the necessity of (i) and (ii).

Since a single probability measure P on $(C, \mathscr{C})$ is tight (Theorem 1.4), it follows by the necessity of (ii) that for each ε and η there is a δ such that $P\{x : w_x(\delta) \geq \varepsilon\} \leq \eta$. If $\{P_n\}$ satisfies condition (ii), therefore, we may ensure that the inequality in (8.3) holds for the finitely many n preceding n_0 by decreasing δ if necessary. Thus we may assume in proving sufficiency that the n_0 in (8.3) is always 1.

Assume that $\{P_n\}$ satisfies (i) and (ii), with n_0 always 1 in (8.3). Given η, choose a so that, if

$$A = \{x : |x(0)| \leq a\},$$

then $P_n(A) \geq 1 - \frac{1}{2}\eta$ for all n, and choose δ_k so that, if

$$A_k = \left\{x : w_x(\delta_k) < \frac{1}{k}\right\},$$

then $P_n(A_k) \geq 1 - \eta/2^{k+1}$ for all n. If K is the closure of $A \cap \bigcap_{k=1}^{\infty} A_k$, then $P_n(K) \geq 1 - \eta$ for all n and, by the Arzelà–Ascoli theorem, K is compact. Hence $\{P_n\}$ is tight.

As the proof shows, we may demand that (8.3) hold for *all* P_n. With this change, the theorem is true if $\{P_n\}$ is replaced by an arbitrary family Π. In applications, we shall often prove (8.3) with δ, ε, and η replaced by (say) $\frac{1}{2}\delta$, 3ε, and 9η, which is just as good.

Theorem 8.2 transforms the concept of tightness in C simply by substituting for compactness its Arzelà–Ascoli characterization. Our next theorem goes only a small step beyond this, but, even so, fills our present needs.

THEOREM 8.3 *The sequence $\{P_n\}$ is tight if these two conditions are satisfied:*

(i) *For each positive η, there exists an a such that*

(8.4) $$P_n\{x:|x(0)| > a\} \le \eta, \qquad n \ge 1.$$

(ii) *For each positive ε and η, there exist a δ, with $0 < \delta < 1$, and an integer n_0 such that*

(8.5) $$\frac{1}{\delta} P_n\left\{x: \sup_{t \le s \le t+\delta} |x(s) - x(t)| \ge \varepsilon\right\} \le \eta, \qquad n \ge n_0,$$

for all t.

Of course, we restrict the t in (8.5) to $0 \le t \le 1$; if $t > 1 - \delta$, we restrict s in the supremum to $t \le s \le 1$. Note that (8.5) is formally satisfied if $\delta > 1/\eta$; but we require $\delta < 1$.

Proof. Fix δ and let

$$A_t = \left\{x: \sup_{t \le s \le t+\delta} |x(s) - x(t)| \ge \varepsilon\right\}.$$

Now s and t each lie in intervals of the form $[i\delta, (i+1)\delta]$. If $|s - t| < \delta$, then these intervals either coincide or abut; it follows that

(8.6) $$P_n\{x: w_x(\delta) \ge 3\varepsilon\} \le P_n\left(\bigcup_{i < \delta^{-1}} A_{i\delta}\right).$$

Since

(8.7) $$P_n\left(\bigcup_{i < \delta^{-1}} A_{i\delta}\right) \le \sum_{i < \delta^{-1}} P_n(A_{i\delta}),$$

(8.5) implies $P_n\{x: w_x(\delta) \ge 3\varepsilon\} \le (1 + [1/\delta])\delta\eta < 2\eta$ (recall that $\delta < 1$).† Therefore condition (ii) here implies the corresponding condition in Theorem 8.2. Since condition (i) is unchanged, the result follows.

The following corollary contains the essential point of the proof.

COROLLARY *If $0 = t_0 < t_1 < \cdots < t_r = 1$, and if*

(8.8) $$t_i - t_{i-1} \ge \delta, \qquad 2 \le i \le r - 1,$$

then

(8.9) $$P\{x: w_x(\delta) \ge 3\varepsilon\} \le \sum_{i=1}^{r} P\left\{x: \sup_{t_{i-1} \le s \le t_i} |x(s) - x(t_{i-1})| \ge \varepsilon\right\}.$$

Note that the inequality in (8.8) need not hold for $i = 1$ and $i = r$.

† We use $[\alpha]$ to denote the integer part of α—the largest integer not larger than α.

Although the inequality (8.6) entails no real loss, (8.7) does, and condition (ii) is not necessary.† Suppose, however, that the sets $A_{i\delta}$ (δ fixed, i varying) are independent under P_n and all have the same measure p, which will be at least approximately true for many of the measures P_n encountered in this chapter. Then

$$P\left(\bigcup_{i<\delta^{-1}} A_{i\delta}\right) = 1 - (1 - p)^{(1+[1/\delta])}.$$

If this quantity is bounded above by η, then (if $\eta \leq \frac{1}{2}$)

$$\frac{p}{\delta} \leq \left(1 + \left[\frac{1}{\delta}\right]\right)p \leq -\log(1 - \eta) \leq 2\eta, \tag{8.10}$$

which is a restriction of the same order as (8.5). Therefore any effective tightness criterion that goes beyond Theorem 8.3 must make some sort of positive use of dependence among the $A_{i\delta}$.

Random Functions

Let X be a mapping from $(\Omega, \mathscr{B}, \mathsf{P})$ into C. For the moment we do not assume X to be measurable ($X^{-1}\mathscr{C} \subset \mathscr{B}$). For each ω in Ω, $X(\omega)$ is an element of C—a continuous function on $[0, 1]$ whose value at t we denote by $X(t, \omega)$. For fixed t, let $X(t)$ denote the real function on Ω with value $X(t, \omega)$ at ω; $X(t)$ is the composition $\pi_t X$. Similarly, let $(X(t_1), \ldots, X(t_k))$ denote the mapping from Ω into R^k with value $(X(t_1, \omega), \ldots, X(t_k, \omega))$ at ω.

If $A = \{x \in C : x(t) \leq \alpha\}$, then $A \in \mathscr{C}$ and $\{\omega : X(t, \omega) \leq \alpha\} = X^{-1}A$. It follows that if X is a random element of C—that is, if $X^{-1}\mathscr{C} \subset \mathscr{B}$—then each $X(t)$ is a random variable ($X(t)^{-1}\mathscr{R}^1 \subset \mathscr{B}$) and hence each $(X(t_1), \ldots, X(t_k))$ is a random vector. Suppose, on the other hand, that each $X(t)$ is a random variable. If B is the closed sphere in C with center x and radius δ, then $X^{-1}B = \{\omega : X(\omega) \in B\} = \bigcap_r \{\omega : x(r) - \delta \leq X(r, \omega) \leq x(r) + \delta\}$, where the intersection extends over the rationals, so that $X^{-1}B \in \mathscr{B}$. Since C is separable, $X^{-1}\mathscr{C} \subset \mathscr{B}$ follows: X is a random element of C. Thus X is a random function (a random element of C) if and only if each $X(t)$ is a random variable. (We have merely proved once more that the finite-dimensional sets generate $\mathscr{C}$.)

Suppose now that $\{X_n\}$ is a sequence of random functions. The sequence is by definition tight when the sequence of corresponding distributions is tight. According to Theorem 8.2, $\{X_n\}$ is tight if and only if the sequence $\{X_n(0)\}$ is tight (on the line) and for each positive ε and η there exist a δ, $0 < \delta < 1$,

† See Problem 1.

and an integer n_0 such that

$$\mathsf{P}\{w(X_n, \delta) \geq \varepsilon\} \leq \eta, \qquad n \geq n_0. \tag{8.11}$$

(We use $w(x, \delta)$ as an alternate notation for $w_x(\delta)$.) This condition stipulates that the random functions X_n do not oscillate too violently.

Theorem 8.3 can be recast in the same way: $\{X_n\}$ is tight if $\{X_n(0)\}$ is tight and if for each positive ε and η there exist a δ, $0 < \delta < 1$, and an integer n_0 such that

$$\frac{1}{\delta}\mathsf{P}\left\{\sup_{t \leq s \leq t+\delta} |X_n(s) - X_n(t)| \geq \varepsilon\right\} \leq \eta \tag{8.12}$$

for $n \geq n_0$ and $0 \leq t \leq 1$ (with s in the supremum restricted to $t \leq s \leq 1$ in case $1 - \delta < t \leq 1$).

As explained in the Introduction, our first interest will be in random functions constructed in the following way. Let $\xi_1, \xi_2, \ldots,$ be random variables on some probability space $(\Omega, \mathscr{B}, \mathsf{P})$. For the present, the ξ_n need not have any special properties such as stationarity and independence. We define $S_n = \xi_1 + \cdots + \xi_n$, with $S_0 = 0$, and construct X_n from the partial sums $S_0, S_1, \ldots, S_n$. For points i/n in $[0, 1]$, we set

$$X_n\left(\frac{i}{n}, \omega\right) = \frac{1}{\sigma\sqrt{n}} S_i(\omega). \tag{8.13}$$

(We confine our attention to norming factors of the form $\sigma\sqrt{n}$—others are possible.) For the remaining points t of $[0, 1]$, we define $X_n(t, \omega)$ by linear interpolation: If $t \in [(i-1)/n, i/n]$, then

$$\begin{aligned} X_n(t, \omega) &= \frac{(i/n) - t}{1/n} X_n\left(\frac{i-1}{n}\right) + \frac{t - (i-1)/n}{1/n} X_n\left(\frac{i}{n}\right) \\ &= \frac{1}{\sigma\sqrt{n}} S_{i-1}(\omega) + n\left(t - \frac{i-1}{n}\right) \frac{1}{\sigma\sqrt{n}} \xi_i(\omega). \end{aligned} \tag{8.14}$$

Since $i - 1 = [nt]$ if $t \in [(i-1)/n, i/n)$, we may define the function more concisely by

$$X_n(t, \omega) = \frac{1}{\sigma\sqrt{n}} S_{[nt]}(\omega) + (nt - [nt]) \frac{1}{\sigma\sqrt{n}} \xi_{[nt]+1}(\omega). \tag{8.15}$$

Since the ξ_i, and hence the S_i, are random variables, it follows by (8.15) that $X_n(t)$ is a random variable for each t. Therefore the X_n are random functions.

When do these random functions form a tight sequence? Since $X_n(0) = 0$, certainly $\{X_n(0)\}$ is tight. By using the definition (8.15) we can translate (8.12) into a restriction on the fluctuations of the partial sums S_i. If $t = k/n$

and $t + \delta = j/n$, with k and j integers, then (8.12) reduces to

$$\frac{1}{\delta}\,\mathsf{P}\left\{\max_{i \le n\delta} \frac{1}{\sigma\sqrt{n}}\,|S_{k+i} - S_k| \ge \varepsilon\right\} \le \eta. \tag{8.16}$$

Although (8.12) and (8.16) in general differ if t and δ are not integral multiples of $1/n$, the discrepancy is, as we shall show, irrelevant to our purposes. Consider the integers j and k defined by the inequalities

$$\frac{k}{n} \le t < \frac{k+1}{n}, \qquad \frac{j-1}{n} \le t + \frac{\delta}{2} < \frac{j}{n}.$$

Because of the polygonal character of X_n, we have

$$\sup_{t \le s \le t + \frac{1}{2}\delta} |X_n(s) - X_n(t)| \le 2 \max_{0 \le i \le j-k} \frac{1}{\sigma\sqrt{n}}\,|S_{k+i} - S_k|.$$

If $n \ge 4/\delta$, then $j - k < n\delta$, so that the maximum on the right does not decrease if the restriction $i \le j - k$ is relaxed to $i \le n\delta$. Therefore, if (8.16) holds for all k and all $n \ge n_0$, then (8.12) holds for all t and all $n \ge \max\{n_0, 4/\delta\}$, provided the ε, η, and δ in (8.12) are replaced by 2ε, 2η, and $\frac{1}{2}\delta$, respectively, which amounts only to a renaming of these quantities.

Thus $\{X_n\}$ is tight if, for each positive ε and η, there exist a δ, with $0 < \delta < 1$, and an integer n_0, such that (8.16) holds for all k and for all $n \ge n_0$. The maximum in (8.16), which extends over $i \le n\delta$, becomes easier to work with if $n\delta$ is replaced by an integer. If $n\delta$ is an integer m—it need not be—then the inequality (8.16) becomes

$$\mathsf{P}\left\{\max_{i \le m} |S_{k+i} - S_k| \ge \frac{\varepsilon}{\sqrt{\delta}}\,\sigma\sqrt{m}\right\} \le \eta\delta.$$

Put $\lambda = \varepsilon/\sqrt{\delta}$ (if δ is small, λ is large); the inequality then further reduces to

$$\mathsf{P}\left\{\max_{i \le m} |S_{k+i} - S_k| \ge \lambda\sigma\sqrt{m}\right\} \le \frac{\eta\varepsilon^2}{\lambda^2}.$$

Since $\eta\varepsilon^2$ is positive if η and ε are, we are led to formulate the following theorem.

THEOREM 8.4 *Suppose $\{X_n\}$ is defined by* (8.15). *The sequence $\{X_n\}$ is tight if for each positive ε there exist a λ, with $\lambda > 1$, and an integer n_0 such that, if $n \ge n_0$, then*

$$\mathsf{P}\left\{\max_{i \le n} |S_{k+i} - S_k| \ge \lambda\sigma\sqrt{n}\right\} \le \frac{\varepsilon}{\lambda^2} \tag{8.17}$$

holds for all k.

We require $\lambda > 1$ ((8.17) is trivial if $\lambda \le \sqrt{\varepsilon}$), which corresponds to the requirement $\delta < 1$ in Theorem 8.3. In concrete cases, proving (8.17) forces large λ's on us.

Proof. Given ε and η, we shall produce a δ $(0 < \delta < 1)$ and an n_0 for which (8.16) holds for all k if $n \ge n_0$. Since (8.16) becomes more stringent as ε and η decrease, we may assume $0 < \varepsilon, \eta < 1$.

By the hypothesis, with $\eta\varepsilon^2$ in place of ε, there exist λ $(\lambda > 1)$ and n_1 such that

$$\mathsf{P}\left\{\max_{i \le n} |S_{k+i} - S_k| \ge \lambda\sigma\sqrt{n}\right\} \le \frac{\eta\varepsilon^2}{\lambda^2} \tag{8.18}$$

for $n \ge n_1$ and $k \ge 1$. Put $\delta = \varepsilon^2/\lambda^2$; since $\lambda > 1 > \varepsilon$, we have $0 < \delta < 1$. Let n_0 be an integer exceeding n_1/δ.

If $n \ge n_0$, then $[n\delta] \ge n_1$, and it follows by (8.18) that

$$\mathsf{P}\left\{\max_{i \le [n\delta]} |S_{k+i} - S_k| \ge \lambda\sigma\sqrt{[n\delta]}\right\} \le \frac{\eta\varepsilon^2}{\lambda^2}.$$

Since $\lambda\sqrt{[n\delta]} \le \varepsilon\sqrt{n}$ and $\eta\varepsilon^2/\lambda^2 = \eta\delta$, (8.16) holds for all k if $n \ge n_0$. This proves Theorem 8.4.

It is not hard to see that it is enough that (8.17) hold for $k \le n\lambda^2/\varepsilon$, but we shall not need this. If $\{\xi_n\}$ is stationary, then (8.17) reduces to

$$\mathsf{P}\left\{\max_{i \le n} |S_i| \ge \lambda\sigma\sqrt{n}\right\} \le \frac{\varepsilon}{\lambda^2}. \tag{8.19}$$

We can absorb σ into λ and require

$$\mathsf{P}\left\{\max_{i \le n} |S_i| \ge \lambda\sqrt{n}\right\} \le \frac{\varepsilon}{\lambda^2} \tag{8.20}$$

with $\lambda > \sigma$.

We shall see later that in some circumstances the hypothesis of Theorem 8.4 is necessary as well as sufficient.

Coordinate Variables

The projection π_t, with value $\pi_t(x) = x(t)$ at $x \in C$, is a random variable on $(C, \mathscr{C})$. We shall often denote this random variable by x_t: For fixed t, x_t is the function on C with value $x(t)$ at x. If there is a probability measure P on $(C, \mathscr{C})$, $\{x_t : 0 \le t \le 1\}$ is a stochastic process. We think of t as a time parameter, and the x_t are commonly called the coordinate variables or functions. The distribution of x_t depends on the measure P, and we shall speak of the distribution of x_t *under* P; we shall often write $P\{x_t \in H\}$ in place of $P\{x: x_t \in H\}$ and $\int x_t\, dP$ in place of $\int x_t\, P(dx)$. Finally, when t is a complicated expression, we shall often revert back to $x(t)$, still intended as a coordinate function.

Remarks. The general theory of weak convergence in C (Theorem 8.2, for example) is due to Prohorov (1956). Theorem 8.4, a convenient reformulation of Theorem 8.3, resulted from discussions with F. Topsøe.

Stone (1963) treats weak convergence in a space of continuous functions on $[0, \infty)$. Lamperti (1962b) discusses spaces of functions satisfying a Hölder condition.

PROBLEMS

1. Let $X(t) = t\xi$, where ξ is a random variable with $\mathsf{P}\{|\xi| \geq \alpha\} \sim \alpha^{-\frac{1}{2}}$ as $\alpha \to \infty$, and, for every n, let P_n be the distribution of X. Then $\{P_n\}$ is tight but does not satisfy condition (ii) of Theorem 8.3.

9. THE EXISTENCE OF WIENER MEASURE

Wiener Measure

Wiener measure, denoted here by W, is a probability measure on $(C, \mathscr{C})$ having the following two properties. First, for each t, the random variable x_t is normally distributed under W with mean 0 and variance t:

$$W\{x_t \leq \alpha\} = \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\alpha} e^{-u^2/2t}\, du. \tag{9.1}$$

If $t = 0$, this is interpreted to mean $W\{x_0 = 0\} = 1$. Second, the stochastic process $\{x_t: 0 \leq t \leq 1\}$ has independent increments under W: If

$$0 \leq t_0 \leq t_1 \leq \cdots \leq t_k \leq 1, \tag{9.2}$$

then the random variables

$$x_{t_1} - x_{t_0},\, x_{t_2} - x_{t_1},\, \ldots,\, x_{t_k} - x_{t_{k-1}} \tag{9.3}$$

are independent under W. In this section, we prove there does exist such a measure W.

If W has these two properties and if $s \leq t$, then x_t (normal with mean 0 and variance t) is the sum of the independent variables x_s (normal with mean 0 and variance s) and $x_t - x_s$, so that $x_t - x_s$ must be normal with mean 0 and variance $t - s$, as may be seen by dividing the characteristic functions. Thus, when (9.3) holds,

$$W\{x_{t_i} - x_{t_{i-1}} \leq \alpha_i,\ i = 1, \ldots, k\} = \prod_{i=1}^{k} \frac{1}{\sqrt{2\pi(t_i - t_{i-1})}} \int_{-\infty}^{\alpha_i} e^{-u^2/2(t_i - t_{i-1})}\, du. \tag{9.4}$$

In particular, the increments are stationary (the distribution of $x_t - x_s$ under W depends only on the difference $t - s$) as well as independent.

If we regard $x(t)$ as (one coordinate of) the position at time t of a moving particle, then x itself gives the history of the particle's motion (relative to this specific coordinate) from time $t = 0$ to time $t = 1$. Wiener measure gives to these paths x a distribution appropriate for the description of Brownian motion—the motion of a pollen grain suspended in water.

In proving the existence of W, we face the problem of proving the existence on $(C, \mathscr{C})$ of a probability measure with specified finite-dimensional distributions. There can for an arbitrary specification be at most one such measure, and for some specifications there is none at all (there is, for example, no P under which the distribution of x_t is a unit mass at 0 for $t < \frac{1}{2}$ and at 1 for $t \geq \frac{1}{2}$).

THEOREM 9.1 *There exists on* $(C, \mathscr{C})$ *a probability measure* W *such that* (9.1) *holds and such that the random variables* (9.3) *are independent under* W *whenever* (9.2) *holds.*

Proof. Let $\xi_1, \xi_2, \ldots$ be independent and normally distributed (on some $(\Omega, \mathscr{B}, \mathsf{P})$) with mean 0 and variance 1. Let X_n be the random function defined by (8.15) with $\sigma = 1$:

$$X_n(t, \omega) = \frac{1}{\sqrt{n}} S_{[nt]}(\omega) + (nt - [nt]) \frac{1}{\sqrt{n}} \xi_{[nt]+1}(\omega). \tag{9.5}$$

Let P_n of the distribution of X_n on C. This measure is well defined because, as we showed in the preceding section, X_n is measurable ($X_n^{-1}\mathscr{C} \subset \mathscr{B}$).

We shall first show that the finite-dimensional distributions of the P_n converge weakly to what we want the finite-dimensional distributions of W to be. After that, we shall prove that the sequence $\{P_n\}$ is tight. It is clear that the limit of any weakly convergent subsequence $\{P_{n'}\}$ will satisfy the requirements placed on W. The idea is that the P_n approximate the putative measure W.

The finite-dimensional distribution $P_n \pi_{t_1 \cdots t_k}^{-1}$ is just the distribution of the random vector $(X_n(t_1), \ldots, X_n(t_k))$. Consider a single time point t. By (9.5) and the assumed normality of the ξ_n, $X_n(t)$ is normal with mean 0 and variance

$$\frac{[nt]}{n} + \frac{(nt - [nt])^2}{n};$$

this variance differs from t by at most $2/n$. Thus $X_n(t) \xrightarrow{\mathscr{D}} N(0, t)$ (see (4.9) for this notation).

Clearly, we can treat two or more time points in the same way; the finite-dimensional distributions therefore converge weakly to those prescribed for W.

To prove that $\{P_n\}$ is tight, we apply Theorem 8.4. To prove (8.20), write

$$E_i = \left\{\max_{j<i} |S_j| < 2\lambda\sqrt{n} \leq |S_i|\right\}. \tag{9.6}$$

By the stationarity and independence of the ξ_n, we have

$$\begin{aligned}
\mathsf{P}\left\{\max_{i\leq n} |S_i| \geq 2\lambda\sqrt{n}\right\} &\leq \mathsf{P}\{|S_n| \geq \lambda\sqrt{n}\} \\
&\quad + \sum_{i=1}^{n-1} \mathsf{P}(E_i \cap \{|S_n - S_i| \geq \lambda\sqrt{n}\}) \\
&= \mathsf{P}\{|S_n| \geq \lambda\sqrt{n}\} + \sum_{i=1}^{n-1} \mathsf{P}(E_i)\mathsf{P}\{|S_{n-i}| \geq \lambda\sqrt{n}\}, \\
&\leq \mathsf{P}\{|S_n| \geq \lambda\sqrt{n}\} + \sum_{i=1}^{n-1} \mathsf{P}(E_i)\mathsf{P}\{|S_{n-i}| \geq \lambda\sqrt{n-i}\}
\end{aligned} \tag{9.7}$$

Since $S_j/\sqrt{j}$ is normally distributed with mean 0 and variance 1 and hence has a finite third absolute moment a independent of j ($a = 2\sqrt{2/\pi}$), it follows by (9.7) and Chebyshev's inequality that

$$\mathsf{P}\left\{\max_{i\leq n} |S_i| \geq 2\lambda\sqrt{n}\right\} \leq \frac{a}{\lambda^3} + \sum_{i=1}^{n-1} \mathsf{P}(E_i)\frac{a}{\lambda^3}.$$

Since the E_i are disjoint, we can conclude

$$\mathsf{P}\left\{\max_{i\leq n} |S_i| \geq 2\lambda\sqrt{n}\right\} \leq \frac{2a}{\lambda^3}, \qquad n = 1, 2, \ldots. \tag{9.8}$$

Given ε, choose λ so that $2a/\lambda < \varepsilon$ and $\lambda > 1$. Then (9.8) implies

$$\mathsf{P}\left\{\max_{i\leq n} |S_i| \geq 2\lambda\sqrt{n}\right\} \leq \frac{\varepsilon}{\lambda^2},$$

which is (8.20), except for the irrelevant factor 2 on the left. By Theorem 8.4, therefore, the $\{P_n\}$ are tight, which completes the proof.

Thus Wiener measure exists. In the sections following, we shall derive some facts about W. For the present, let us only show that the elements of C are of unbounded variation, except for a set of Wiener measure 0. This is interesting because it exhibits a sense in which Wiener paths (elements of C chosen according to the probability measure W) are highly irregular—they are continuous, but only just.

An element x of C is of bounded variation if and only if there exists a number M_x such that

$$\sum_{i=1}^{k} |x(t_i) - x(t_{i-1})| \leq M_x$$

for all $t_0, t_1, \ldots, t_k$ with $0 \le t_0 \le t_1 \le \cdots \le t_k \le 1$. Therefore it will suffice to show that, if

$$f_n(x) = \sum_{i=1}^{2^n} \left| x\left(\frac{i}{2^n}\right) - x\left(\frac{i-1}{2^n}\right) \right|,$$

then $f_n(x) \to \infty$ except for x in a set of W-measure 0. Now $|N|$ (see (4.10)) has positive mean b and variance c ($b = \sqrt{2/\pi}$ and $c = 1 - 2/\pi$). Since the $|x(i2^{-n}) - x((i-1)2^{-n})|$ are, under W, independent with mean $2^{-n/2}b$ and variance $2^{-n}c$, the mean and variance of $f_n(x)$ are $2^{n/2}b$ and c. It follows by Chebyshev's inequality that $W\{x : f_n(x) > \alpha\} \to 1$ for each α. Since $f_{n+1}(x) \ge f_n(x)$, it follows further that $f_n(x) \to \infty$ except on a set of Wiener measure 0.

We shall also use the symbol W to denote a random element with values in C and with Wiener measure as its distribution; that is, W is a measurable mapping from some $(\Omega, \mathscr{B}, \mathsf{P})$ to C with the property that

$$\mathsf{P}\{\omega : W(\omega) \in A\} = W(A), \qquad A \in \mathscr{C},$$

where the W on the right is Wiener measure as already defined. That there exists such a random element follows by the argument centering on (4.4). This dual use of W should cause no confusion.

We shall denote by $W_t(\omega)$ or $W(t, \omega)$ the value at t of the random function $W(\omega)$. Thus $\{W_t : 0 \le t \le 1\}$ is a stochastic process whose paths are continuous for *all* ω. This process we shall call the *Wiener process* or the *Brownian motion process*. The relation $X_n \xrightarrow{\mathscr{D}} W$ can be interpreted in accordance with (4.5) or in accordance with (4.7), depending on whether W is construed as a random function or as a measure—the meaning is the same for the two interpretations.

The Brownian Bridge

A random element X of C is *Gaussian* if all its finite-dimensional distributions are normal. The distribution in C of a Gaussian random element is completely specified by the means $\mathsf{E}\{X(t)\}$ and product moments $\mathsf{E}\{X(s)X(t)\}$ $0 \le s,\ t \le 1$, because these determine the finite-dimensional distributions. For W, the moments are

(9.9) $$\mathsf{E}\{W_t\} = 0$$

and

(9.10) $$\mathsf{E}\{W_s W_t\} = s \qquad \text{if} \quad s \le t.$$

In studying the behavior of empirical distribution functions, we shall need a Gaussian random element W° of C whose distribution is specified by the requirements

$$\mathsf{E}\{W_t^\circ\} = 0 \tag{9.11}$$

and

$$\mathsf{E}\{W_s^\circ W_t^\circ\} = s(1-t) \quad \text{if} \quad s \le t. \tag{9.12}$$

Although we could prove the existence of such a W° by the methods of Theorem 9.1, it is simpler to construct W° from W by setting

$$W_t^\circ = W_t - tW_1, \qquad 0 \le t \le 1. \tag{9.13}$$

Certainly W°, thus defined, is a Gaussian random element of C, and (9.11) and (9.12) follow from (9.9) and (9.10).

The random element W° is called the *Brownian bridge* (or tied-down Brownian motion). Note that $W_0^\circ = W_1^\circ = 0$ with probability 1. From (9.11) and (9.12) it follows that

$$\mathsf{E}\{(W_t^\circ - W_s^\circ)^2\} = (t-s)(1-(t-s)) \quad \text{if} \quad s \le t \tag{9.14}$$

and that

$$\begin{aligned} \mathsf{E}\{(W_{s_2}^\circ - W_{s_1}^\circ)(W_{t_2}^\circ - W_{t_1}^\circ)\} \\ = -(s_2 - s_1)(t_2 - t_1) \quad \text{if} \quad s_1 \le s_2 \le t_1 \le t_2. \end{aligned} \tag{9.15}$$

We shall also use W° to denote the distribution on C of the random element W°. If $h: C \to C$ takes the function x to the function with value $x(t) - tx(1)$ at t, then the measures W° and W are related by $W^\circ = Wh^{-1}$.

Separable Stochastic Processes†

Let us connect our construction of Wiener measure with the notion of a separable stochastic process. According to Kolmogorov's existence theorem (p. 230), there exists a stochastic process $\{\xi_t: 0 \le t \le 1\}$, on some probability space $(\Omega, \mathscr{B}, \mathsf{P})$, for which the distribution of each finite collection of the ξ_t coincides with the corresponding finite-dimensional distribution prescribed for W. In the standard construction, $(\Omega, \mathscr{B})$ is the product of a collection of copies of $(R^1, \mathscr{R}^1)$, one copy for each t in $[0, 1]$, and the ξ_t are the coordinate functions.

The set Ω_0 of ω for which $\xi_t(\omega)$ is continuous in t, $0 \le t \le 1$, need not satisfy $\mathsf{P}(\Omega_0) = 1$; indeed, in the standard construction Ω_0 does not even lie in $\mathscr{B}$. It is possible, however, by another procedure, to find a *separable* process with the prescribed finite-dimensional distributions. The process $\{\xi_t: 0 \le t \le 1\}$ on $(\Omega, \mathscr{B}, \mathsf{P})$ is separable‡ if $\mathscr{B}$ is complete relative to P and

† The rest of this section may be omitted.

‡ See Doob (1953, p. 51).

if there exists in $\mathscr{B}$ a set E of measure 0 and in $[0, 1]$ a countable set T_0 such that, for each α and β and for each open interval I, we have

(9.16)
$$\{\omega : \alpha \leq \xi_t(\omega) \leq \beta, t \in I \cap T_0\} - \{\omega : \alpha \leq \xi_t(\omega) \leq \beta, t \in I \cap T\} \subset E,$$

where we have written T in place of $[0, 1]$. The left-hand set in the difference here lies in $\mathscr{B}$ because T_0 is countable; since $\mathsf{P}(E) = 0$ and $\mathscr{B}$ is complete, the right-hand set in the difference must lie in $\mathscr{B}$ also and have the same probability.

If $\{\xi_t\}$ is separable and has the finite-dimensional distributions appropriate to Brownian motion, then† the set Ω_0 of ω for which the sample path $\big(\xi_t(\omega)$ as a function of $t\big)$ is continuous lies in $\mathscr{B}$ and satisfies $\mathsf{P}(\Omega_0) = 1$. If we map Ω_0 into C by carrying ω into its sample path, and if we carry P (restricted to Ω_0) over to $(C, \mathscr{C})$ via this mapping, we arrive at W, which gives another method of constructing Wiener measure.

The problem of proving the continuity of the sample paths of separable processes and the problem of constructing measures on $(C, \mathscr{C})$ are effectively the same. If each separable process having certain specified finite-dimensional distributions also has continuous sample paths with probability 1, then, as the above construction shows, there exists on $(C, \mathscr{C})$ a probability measure with these finite-dimensional distributions. Let us prove the converse.

THEOREM 9.2 *If* $\{\xi_t : 0 \leq t \leq 1\}$ *is a separable stochastic process, and if there exists on* $(C, \mathscr{C})$ *a probability measure with the same finite-dimensional distributions as* $\{\xi_t\}$, *then the sample paths of the process are continuous with probability* 1.

Proof. In proving this result it is no restriction to assume that T_0 contains all the rationals in the unit interval, in which case (9.16) holds if I is a closed interval with rational endpoints. We shall use the relations

(9.17) $$\left(\bigcup_\theta E_\theta\right) + \left(\bigcup_\theta E'_\theta\right) \subset \bigcup_\theta \left(E_\theta + E'_\theta\right)$$

and

(9.18) $$\left(\bigcap_\theta E_\theta\right) + \left(\bigcap_\theta E'_\theta\right) \subset \bigcup_\theta \left(E_\theta + E'_\theta\right),$$

each of which is valid whatever the range of the index θ.‡

Let V denote the general system

(9.19) $$V: k; r_0, \ldots, r_k; \alpha_1, \ldots, \alpha_k,$$

where k is an arbitrary integer, the r_i and the α_i are rational, and $0 = r_0 < \cdots < r_k = 1$. There are countably many such systems V. For V given

† See Doob (1953, p. 393).

‡ Recall that $+$ stands for symmetric difference.

by (9.19) and ε positive, define

$$(9.20) \qquad \Omega_{T_0}(V, \varepsilon) = \bigcap_{i=1}^{k} \{\omega : \alpha_i \le \xi_t(\omega) \le \alpha_i + \varepsilon,\ t \in [r_{i-1}, r_i] \cap T_0\}$$

and

$$(9.21) \qquad \Omega_{T_0} = \bigcap_{\varepsilon} \bigcup_{V} \Omega_{T_0}(V, \varepsilon),$$

where the intersection extends over positive, rational ε. Let $\Omega_T(V, \varepsilon)$ and Ω_T be the same sets but with T_0 replaced by $T = [0, 1]$ in (9.20) and (9.21).

Since $\Omega_T(V, \varepsilon) + \Omega_{T_0}(V, \varepsilon) \subset E$, (9.17) and (9.18) imply

$$(9.22) \qquad \Omega_T + \Omega_{T_0} \subset E.$$

Now $\Omega_{T_0}(V, \varepsilon)$ is a countable intersection of sets in $\mathscr{B}$ and hence itself lies in $\mathscr{B}$. Therefore $\Omega_{T_0} \in \mathscr{B}$. It follows by (9.22) that Ω_T lies in $\mathscr{B}$ and $\mathsf{P}(\Omega_T) = \mathsf{P}(\Omega_{T_0})$. Since Ω_T is exactly the set of ω with continuous paths, we need only prove

$$(9.23) \qquad \mathsf{P}(\Omega_{T_0}) = 1.$$

Let $(t_1, t_2, \ldots)$ be an enumeration of the points in T_0. For V given by (9.19) and ε positive, let $H_{T_0}(V, \varepsilon)$ denote the set of points $z = (z_1, z_2, \ldots)$ in R^∞ such that, for each $i = 1, \ldots, k$, the inequality $\alpha_i \le z_u \le \alpha_i + \varepsilon$ holds for every coordinate index u for which $t_u \in [r_{i-1}, r_i]$. Then $H_{T_0}(V, \varepsilon) \in \mathscr{R}^\infty$. If the mapping $\varphi : \Omega \to R^\infty$ is defined by $\varphi(\omega) = (\xi_{t_1}(\omega), \xi_{t_2}(\omega), \ldots)$, then $\varphi^{-1} H_{T_0}(V, \varepsilon) = \Omega_{T_0}(V, \varepsilon)$. If $H_{T_0} = \bigcap_\varepsilon \bigcup_V H_{T_0}(V, \varepsilon)$, then $H_{T_0} \in \mathscr{R}^\infty$ and

$$(9.24) \qquad \varphi^{-1} H_{T_0} = \Omega_{T_0}.$$

Now define $\psi : C \to R^\infty$ by $\psi(x) = (x(t_1), x(t_2), \ldots)$. Let P be the probability measure on $(C, \mathscr{C})$ whose finite-dimensional distributions coincide with those of $\{\xi_t\}$. If H is a finite-dimensional set in R^∞, then

$$(9.25) \qquad \mathsf{P}(\varphi^{-1} H) = P(\psi^{-1} H);$$

since the finite-dimensional sets form a field generating $\mathscr{R}^\infty$, (9.25) holds also for every H in $\mathscr{R}^\infty$. By (9.24), therefore, $\mathsf{P}(\Omega_{T_0}) = P(\psi^{-1} H_{T_0})$. But $\psi^{-1} H_{T_0} = C$, from which (9.23) follows.

Thus proving continuity for separable processes and constructing measures on C are the same activity.

Remarks. See Itô and McKean (1965) for other constructions of Wiener measure and for some history. For an interesting account of Wiener's early work, see the articles in "Norbert Wiener, 1894–1964," *Bull. Amer. Math. Soc.* **72** (1966), number 1, part II.

PROBLEMS

1. Let ξ_r, with r ranging over the rationals in $[0, 1]$, be random variables with the finite-dimensional distributions appropriate to Brownian motion. By adapting the arguments in the proof of Theorem 9.1 but avoiding the tightness concept (and indeed all mention of the space C), show that $\{\xi_r\}$ is, with probability 1, uniformly continuous in r. For t irrational, define $\xi_t = \lim_{r \to t} \xi_r$, and, by the argument preceding Theorem 9.2, construct Wiener measure on C anew.

2. Show that Wiener measure has no locally compact support.

3. For $0 \leq t < \infty$, define $V_t = (1 + t) W^\circ_{t/(1+t)}$, where W° is the Brownian bridge. The process $\{V_t : t \geq 0\}$ has sample paths that are continuous for all ω, the finite-dimensional distributions are Gaussian, and the moments are $\mathsf{E}\{V_t\} = 0$ and $\mathsf{E}\{V_s V_t\} = \min(s, t)$. The process thus represents a Brownian motion over the time interval $[0, \infty)$.

10. DONSKER'S THEOREM

The Theorem

Given random variables $\xi_1, \xi_2, \ldots$ defined on $(\Omega, \mathscr{B}, \mathsf{P})$, let $S_n = \xi_1 + \cdots + \xi_n$ be the partial sums and define a random element X_n of C by (8.15):

$$X_n(t, \omega) = \frac{1}{\sigma\sqrt{n}} S_{[nt]}(\omega) + (nt - [nt]) \frac{1}{\sigma\sqrt{n}} \xi_{[nt]+1}(\omega). \tag{10.1}$$

The following theorem on the convergence in distribution of the X_n is due to Donsker. The Introduction has examples of its use.

THEOREM 10.1 *Suppose the random variables ξ_n are independent and identically distributed with mean* 0 *and finite, positive variance σ^2:*

$$\mathsf{E}\{\xi_n\} = 0, \qquad \mathsf{E}\{\xi_n^2\} = \sigma^2. \tag{10.2}$$

Then the random functions X_n defined by (10.1) *satisfy*

$$X_n \xrightarrow{\mathscr{D}} W. \tag{10.3}$$

Proof. We first show that the finite-dimensional distributions of the X_n converge to those of W. Consider first a single time point s; we must prove

$$X_n(s) \xrightarrow{\mathscr{D}} W_s. \tag{10.4}$$

Since

$$\left| X_n(s) - \frac{1}{\sigma\sqrt{n}} S_{[ns]} \right| \leq \frac{1}{\sigma\sqrt{n}} \xi_{[ns]+1} \xrightarrow{P} 0 \tag{10.5}$$

by Chebyshev's inequality, (10.4) will follow by Theorem 4.1 if we prove

$$\frac{1}{\sigma\sqrt{n}} S_{[ns]} \xrightarrow{\mathscr{D}} W_s.$$

But this is a direct consequence of the Lindeberg–Lévy central limit theorem (Theorem 7.4) and the fact that $[ns]/n \to s$.

Consider now two time points s and t with $s < t$. We are to prove

$$(X_n(s), X_n(t)) \xrightarrow{\mathscr{D}} (W_s, W_t),$$

which will follow by Corollary 1 to Theorem 5.1 if we prove

$$(X_n(s), X_n(t) - X_n(s)) \xrightarrow{\mathscr{D}} (W_s, W_t - W_s).$$

Because of (10.5) and the same relation with t in place of s, it is enough to prove

$$\left(\frac{1}{\sigma\sqrt{n}} S_{[ns]}, \frac{1}{\sigma\sqrt{n}} (S_{[nt]} - S_{[ns]})\right) \xrightarrow{\mathscr{D}} (W_s, W_t - W_s). \tag{10.6}$$

Since the components on the left are independent, by the independence of the ξ_n, (10.6) follows by the central limit theorem and Theorem 3.2. A set of three or more time points can be treated in the same way, and hence the finite-dimensional distributions converge properly.

We prove tightness via the following lemma, which is slightly more general than presently required.

LEMMA *Let $\xi_1, \ldots, \xi_m$ be independent random variables with mean 0 and finite variances σ_i^2; put $S_i = \xi_1 + \cdots + \xi_i$ and $s_i^2 = \sigma_1^2 + \cdots + \sigma_i^2$. Then*

$$\mathsf{P}\left\{\max_{i \le m} |S_i| \ge \lambda s_m\right\} \le 2\mathsf{P}\left\{|S_m| \ge (\lambda - \sqrt{2})s_m\right\}. \tag{10.7}$$

To prove (10.7)—note that it is trivial if $\lambda \le \sqrt{2}$—consider the sets

$$E_i = \left\{\max_{j < i} |S_j| < \lambda s_m \le |S_i|\right\}.$$

Clearly,

$$\mathsf{P}\left\{\max_{i \le m} |S_i| \ge \lambda s_m\right\} \le \mathsf{P}\{|S_m| \ge (\lambda - \sqrt{2})s_m\} + \sum_{i=1}^{m-1} \mathsf{P}(E_i \cap \{|S_m| < (\lambda - \sqrt{2})s_m\}). \tag{10.8}$$

Since $|S_i| \ge \lambda s_m$ and $|S_m| < (\lambda - \sqrt{2})s_m$ together imply $|S_m - S_i| \ge \sqrt{2}s_m$, it follows by Chebyshev's inequality and the assumed independence of the ξ_i that the sum in (10.8) is at most

$$\begin{aligned}\sum_{i=1}^{m-1} \mathsf{P}(E_i)\mathsf{P}\{|S_m - S_i| \ge \sqrt{2}s_m\} &\le \sum_{i=1}^{m-1} \mathsf{P}(E_i) \frac{1}{2s_m^2} \sum_{k=i+1}^{m} \sigma_k^2 \\ &\le \tfrac{1}{2} \sum_{i=1}^{m-1} \mathsf{P}(E_i) \le \tfrac{1}{2}\mathsf{P}\left\{\max_{i \le m} |S_i| \ge \lambda s_m\right\}.\end{aligned} \tag{10.9}$$

And now (10.8) and (10.9) combine to give (10.7).

Applying the lemma to the random variables involved in Theorem 10.1, we have, for $\lambda > 2\sqrt{2}$,

$$\mathsf{P}\left\{\max_{i\leq n} |S_i| \geq \lambda\sigma\sqrt{n}\right\} \leq 2\mathsf{P}\{|S_n| \geq \tfrac{1}{2}\lambda\sigma\sqrt{n}\}.$$

By the central limit theorem,

$$\mathsf{P}\{|S_n| \geq \tfrac{1}{2}\lambda\sigma\sqrt{n}\} \to \mathsf{P}\{|N| \geq \tfrac{1}{2}\lambda\} < \frac{8}{\lambda^3}\mathsf{E}\{|N|^3\}.$$

Therefore, if ε is positive, we have

$$\limsup_{n\to\infty} \mathsf{P}\left\{\max_{i\leq n} |S_i| \geq \lambda\sigma\sqrt{n}\right\} < \frac{\varepsilon}{\lambda^2} \tag{10.10}$$

for λ sufficiently large. Tightness now follows by Theorem 8.4.

An Application

As pointed out in the Introduction, Donsker's theorem has this qualitative interpretation: $X_n \xrightarrow{\mathscr{D}} W$ says that, if τ is small, then a particle subject to independent displacements $\xi_1, \xi_2, \ldots$ at successive times $\tau, 2\tau, \ldots$ will, viewed from afar, appear to perform approximately a Brownian motion.

More important than this qualitative interpretation is the use of Donsker's theorem to prove limit theorems for various functions of the partial sums S_n. The Introduction indicates how to use the relation $X_n \xrightarrow{\mathscr{D}} W$ to derive the limiting distribution of $\max_{i\leq n} S_i$; let us now carry this through in detail.

Since $h(x) = \sup_t x(t)$ is a continuous function on C, $X_n \xrightarrow{\mathscr{D}} W$ implies, by Corollary 1 to Theorem 5.1, that

$$\sup_t X_n(t) \xrightarrow{\mathscr{D}} \sup_t W_t.$$

The obvious relation

$$\sup_{0\leq t\leq 1} X_n(t) = \max_{i\leq n} \frac{1}{\sigma\sqrt{n}} S_i$$

now implies

$$\frac{1}{\sigma\sqrt{n}} \max_{i\leq n} S_i \xrightarrow{\mathscr{D}} \sup_t W_t. \tag{10.11}$$

Thus we would have the limiting distribution of $\max_{i\leq n} S_i$ (properly normalized) if we knew the distribution of $\sup_t W_t$. (We still assume the hypotheses of Theorem 10.1, of course.) The technique we shall use to find this latter distribution is to compute the limiting distribution of $\max_{i\leq n} S_i$ in an easy special case.

Suppose that $S_0, S_1, \ldots$ are the random variables for a symmetric random walk starting from the origin. Suppose, that is, that the ξ_n are independent and satisfy

(10.12) $$\mathsf{P}\{\xi_n = 1\} = \mathsf{P}\{\xi_n = -1\} = \tfrac{1}{2}.$$

Let us show that, if a is a nonnegative integer, then

(10.13) $$\mathsf{P}\left\{\max_{0\le i\le n} S_i \ge a\right\} = 2\mathsf{P}\{S_n > a\} + \mathsf{P}\{S_n = a\}.$$

The case $a = 0$ is easy. Assume $a > 0$ and put $M_i = \max_{0\le j\le i} S_j$. Since

$$\mathsf{P}\{M_n \ge a\} - \mathsf{P}\{S_n = a\} = \mathsf{P}\{M_n \ge a, S_n < a\} + \mathsf{P}\{M_n \ge a, S_n > a\}$$

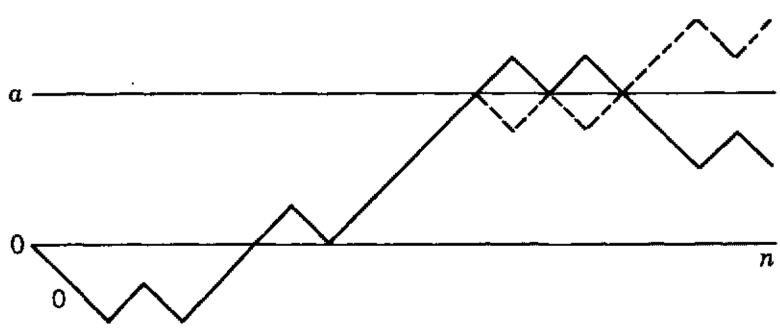
and

$$\mathsf{P}\{M_n \ge a, S_n > a\} = \mathsf{P}\{S_n > a\},$$

(10.13) will follow if we prove

(10.14) $$\mathsf{P}\{M_n \ge a, S_n < a\} = \mathsf{P}\{M_n \ge a, S_n > a\}.$$

Because of (10.12), all 2^n possible paths $(S_1, S_2, \ldots, S_n)$ have the same probability 2^{-n}. Therefore (10.14) will follow if we show that the number of paths contributing to the left-hand event is the same as the number of paths



contributing to the right-hand event, and to show this it suffices to match the paths in a one-to-one manner.† Given a path $(S_1, S_2, \ldots, S_n)$ contributing to the left-hand event in (10.14), match it with the path obtained by reflecting through a all the partial sums after the first one that achieves the height a. Since the correspondence is one-to-one, (10.14) follows. This argument is an example of the *reflection principle.*

Let α be an arbitrary nonnegative number, and let $a_n = -[-\alpha n^{\frac{1}{2}}]$. By (10.14), we have

(10.15) $$\mathsf{P}\left\{\max_{i\le n} \frac{1}{\sqrt{n}} S_i \ge \alpha\right\} = 2\mathsf{P}\{S_n > a_n\} + \mathsf{P}\{S_n = a_n\}.$$

By the central limit theorem,

$$\mathsf{P}\{S_n \ge a_n\} \to \mathsf{P}\{N \ge \alpha\}$$

† The new mathematics.

($\sigma^2 = 1$ in this case, in view of (10.12)). Since the largest term in the symmetric binomial distribution goes to 0, the term $\mathsf{P}\{S_n = a_n\}$ in (10.15) is negligible. Thus

$$\mathsf{P}\left\{\max_{i \le n} \frac{1}{\sqrt{n}} S_i \ge \alpha\right\} \to 2\mathsf{P}\{N \ge \alpha\}, \qquad \alpha \ge 0. \tag{10.16}$$

Combining (10.16) with (10.11) (still assuming (10.12)), we conclude

$$\mathsf{P}\{\sup_t W_t \le \alpha\} = \frac{2}{\sqrt{2\pi}} \int_0^\alpha e^{-\frac{1}{2}u^2}\, du, \qquad \alpha \ge 0. \tag{10.17}$$

Of course, the left side of (10.17) vanishes if $\alpha < 0$. We have derived a fact about Brownian motion by combining Donsker's theorem with a computation involving random walk, a computation that was simple partly because it reduced to enumeration and partly because a random walk cannot pass above a positive integer a without passing through it.

Let us now drop the assumption (10.12). If the ξ_n are independent and identically distributed and satisfy (10.2), then (10.11) holds, and from (10.17) we can now conclude

$$\mathsf{P}\left\{\frac{1}{\sigma\sqrt{n}} \max_{i \le n} S_i \le \alpha\right\} \to \frac{2}{\sqrt{2\pi}} \int_0^\alpha e^{-\frac{1}{2}u^2}\, du, \qquad \alpha \ge 0. \tag{10.18}$$

Thus we have derived the limiting distribution of $\max_{i \le n} S_i$ under the hypotheses of the Lindeberg–Lévy theorem.

This argument follows a general pattern. If h is continuous on C—or continuous except at points forming a set of Wiener measure 0—then $X_n \xrightarrow{\mathscr{D}} W$ implies

$$h(X_n) \xrightarrow{\mathscr{D}} h(W). \tag{10.19}$$

(In the case just analyzed, $h(x) = \sup_t x(t)$.) We can find the limiting distribution of $h(X_n)$ if we know the distribution of $h(W)$, and we can often find the distribution of $h(W)$ by finding the limiting distribution of $h(X_n)$ in some special case and then using (10.19) in the other direction. The next section contains further examples in this pattern.

Therefore, if the ξ_n are independent and identically distributed with $\mathsf{E}\{\xi_n\} = 0$ and $\mathsf{E}\{\xi_n^2\} = \sigma^2$, then the limiting distribution of $h(X_n)$ does not depend on any further properties of the ξ_n. For this reason, Donsker's theorem is often called *the* (or *an*) *invariance principle*. Here we shall call it instead *the* (or *a*) *functional central limit theorem*. This term will be used for a variety of similar theorems in what follows, just as the term *central limit theorem* itself is used for a whole class of theorems.

A Necessary Condition for Tightness

Let us drop the assumption that the ξ_n are independent and directly assume that $X_n \xrightarrow{\mathscr{D}} W$ (with X_n still defined by (10.1)). We shall show that, in this circumstance, for each positive ε there exists a λ exceeding 1 such that

$$\mathsf{P}\left\{\max_{i\leq n} |S_i| \geq \lambda\sigma\sqrt{n}\right\} \leq \frac{\varepsilon}{\lambda^2} \tag{10.20}$$

holds for all n exceeding some n_0. In case $\{\xi_n\}$ is stationary, this is just the hypothesis in Theorem 8.4, which we verified in proving Donsker's theorem.

Let $Y = \sup_t |W_t|$. Because of (10.17) and the symmetry of W under reflection through 0, Y has a finite second moment. From (10.19) with $h(x) = \sup_t |x(t)|$, it follows that, for each positive λ,

$$\mathsf{P}\left\{\max_{i\leq n} |S_i| \geq \lambda\sigma\sqrt{n}\right\} \to \mathsf{P}\{Y \geq \lambda\} \leq \frac{1}{\lambda^2}\int_{\{Y\geq\lambda\}} Y^2\, d\mathsf{P}. \tag{10.21}$$

Given ε, choose λ so large that the integral here is less than ε; it follows from (10.21) that (10.20) holds for all sufficiently large n.

Thus we see, after the fact, that (10.20) is "the" condition to verify in proving tightness in Donsker's theorem. In Chapter 4 we shall use the same condition in proving functional central limit theorems for sequences of dependent random variables.

The preceding argument really shows this: Suppose that $\{\xi_n\}$ is stationary and that the finite-dimensional distributions of the X_n defined by (10.1) converge to those of a random function X, where $\sup_t |X(t)|$ has a finite second moment. Under these conditions, if $\{X_n\}$ is tight, then the hypothesis of Theorem 8.4 is satisfied. To this extent, this hypothesis is necessary; compare the remarks centering on (8.10).

Another Proof of Donsker's Theorem†

The rest of this section is devoted to a second proof of Theorem 10.1. The proof is interesting because it makes little use of the preceding theory. In fact, granted the existence of Wiener measure,‡ we need only Theorem 2.2 and the central limit theorem.

We are to prove $P_n \Rightarrow W$, where P_n is the distribution of X_n. According to Corollary 2 of Theorem 2.2, it is enough to prove that $P_n(A) \to W(A)$ whenever A is a finite intersection of open spheres and satisfies $W(\partial A) = 0$.

---

† The rest of this section may be omitted.

‡ Which can itself be established by a direct argument; see Problem 1 of Section 9.

Now if A is the intersection of spheres with centers $x_1, \ldots, x_k$ and radii $\varepsilon_1, \ldots, \varepsilon_k$, and if

$$y(t) = \max_{1 \le i \le k} \left(x_i(t) - \varepsilon_i\right)$$

and

$$z(t) = \min_{1 \le i \le k} \left(x_i(t) + \varepsilon_i\right),$$

then A has the form

$$A = \{x : y(t) < x(t) < z(t), 0 \le t \le 1\}. \tag{10.22}$$

We are to prove

$$P_n(A) \to W(A) \tag{10.23}$$

under the assumptions that A is defined by (10.22) with $y, z \in C$ and that $W(\partial A) = 0$. We may also assume that $y(t) < z(t)$ for all t, since A is otherwise empty. (If $y(0) \ne 0$ and $z(0) \ne 0$, then $W(\partial A) = 0$ holds automatically, but this need not concern us.) From here on, y and z are fixed.

In our first proof of Donsker's theorem, we showed that the finite-dimensional distributions of the P_n converge weakly to those of W. This fact, which does not really involve the space C at all, we shall assume here. Thus

$$\lim_{n\to\infty} P_n\{x : (x(t_1), \ldots, x(t_k)) \in H\} = W\{x : (x(t_1), \ldots, x(t_k)) \in H\} \tag{10.24}$$

holds for each k-dimensional Borel set H satisfying

$$W\{x : (x(t_1), \ldots, x(t_k)) \in \partial H\} = 0.$$

If, for integral v,

$$A_v = \left\{x : y\left(\frac{i}{v}\right) < x\left(\frac{i}{v}\right) < z\left(\frac{i}{v}\right), i = 0, 1, \ldots, v\right\},$$

then, as u increases to infinity, A_{2^u} decreases to the set of x for which $y(t) < x(t) < z(t)$ holds for all dyadic rationals t. But if $W(\partial A) = 0$, then this last set differs from A by a set of W-measure 0. Therefore for each positive η there exists a v such that $W(A) \le W(A_v) \le W(A) + \eta$. Since $P_n(A) \le P_n(A_v)$, and, since $\lim_n P_n(A_v) = W(A_v)$ by (10.24), we have $\limsup_n P_n(A) \le W(A) + \eta$. Since η was arbitrary,

$$\limsup_n P_n(A) \le W(A). \tag{10.25}$$

We complete the proof by showing that

$$\liminf_n P_n(A) \ge W(A), \tag{10.26}$$

which is the harder part. Given η, choose ε so that, if

$$D(\varepsilon) = \{x : y(t) + 3\varepsilon < x(t) < z(t) - 3\varepsilon, 0 \le t \le 1\}, \tag{10.27}$$

then $W(D(\varepsilon)) > W(A) - \eta$. (Clearly, the set (10.27) increases to A as ε decreases to 0.) Now define

$$D_v(\varepsilon) = \left\{x: y\left(\frac{k}{v}\right) + 3\varepsilon < x\left(\frac{k}{v}\right) < z\left(\frac{k}{v}\right) - 3\varepsilon, k \doteq 0, 1, \ldots, v\right\}.$$

For each v, we have $D(\varepsilon) \subset D_v(\varepsilon)$ and hence

(10.28) $$W(D_v(\varepsilon)) > W(A) - \eta.$$

Choose and fix a v so large that

(10.29) $$\frac{9}{\varepsilon^2 v} < \eta, \qquad w_y\left(\frac{1}{v}\right) < \varepsilon, \qquad w_z\left(\frac{1}{v}\right) < \varepsilon,$$

where w denotes the modulus of continuity (8.1). Then $n \geq v$ implies

(10.30) $$w_y\left(\frac{1}{n}\right) < \varepsilon, \qquad w_z\left(\frac{1}{n}\right) < \varepsilon.$$

Let L_n denote the set of elements of C that are linear on each interval $[(i - 1)/n, i/n]$, $i = 1, \ldots, n$, and let G_n be the set of x in C for which

(10.31) $$y\left(\frac{i}{n}\right) + \varepsilon \leq x\left(\frac{i}{n}\right) \leq z\left(\frac{i}{n}\right) - \varepsilon$$

is violated for some $i = 0, 1, \ldots, n$. If $x \in L_n$, if $x \in G_n{}^c$ (so that (10.31) holds for all $i = 0, 1, \ldots, n$), and if $n \geq v$, then, by (10.30), $y(t) < x(t) < z(t)$ for all $t \in [0, 1]$, so that $x \in A$. Thus $L_n \cap G_n{}^c \subset A$. Let $G_{n,r}$ be the set of x for which the double inequality (10.31) holds for $i < r$ but not for $i = r$. Since $P_n(L_n) = 1$, and since the $G_{n,r}$ are disjoint and add to G, we have

(10.32) $$P_n(A^c) \leq \sum_{r=0}^{n} P_n(G_{n,r}), \qquad n \geq v.$$

For $1 \leq r \leq n$, let $k_{n,r}$ denote that integer k $(1 \leq k \leq v)$ such that $(k - 1)/v < r/n \leq k/v$ and let $k_{n,0} = 0$. By (10.32),

(10.33) $$P_n(A^c) \leq \sum_{r=1}^{n} P_n\left(G_{n,r} \cap \left\{x: \left|x\left(\frac{r}{n}\right) - x\left(\frac{k_{n,r}}{v}\right)\right| < \varepsilon\right\}\right)$$
$$+ \sum_{r=1}^{n} P_n\left(G_{n,r} \cap \left\{x: \left|x\left(\frac{r}{n}\right) - x\left(\frac{k_{n,r}}{v}\right)\right| \geq \varepsilon\right\}\right).$$

If $x \in G_{n,r}$ and $|x(r/n) - x(k_{n,r}/v)| < \varepsilon$, then, by (10.30),

$$x\left(\frac{k_{n,r}}{v}\right) \notin \left[y\left(\frac{k_{n,r}}{v}\right) + 3\varepsilon, z\left(\frac{k_{n,r}}{v}\right) - 3\varepsilon\right],$$

so that $x \notin D_v(\varepsilon)$. Thus the first sum on the right in (10.33) is at most $P_n(D_v(\varepsilon)^c)$:

$$(10.34)\quad P_n(A^c) \leq P_n(D_v(\varepsilon)^c) + \sum_{r=1}^{n} P_n\left(G_{n,r} \cap \left\{x: \left|x\left(\frac{r}{n}\right) - x\left(\frac{k_{n,r}}{v}\right)\right| \geq \varepsilon\right\}\right)$$

for $n \geq v$.

We shall now show that the sum in (10.34) is small. Observe first that

$$(10.35)\qquad \int (x_t - x_s)^2\, dP_n \leq 9\,|t - s|$$

for all s, t, and n. In fact, if s and t are of the form i/n and j/n, then the left side of (10.35) is $\mathsf{E}\{(S_i - S_j)^2/n\sigma^2\} = |t - s|$; if s and t lie in the same subinterval $[(i-1)/n, i/n]$, then the left side of (10.35) is $n(t-s)^2$, which is at most $|t - s|$; and the general result follows by Minkowski's inequality. By the independence of the ξ_n, therefore,

$$(10.36)\quad P_n\left(G_{n,r} \cap \left\{x: \left|x\left(\frac{r}{n}\right) - x\left(\frac{k_{n,r}}{v}\right)\right| \geq \varepsilon\right\}\right)$$
$$= P_n(G_{n,r})P_n\left\{x: \left|x\left(\frac{r}{n}\right) - x\left(\frac{k_{n,r}}{v}\right)\right| \geq \varepsilon\right\}$$
$$\leq P_n(G_{n,r})\left(\frac{9}{\varepsilon^2}\right)\left(\frac{r}{n} - \frac{k_{n,r}}{v}\right).$$

Since $r/n - k_{n,r}/v \leq 1/v$, it follows by (10.29) that the last member of (10.36) is at most $\eta P_n(G_{n,r})$. Using (10.34) and the fact that the $P(G_{n,r})$ sum to at most 1, we arrive at

$$P_n(A^c) \leq P_n(D_v(\varepsilon)^c) + \eta, \qquad n \geq v.$$

But $\lim_n P_n(D_v(\varepsilon)^c) = W(D_v(\varepsilon)^c)$ by (10.24), and so, by (10.28),

$$\limsup_n P_n(A^c) \leq W(A^c) + 2\eta.$$

Since η was arbitrary, this establishes (10.26), which combines with (10.25) to yield (10.23), proving Donsker's theorem once more.

Remarks. Erdös and Kac (1946) first conceived the invariance principle itself—the idea of computing the limiting distribution of a quantity such as $\max_{k \leq n} S_k$ first in a special case and then passing to the general case by connecting the result with Brownian motion (which is how it all started). An early paper in this direction was Kolmogorov (1931).

The original assertion of Donsker (1951) was not that $X_n \xrightarrow{\mathscr{D}} W$ or that $P_n \Rightarrow W$, but, what is by Theorem 5.2 the same thing, that $h(X_n) \xrightarrow{\mathscr{D}} h(W)$ for all real, continuous functions h on C. The arguments leading to (10.23) are his. Kolmogorov and Prohorov (1954) first pointed out that weak convergence is relevant and that $P_n \Rightarrow W$ follows from (10.23) and Theorem 2.2. Prohorov (1953 and 1956) brought the tightness concept to bear. For a very different approach to these problems, see Knight (1962), and for a very different kind of application, see Strassen (1964). Lamperti (1962a) has a result differing from Theorem 10.1 in that the limiting process is not Brownian motion.

PROBLEMS

1. Let $\xi_{n1}, \ldots, \xi_{nk_n}$ be independent with mean 0 and variances σ^2_{ni}; put $S_{ni} = \xi_{n1} + \cdots + \xi_{ni}$, $s^2_{ni} = \sigma^2_{n1} + \cdots + \sigma^2_{ni}$, and $s_n{}^2 = s^2_{nk_n}$. Let X_n be the random function that is linear on each interval $[s^2_{n,i-1}/s_n{}^2, s^2_{ni}/s_n{}^2]$ and has values $X_n(s^2_{ni}/s_n{}^2) = S_{ni}/s_n$ at the points of division. Assume Lindeberg's condition (7.3) holds and generalize Donsker's theorem by using Lindeberg's theorem, the inequality (10.7), and the corollary to Theorem 8.3 to prove $X_n \xrightarrow{\mathcal{D}} W$. (This result is due to Prohorov (1956).)

11. FUNCTIONS OF BROWNIAN MOTION PATHS

In the preceding section we used Donsker's theorem and properties of random walk to find the distribution of $\sup_t W_t$. Having found this distribution, we used Donsker's theorem once more to find the limiting distribution of $\max_{i \le n} S_i$ in general. Here we shall apply the same technique to other functions of Brownian motion paths and of partial sums. We also compute some distributions associated with the Brownian bridge.†

We shall say we are in the Lindeberg–Lévy case when dealing with partial sums S_n of independent, identically distributed random variables ξ_n with mean 0 and finite, positive variance σ^2. In this case X_n will always denote the random function (10.1). We shall say we are in the random walk case if each ξ_n assumes the values $+1$ and -1 with probability $\frac{1}{2}$ each. In this case $\sigma^2 = 1$.

Maximum and Minimum

Let $m = \inf_t W_t$ and $M = \sup_t W_t$, and let

$$m_n = \min_{0 \le i \le n} S_i, \qquad M_m = \max_{0 \le i \le n} S_i \tag{11.1}$$

be the corresponding quantities for partial sums. The mapping carrying the point x of C to the point $\big(\inf_t x(t), \sup_t x(t), x(1)\big)$ of R^3 is everywhere continuous, so that, by Theorems 10.1 and 5.1,

$$\frac{1}{\sigma\sqrt{n}}(m_n, M_n, S_n) \xrightarrow{\mathcal{D}} (m, M, W_1) \tag{11.2}$$

in the Lindeberg–Lévy case.

---

† Although weak-convergence results in C would have little point if it were not possible to perform computations of the sort carried through in this section, the computations are not themselves essential to an understanding of the general theory.

We shall first find an explicit formula for

$$p_n(a, b, v) = \mathsf{P}\{a < m_n \le M_n < b, S_n = v\} \tag{11.3}$$

in the random walk case. We shall show that, if

$$p_n(j) = \mathsf{P}\{S_n = j\}, \tag{11.4}$$

then

$$p_n(a, b, v) = \sum_{k=-\infty}^{\infty} p_n(v + 2k(b - a)) - \sum_{k=-\infty}^{\infty} p_n(2b - v + 2k(b - a)) \tag{11.5}$$

for integers a, b, and v satisfying

$$a \le 0 \le b, \qquad a < b, \qquad a \le v \le b. \tag{11.6}$$

If $a < b$, then the series in (11.5) are really finite sums. Notice that both sides of (11.5) vanish if n and v have opposite parity.

For particular values of n, a, b, and v, let us denote the equation (11.5) by $[n, a, b, v]$. We shall prove by induction on n that $[n, a, b, v]$ is valid if (11.6) holds. For $n = 0$, this follows by a straightforward examination of cases. Assume as induction hypothesis that $[n - 1, a, b, v]$ holds for a, b, v satisfying (11.6). If $a = 0$, then (11.3) vanishes (note that i starts at 0 in the minimum in (11.1)), and the sums on the right in (11.5) cancel because $p_n(j) = p_n(-j)$. Thus $[n, a, b, v]$ is valid if (11.6) holds and $a = 0$; we may dispose of the case $b = 0$ in the same way. To complete the induction step we must verify $[n, a, b, v]$ under the assumption that $a < 0 < b$ and $a \le v \le b$. But in this case $a + 1 \le 0$ and $b - 1 \ge 0$, so that $[n - 1, a - 1, b - 1, v - 1]$ and $[n - 1, a + 1, b + 1, v + 1]$ both come under the induction hypothesis and hence are valid. And now $[n, a, b, v]$ follows by the probabilistically obvious recursions

$$p_n(j) = \tfrac{1}{2}p_{n-1}(j - 1) + \tfrac{1}{2}p_{n-1}(j + 1)$$

and

$$p_n(a, b, v) = \tfrac{1}{2}p_{n-1}(a - 1, b - 1, v - 1) + \tfrac{1}{2}p_{n-1}(a + 1, b + 1, v + 1).\dagger$$

From (11.5) it follows by summation over v that, if

$$a \le 0 \le b, \qquad a \le u < v \le b, \tag{11.7}$$

then

$$\begin{aligned} \mathsf{P}\{a < m_n \le M_n < b, u < S_n < v\} &= \sum_{k=-\infty}^{\infty} \mathsf{P}\{u + 2k(b - a) < S_n < v + 2k(b - a)\} \\ &\quad - \sum_{k=-\infty}^{\infty} \mathsf{P}\{2b - v + 2k(b - a) < S_n < 2b - u + 2k(b - a)\}. \end{aligned} \tag{11.8}$$

† Problem 2 outlines how (11.5) can be derived (as opposed to verified).

Taking $a = -n - 1$ in this formula leads to

$$(11.9)\quad \mathsf{P}\{M_n < b, u < S_n < v\} = \mathsf{P}\{u < S_n < v\} - \mathsf{P}\{2b - v < S_n < 2b - u\},$$

valid for $-n - 1 \leq u < v \leq b$, $b \geq 0$.† From (11.9) it is possible to retrieve (10.13).

Now (11.8) holds in the random walk case and, because of (11.2), we can find the distribution of (m, M, W_1) by passing to the limit. If a, b, u, and v are real numbers satisfying (11.7), replace them in (11.8) by the integers $[an^{\frac{1}{2}}]$, $-[-bn^{\frac{1}{2}}]$, $[un^{\frac{1}{2}}]$, and $-[-vn^{\frac{1}{2}}]$, respectively. Because of the central limit theorem and the continuity of the normal distribution, a termwise passage to the limit in (11.8) yields

$$(11.10)\quad \mathsf{P}\{a < m \leq M < b, u < W_1 < v\} = \sum_{k=-\infty}^{\infty} \mathsf{P}\{u + 2k(b - a) < N < v + 2k(b - a)\} - \sum_{k=-\infty}^{\infty} \mathsf{P}\{2b - v + 2k(b - a) < N < 2b - u + 2k(b - a)\}.$$

The interchange of limit with summation over k can be justified by the series form of Scheffé's theorem (p. 224).

The joint distribution of M and W_1 alone could be obtained by letting a tend to $-\infty$ in (11.10), but it is simpler to return to the random walk case and pass to the limit in (11.9), which yields

$$(11.11)\quad \mathsf{P}\{M < b, u < W_1 < v\} = \mathsf{P}\{u < N < v\} - \mathsf{P}\{2b - v < N < 2b - u\},$$

valid for $u < v \leq b$, $b \geq 0$. Taking $v = b$ and letting $u \to -\infty$ leads back to (10.17).

From (11.10) with $u = a$ and $v = b$ we have

$$(11.12)\quad \mathsf{P}\{a < m \leq M < b\} = \sum_{k=-\infty}^{\infty} (-1)^k \mathsf{P}\{a + k(b - a) < N < b + k(b - a)\},$$

valid for $a \leq 0 \leq b$. And this result with $a = -b$ gives

$$(11.13)\quad \mathsf{P}\{\sup_t |W_t| < b\} = \sum_{k=-\infty}^{\infty} (-1)^k \mathsf{P}\{(2k - 1)b < N < (2k + 1)b\}$$

† See Problem 1 for another proof.

for $b \geq 0$. By continuity, the strict inequalities in all these formulas can be relaxed to allow equality. And the right sides can all be written out as sums of normal integrals. It is possible† to transform the series in (11.13) to

$$1 - \frac{4}{\pi} \sum_{k=1}^{\infty} \frac{(-1)^k}{2k+1} e^{-\pi^2(2k+1)^2/8b^2}.$$

Although we derived (11.10) through (11.13) by passing to the limit in the random walk case, we have the limiting distributions for (m_n, M_n, S_n), (M_n, S_n), (m_n, M_n), and $\max_{i \leq n} |S_i|$ (properly normalized) in the more general Lindeberg–Lévy case because (11.2) holds there.

The Arc Sine Law

For x in C, let $h_1(x)$ be the supremum of those t in $[0, 1]$ for which $x(t) = 0$; let $h_2(x)$ be the Lebesgue measure of those t in $[0, 1]$ for which $x(t) > 0$; and let $h_3(x)$ be the Lebesgue measure of those t in $[0, h_1(x)]$ for which $x(t) > 0$. Then $T = h_1(W)$ is the time at which the Wiener path W last passes through 0, $U = h_2(W)$ is the total amount of time W spends above 0, and $V = h_3(W)$ is the amount of time W spends above 0 in the interval $[0, T]$. We shall find the joint distribution of (T, U, V, W_1).

In Appendix II (pp. 230 ff.) it is shown that each of the mappings h_1, h_2, and h_3 is measurable and is continuous except on a set of Wiener measure 0. Therefore

$$(11.14) \qquad \big(h_1(X_n), h_2(X_n), h_3(X_n), X_n(1)\big) \xrightarrow{\mathscr{D}} (T, U, V, W_1)$$

in the Lindeberg–Lévy case. In the random walk case, the vector on the left has a simple interpretation: $T_n = nh_1(X_n)$ is the maximum i, $1 \leq i \leq n$, for which $S_i = 0$; $U_n = nh_2(X_n)$ is the number of i, $1 \leq i \leq n$, for which S_{i-1} and S_i are both nonnegative; $V_n = nh_3(X_n)$ is the number of i, $1 \leq i \leq T_n$, for which S_{i-1} and S_i are both nonnegative; and, of course, $X_n(1) = S_n/\sqrt{n}$.

With these definitions we therefore have

$$(11.15) \qquad \left(\frac{1}{n} T_n, \frac{1}{n} U_n, \frac{1}{n} V_n, \frac{1}{\sqrt{n}} S_n\right) \xrightarrow{\mathscr{D}} (T, U, V, W_1)$$

in the random walk case; we shall find the distribution of (T, U, V, W_1) by passing to the limit. In the general Lindeberg–Lévy case, the left side of (11.14) is a somewhat more complicated function of the partial sums $S_1, \ldots,$ S_n. After we have found the distribution of (T, U, V, W_1), we shall show how (11.14) leads even in the general case to limit theorems for quantities associated in a natural way with the partial sums.

† See Feller (1966, pp. 330 and 594).

Since the random vector (T, U, V, W_1) is constrained by

$$(11.16) \qquad U = \begin{cases} 1 - T + V & \text{if} \quad W_1 \geq 0, \\ V & \text{if} \quad W_1 \leq 0, \end{cases}$$

it suffices to consider (T, V, W_1) and the related vector (T_n, V_n, S_n). The distribution of the latter quantity in the random walk case we shall derive from three facts which admit of elementary proofs we shall not carry through here.

First, we shall use the local limit theorem for random walk: If m tends to infinity and j varies with m in such a way that j and m have the same parity and $j/\sqrt{m} \to y$, then†

$$(11.17) \qquad \frac{\sqrt{m}}{2}\, p_m(j) \to \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}y^2}.$$

Second, we shall use the fact that

$$(11.18) \qquad \mathsf{P}\{S_1 > 0, \ldots, S_{m-1} > 0, S_m = j\} = \frac{j}{m}\, p_m(j)$$

for j positive.‡ If $S_{2m} = 0$, then $U_{2m} = V_{2m}$ assumes one of the values 0, 2, ..., $2m$; the third fact§ we shall use is that these $m + 1$ values all have the same conditional probability:

$$(11.19) \qquad \mathsf{P}\{V_{2m} = 2i \mid S_{2m} = 0\} = \frac{1}{m+1}, \qquad i = 0, 1, \ldots, m.$$

To compute the probability that $T_n = 2k$, $V_n = 2i$, and $S_n = j$, condition on the event $S_{2k} = 0$. Conditionally on this event, $(S_0, \ldots, S_{2k})$ and $(S_{2k+1}, \ldots, S_n)$ are independent, V_n depends only on the first sequence, and $T_n = 2k$ and $S_n = j$ if and only if the elements of the second sequence are nonzero and the last one is j. By (11.18) and (11.19) we conclude that

$$(11.20) \qquad \mathsf{P}\{T_n = 2k, V_n = 2i, S_n = j\} = p_{2k}(0)\, \frac{1}{k+1}\, \frac{j}{n-2k}\, p_{n-2k}(j)$$

if

$$(11.21) \qquad 0 \leq 2i \leq 2k < n, \qquad j > 0.$$

Both sides of (11.20) vanish if n and j have opposite parity. For j negative, the same formula holds with $|j|$ in place of j on the right.

---

† Feller (1957, p. 170) has the local theorem for the binomial distribution, and (11.17) follows because $\frac{1}{2}(S_m + m)$ is binomially distributed.

‡ See Feller (1957, p. 70). This also follows from (11.9).

§ See Feller (1957, p. 72).

We shall apply Theorem 7.8 to the lattice of points $(2k/n, 2i/n, j/\sqrt{n})$, where j and n have the same parity. Suppose k, i, and j tend to infinity with n in such a way that

$$\frac{2k}{n} \to t, \qquad \frac{2i}{n} \to v, \qquad \frac{j}{\sqrt{n}} \to x,$$

where $0 < v < t < 1$ and $x > 0$. Then (11.21) holds for large n, and it follows by (11.20) and (11.17) that

$$\left(\frac{2}{n} \cdot \frac{2}{n} \cdot \frac{2}{\sqrt{n}}\right)^{-1} \mathsf{P}\{T_n = 2k, V_n = 2i, S_n = j\} \to g(t, x),$$

where

$$g(t, x) = \frac{1}{2\pi} \frac{|x|}{[t(1-t)]^{\frac{3}{2}}} e^{-\frac{1}{2}x^2/(1-t)}, \qquad 0 < t < 1. \tag{11.22}$$

The same result holds for negative x by symmetry. Therefore

$$\left(\frac{1}{n} T_n, \frac{1}{n} V_n, \frac{1}{\sqrt{n}} S_n\right) \tag{11.23}$$

has (in the random walk case) the limiting distribution in R^3 specified by the density

$$f(t, v, x) = \begin{cases} g(t, x) & \text{if} \quad 0 < v < t < 1, \\ 0 & \text{otherwise.} \end{cases} \tag{11.24}$$

By (11.15), (T, V, W_1) has this density. Because of (11.16), the distribution of (T, U, V, W_1) can be written out explicitly also.

From (11.24) it follows that the conditional distribution of V given T and W_1 is uniform on $[0, T]$; this corresponds to (11.19). By (11.16), if $T = t$ and $W_1 = x$, then U is uniformly distributed over $[1 - t, 1]$ for $x > 0$ and over $[0, t]$ for $x < 0$. Using (11.24) to account for the possible values of t and x, we find for the density of U alone

$$\iint\limits_{\substack{x>0 \\ 1-u<t}} g(t, x)\, dt\, dx + \iint\limits_{\substack{x<0 \\ u<t}} g(t, x)\, dt\, dx. \tag{11.25}$$

Now the integral of $g(t, x)$ over the range $x > 0$ is $1/[2\pi t^{\frac{3}{2}}(1-t)^{\frac{1}{2}}]$, which is the derivative of $-((1-t)/t)^{\frac{1}{2}}/\pi$, and hence (11.25) reduces to $1/[\pi u^{\frac{1}{2}}(1-u)^{\frac{1}{2}}]$. Therefore

$$\mathsf{P}\{U \le u\} = \frac{1}{\pi} \int_0^u \frac{ds}{\sqrt{s(1-s)}} = \frac{2}{\pi} \arcsin \sqrt{u}, \qquad 0 < u < 1. \tag{11.26}$$

This is Lévy's arc sine distribution. A similar computation shows that T also follows the arc sine law:

(11.27) $$\mathsf{P}\{T \leq t\} = \frac{2}{\pi} \arc \sin \sqrt{t}, \qquad 0 < t < 1.$$

Let us now combine (11.14) with the facts just derived to obtain a limit theorem for the general Lindeberg–Lévy case. Let us agree to say that a 0-crossing takes place at i if the event

(11.28) $$E_i = \{S_i = 0\} \cup \{S_{i-1} > 0 > S_i\} \cup \{S_{i-1} < 0 < S_i\}$$

occurs (which in the random walk case is to say that $S_i = 0$). Let T_n' be the maximum i, $1 \leq i \leq n$, for which a 0-crossing takes place at i; let U_n' be the number of i, $1 \leq i \leq n$, for which $S_i > 0$; and let V_n' be the number of i, $1 \leq i \leq T_n'$, for which $S_i > 0$. We shall prove that

(11.29) $$\left(\frac{1}{n} T_n', \frac{1}{n} U_n', \frac{1}{n} V_n', \frac{1}{\sigma\sqrt{n}} S_n\right) \xrightarrow{\mathscr{D}} (T, U, V, W_1)$$

by showing that the quantity on the left here approximates the left side of (11.14).

Clearly, T_n'/n is within $1/n$ of $h_1(X_n)$. If γ_n is the number of i, $1 \leq i \leq n$, for which E_i occurs—the number of 0-crossings—then U_n'/n and V_n'/n are within γ_n/n of $h_2(X_n)$ and $h_3(X_n)$, respectively. Therefore (11.29) will follow from (11.14) and Theorem 4.1 if we prove that $\gamma_n/n \xrightarrow{P} 0$, and for this it is enough to show that

(11.30) $$\mathsf{E}\left\{\frac{\gamma_n}{n}\right\} = \frac{1}{n}\sum_{i=1}^{n} \mathsf{P}(E_i) \to 0.$$

But

$$\mathsf{P}(E_i) \leq \mathsf{P}\{|\xi_i| \geq \varepsilon\sqrt{i}\} + \mathsf{P}\{|S_{i-1}| \leq \varepsilon\sqrt{i}\}$$

for each positive ε, and hence, by the central limit theorem, $\mathsf{P}(E_i) \to 0$. And now (11.30) is a consequence of the theorem on arithmetic means of convergent sequences.

From (11.29) we may conclude for example that U_n'/n and T_n'/n have arc sine distributions in the limit.

The Brownian Bridge

The Brownian bridge W° behaves like a Wiener path W conditioned by the requirement $W_1 = 0$. With an appropriate passage to the limit to take account of the fact that $\{W_1 = 0\}$ is an event of probability 0, this observation can be used to derive distributions associated with W°.

Let P_ε be the probability measure on C defined by

$$P_\varepsilon(A) = \mathsf{P}\{W \in A \mid 0 \le W_1 \le \varepsilon\}, \qquad A \in \mathscr{C}.$$

We shall prove that

$$P_\varepsilon \Rightarrow W^\circ \tag{11.31}$$

as ε tends to 0.† Take the random function W as defined on some probability space and take the random function W° to be defined on the same space by $W_t^\circ = W_t - tW_1$ (see the construction in Section 9). If we prove that

$$\limsup_{\varepsilon \to 0} \mathsf{P}\{W \in F \mid 0 \le W_1 \le \varepsilon\} \le \mathsf{P}\{W^\circ \in F\} \tag{11.32}$$

for each closed F in C, then (11.31) will follow by Theorem 2.1.

For arbitrary $t_1, \ldots, t_k$, W_1 is independent of $(W_{t_1}^\circ, \ldots, W_{t_k}^\circ)$ because it is uncorrelated with each component. Therefore

$$\mathsf{P}\{W^\circ \in A,\ W_1 \in B\} = \mathsf{P}\{W^\circ \in A\}\mathsf{P}\{W_1 \in B\} \tag{11.33}$$

if A is a finite-dimensional set in C and B lies in $\mathscr{R}^1$. But for B fixed the set of A in $\mathscr{C}$ that satisfy (11.33) is a monotone class and hence‡ coincides with $\mathscr{C}$. Thus (11.33) holds for $A \in \mathscr{C}$ and $B \in \mathscr{R}^1$. In particular,

$$\mathsf{P}\{W^\circ \in A \mid 0 \le W_1 \le \varepsilon\} = \mathsf{P}\{W^\circ \in A\}.$$

Since $\rho(W, W^\circ) = |W_1|$, where ρ is the metric on C, $|W_1| \le \delta$ and $W \in F$ imply $W^\circ \in F_\delta = \{x : \rho(x, F) \le \delta\}$. Therefore, if $\varepsilon < \delta$,

$$\mathsf{P}\{W \in F \mid 0 \le W_1 \le \varepsilon\} \le \mathsf{P}\{W^\circ \in F_\delta \mid 0 \le W_1 \le \varepsilon\} = \mathsf{P}\{W^\circ \in F_\delta\}.$$

The limit superior in (11.32) is thus at most $\mathsf{P}\{W^\circ \in F_\delta\}$, which decreases to $\mathsf{P}\{W^\circ \in F\}$ as $\delta \downarrow 0$ if F is closed. This proves (11.32) and hence (11.31).§

Suppose now that h is a measurable mapping from C to R^k and that W° has probability 0 of lying in the set D_h of discontinuities of h. It follows by (11.31) and Theorem 5.1 that

$$\mathsf{P}\{h(W^\circ) \le \alpha\} = \lim_{\varepsilon \to 0} \mathsf{P}\{h(W) \le \alpha \mid 0 \le W_1 \le \varepsilon\} \tag{11.34}$$

holds for each α at which the left side is continuous (as a function of α ranging over R^k). From (11.34) we can find explicit forms for some distributions connected with W°. Sometimes an alternative form of (11.34) is more

† We can let ε tend continuously to 0 and use the definition involving (2.4). Alternatively, we can let ε tend to 0 through some sequence (say the reciprocals of integers) fixed throughout the discussion.

‡ See Halmos (1950, p. 26).

§ This part of the argument merely repeats the proof of Theorem 4.1.

convenient:

$$\text{(11.35)}\qquad \mathsf{P}\{h(W^\circ) \le \alpha\} = \lim_{\varepsilon \to 0} \mathsf{P}\{h(W) \le \alpha \mid -\varepsilon \le W_1 \le 0\}.$$

This is established in the same way (in place of $[0, \varepsilon]$ we could use any subset of $[-\varepsilon, \varepsilon]$ with positive measure).

Let

$$m^\circ = \inf_t W_t^\circ, \qquad M^\circ = \sup_t W_t^\circ.$$

Suppose that $a < 0 < b$ and that $0 < \varepsilon < b$; by (11.10) we have, if $c = b - a$,

$$\text{(11.36)}\quad \mathsf{P}\{a < m \le M < b, 0 < W_1 < \varepsilon\}$$

$$= \sum_{k=-\infty}^{\infty} \mathsf{P}\{2kc < N < 2kc + \varepsilon\} - \sum_{k=-\infty}^{\infty} \mathsf{P}\{2kc + 2b - \varepsilon < N < 2kc + 2b\}.$$

Since

$$\text{(11.37)}\qquad \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \mathsf{P}\{x < N < x + \varepsilon\} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2},$$

if we take the limit inside the sums in (11.36), then (11.34) yields

$$\text{(11.38)}\qquad \mathsf{P}\{a < m^\circ \le M^\circ \le b\} = \sum_{k=-\infty}^{\infty} e^{-2(kc)^2} - \sum_{k=-\infty}^{\infty} e^{-2(b+kc)^2}.$$

Since the series converge uniformly in ε, the interchange is all right. Thus we have the distribution of (m°, M°). Taking $-a = b$ here gives

$$\text{(11.39)}\qquad \mathsf{P}\{\sup_t |W_t^\circ| \le b\} = 1 + 2\sum_{k=1}^{\infty}(-1)^k e^{-2k^2b^2}, \qquad b > 0.$$

By an entirely similar analysis applied to (11.11),

$$\text{(11.40)}\qquad \mathsf{P}\{M^\circ < b\} = 1 - e^{-2b^2}, \qquad b > 0.$$

Let U° be the Lebesgue measure of those t in $[0, 1]$ for which $W_t^\circ > 0$. We shall show that U° is uniformly distributed over $[0, 1]$ by showing that

$$\text{(11.41)}\qquad \lim_{\varepsilon \to 0} \mathsf{P}\{U \le \alpha \mid -\varepsilon \le W_1 \le 0\} = \alpha, \qquad 0 < \alpha < 1.$$

Because of (11.16), the conditional probability here is

$$\mathsf{P}\{V \le \alpha \mid -\varepsilon \le W_1 \le 0\}.$$

From the form of the density (11.24) we saw that the distribution of V given T and W_1 is uniform on $(0, T)$. In other words, if $L = V/T$, then L is uniformly distributed on $(0, 1)$ and is independent of (T, W_1). Therefore the conditional probability in (11.41) is

$$\mathsf{P}\{TL \le \alpha \,|\, -\varepsilon \le W_1 \le 0\} = \int_0^1 \mathsf{P}\left\{T \le \frac{\alpha}{s} \,\middle|\, -\varepsilon < W_1 \le 0\right\} ds,$$

and (11.41) will follow by the bounded convergence theorem if we prove the intuitively obvious relation

$$\mathsf{P}\{T \le \theta \,|\, -\varepsilon \le W_1 \le 0\} \to 0, \qquad \theta < 1.$$

But this follows by (11.37) and the form of the density (11.24). Therefore

$$\mathsf{P}\{U^\circ \le \alpha\} = \alpha, \qquad 0 < \alpha < 1. \tag{11.42}$$

Remarks. For further evaluation of distributions of functions of Brownian motion paths, see Itô and McKean (1965), Karlin (1966, Chapter 10), Erdös and Kac (1946 and 1947), Mark (1949), Darling and Erdös (1956), and Section 16 below.

PROBLEMS

1. Show by reflection in the random walk case that

$$\mathsf{P}\{M_n \ge b,\, S_n = v\} = \begin{cases} p_n(v) & \text{if } v \ge b, \\ p_n(2b - v) & \text{if } v \le b \end{cases} \tag{11.43}$$

for $b \ge 0$. Derive (11.9) from this.

2. For nonnegative integers c_i, let $\pi(c_1, \dots, c_k; v)$ be the probability that an n-step random walk (n fixed) meets c_1 (one or more times), then meets $-c_2$, then meets $c_3, \dots$, then meets $(-1)^{k+1}c_k$, and ends at v. Use (11.43) and induction on k to show that

$$\pi(c_1, \dots, c_k; v) = \begin{cases} p_n(2c_1 + \cdots + 2c_{k-1} + (-1)^{k+1}v) & \text{if } \ (-1)^{k+1}v \ge c_k, \\ p_n(2c_1 + \cdots + 2c_k - (-1)^{k+1}v) & \text{if } \ (-1)^{k+1}v \le c_k. \end{cases}$$

[Reflect through $(-1)^k c_{k-1}$ the part of the path to the right of the first passage through that point following successive passages through $c_1, -c_2, \dots, (-1)^{k-2}c_{k-2}$.] Derive (11.5) by showing that $p_n(a, b, v)$ is

$$\begin{aligned} p_n(v) &- \pi(b; v) + \pi(b, a; v) - \pi(b, a, b; v) + \cdots \\ &- \pi(a; v) + \pi(a, b; v) - \pi(a, b, a; v) + \cdots, \end{aligned}$$

3. For x in c let $h(x)$ be the smallest t for which $x(t) = \sup_s x(s)$. Show that h is measurable and is continuous on a set of W-measure 1. Let τ_n be the smallest k for which $S_k = \max_{i \le n} S_i$ and prove

$$\mathsf{P}\left\{\frac{\tau_n}{n} \le \alpha\right\} \to \frac{2}{\pi} \arcsin \sqrt{\alpha}, \qquad 0 < \alpha < 1,$$

in the Lindeberg-Lévy case. [See Feller (1957, p. 86) for the random walk case.]

4. Derive the joint limiting distribution of the maximum and minimum of $S_i - in^{-1}S_n$, $0 < i < n$, in the Lindeberg-Lévy case. [Consider $Y_n(t) = X_n(t) - tX_n(1)$ with X_n defined by (10.1).]

12. FLUCTUATIONS OF PARTIAL SUMS

In Sections 9 and 10 we established tightness for sequences of random functions by finding bounds for the distribution of the maximum of certain partial sums. Here we shall derive such bounds under fairly general conditions; the results lead to practical tightness criteria which will be used throughout the rest of the book.

Maxima

Let $\xi_1, \ldots, \xi_m$ be random variables; they need not be independent or identically distributed. Let $S_k = \xi_1 + \cdots + \xi_k$ ($S_0 = 0$), and put

(12.1) $$M_m = \max_{0 \le k \le m} |S_k|.$$

We shall obtain upper bounds for $\mathsf{P}\{M_m \ge \lambda\}$ by an indirect approach.

If

(12.2) $$M'_m = \max_{0 \le k \le m} \min\{|S_k|, |S_m - S_k|\},$$

then

(12.3) $$M'_m \le M_m.$$

If $\xi_1, \ldots, \xi_{k-1}, \xi_{k+1}, \ldots, \xi_m$ are identically 0, then $M'_m = 0$ and $M_m = |\xi_k|$. Thus there can be no inequality opposite to (12.3).

On the other hand, we do have

$$|S_k| \le \min\{|S_m| + |S_k|, |S_m| + |S_m - S_k|\} = |S_m| + \min\{|S_k|, |S_m - S_k|\},$$

so that

(12.4) $$M_m \le M'_m + |S_m|.$$

(There is equality here if $S_m = 0$.) Therefore

(12.5) $$\mathsf{P}\{M_m \ge \lambda\} \le \mathsf{P}\left\{M'_m \ge \frac{\lambda}{2}\right\} + \mathsf{P}\left\{|S_m| \ge \frac{\lambda}{2}\right\}.$$

If we can find separate bounds for the terms on the right in (12.5), we shall have a bound for the term on the left.

If we bound $\mathsf{P}\{M_m \ge \lambda\}$ via (12.5) instead of by a direct analysis of some kind, there may be some loss of accuracy. On the other hand, since the right side of (12.5) is at most $2\mathsf{P}\{M_m \ge \frac{1}{2}\lambda\}$, the loss need not be great: an extra factor of 2 will not matter for our purposes, and the bounds we derive will decrease with increasing λ slowly enough that passing from λ to $\lambda/2$ will have no important effect.

There is a second way of deriving bounds on the distribution of M_m from bounds on the distribution of M'_m. We shall show that

$$M_m \leq 3M'_m + \max_{1 \leq i \leq m} |\xi_i|, \tag{12.6}$$

from which it will follow that

$$\mathsf{P}\{M_m \geq \lambda\} \leq \mathsf{P}\left\{M'_m \geq \frac{\lambda}{4}\right\} + \mathsf{P}\left\{\max_{1 \leq i \leq m} |\xi_i| \geq \frac{\lambda}{4}\right\}. \tag{12.7}$$

Again, bounding the terms on the right will yield a bound for the term on the left. And the right side of (12.7) is at most $2\mathsf{P}\{M_m \geq \frac{1}{8}\lambda\}$, so that using (12.7) can result in no substantial loss of accuracy.

To prove (12.6), consider the set I consisting of those k, $0 \leq k \leq m$, for which $|S_k| \leq |S_m - S_k|$; certainly, $0 \in I$. If $S_m = 0$, then $M_m = M'_m$, so that (12.6) holds. If $S_m \neq 0$, then $m \notin I$ and hence there is a k, $0 < k \leq m$, for which $k - 1 \in I$ and $k \notin I$: $|S_{k-1}| \leq |S_m - S_{k-1}|$, $|S_{k-1}| \leq M'_m$, $|S_m - S_k| < |S_k|$, and $|S_m - S_k| \leq M'_m$. For this k we have

$$|S_m| \leq |S_{k-1}| + |\xi_k| + |S_m - S_k| \leq 2M'_m + |\xi_k|, \tag{12.8}$$

from which (12.6) follows via (12.4). (There is equality in (12.6) if $m = 5$ and $\xi_1 = \xi_2 = \xi_3 = \xi_4 = -\xi_5$.)

Product Moments

In view of (12.5) and (12.7), there is profit in bounding $\mathsf{P}\{M'_m \geq \lambda\}$. Let us suppose that there exist nonnegative numbers $u_1, \ldots, u_m$ such that

$$\mathsf{E}\{|S_j - S_i|^\gamma \cdot |S_k - S_j|^\gamma\} \leq \left(\sum_{i<l\leq j} u_l\right)^\alpha \left(\sum_{j<l\leq k} u_l\right)^\alpha, \qquad 0 \leq i \leq j \leq k \leq m, \tag{12.9}$$

where γ and α are positive. For example, this will hold with $\gamma = 2$ and $\alpha = 1$ if the ξ_i are independent, if $\mathsf{E}\{\xi_i\} = 0$, and if u_i is taken to be the variance of ξ_i.

Since $xy \leq (x + y)^2$ for nonnegative x and y, (12.9) implies

$$\mathsf{E}\{|S_j - S_i|^\gamma \cdot |S_k - S_j|^\gamma\} \leq \left(\sum_{i<l\leq k} u_l\right)^{2\alpha}, \qquad 0 \leq i \leq j \leq k \leq m. \tag{12.10}$$

Moreover, if $|S_j - S_i| \geq \lambda$ and $|S_k - S_j| \geq \lambda$, where λ is positive, then $|S_j - S_i|^\gamma \cdot |S_k - S_j|^\gamma \geq \lambda^{2\gamma}$, so that (12.10) implies

$$\mathsf{P}\{|S_j - S_i| \geq \lambda, |S_k - S_j| \geq \lambda\} \leq \frac{1}{\lambda^{2\gamma}}\left(\sum_{i<l\leq k} u_l\right)^{2\alpha}, \qquad 0 \leq i \leq j \leq k \leq m. \tag{12.11}$$

(The inequality is trivial if $i = j$ or if $j = k$.) Note that (12.9) implies (12.10) and (12.10) implies (12.11) also in the case $\gamma = 0$, provided we take $|x|^0$ as 1 if $x \neq 0$ and as 0 if $x = 0$.

THEOREM 12.1 *If $\gamma \geq 0$ and $\alpha > \frac{1}{2}$, and if* (12.11) *holds for all positive λ, then, for all positive λ,*

$$\mathsf{P}\{M'_m \geq \lambda\} \leq \frac{K_{\gamma,\alpha}}{\lambda^{2\gamma}}(u_1 + \cdots + u_m)^{2\alpha}, \tag{12.12}$$

where $K_{\gamma,\alpha}$ is a constant depending only on γ and α.

Although its specific value will have no importance for us, the constant $K_{\gamma,\alpha}$ may be taken as

$$K_{\gamma,\alpha} = \left[\frac{1}{2^{1/(2\gamma+1)}} - \left(\frac{1}{2^{1/(2\gamma+1)}}\right)^{2\alpha}\right]^{-(2\gamma+1)} \tag{12.13}$$

Theorem 12.1 is of purely theoretical value: $K_{2,1}$, for example, is 55,021 to the nearest integer.

Applications

Before proving the theorem, let us examine several of its consequences. If the ξ_i are independent with $\mathsf{E}\{\xi_i\} = 0$ and $\mathsf{E}\{\xi_i^2\} = \sigma_i^2$, then (12.9) holds with $\gamma = 2$ and $\alpha = 1$, provided u_i is taken to be σ_i^2. Now $\sigma_1^2 + \cdots + \sigma_m^2 = s_m^2$ is the variance of S_m, and (12.12) implies

$$\mathsf{P}\{M'_m \geq \lambda\} \leq \frac{Ks_m^4}{\lambda^4}, \tag{12.14}$$

with $K = K_{2,1}$. Replacing λ by λs_m and using (12.5) and (12.7), we obtain

$$\mathsf{P}\{M_m \geq \lambda s_m\} \leq \frac{2^4 K}{\lambda^4} + \mathsf{P}\{|S_m| \geq \tfrac{1}{2}\lambda s_m\} \tag{12.15}$$

and

$$\mathsf{P}\{M_m \geq \lambda s_m\} \leq \frac{4^4 K}{\lambda^4} + \mathsf{P}\left\{\max_{1 \leq i \leq m} |\xi_i| \geq \tfrac{1}{4}\lambda s_m\right\}. \tag{12.16}$$

According to (10.7),

$$\mathsf{P}\{M_m \geq \lambda s_m\} \leq 2\mathsf{P}\{|S_m| \geq (\lambda - \sqrt{2})s_m\}. \tag{12.17}$$

The factor 2 on the right here is of no importance and, if λ is large, $\frac{1}{2}\lambda$ and $\lambda - \sqrt{2}$ are about the same. Thus (12.15) represents no advance over (12.17) in the independent case. But (12.16) does, because

$$\begin{aligned} \mathsf{P}\left\{\max_{1 \leq i \leq m} |\xi_i| \geq \tfrac{1}{4}\lambda s_m\right\} &\leq \sum_{i=1}^{m} \mathsf{P}\{|\xi_i| \geq \tfrac{1}{4}\lambda s_m\} \\ &\leq \frac{4^2}{\lambda^2}\frac{1}{s_m^2}\sum_{i=1}^{m}\int_{\{|\xi_i| \geq \frac{1}{4}\lambda s_m\}} \xi_i^2 \, d\mathsf{P}, \end{aligned} \tag{12.18}$$

so that (12.16) implies

$$\mathsf{P}\{M_m \geq \lambda s_m\} \leq \frac{4^4 K}{\lambda^4} + \frac{4^2}{\lambda^2}\frac{1}{s_m^2}\sum_{i=1}^m \int_{\{|\xi_i| \geq \frac{1}{4}\lambda s_m\}} \xi_i^2 \, d\mathsf{P}. \tag{12.19}$$

The sum on the right is the one involved in Lindeberg's condition (7.3).

It is possible to deduce further results. For example, from (12.14) it follows (see (3) on p. 223) that

$$\int_{\{(M_m')^2 \geq \alpha s_m^2\}} \left(\frac{M'}{s_m}\right)^2 d\mathsf{P} \leq 2\frac{K}{\alpha}.$$

Since the ξ_i are independent, we have (the indices in the sums are constrained by $1 \leq i, k \leq m$)

$$\begin{aligned}
\int_{\{\max_k \xi_k^2 \geq \alpha s_m^2\}} \max_i \frac{\xi_i^2}{s_m^2}\, d\mathsf{P} &\leq \sum_{k,i} \int_{\{\xi_k^2 \geq \alpha s_m^2\}} \frac{\xi_i^2}{s_m^2}\, d\mathsf{P} \\
&= \sum_i \int_{\{\xi_i^2 \geq \alpha s_m^2\}} \frac{\xi_i^2}{s_m^2}\, d\mathsf{P} + \sum_{k \neq i} \mathsf{P}\{\xi_k^2 \geq \alpha s_m^2\} \frac{\sigma_i^2}{s_m^2} \\
&\leq \left(1 + \frac{1}{\alpha}\right) \sum_i \frac{1}{s_m^2} \int_{\{\xi_i^2 \geq \alpha s_m^2\}} \xi_i^2 \, d\mathsf{P}.
\end{aligned}$$

For nonnegative U and V, we have

$$\int_{\{U+V \geq \alpha\}} (U+V)\, d\mathsf{P} \leq 2\int_{\{U \geq \frac{1}{2}\alpha\}} U\, d\mathsf{P} + 2\int_{\{V \geq \frac{1}{2}\alpha\}} V\, d\mathsf{P}.$$

That

$$\int_{\{M_m^2 \geq \alpha s_m^2\}} \left(\frac{M_m^2}{s_m^2}\right) d\mathsf{P} \leq K'\left[\frac{1}{\alpha} + \sum_{i=1}^m \frac{1}{s_m^2}\int_{\{|\xi_i| \geq \frac{1}{4}\alpha s_m\}} \xi_i^2 \, d\mathsf{P}\right] \tag{12.20}$$

for some universal constant K' now follows by (12.6) and the fact that the sum on the right here is at most 1. The inequality persists if M_m^2 is replaced by S_m^2. This proves in particular the uniform integrability of the squares of the normalized row sums for a triangular array satisfying Lindeberg's condition.

Suppose now that the ξ_i are independent and identically distributed with $\mathsf{E}\{\xi_i\} = 0$ and $\mathsf{E}\{\xi_i^2\} = \sigma^2$. Then (12.15) implies

$$\mathsf{P}\{M_m \geq \lambda\sigma\sqrt{m}\} \leq \frac{2^4 K}{\lambda^4} + \mathsf{P}\{|S_m| \geq \tfrac{1}{2}\lambda\sigma\sqrt{m}\}; \tag{12.21}$$

(12.16) implies

$$\mathsf{P}\{M_m \geq \lambda\sigma\sqrt{m}\} \leq \frac{4^4 K}{\lambda^4} + \mathsf{P}\left\{\max_{1 \leq i \leq m} |\xi_i| \geq \tfrac{1}{4}\lambda\sigma\sqrt{m}\right\}; \tag{12.22}$$

and (12.19) implies

$$\mathsf{P}\{M_m \geq \lambda\sigma\sqrt{m}\} \leq \frac{4^4 K}{\lambda^4} + \frac{4}{\lambda^2\sigma^2}\int_{\{|\xi_1| \geq \frac{1}{4}\lambda\sigma\sqrt{m}\}} \xi_1^2 \, d\mathsf{P}. \tag{12.23}$$

Since the integral here goes to 0 as $\lambda \to \infty$, for sufficiently large λ we have

$$\mathsf{P}\{M_m \ge \lambda\sigma\sqrt{m}\} \le \frac{\varepsilon}{\lambda^2}, \qquad m = 1, 2, \ldots .$$

Thus we have another proof of the tightness of the random functions involved in Donsker's theorem (see (10.10)), a proof which does not depend on the central limit theorem.

Although the random variables are independent in the applications just given, Theorem 12.1 does not require independence. This is a great advantage: Most later applications will involve dependent sequences.

Proof of Theorem 12.1

The assumption in Theorem 12.1 is that

$$\mathsf{P}\{|S_j - S_i| \ge \lambda, |S_k - S_j| \ge \lambda\} \le \frac{1}{\lambda^{2\gamma}}\Big(\sum_{i<l\le k} u_l\Big)^{2\alpha}, \qquad 0 \le i \le j \le k \le m, \tag{12.24}$$

for all positive λ; we are to find a constant K, depending only on γ and α, for which

$$\mathsf{P}\{M'_m \ge \lambda\} \le \frac{K}{\lambda^{2\gamma}}(u_1 + \cdots + u_m)^{2\alpha}. \tag{12.25}$$

Write

$$\delta = \frac{1}{2\gamma + 1}, \tag{12.26}$$

so that $0 < \delta \le 1$. For large K we have

$$2^\delta\left[\frac{1}{2^{2\alpha\delta}} + \frac{1}{K^\delta}\right] \le 1, \tag{12.27}$$

because the left-hand member approaches $1/2^{(2\alpha-1)\delta}$ as $K \to \infty$ and ($2\alpha > 1$, $\delta > 0$) this limit is less than 1. We shall prove that (12.25) holds if K satisfies (12.27) and if

$$K \ge 1. \tag{12.28}$$

(In point of fact, (12.27) implies (12.28). The smallest K satisfying (12.27) is given by (12.13).)

The proof goes by induction on m. The result being trivial for $m = 1$, consider the case $m = 2$. Since $M'_2 = \min\{|S_1|, |S_2 - S_1|\}$, (12.24) and (12.28) imply

$$\mathsf{P}\{M'_2 \ge \lambda\} \le \frac{1}{\lambda^{2\gamma}}(u_1 + u_2)^{2\alpha} \le \frac{K}{\lambda^{2\gamma}}(u_1 + u_2)^{2\alpha}.$$

Assume now as induction hypothesis that the result holds for each integer less than m. We shall prove it for m itself. Write $u = u_1 + \cdots + u_m$; we may assume $u > 0$. There exists an integer h, $1 \le h \le m$, such that

$$\frac{u_1 + \cdots + u_{h-1}}{u} \le \frac{1}{2} \le \frac{u_1 + \cdots + u_h}{u}, \tag{12.29}$$

where the sum on the left is 0 if $h = 1$.

Consider the four quantities

$$\begin{aligned}
U_1 &= \max_{0 \le i \le h-1} \min \{|S_i|, |S_{h-1} - S_i|\},\\
U_2 &= \max_{h \le j \le m} \min \{|S_j - S_h|, |(S_m - S_h) - (S_j - S_h)|\}\\
&= \max_{h \le j \le m} \min \{|S_j - S_h|, |S_m - S_j|\},\\
D_1 &= \min \{|S_{h-1}|, |S_m - S_{h-1}|\},\\
D_2 &= \min \{|S_h|, |S_m - S_h|\}.
\end{aligned}$$

Since (12.24) holds if m is replaced by $h - 1$, and since $h - 1 < m$, we may apply the induction hypothesis to the random variables $\xi_1, \ldots, \xi_{h-1}$ and the quantities $u_1, \ldots, u_{h-1}$ and conclude that

$$\mathsf{P}\{U_1 \ge \lambda\} \le \frac{K}{\lambda^{2\gamma}}(u_1 + \cdots + u_{h-1})^{2\alpha} \le \frac{u^{2\alpha}}{\lambda^{2\gamma}}\frac{K}{2^{2\alpha}}, \tag{12.30}$$

the last inequality following by (12.29). (If $h = 1$, (12.30) is trivial.)

If the indices in (12.24) are restricted to $h \le i \le j \le k \le m$, then only the random variables $\xi_{h+1}, \ldots, \xi_m$ are involved. Since $m - h < m$, the induction hypothesis applies to $\xi_{h+1}, \ldots, \xi_m$ and $u_{h+1}, \ldots, u_m$; hence

$$\mathsf{P}\{U_2 \ge \lambda\} \le \frac{K}{\lambda^{2\gamma}}(u_{h+1} + \cdots + u_m)^{2\alpha} \le \frac{u^{2\alpha}}{\lambda^{2\gamma}}\frac{K}{2^{2\alpha}}, \tag{12.31}$$

the last inequality following by (12.29) again. (If $h = m$, (12.31) is trivial.)

Now (12.24) implies

$$\mathsf{P}\{D_1 \ge \lambda\} \le \frac{1}{\lambda^{2\gamma}}(u_1 + \cdots + u_m)^{2\alpha} = \frac{u^{2\alpha}}{\lambda^{2\gamma}} \tag{12.32}$$

and

$$\mathsf{P}\{D_2 \ge \lambda\} \le \frac{u^{2\alpha}}{\lambda^{2\gamma}}. \tag{12.33}$$

(If $h = 1$, then (12.32) is trivial; if $h = m$, then (12.33) is trivial.)

We shall show that

$$\min \{|S_i|, |S_m - S_i|\} \le U_1 + D_1 \qquad \text{if} \qquad 0 \le i \le h - 1. \tag{12.34}$$

Let μ_i be the minimum on the left. If $|S_i| \le U_1$, then

$$\mu_i \le |S_i| \le U_1 \le U_1 + D_1.$$

Suppose $|S_{h-1} - S_i| \le U_1$; if $|S_{h-1}| = D_1$, then

$$\mu_i \le |S_i| \le |S_{h-1} - S_i| + |S_{h-1}| \le U_1 + D_1.$$

Suppose $|S_{h-1} - S_i| \le U_1$ and $|S_m - S_{h-1}| = D_1$; then

$$\mu_i \le |S_m - S_i| \le |S_{h-1} - S_i| + |S_m - S_{h-1}| \le U_1 + D_1.$$

This proves (12.34). The same sort of argument shows that

$$\min\{|S_j|, |S_m - S_j|\} \le U_2 + D_2 \qquad \text{if} \quad h \le j \le m. \tag{12.35}$$

By (12.34) and (12.35),

$$M'_m \le \max\{U_1 + D_1, U_2 + D_2\}, \tag{12.36}$$

and hence

$$\mathsf{P}\{M'_m \ge \lambda\} \le \mathsf{P}\{U_1 + D_1 \ge \lambda\} + \mathsf{P}\{U_2 + D_2 \ge \lambda\}. \tag{12.37}$$

If λ_0 and λ_1 are positive and $\lambda_0 + \lambda_1 = \lambda$, then, by (12.30) and (12.32),

$$\mathsf{P}\{U_1 + D_1 \ge \lambda\} \le \mathsf{P}\{U_1 \ge \lambda_0\} + \mathsf{P}\{D_1 \ge \lambda_1\} \le \frac{u^{2\alpha}}{\lambda_0^{2\gamma}}\frac{K}{2^{2\alpha}} + \frac{u^{2\alpha}}{\lambda_1^{2\gamma}}. \tag{12.38}$$

Calculus shows that, if C_0, C_1, and λ are positive numbers, then

$$\min_{\substack{\lambda_0, \lambda_1 > 0 \\ \lambda_0 + \lambda_1 = \lambda}} \left[\frac{C_0}{\lambda_0^{2\gamma}} + \frac{C_1}{\lambda_1^{2\gamma}}\right] = \frac{1}{\lambda^{2\gamma}}\,[C_0^{\delta} + C_1^{\delta}]^{1/\delta}, \tag{12.39}$$

with δ defined by (12.26). Minimizing the right-most member of (12.38), we arrive at

$$\mathsf{P}\{U_1 + D_1 \ge \lambda\} \le \frac{u^{2\alpha}}{\lambda^{2\gamma}}\left[\left(\frac{K}{2^{2\alpha}}\right)^{\delta} + 1\right]^{1/\delta}. \tag{12.40}$$

The same inequality holds for $U_2 + D_2$ (use (12.31) and (12.33)). By (12.37), therefore

$$\mathsf{P}\{M'_m \ge \lambda\} \le \frac{u^{2\alpha}}{\lambda^{2\gamma}}\,2\left[\left(\frac{K}{2^{2\alpha}}\right)^{\delta} + 1\right]^{1/\delta}. \tag{12.41}$$

By the choice (12.27) of K, the right member here is at most $Ku^{2\alpha}/\lambda^{2\gamma}$. This completes the induction step and the proof of Theorem 12.1.

Moments†

Let us now replace (12.9) by the assumption that

$$\mathsf{E}\{|S_j - S_i|^\gamma\} \le \Big(\sum_{i<l\le j} u_l\Big)^\alpha, \qquad 0 \le i \le j \le m. \tag{12.42}$$

It follows from this that, if λ is positive, then

$$\mathsf{P}\{|S_j - S_i| \ge \lambda\} \le \frac{1}{\lambda^\gamma}\Big(\sum_{i<l\le j} u_l\Big)^\alpha, \qquad 0 \le i \le j \le m. \tag{12.43}$$

In place of $\alpha > \frac{1}{2}$ we make the stronger requirement that $\alpha > 1$.

THEOREM 12.2 *If* $\gamma \ge 0$ *and* $\alpha > 1$, *and if* (12.43) *holds for all positive* λ, *then, for all positive* λ,

$$\mathsf{P}\{M_m \ge \lambda\} \le \frac{K'_{\gamma,\alpha}}{\lambda^\gamma}(u_1 + \cdots + u_m)^\alpha, \tag{12.44}$$

where $K'_{\gamma,\alpha}$ *depends only on* γ *and* α.

We may define $K'_{\gamma,\alpha}$ by

$$K'_{\gamma,\alpha} = 2^\gamma(1 + K_{\frac{1}{2}\gamma,\frac{1}{2}\alpha}). \tag{12.45}$$

Proof. By Schwarz's inequality, $\mathsf{P}(E_1 \cap E_2) \le \mathsf{P}^{\frac{1}{2}}(E_1)\mathsf{P}^{\frac{1}{2}}(E_2)$. Therefore (12.43) implies (recall $xy \le (x+y)^2$)

$$\begin{aligned}\mathsf{P}\{|S_j - S_i| \ge \lambda, |S_k - S_j| \ge \lambda\} &\le \frac{1}{\lambda^{\gamma/2}}\Big(\sum_{i<l\le j} u_l\Big)^{\alpha/2}\frac{1}{\lambda^{\gamma/2}}\Big(\sum_{j<l\le k} u_l\Big)^{\alpha/2} \\ &\le \frac{1}{\lambda^\gamma}\Big(\sum_{i<l\le k} u_l\Big)^\alpha.\end{aligned}$$

Thus (12.11) holds with γ and α replaced by $\frac{1}{2}\gamma$ and $\frac{1}{2}\alpha$, and Theorem 12.1 implies

$$\mathsf{P}\{M'_m \ge \lambda\} \le \frac{K}{\lambda^\gamma}(u_1 + \cdots + u_m)^\alpha \tag{12.46}$$

with $K = K_{\frac{1}{2}\gamma,\frac{1}{2}\alpha}$. By (12.43) we have

$$\mathsf{P}\{|S_m| \ge \lambda\} \le \frac{1}{\lambda^\gamma}(u_1 + \cdots + u_m)^\alpha, \tag{12.47}$$

and (12.44) follows by (12.5). (Note that, since the right-hand members of the last two inequalities are of the same order, (12.44) cannot be essentially improved by using extra information about the distribution of $|S_m|$.)

† The remaining results in this section are needed in Chapters 3 and 4 but not in Section 13, the last in this chapter.

As an application of Theorem 12.2, consider again independent, identically distributed variables ξ_i with mean 0 and variance σ^2; this time assume there exists a finite fourth moment $\tau^4 = \mathsf{E}\{\xi_i^4\}$. Now

$$\mathsf{E}\{S_r^4\} = \sum \mathsf{E}\{\xi_i \xi_j \xi_k \xi_l\},$$

where the indices range independently from 1 to r. Since the ξ_i are independent and their means vanish, if the value of some index in the summand differs from those of the other three, the term vanishes. Since $\sigma^4 \leq \tau^4$,

$$\mathsf{E}\{S_r^4\} = r\tau^4 + 3r(r-1)\sigma^4 \leq 4r^2\tau^4,$$

and hence (12.42) holds with $\gamma = 4$, $\alpha = 2$, and $u_1 = \cdots = u_m = 2\tau^2$.

Theorem 12.2 now implies

$$\mathsf{P}\{M_m \geq \lambda\} \leq 4\tau^4 K \frac{m^2}{\lambda^4}, \tag{12.48}$$

where $K = K'_{4,2}$. Replacing λ by $\lambda\sigma\sqrt{m}$ yields

$$\mathsf{P}\{M_m \geq \lambda\sigma\sqrt{m}\} \leq 4\tau^4 K \frac{1}{\lambda^4\sigma^4}. \tag{12.49}$$

From this and Theorem 8.4, the tightness of the random functions (10.1) involved in Donsker's theorem follows easily. The argument involving (12.23) is more powerful than this one because it requires only second moments.

A Tightness Criterion

Theorem 12.2 leads to a condition for tightness of a sequence $\{X_n\}$ of random elements of C.

THEOREM 12.3 *The sequence $\{X_n\}$ is tight if it satisfies these two conditions:*

(i) *The sequence $\{X_n(0)\}$ is tight.*

(ii) *There exist constants $\gamma \geq 0$ and $\alpha > 1$ and a nondecreasing, continuous function F on $[0, 1]$ such that*

$$\mathsf{P}\{|X_n(t_2) - X_n(t_1)| \geq \lambda\} \leq \frac{1}{\lambda^\gamma} |F(t_2) - F(t_1)|^\alpha \tag{12.50}$$

holds for all t_1, t_2, and n and all positive λ.

The moment condition

$$\mathsf{E}\{|X_n(t_2) - X_n(t_1)|^\gamma\} \leq |F(t_2) - F(t_1)|^\alpha \tag{12.51}$$

implies (12.50).

Proof. By Theorem 8.2 it suffices to produce, given ε and η, a δ $(0 < \delta < 1)$ for which

$$\mathsf{P}\{w(X_n, \delta) \geq 3\varepsilon\} \leq \eta \tag{12.52}$$

for all n, and, by the corollary to Theorem 8.3, this will hold if δ^{-1} is an integer and

$$\sum_{j<\delta^{-1}} \mathsf{P}\left\{\sup_{j\delta \leq s \leq (j+1)\delta} |X_n(s) - X_n(j\delta)| \geq \varepsilon\right\} \leq \eta. \tag{12.53}$$

Fix n, δ, and j for the moment and, for a positive integer m, consider the random variables

$$\xi_i = X_n\left(j\delta + \frac{i}{m}\,\delta\right) - X_n\left(j\delta + \frac{i-1}{m}\,\delta\right), \qquad i = 1, 2, \ldots, m. \tag{12.54}$$

By (12.50), these random variables satisfy (12.43) with $u_i = F(j\delta + i\delta m^{-1}) - F(j\delta + (i-1)\delta m^{-1})$. By Theorem 12.2, therefore,

$$\begin{aligned}\mathsf{P}\left\{\max_{0 \leq i \leq m}\left| X_n\left(j\delta + \frac{i}{m}\,\delta\right) - X_n(j\delta)\right| \geq \varepsilon\right\}& \\ \leq \frac{K}{\varepsilon^{\gamma}}[F((j+1)\delta) - F(j\delta)]^{\alpha}&,\end{aligned} \tag{12.55}$$

where $K = K'_{\gamma,\alpha}$. Since X_n lies in C, letting $m \to \infty$ leads to

$$\mathsf{P}\left\{\sup_{j\delta \leq s \leq (j+1)\delta} |X_n(s) - X_n(j\delta)| \geq \varepsilon\right\} \leq \frac{K}{\varepsilon^{\gamma}}[F((j+1)\delta) - F(j\delta)]^{\alpha}. \tag{12.56}$$

Therefore

$$\begin{aligned}\sum_{j<\delta^{-1}} \mathsf{P}\left\{\sup_{j\delta \leq s \leq (j+1)\delta} |X_n(s) - X_n(j\delta)| \geq \varepsilon\right\}& \\ \leq \frac{K}{\varepsilon^{\gamma}}[F(1) - F(0)]\left[\max_{j<\delta^{-1}} |F((j+1)\delta) - F(j\delta)|\right]^{\alpha-1}&\end{aligned} \tag{12.57}$$

if δ^{-1} is integral. Since F is continuous and $\alpha > 1$, we may make this last quantity small by taking δ the reciprocal of a large integer, which completes the proof.

The ideas in this proof lead to a condition for the existence in C of a random element with specified finite-dimensional distributions. For each k-tuple of points of $[0, 1]$, let $\mu_{t_1 \cdots t_k}$ be a probability measure on $(R^k, \mathscr{R}^k)$. Assume these measures satisfy the consistency conditions of Kolmogorov's existence theorem (in its general form; see p. 230).

THEOREM 12.4 *There exists in C a random element with finite-dimensional distributions* $\mu_{t_1 \cdots t_k}$, *provided these distributions are consistent and provided* *there exist constants* $\gamma \geq 0$ *and* $\alpha > 1$ *and a nondecreasing, continuous function* F *on* $[0, 1]$ *such that*

$$\mu_{t_1, t_2}\{(\beta_1, \beta_2) : |\beta_1 - \beta_2| \geq \lambda\} \leq \frac{1}{\lambda^\gamma} |F(t_2) - F(t_1)|^\alpha \tag{12.58}$$

holds for all t_1 *and* t_2 *and all positive* λ.

If the $\mu_{t_1 \cdots t_k}$ satisfy

$$\int_{R^2} |\beta_1 - \beta_2|^\gamma \, d\mu_{t_1, t_2}(\beta_1, \beta_2) \leq |F(t_2) - F(t_1)|^\alpha, \tag{12.59}$$

then (12.58) follows.

Proof. For each n, construct a polygonal random function X_n that is linear over each interval $[(i-1)2^{-n}, i2^{-n}]$ and for which the joint distribution of

$$X_n(0), X_n\left(\frac{1}{2^n}\right), \ldots, X_n\left(\frac{2^n - 1}{2^n}\right), X_n(1)$$

is $\mu_{t_0, \ldots, t_{2^n}}$ with $t_i = i/2^n$. If t_1 and t_2 are integral multiples of 2^{-n}, then, by (12.58),

$$\mathsf{P}\{|X_n(t_2) - X_n(t_1)| \geq \lambda\} \leq \frac{1}{\lambda^\gamma} |F(t_2) - F(t_1)|^\alpha. \tag{12.60}$$

As in the preceding proof, consider fixed n, δ, and j, where δ^{-1} is assumed integral. If the points

$$j\delta + i\delta m^{-1}, \qquad i = 0, 1, \ldots, m \tag{12.61}$$

involved in (12.54) are all integral multiples if 2^{-n}, then (12.55) follows as before. Suppose now that $\delta 2^n$ is an integer. If we take $m = \delta 2^n$, then the points (12.61) are indeed integral multiples of 2^{-n}, so that (12.55) holds. Moreover the points (12.61) are in this case exactly the integral multiples of 2^{-n} in the interval $[j\delta, (j+1)\delta]$, so that, because of the polygonal character of X_n, (12.56) and (12.57) again hold.

Thus (12.57) holds if δ^{-1} and $\delta 2^n$ are both integers. If we take $\delta = 2^{-v}$, where v is an integer large enough that the right side of (12.57) is less than η, then (12.52) holds for all $n \geq v$, which, since the $X_n(0)$ all have the same distribution, is enough for tightness.

By Prohorov's theorem, therefore, there is a random element X of C and a subsequence $\{X_{n'}\}$ such that $X_{n'} \xrightarrow{\mathscr{D}} X$. If $t_1, \ldots, t_k$ are dyadic rationals, then, by the consistency assumption, the distribution of $(X_n(t_1), \ldots, X_n(t_k))$ is exactly $\mu_{t_1 \cdots t_k}$ for all sufficiently large n, and it follows that $(X(t_1), \ldots, X(t_k))$ has distribution $\mu_{t_1 \cdots t_k}$. For general points $t_1, \ldots, t_k$ of $[0, 1]$, there are dyadic rationals $s_1^{(v)}, \ldots, s_k^{(v)}$ with $\lim_v s_i^{(v)} = t_i$; since

$(X(s_1^{(v)}), \ldots, X(s_k^{(v)}))$ converges in distribution to $(X(t_1), \ldots, X(t_k))$, and since, by (12.58) and the continuity of F, $\mu_{s_1^{(v)} \cdots s_k^{(v)}}$ converges weakly to $\mu_{t_1 \cdots t_k}$, $(X(t_1), \ldots, X(t_k))$ has $\mu_{t_1 \cdots t_k}$ as its distribution. Thus X is the random function required.

The existence of W and W° follow anew from Theorem 12.4.†

Further Inequalities

For Chapter 3, we shall need a strengthening of Theorem 12.1. If

(12.62) $$M''_m = \max_{0 \le i \le j \le k \le m} \min \{|S_j - S_i|, |S_k - S_j|\},$$

then

(12.63) $$M'_m \le M''_m.$$

(If $m = 3$ and $\xi_1 = -\xi_2 = \xi_3$, then $M'_m = 0$, and $M''_m = |\xi_1|$.) Therefore bounds on the tail of the distribution of M''_m are stronger than bounds on the tail of the distribution of M'_m, and the following result sharpens Theorem 12.1.

THEOREM 12.5 *If $\gamma \ge 0$ and $\alpha > \frac{1}{2}$, and if*

(12.64) $$\mathsf{P}\{|S_j - S_i| \ge \lambda, |S_k - S_j| \ge \lambda\} \le \frac{1}{\lambda^{2\gamma}} \Big(\sum_{i < l \le k} u_l \Big)^{2\alpha}, \quad 0 \le i \le j \le k \le m,$$

holds for all positive λ, then, for all positive λ,

(12.65) $$\mathsf{P}\{M''_m \ge \lambda\} \le \frac{K''_{\gamma,\alpha}}{\lambda^{2\gamma}} (u_1 + \cdots + u_m)^{2\alpha},$$

where $K''_{\gamma,\alpha}$ is a constant depending only on γ and α.

Proof. Put

(12.66) $$N_m = \min_{1 \le l \le m} \max \Big\{ \max_{0 \le i < l} |S_i|, \max_{l \le i \le m} |S_m - S_i| \Big\}.$$

If l is the index that achieves the minimum here and $i \le j \le k$, then either $i, j < l$, or else $l \le j, k$, from which it follows that

(12.67) $$M''_m \le 2N_m.$$

Hence it will suffice to find a K, depending only on γ and α, such that (12.64)

† Theorem 12.4 combines with Theorem 9.2 to give a condition for the continuity of sample paths of separable processes. For extensions and variations of Theorems 12.3 and 12.4, see Problems 7 and 8 in this section and Problem 4 in Section 15.

implies

$$\mathsf{P}\{N_m \geq \lambda\} \leq \frac{K}{\lambda^{2\gamma}}(u \ + \cdots + u_m)^{2\alpha}. \tag{12.68}$$

(Since $N_m \leq 2M''_m$, (12.68) and (12.65) have the same strength.)

For sufficiently large K, $K \geq 1$ and

$$\left(\frac{2K}{2^{2\alpha}}\right)^{\delta} + (4 \cdot 2^{2\gamma})^{\delta} \leq K^{\delta}, \tag{12.69}$$

where $\delta = 1/(2\gamma + 1)$, as before. We shall prove by induction on m that such a K works. The cases $m = 1$ and $m = 2$ are easy.

Assume as induction hypothesis that the result holds for integers smaller than m. As in the proof of Theorem 12.1, choose h, $1 \leq h \leq m$, so that

$$\frac{u_1 + \cdots + u_{h-1}}{u} \leq \frac{1}{2} \leq \frac{u_1 + \cdots + u_h}{u},$$

where $u = u_1 + \cdots + u_m$.

Write A_1 for N_{h-1}:

$$A_1 = \min_{1 \leq l_1 < h} \max \left\{ \max_{0 \leq i < l_1} |S_i|, \max_{l_1 \leq i < h} |S_{h-1} - S_i| \right\}. \tag{12.70}$$

Write A_2 for the value of N_{m-h} based on the quantities $\xi_{h+1}, \ldots, \xi_m$:

$$A_2 = \min_{h < l_2 \leq m} \max \left\{ \max_{h \leq i < l_2} |S_i - S_h|, \max_{l_2 \leq i \leq m} |S_m - S_i| \right\}. \tag{12.71}$$

By the induction hypothesis applied to $\xi_1, \ldots, \xi_{h-1}$, we have

$$\mathsf{P}\{A_1 \geq \lambda\} \leq \frac{K}{\lambda^{2\gamma}}(u_1 + \cdots + u_{h-1})^{2\alpha} \leq \frac{u^{2\alpha}}{\lambda^{2\gamma}} \frac{K}{2^{2\alpha}}. \tag{12.72}$$

By the induction hypothesis applied to $\xi_{h+1}, \ldots, \xi_m$, we have

$$\mathsf{P}\{A_2 \geq \lambda\} \leq \frac{K}{\lambda^{2\gamma}}(u_{h+1} + \cdots + u_m)^{2\alpha} \leq \frac{u^{2\alpha}}{\lambda^{2\gamma}} \frac{K}{2^{2\alpha}}. \tag{12.73}$$

Write

$$\mu(i, j, k) = \min \{|S_j - S_i|, |S_k - S_j|\}$$

and define

(12.74)

$$B = \max \{\mu(0, h-1, m); \mu(0, h-1, h); \mu(h-1, h, m); \mu(0, h, m)\}.$$

By (12.64),

$$\mathsf{P}\{B \geq \lambda\} \leq 4 \frac{u^{2\alpha}}{\lambda^{2\gamma}}. \tag{12.75}$$

We next show that

$$N_m \leq \max\{A_1, A_2\} + 2B. \tag{12.76}$$

Let l_1 and l_2 be the specific values of the indices that achieve the minima in (12.70) and (12.71). To prove (12.76), we must produce an l, $1 \leq l \leq m$, for which

$$\max_{0<i<l} |S_i| \leq \max\{A_1, A_2\} + 2B \tag{12.77}$$

and

$$\max_{l\leq i\leq m} |S_m - S_i| \leq \max\{A_1, A_2\} + 2B. \tag{12.78}$$

Suppose first that

$$|S_{h-1}| \leq B \tag{12.79}$$

and

$$|S_m - S_h| \leq B. \tag{12.80}$$

We shall show that the choice $l = h$ satisfies (12.77) and (12.78). Indeed, $0 \leq i < l_1$ implies $|S_i| \leq A_1$, and $l_1 \leq i < l = h$ implies

$$|S_i| \leq |S_i - S_{h-1}| + |S_{h-1}| \leq A_1 + B$$

because of (12.79), so that (12.77) holds; and (12.78) is established in the same way.

Suppose now that one or the other of (12.79) and (12.80) fails. For definiteness, suppose (12.79) fails. We shall show that the choice $l = l_1$ satisfies (12.77) and (12.78). Since (12.79) fails, it follows by the definition (12.74) of B that $|S_m - S_{h-1}| \leq B$ and $|\xi_h| \leq B$. If $0 \leq i < l = l_1$, then $|S_i| \leq A_1$. Hence (12.77) holds. If $l = l_1 \leq i < h$, then

$$|S_m - S_i| \leq |S_{h-1} - S_i| + |S_m - S_{h-1}| \leq A_1 + B;$$

if $h \leq i < l_2$, then

$$|S_m - S_i| \leq |S_i - S_h| + |\xi_h| + |S_m - S_{h-1}| \leq A_2 + 2B;$$

if $l_2 \leq i \leq m$, then $|S_m - S_i| \leq A_2$. Hence (12.78) holds. If (12.80) fails instead of (12.79), then (12.77) and (12.78) hold with $l = l_2$. This proves (12.76).

For positive numbers λ_0 and λ_1 adding to λ, (12.76) implies

$$\mathsf{P}\{N_m \geq \lambda\} \leq \mathsf{P}\{A_1 \geq \lambda_0\} + \mathsf{P}\{A_2 \geq \lambda_0\} + \mathsf{P}\{B \geq \tfrac{1}{2}\lambda_1\}. \tag{12.81}$$

Applying the inequalities (12.72), (12.73), and (12.75), we get

$$\mathsf{P}\{N_m \geq \lambda\} \leq \frac{u^{2\alpha}}{\lambda_0^{2\gamma}} \frac{2K}{2^{2\alpha}} + 4 \cdot 2^{2\gamma} \frac{u^{2\alpha}}{\lambda_1^{2\gamma}}.$$

And now (12.39) implies

$$\mathsf{P}\{N_m \geq \lambda\} \leq \frac{u^{2\alpha}}{\lambda^{2\gamma}}\left[\left(\frac{2K}{2^{2\alpha}}\right)^{\delta} + (4\cdot 2^{2\gamma})^{\delta}\right]^{1/\delta},$$

from which (12.68) follows by (12.69), which completes the proof.

The hypotheses of Theorem 12.5 are satisfied if the ξ_i are independent, $\mathsf{E}\{\xi_i\} = 0$, $\mathsf{E}\{\xi_i^2\} = u_i$, $\gamma = 2$, and $\alpha = 1$. In this case, the ξ_i and the u_i also satisfy the stronger hypotheses of the following theorem and hence its stronger conclusion. If all but one of the variances u_i vanish, then $\mathsf{P}\{M''_m \geq \lambda\} = 0$ for $\lambda > 0$; although the right side of (12.65) is positive, the right side of (12.83) vanishes.

THEOREM 12.6 *If $\gamma \geq 0$ and $\alpha > \frac{1}{2}$, and if*

$$\mathsf{P}\{|S_j - S_i| \geq \lambda, |S_k - S_j| \geq \lambda\} \leq \frac{1}{\lambda^{2\gamma}}\Big(\sum_{i<l\leq j} u_l\Big)^{\alpha}\Big(\sum_{j<l\leq k} u_l\Big)^{\alpha} \tag{12.82}$$

for all positive λ, then, for all positive λ,

$$\mathsf{P}\{M''_m \geq \lambda\} \leq \frac{K'''_{\gamma,\alpha}}{\lambda^{2\gamma}}(u_1 + \cdots + u_m)^{2\alpha} \min_{1\leq h\leq m}\left[1 - \frac{u_h}{u_1 + \cdots + u_m}\right]^{\alpha}, \tag{12.83}$$

where $K'''_{\gamma,\alpha}$ depends only on γ and α.

The moment form of (12.82), which implies it, is the inequality (12.9) with which the whole discussion began.

Proof. Choose h to minimize the final factor in (12.83); write $u = u_1 + \cdots + u_m$, $p = (u_1 + \cdots + u_{h-1})/u$, $p_h = u_h/u$, and $q = (u_{h+1} + \cdots + u_m)/u$. As in the preceding proof, define A_1 by (12.70), A_2 by (12.71), and B by (12.74). By (12.82),

$$\mathsf{P}\{\mu(0, h-1, m) \geq \lambda\} \leq \frac{u^{2\alpha}}{\lambda^{2\gamma}} p^{\alpha}(p_h + q)^{\alpha} \leq \frac{u^{2\alpha}}{\lambda^{2\gamma}}(1 - p_h)^{\alpha},$$

and there is a similar inequality for each of the other μ's in (12.74). Therefore

$$\mathsf{P}\{B \geq \lambda\} \leq 4\frac{u^{2\alpha}}{\lambda^{2\gamma}}(1 - p_h)^{\alpha}.$$

Write $K = K''_{\gamma,\alpha}$. Since (12.82) implies (12.64), it follows by Theorem 12.5 (in the stronger form (12.68)) that

$$\mathsf{P}\{A_1 \geq \lambda\} \leq \frac{Ku^{2\alpha}}{\lambda^{2\gamma}} p^{2\alpha} \leq \frac{Ku^{2\alpha}}{\lambda^{2\gamma}}(1 - p_h)^{\alpha}$$

and

$$\mathsf{P}\{A_2 \geq \lambda\} \leq \frac{Ku^{2\alpha}}{\lambda^{2\gamma}} q^{2\alpha} \leq \frac{Ku^{2\alpha}}{\lambda^{2\gamma}}(1 - p_h)^{\alpha}.$$

Now (12.76) holds just as before, and hence

$$\mathsf{P}\{N_m \geq \lambda\} \leq \mathsf{P}\{A_1 \geq \tfrac{1}{2}\lambda\} + \mathsf{P}\{A_2 \geq \tfrac{1}{2}\lambda\} + \mathsf{P}\{B \geq \tfrac{1}{4}\lambda\}.$$

Combining this inequality with the three preceding ones, we arrive at

$$\mathsf{P}\{N_m \geq \lambda\} \leq 4^{2\gamma}(2K + 4)\frac{u^{2\alpha}}{\lambda^{2\gamma}}(1 - p_h)^\alpha.$$

And now (12.83), for an appropriate $K'''_{\gamma,\alpha}$, follows easily by (12.67).

Remarks. The results of this section, new as such, stem from the work of Kolmogorov (see Slutsky (1937)) and Chentsov (1956).

PROBLEMS

1. Do Problem 1 of Section 10 once more, using (12.19) in place of (10.7).

2. If we weaken (12.42) by assuming only that the inequality holds for the special pair of indices $i = 0$, $j = m$, but compensate by assuming that $S_1, \ldots, S_m$ forms a martingale and that $\gamma \geq 1$, then [Doob (1953, p. 314)] $\mathsf{P}\{M_m \geq \lambda\} \leq (u_1 + \cdots + u_m)^\alpha/\lambda^\gamma$, an inequality of essentially the same strength as (12.44).

3. Assume $\xi_1, \ldots, \xi_m$ independent with mean 0 and finite moments of order $2k$. Use the multinomial theorem to find a constant C_k, depending on k alone, such that

$$\mathsf{E}\{S_m^{2k}\} \leq C_k\left[\sum_{i=1}^{m} \mathsf{E}^{1/k}\{\xi_i^{2k}\}\right]^k.$$

Generalize (12.48).

4. From (12.12) deduce that for positive ε

$$\mathsf{E}\{(M'_m)^{2\gamma-\varepsilon}\} \leq \frac{2\gamma}{\varepsilon} K_{\gamma,\alpha}(u_1 + \cdots + u_m)^{2\alpha(1-\varepsilon/2\gamma)}.$$

5. Adapt the proof of Menshov's inequality [Doob (1953, p. 156)] to show that, if (12.10) holds with $\alpha \geq \frac{1}{2}$ and $\gamma \geq \frac{1}{2}$, then

$$\mathsf{E}\{(M'_m)^{2\gamma}\} \leq (\log_2 2m)^{2\gamma}(u_1 + \cdots + u_m)^{2\alpha}.$$

Now deduce that, if (12.42) holds with $\alpha \geq 1$ and $\gamma \geq 1$, then

$$\mathsf{E}\{M_m{}^\gamma\} \leq (\log_2 4m)^\gamma\,(u_1 + \cdots + u_m)^\alpha.$$

(If $\gamma = 2$ and $\alpha = 1$, this gives Menshov's inequality again.)

6. Use Theorem 12.2 to show that, if $\xi_1, \xi_2, \ldots$ satisfies

$$\mathsf{E}\left\{\left|\sum_{i<l\leq j} \xi_l\right|^\gamma\right\} \leq \left(\sum_{i<l\leq j} u_l\right)^\alpha, \qquad 0 \leq i \leq j < \infty,$$

with $\gamma \geq 0$, $\alpha > 1$, and $\Sigma u_l < \infty$, then $\Sigma \xi_l$ converges with probability 1. Find a result connected in an analogous way with Theorem 12.6.

7. Theorem 12.3 is still true if we replace (12.50) by the assumption that

$$(12.84) \qquad \mathsf{P}\{|X_n(t) - X_n(t_1)| \geq \lambda, |X_n(t_2) - X_n(t)| \geq \lambda\} \leq \frac{1}{\lambda^{2\gamma}}|F(t_2) - F(t_1)|^{2\alpha}$$

holds for $t_1 \leq t \leq t_2$ and for all n, where now $\alpha > \frac{1}{2}$.

8. Theorem 12.4 is false if (12.58) is weakened to the analogue of (12.84); it is also false if the assumption $\alpha > 1$ is weakened to $\alpha \geq 1$. [Suppose μ_t is a unit mass at 0 or at 1 according as $t < \frac{1}{2}$ or $t \geq \frac{1}{2}$.] See, however, Problem 4 in Section 15.

9. Let

$$x_{nk}(t) = \begin{cases} 2n\left(t - \dfrac{k-1}{n}\right) & \text{for } \dfrac{k-1}{n} \leq t \leq \dfrac{k-1}{n} + \dfrac{1}{2n} \\ 2n\left(\dfrac{k}{n} - t\right) & \text{for } \dfrac{k-1}{n} + \dfrac{1}{2n} \leq t \leq \dfrac{k}{n} \\ 0 & \text{elsewhere,} \end{cases}$$

and let X_n take the value x_{nk} with probability $1/n$, $k = 1, \ldots, n$. Then

$$\mathsf{E}\{|X_n(t_2) - X_n(t_1)|\} \leq 4\,|t_2 - t_1|$$

for all n, t_1, and t_2. Since $\{X_n\}$ is not tight, we cannot take $\alpha = 1$ in Theorem 12.3. Similarly, we cannot take $\alpha = \frac{1}{2}$ in Problem 7.

13. EMPIRICAL DISTRIBUTION FUNCTIONS

Let $\xi_1, \xi_2, \ldots$ be random variables, on some $(\Omega, \mathscr{B}, \mathsf{P})$. We shall assume that

$$(13.1) \qquad 0 \leq \xi_n(\omega) \leq 1,$$

which can always be arranged by a transformation.

The *empirical* (or sample) *distribution function* $F_n(t, \omega)$ corresponding to the points $\xi_1(\omega), \ldots, \xi_n(\omega)$ is defined, for $0 \leq t \leq 1$, as $1/n$ times the number of $i \leq n$ for which $\xi_i(\omega) \leq t$.

If the ξ_n are independent and have a common distribution function $F(t)$ (by (13.1), only values of t in $[0, 1]$ matter), then, for large n, $F_n(t, \omega)$ should approximate $F(t)$—the difference

$$(13.2) \qquad F_n(t, \omega) - F(t)$$

should be small. According to the Glivenko–Cantelli theorem,

$$(13.3) \qquad \sup_{0 \leq t \leq 1} |F_n(t, \omega) - F(t)|$$

converges to 0 with probability 1.† If F is continuous, it is possible to find the limiting distribution of (13.3), properly normalized; this limit theorem, due to Kolmogorov, stands to the Glivenko–Cantelli theorem as the central limit theorem does to the law of large numbers.

We shall consider in this section only the case $F(t) \equiv t$, so that the ξ_n, assumed independent, have a common uniform distribution over $[0, 1]$.

† Loève (1960, p. 20).

Kolmogorov's result is that

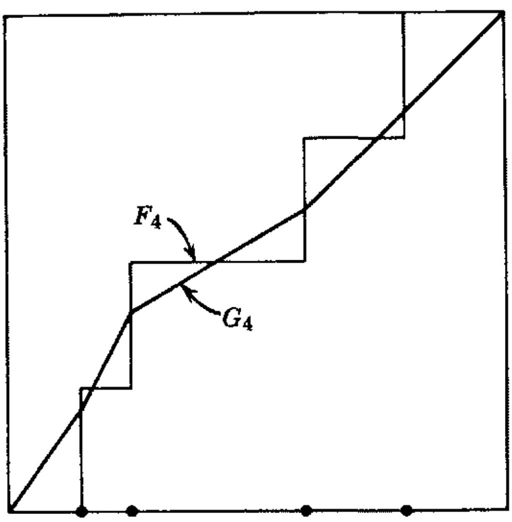
$$(13.4)\quad \mathsf{P}\{\omega : \sup_t |\sqrt{n}(F_n(t,\omega) - t)| \le \alpha\} \to 1 - 2\sum_{k=1}^{\infty}(-1)^{k+1}e^{-2k^2\alpha^2}, \quad \alpha \ge 0.$$

We shall derive (13.4) by using the theory of weak convergence in C.

Consider the function $Y_n(\omega)$ with value

$$(13.5)\qquad Y_n(t,\omega) = \sqrt{n}(F_n(t,\omega) - t)$$

at t. Although $Y_n(\omega)$ is a function on $[0, 1]$ produced at random, it is not an element of C, being obviously discontinuous. If we could establish the convergence in distribution of Y_n as a random element of an appropriate space of discontinuous functions with an appropriate metric, we would be in a position to derive (13.4) by the techniques of Sections 10 and 11, because



$$\sup_t |\sqrt{n}(F_n(t,\omega) - t)| = h(Y_n(\omega))$$

with $h(x) = \sup_t |x(t)|$. Although in the next chapter we shall analyze Y_n as a random element of a metric space of discontinuous functions, here we shall circumvent the discontinuity problems by adopting a different definition of empirical distribution function. (We shall still be able to derive (13.4) itself—not some perturbation of it.)

Let $G_n(t,\omega)$ be, as a function of t ranging over $[0, 1]$, the distribution function corresponding to a uniform distribution of mass $(n+1)^{-1}$ over each of the $(n+1)$ intervals $[\xi_{(i-1)}(\omega), \xi_{(i)}(\omega)]$, where $\xi_{(0)} = 0$, $\xi_{(n+1)} = 1$, and $\xi_{(1)}, \ldots, \xi_{(n)}$ are the values $\xi_1, \ldots, \xi_n$ ranged in increasing order. (Note that $F_n(t,\omega)$ corresponds to a distribution of mass n^{-1} to each of the n *points* $\xi_1(\omega), \ldots, \xi_n(\omega)$.) The functions $F_n(t,\omega)$ and $G_n(t,\omega)$ are close:

$$(13.6)\qquad |F_n(t,\omega) - G_n(t,\omega)| \le \frac{1}{n}, \qquad 0 \le t \le 1.$$

Now let $Z_n(\omega)$ be the element of C with value

$$(13.7)\qquad Z_n(t,\omega) = \sqrt{n}(G_n(t,\omega) - t)$$

at t. Since each $Z_n(t)$ is a random variable, Z_n is a random element of C ($Z_n^{-1}\mathscr{C} \subset \mathscr{B}$). By (13.6), we have

$$(13.8)\qquad \sup_t |Y_n(t,\omega) - Z_n(t,\omega)| \le \frac{1}{\sqrt{n}}.$$

THEOREM 13.1 *If the* ξ_n *are independent and uniformly distributed on* $[0, 1]$, *and if* Z_n *is defined by* (13.7), *then*

$$Z_n \xrightarrow{\mathscr{D}} W^\circ, \tag{13.9}$$

where W° *is the Brownian bridge.*

Before proving the theorem, let us see how it implies (13.4). If $h(x) = \sup_t |x(t)|$, then h is continuous on C and hence (13.9) implies $h(Z_n) \xrightarrow{\mathscr{D}} h(W^\circ)$. By (11.39),

$$P\{\sup_t |Z_n(t)| \le \alpha\} \to 1 - 2\sum_{k=1}^{\infty}(-1)^{k+1}e^{-2k^2\alpha^2}, \qquad \alpha \ge 0,$$

so that (13.4) follows from (13.8) and Theorem 4.1.

Using (11.40), we can in the same way derive the relation

$$P\{\sqrt{n}\,\sup_t (F_n(t, \omega) - t) \le \alpha\} \to 1 - e^{-2\alpha^2}, \qquad \alpha \ge 0. \tag{13.10}$$

If $h(x)$ is the Lebesgue measure of the set of t for which $x(t) > 0$, then, by (11.42),

$$P\{h(Z_n) \le \alpha\} \to \alpha, \qquad 0 \le \alpha \le 1; \tag{13.11}$$

$h(Z_n)$ is approximately n^{-1} times the number of i, $1 \le i \le n$, for which $\xi_{(i)} > i/n$.

Proof. To prove (13.9), we first show that the finite-dimensional distributions of Z_n converge to those of W°. Let $U_n(t, \omega) = nF_n(t, \omega)$ be the number of points among $\xi_1(\omega), \ldots, \xi_n(\omega)$ that satisfy $\xi_i(\omega) \le t$. If

$$0 = t_0 < t_1 < \cdots < t_k = 1,$$

then the random variables

$$U_n(t_i) - U_n(t_{i-1}), \qquad i = 1, 2, \ldots, k, \tag{13.12}$$

are multinomially distributed with parameters n and $p_i = t_i - t_{i-1}$, $i = 1, 2, \ldots, k$. It follows by the central limit theorem for multinomial trials that the random vector with components

$$Y_n(t_i) - Y_n(t_{i-1}) = \frac{1}{\sqrt{n}}\big((U_n(t_i) - U_n(t_{i-1})) - np_i\big), \qquad i = 1, 2, \ldots, k, \tag{13.13}$$

converges in distribution to the random vector with components $W^\circ_{t_i} - W^\circ_{t_{i-1}}$ (by (9.14) and (9.15), these normally distributed random variables have the right variances $p_i(1 - p_i)$ and covariances $-p_ip_j$). Because of (13.8), the same is true if we replace (13.13) by

$$Z_n(t_i) - Z_n(t_{i-1}), \qquad i = 1, 2, \ldots, k.$$

The finite-dimensional distributions of the Z_n thus converge properly. If we prove $\{Z_n\}$ tight, (13.9) will follow. By Theorem 8.3, it suffices to show that, for each positive ε and η, there exists a δ, $0 < \delta < 1$, and an n_0 such that, if $n \geq n_0$, then

$$\mathsf{P}\left\{\sup_{t\leq s\leq t+\delta} |Z_n(s) - Z_n(t)| \geq \varepsilon\right\} \leq \delta\eta \tag{13.14}$$

holds for all t.

Because of (13.8), we may replace (13.14) by

$$\mathsf{P}\left\{\sup_{t\leq s\leq t+\delta} |Y_n(s) - Y_n(t)| \geq \varepsilon\right\} \leq \delta\eta. \tag{13.15}$$

The event whose probability appears in (13.15) is easily shown to lie in the σ-field generated by $\xi_1, \ldots, \xi_n$; (13.15) is simply an inequality involving $\xi_1, \ldots, \xi_n$—we have not embarked on an analysis of Y_n as a random element of a space of discontinuous functions. For notational convenience, we shall take $t = 0$ in (13.15):

$$\mathsf{P}\left\{\sup_{s\leq\delta} |Y_n(s)| \geq \varepsilon\right\} \leq \delta\eta. \tag{13.16}$$

Since the distributions of the increments of $Y_n(t)$ are stationary, this is really no restriction anyway.

We shall bound the probability in (13.16) by using Theorem 12.1. Let us show that

(13.17)

$$\mathsf{E}\{|Y_n(s + p_1) - Y_n(s)|^2 \cdot |Y_n(s + p_1 + p_2) - Y_n(s + p_1)|^2\} \leq 6p_1p_2.$$

For $1 \leq i \leq n$, let α_i be $1 - p_1$ or $-p_1$ according as ξ_i lies in $(s, s + p_1]$ or not, and let β_i be $1 - p_2$ or $-p_2$ according as ξ_i lies in $(s + p_1, s + p_1 + p_2]$ or not. Then (13.17) is equivalent to

$$\mathsf{E}\left\{\left(\sum_{i=1}^n \alpha_i\right)^2 \left(\sum_{i=1}^n \beta_i\right)^2\right\} \leq 6n^2p_1p_2. \tag{13.18}$$

Since the ξ_i are independent, so are the random vectors (α_i, β_i). Since ξ_i is uniformly distributed, (α_i, β_i) takes on the values $(1 - p_1, -p_2)$, $(-p_1, 1 - p_2)$, and $(-p_1, -p_2)$ with respective probabilities p_1, p_2, and $p_3 = 1 - p_1 - p_2$. Now $\mathsf{E}\{\alpha_i\} = \mathsf{E}\{\beta_i\} = 0$, and considerations of symmetry lead to

$$\mathsf{E}\left\{\left(\sum_{i=1}^n \alpha_i\right)^2 \left(\sum_{i=1}^n \beta_i\right)^2\right\} = n\mathsf{E}\{\alpha_1^2\beta_1^2\} + n(n - 1)\mathsf{E}\{\alpha_1^2\}\mathsf{E}\{\beta_2^2\}$$
$$+ 2n(n - 1)\mathsf{E}\{\alpha_1\beta_1\}\mathsf{E}\{\alpha_2\beta_2\}.$$

Now (13.18) follows from

$$\mathsf{E}\{\alpha_1^2\beta_1^2\} = p_1(1-p_1)^2p_2^2 + p_2p_1^2(1-p_2)^2 + p_3p_1^2p_2^2 \le 3p_1p_2,$$

$$\mathsf{E}\{\alpha_1^2\}\mathsf{E}\{\beta_2^2\} = p_1(1-p_1)p_2(1-p_2) \le p_1p_2,$$

and

$$\mathsf{E}\{\alpha_1\beta_1\}\mathsf{E}\{\alpha_2\beta_2\} = p_1^2p_2^2 \le p_1p_2.$$

For fixed δ, consider the variables

$$Y_n\left(\frac{i}{m}\,\delta\right) - Y_n\left(\frac{i-1}{m}\,\delta\right), \qquad i = 1, \ldots, m. \tag{13.19}$$

If $\gamma = 2$ and $\alpha = 1$, and if $u_i = 6^{\frac{1}{2}}\delta/m$, then, by (13.17), the hypotheses of Theorem 12.1 are satisfied, with the variables (13.19) in the role of the ξ_i in that theorem. If

$$M'_m = \max_{1\le i\le m} \min\left\{\left|Y_n\left(\frac{i}{m}\,\delta\right)\right|, \left|Y_n(\delta) - Y_n\left(\frac{i}{m}\,\delta\right)\right|\right\},$$

it follows that

$$\mathsf{P}\{M'_m \ge \varepsilon\} \le \frac{6K}{\varepsilon^4}\,\delta^2, \tag{13.20}$$

with $K = K_{2,1}$.

If

$$M_m = \max_{1\le i\le m}\left|Y_n\left(\frac{i}{m}\,\delta\right)\right|,$$

then, by (12.4),

$$M_m \le M'_m + |Y_n(\delta)|.$$

From this and (13.20) we have

$$\mathsf{P}\{M_m \ge \varepsilon\} \le \frac{2^4\cdot 6K}{\varepsilon^4}\,\delta^2 + \mathsf{P}\left\{|Y_n(\delta)| \ge \frac{\varepsilon}{2}\right\}. \tag{13.21}$$

Now, for each ω, $Y_n(s, \omega)$ is right-continuous in s. As $m \to \infty$, therefore, M_m converges to $\sup_{s\le\delta}|Y_n(s,\omega)|$ for each ω. Hence (13.21) implies

$$\mathsf{P}\left\{\sup_{s\le\delta}|Y_n(s)| > \varepsilon\right\} \le \frac{2^4\cdot 6K}{\varepsilon^4}\,\delta^2 + \mathsf{P}\left\{|Y_n(\delta)| \ge \frac{\varepsilon}{2}\right\}. \tag{13.22}$$

Because of the asymptotic normality of (13.13), $Y_n(\delta) \xrightarrow{\mathscr{D}} \sqrt{\delta(1-\delta)}\,N$ as $n \to \infty$ (δ fixed), and hence

$$\mathsf{P}\left\{|Y_n(\delta)| \ge \frac{\varepsilon}{2}\right\} \to \mathsf{P}\left\{N \ge \frac{\varepsilon/2}{\sqrt{\delta(1-\delta)}}\right\} \le \frac{2^4\delta^2}{\varepsilon^4}\,\mathsf{E}\{N^4\} = \frac{3\cdot 2^4}{\varepsilon^4}\,\delta^2.$$

For n exceeding some n_δ, therefore,

$$\mathsf{P}\left\{|Y_n(\delta)| \geq \frac{\varepsilon}{2}\right\} < \frac{6 \cdot 2^4}{\varepsilon^4}\,\delta^2;$$

hence, by (13.22),

$$\mathsf{P}\left\{\sup_{s \leq \delta} |Y_n(s)| > \varepsilon\right\} \leq \frac{6 \cdot 2^4(K+1)}{\varepsilon^4}\,\delta^2. \tag{13.23}$$

Given ε and η, choose δ so that $6 \cdot 2^4(K+1)\delta^2/\varepsilon^4 < \delta\eta$. For $n > n_\delta$, (13.16) follows by (13.23).

This completes the proof of Theorem 13.1. The treatment is evasive in that we really analyze Y_n, replacing Y_n by Z_n only in order to stay in C. In Section 16 we treat Y_n in the space natural to it and prove $Y_n \xrightarrow{\mathscr{D}} W^\circ$, and generalizations, in that space. Although the proofs in Section 16 depend on a considerable body of theory, they are much more transparent than the one just given.

Remarks. Doob (1949) conceived the idea of proving (13.4) by passing from the random function (13.5) to the Brownian bridge; Donsker (1952) justified the passage. The proof here derives from Chentsov (1956). Kac (1949) originally proved (13.11). See Darling (1957) for an account of limit theorems connected with (13.5) and for a large bibliography. See also the more recent papers: Bickel (1968a and 1968b), Birnbaum and Pyke (1958), Pyke (1965 and 1968), and Pyke and Shorack (1968).

CHAPTER 3

The Space D

14. THE GEOMETRY OF D

The space C is unsuitable for the description of processes that, like the Poisson process and unlike Brownian motion, must contain jumps. In this chapter we study weak convergence in a space that includes certain discontinuous functions.

The Space D

Let $D = D[0, 1]$ be the space of functions x on $[0, 1]$ that are right-continuous and have left-hand limits:

(i) For $0 \le t < 1$, $x(t+) = \lim_{s \downarrow t} x(s)$ exists and $x(t+) = x(t)$.
(ii) For $0 < t \le 1$, $x(t-) = \lim_{s \uparrow t} x(s)$ exists.

A function x is said to have a *discontinuity of the first kind* at t if $x(t-)$ and $x(t+)$ exist but differ and $x(t)$ lies between them. Any discontinuities of an element of D are of the first kind; the requirement $x(t) = x(t+)$ is a convenient normalization. Of course, C is a subset of D.

For $x \in D$ and $T_0 \subset [0, 1]$, put

$$w_x(T_0) = \sup \{|x(s) - x(t)| : s, t \in T_0\}. \tag{14.1}$$

The modulus of continuity of x, defined by (8.1), may be expressed as

$$w_x(\delta) = \sup_{0 \le t \le 1-\delta} w_x[t, t + \delta]. \tag{14.2}$$

A continuous function on $[0, 1]$ is uniformly continuous. The following lemma gives the corresponding uniformity idea for elements of D.

LEMMA 1 *For each x in D and each positive ε, there exist points t_0, $t_1, \ldots, t_r$ such that*

$$0 = t_0 < t_1 < \cdots < t_r = 1 \quad . \tag{14.3}$$

and

$$w_x[t_{i-1}, t_i) < \varepsilon, \qquad i = 1, 2, \ldots, r. \tag{14.4}$$

Proof. Let τ be the supremum of those t in $[0, 1]$ for which $[0, t)$ can be decomposed into finitely many subintervals $[t_{i-1}, t_i)$ satisfying (14.4). Since $x(0) = x(0+)$, we have $\tau > 0$; since $x(\tau-)$ exists, $[0, \tau)$ can itself be so decomposed; $\tau < 1$ is impossible because $x(\tau) = x(\tau+)$ in this case.

From this lemma it follows that there can be at most finitely many points t at which the jump (saltus) $|x(t) - x(t-)|$ exceeds a given positive number; in particular, x has at most countably many discontinuities. It follows also that x is bounded:

$$\sup_t |x(t)| < \infty. \tag{14.5}$$

Finally, it follows that x can be uniformly approximated by simple functions constant over intervals, so that x is Borel measurable.

We shall need a modulus that plays in D the role the modulus of continuity plays in C. For $0 < \delta < 1$, put

$$w'_x(\delta) = \inf_{\{t_i\}} \max_{0 < i \leq r} w_x[t_{i-1}, t_i), \tag{14.6}$$

where the infimum extends over the finite sets $\{t_i\}$ of points satisfying

$$\begin{cases} 0 = t_0 < t_1 < \cdots < t_r = 1, \\ t_i - t_{i-1} > \delta, \qquad i = 1, 2, \ldots, r. \end{cases} \tag{14.7}$$

Lemma 1 is equivalent to the assertion that

$$\lim_{\delta \to 0} w'_x(\delta) = 0 \tag{14.8}$$

holds for every x in D.

Even if x does not lie in D, the definition of $w'_x(\delta)$ makes sense. Just as $\lim_{\delta \to 0} w_x(\delta) = 0$ is necessary and sufficient for an arbitrary function x on $[0, 1]$ to lie in C, (14.8) is necessary and sufficient for x to lie in D.

Since $[0, 1)$ can, for each $\delta < \frac{1}{2}$, be split into subintervals $[t_{i-1}, t_i)$ with $\delta < t_i - t_{i-1} \leq 2\delta$, we have

$$w'_x(\delta) \leq w_x(2\delta), \qquad \text{if} \quad \delta < \tfrac{1}{2}. \tag{14.9}$$

There can be no general inequality in the opposite direction because of (14.8) and the fact that $w_x(\delta)$ does not go to 0 with δ if x has discontinuities.

Suppose, however, that $x \in C$. Given ε, choose points $\{t_i\}$ satisfying (14.7) and

$$\max_{0<i\leq r} w_x\,[t_{i-1}, t_i) < w_x'(\delta) + \varepsilon. \tag{14.10}$$

If $|s - t| < \delta$, then s and t lie in the same subinterval $[t_{i-1}, t_i)$ or in abutting ones, and it follows by (14.10) and the assumed continuity of x that $w_x(\delta) \leq 2w_x'(\delta) + 2\varepsilon$. Since ε was arbitrary,

$$w_x(\delta) \leq 2w_x'(\delta) \qquad \text{if} \quad x \in C. \tag{14.11}$$

By (14.9) and (14.11), the moduli $w_x(\delta)$ and $w_x'(\delta)$ are essentially the same for continuous functions x.

The Skorohod Topology

Two functions x and y are near one another in the uniform topology used for C if the graph of $x(t)$ can be carried onto the graph of $y(t)$ by a uniformly small perturbation of the ordinates, with the abscissas kept fixed. In D, we shall allow also a uniformly small deformation of the time scale. Physically, this amounts to the admission that we cannot measure time with perfect accuracy any more than we can position. The following topology, devised by Skorohod, embodies this idea.

Let Λ denote the class of strictly increasing, continuous mappings of $[0, 1]$ onto itself. If $\lambda \in \Lambda$, then $\lambda(0) = 0$ and $\lambda(1) = 1$. For x and y in D, define $d(x, y)$ to be the infimum of those positive ε for which there exists in Λ a λ such that

$$\sup_t |\lambda t - t| \leq \varepsilon \tag{14.12}$$

and

$$\sup_t |x(t) - y(\lambda t)| \leq \varepsilon. \tag{14.13}$$

By (14.5), $d(x, y)$ is finite (take $\lambda t \equiv t$). Clearly, $d(x, y) \geq 0$; and $d(x, y) = 0$ implies that for each t either $x(t) = y(t)$ or $x(t) = y(t-)$, which in turn implies $x = y$. If λ lies in Λ, so does λ^{-1}; $d(x, y) = d(y, x)$ follows from

$$\sup_t |\lambda^{-1}t - t| = \sup_t |\lambda t - t|$$

and

$$\sup_t |x(\lambda^{-1}t) - y(t)| = \sup_t |x(t) - y(\lambda t)|.$$

If λ_1 and λ_2 lie in D, so does their composition $\lambda_2\lambda_1$; the triangle inequality follows from

$$\sup_t |\lambda_2\lambda_1 t - t| \leq \sup_t |\lambda_1 t - t| + \sup_t |\lambda_2 t - t|$$

and

$$\sup_t |x(t) - z(\lambda_2\lambda_1 t)| \leq \sup_t |x(t) - y(\lambda_1 t)| + \sup_t |y(t) - z(\lambda_2 t)|.$$

Thus d is a metric.

This metric defines the Skorohod topology. The uniform distance between x and y may be defined as the infimum of those positive ε for which $\sup_t |x(t) - y(t)| \leq \varepsilon$. The λ in (14.12) and (14.13) represents the uniformly small deformation of the time scale mentioned above.

Elements x_n of D converge to a limit x in the Skorohod topology if and only if there exist functions λ_n in Λ such that

$$\lim_n x_n(\lambda_n t) = x(t)$$

uniformly in t and

$$\lim_n \lambda_n t = t$$

uniformly in t. If x_n goes uniformly to x, then there is convergence in the Skorohod topology (take $\lambda_n t \equiv t$). On the other hand, there is convergence

$$x_n = I_{[0,\frac{1}{2}+1/n)} \to x = I_{[0,\frac{1}{2})}$$

in the Skorohod topology, whereas $x_n(t) \to x(t)$ fails in this case at $t = \frac{1}{2}$. Since

$$|x_n(t) - x(t)| \leq |x_n(t) - x(\lambda_n^{-1}t)| + |x(\lambda_n^{-1}t) - x(t)|,$$

Skorohod convergence does imply $x_n(t) \to x(t)$ for continuity points t of x, and if x is (uniformly) continuous on all of $[0, 1]$, then Skorohod convergence implies uniform convergence. In particular, *the Skorohod topology relativized to C coincides with the uniform topology there.*

The space D is separable; one countable, dense set consists of those x that have a rational value at $t = 1$ and have, for some integer k, a constant, rational value over each subinterval $[(i - 1)/k, i/k)$, $0 < i \leq k$ (use Lemma 1 and the definition of d). But D is not complete under the metric d: If

$$x_n = I_{[\frac{1}{2},\frac{1}{2}+1/n)}, \tag{14.14}$$

then

$$d(x_n, x_m) = \left|\frac{1}{n} - \frac{1}{m}\right|, \tag{14.15}$$

and hence $\{x_n\}$ is fundamental in the metric d, even though it is not convergent. We shall introduce in D another metric d_0—a metric which is equivalent to d (gives the Skorohod topology) but under which D is complete. Completeness facilitates characterizing compact sets.

The idea in defining d_0 is to require that the time-deformation λ that intervenes in the definition of d be near the identity function in a sense which at first appears more stringent than (14.12); namely, we require that the slope $(\lambda t - \lambda s)/(t - s)$ of each chord be nearly 1 or, what is the same thing and analytically more convenient, that its logarithm be nearly 0.

If λ is a nondecreasing function on $[0, 1]$ with $\lambda 0 = 0$ and $\lambda 1 = 1$, put

$$\|\lambda\| = \sup_{s \neq t} \left| \log \frac{\lambda t - \lambda s}{t - s} \right|. \tag{14.16}$$

If $\|\lambda\|$ is finite, then the slopes of the chords of λ are bounded away from 0 and infinity; therefore it is both continuous and strictly increasing and hence is a member of Λ. Although $\|\lambda\|$ may be infinite even if λ does lie in Λ, such elements of Λ do not enter into the following definition.

Let $d_0(x, y)$ be the infimum of those positive ε for which Λ contains some λ with

$$\|\lambda\| \leq \varepsilon \tag{14.17}$$

and

$$\sup_t |x(t) - y(\lambda t)| \leq \varepsilon. \tag{14.18}$$

By (14.5), $d_0(x, y)$ is finite (take $\lambda t \equiv t$). That d_0 is a metric follows from the relations

$$\|\lambda^{-1}\| = \|\lambda\| \tag{14.19}$$

and

$$\|\lambda_2\lambda_1\| \leq \|\lambda_1\| + \|\lambda_2\|. \tag{14.20}$$

We shall show presently that d and d_0 are equivalent metrics. We shall also show that D is complete under d_0; for the present, note that, if x_n is defined by (14.14), then

$$d_0(x_n, x_m) = \min\left\{1, \left|\log\frac{m}{n}\right|\right\}, \qquad m, n > 3, \tag{14.21}$$

so that at least the (nonconvergent) sequence $\{x_n\}$ is not d_0-fundamental.

If $d_0(x, y) < \varepsilon$, then (14.17) and (14.18) hold for some $\lambda \in \Lambda$. If $\varepsilon < \frac{1}{4}$, then, since $\lambda 0 = 0$,

$$\log(1 - 2\varepsilon) < -\varepsilon \leq \log\frac{\lambda t}{t} \leq \varepsilon < \log(1 + 2\varepsilon),\dagger$$

so that $|\lambda t - t| \leq 2\varepsilon$. Therefore

$$d(x, y) \leq 2d_0(x, y) \qquad \text{if} \quad d_0(x, y) < \tfrac{1}{4}. \tag{14.22}$$

A comparison of (14.15) and (14.21) shows that there can be no general inequality in the direction opposite to (14.22). The following lemma shows, however, that $d_0(x, y)$ is small if $d(x, y)$ and $w'_x(\delta)$ (or $w'_y(\delta)$) are both small.

LEMMA 2 *If $d(x, y) < \delta^2$, where $0 < \delta < \frac{1}{4}$, then $d_0(x, y) \leq 4\delta + w'_x(\delta)$.*

Proof. Choose points t_i satisfying (14.7) and

$$w_x[t_{i-1}, t_i) < w'_x(\delta) + \delta, \qquad i = 1, 2, \ldots, r. \tag{14.23}$$

† If $|s| \leq \frac{1}{2}$, then $s - s^2 \leq \log(1 + s)$, whereas $\log(1 + s) \leq s$ for all $s > -1$.

Now choose from Λ a μ such that

$$\sup_t |x(t) - y(\mu t)| = \sup_t |x(\mu^{-1}t) - y(t)| < \delta^2 \tag{14.24}$$

and

$$\sup_t |\mu t - t| < \delta^2. \tag{14.25}$$

We want to define in Λ a λ that will be near μ but will not, as μ may, have chords with slopes far removed from 1. Take λ to agree with μ at the points t_i and to be linear in between. Since the composition $\mu^{-1}\lambda$ fixes the t_i and is increasing, t and $\mu^{-1}\lambda t$ always lie in the same subinterval $[t_{i-1}, t_i)$. By (14.23) and (14.24), therefore,

$$\begin{aligned} |x(t) - y(\lambda t)| &\le |x(t) - x(\mu^{-1}\lambda t)| + |x(\mu^{-1}\lambda t) - y(\lambda t)| \\ &< w'_x(\delta) + \delta + \delta^2 < 4\delta + w'_x(\delta). \end{aligned}$$

It is now enough to prove $\|\lambda\| \le 4\delta$. Since λ agrees with μ at the t_i, it follows by (14.25) and the inequality $t_i - t_{i-1} > \delta$ that

$$|(\lambda t_i - \lambda t_{i-1}) - (t_i - t_{i-1})| < 2\delta^2 < 2\delta(t_i - t_{i-1}).$$

From the polygonal character of λ, it now follows that

$$|(\lambda t - \lambda s) - (t - s)| \le 2\delta\, |t - s|$$

holds for all s and t. Therefore

$$\log(1 - 2\delta) \le \log \frac{\lambda t - \lambda s}{t - s} \le \log(1 + 2\delta);$$

since $\delta < \frac{1}{4}$, $\|\lambda\| \le 4\delta$ follows.

THEOREM 14.1 *The metrics d and d_0 are equivalent.*

Proof. Let us denote an open d-sphere by $S_d(x, \varepsilon)$ and an open d_0-sphere by $S_{d_0}(x, \varepsilon)$. It follows by (14.22) that inside an arbitrary $S_d(x, \varepsilon)$ we can find an $S_{d_0}(x, \delta)$. (The choice of the new radius does not depend on the center x.)

Lemma 2 implies that, if

$$\delta < \tfrac{1}{4}, \qquad 4\delta + w'_x(\delta) < \varepsilon, \tag{14.26}$$

then $S_d(x, \delta^2) \subset S_{d_0}(x, \varepsilon)$. Given x and ε, we can, by (14.8), find a δ satisfying (14.26). Inside an arbitrary d_0-sphere we can thus find a d-sphere with the same center. (This time the choice of the new radius does depend on the center x, as must be true if D is to be complete under d_0—if d and d_0 are to give different classes of fundamental sequences.) Therefore d and d_0 are equivalent metrics.

Completeness of D

THEOREM 14.2 *The space D is complete in the metric d_0.*

Proof. It is enough to show that each d_0-fundamental sequence contains a d_0-convergent subsequence. If $\{x_k\}$ is a d_0-fundamental sequence, it contains a subsequence $\{y_n\} = \{x_{k_n}\}$ such that

$$d_0(y_n, y_{n+1}) < \frac{1}{2^n}. \tag{14.27}$$

We shall prove that $\{y_n\}$ is d_0-convergent to some limit.

By (14.27), Λ contains a μ_n such that

$$\sup_t |y_n(t) - y_{n+1}(\mu_n t)| < \frac{1}{2^n} \tag{14.28}$$

and

$$\|\mu_n\| < \frac{1}{2^n}. \tag{14.29}$$

This implies, for $m \geq 1$,

$$\begin{aligned}\sup_t |\mu_{n+m+1}\mu_{n+m} \cdots \mu_{n+1}\mu_n t - \mu_{n+m} \cdots \mu_{n+1}\mu_n t| \\ = \sup_s |\mu_{n+m+1}s - s| \leq \frac{1}{2^{n+m}}.\end{aligned}$$

(Here $\mu_{n+m} \cdots \mu_{n+1}\mu_n$ denotes iterated composition.) For fixed n the functions

$$\mu_{n+m} \cdots \mu_{n+1}\mu_n t \tag{14.30}$$

are thus uniformly fundamental as $m \to \infty$. Therefore (14.30) converges uniformly to a limit

$$\lambda_n t = \lim_m \mu_{n+m} \cdots \mu_{n+1}\mu_n t. \tag{14.31}$$

The function λ_n must be continuous and nondecreasing and must satisfy $\lambda_n 0 = 0$, $\lambda_n 1 = 1$. If we prove that $\|\lambda_n\|$ is finite, it will follow that λ_n is strictly increasing and hence is a member of Λ. By (14.20),

$$\begin{aligned}&\left| \log \frac{\mu_{n+m} \cdots \mu_{n+1}\mu_n t - \mu_{n+m} \cdots \mu_{n+1}\mu_n s}{t - s} \right| \\ &\leq \|\mu_{n+m} \cdots \mu_{n+1}\mu_n\| \leq \|\mu_n\| + \|\mu_{n+1}\| + \cdots + \|\mu_{n+m}\| \leq \frac{1}{2^{n-1}}.\end{aligned}$$

Letting $m \to \infty$ in the first member of this inequality, we find that $\|\lambda_n\| \leq 1/2^{n-1}$; in particular, $\|\lambda_n\|$ is finite and $\lambda_n \in \Lambda$.

By (14.31), $\lambda_n = \lambda_{n+1}\mu_n$. Therefore, by (14.28),

$$\sup_t |y_n(\lambda_n^{-1}t) - y_{n+1}(\lambda_{n+1}^{-1}t)| = \sup_s |y_n(s) - y_{n+1}(\mu_n s)| < \frac{1}{2^n}.$$

In consequence, the functions $y_n(\lambda_n^{-1}t)$, which are elements of D, are uniformly fundamental and hence converge uniformly to a limit function $x(t)$. It is easy to show that x must also be an element of D. Since $\sup_t |y_n(\lambda_n^{-1}t) - x(t)| \to 0$ and $\|\lambda_n\| \to 0$, we have $d_0(y_n, x) \to 0$, which completes the proof.

Compactness in D

We turn now to the problem of characterizing the compact subsets of D. With the modulus $w'_x(\delta)$ defined by (14.6), we have an analogue of the Arzelà–Ascoli theorem (p. 221).

THEOREM 14.3 *A set A has compact closure in the Skorohod topology if and only if*

$$\sup_{x\in A} \sup_t |x(t)| < \infty \tag{14.32}$$

and

$$\lim_{\delta\to 0} \sup_{x\in A} w'_x(\delta) = 0. \tag{14.33}$$

Note that, because of (14.9), (14.33) is weaker than

$$\lim_{\delta\to 0} \sup_{x\in A} w_x(\delta) = 0. \tag{14.34}$$

Proof. To prove the sufficiency of these conditions, it is enough, since D is d_0-complete, to show they imply that A is totally bounded with respect to d_0. Let us first show that (14.32) and (14.33) imply that A is totally bounded with respect to the metric d.

Given a positive ε, choose an integer k such that $1/k < \varepsilon$ and $w'_x(1/k) < \varepsilon$ for all x in A. Take H to be a finite ε-net in the linear interval $[-\alpha, \alpha]$, where

$$\alpha = \sup_{x\in A} \sup_t |x(t)|.$$

Let B be the finite set of y in D that assume on each of the intervals $[(u-1)/k, u/k)$ a constant value from H and satisfy $y(1) \in H$. We shall show that B is a 2ε-net with respect to d.

Given x in A, use the inequality $w'_x(1/k) < \varepsilon$ to choose points t_i, $0 = t_0 < t_1 < \cdots < t_r = 1$, such that

$$t_i - t_{i-1} > \frac{1}{k} \tag{14.35}$$

and

$$w'_x[t_{i-1}, t_i) < \varepsilon, \qquad 0 < i \le r. \tag{14.36}$$

Choose integers u_i such that $u_i/k \le t_i < (u_i + 1)/k$, $i = 0, 1, \ldots, r$. Since he u_i must be distinct by (14.35), there is in Λ a λ that carries u_i/k to t_i .nd is linear in between these points. Choose in B a point y such that

$$\left| y\left(\frac{u}{k}\right) - x\left(\lambda \frac{u}{k}\right) \right| < \varepsilon, \qquad 0 \le u \le k. \tag{14.37}$$

Since an interval

$$\left[\lambda \frac{u}{k}, \lambda \frac{u+1}{k}\right)$$

nust be inside one of the intervals

$$\left[\lambda \frac{u_i}{k}, \lambda \frac{u_{i+1}}{k}\right) = [t_i, t_{i+1}),$$

:he function $x(\lambda t)$ cannot, by (14.36), vary by more than ε as t varies over an nterval $[u/k, (u+1)/k)$. Since y is constant over intervals of this form, (14.37) implies $|y(t) - x(\lambda t)| < 2\varepsilon$ for all t. By construction,

$$\left| \lambda \frac{u_i}{k} - \frac{u_i}{k} \right| = \left| t_i - \frac{u_i}{k} \right| < \frac{1}{k} < \varepsilon;$$

by linearity, $\sup_t |\lambda t - t| < \varepsilon$. Thus $d(x, y) < 2\varepsilon$. Therefore B is a 2ε-net in the sense of d.

Given a positive η, choose δ so that $0 < \delta < \frac{1}{4}$ and so that $4\delta + w'_x(\delta) < \eta$ holds for all x in A. Then choose ε so that $0 < 2\varepsilon < \delta^2$. If B is the set just constructed—a finite 2ε-net for A in the sense of the metric d—then, by Lemma 2, B is an η-net for A in the sense of the metric d_0. Thus A is d_0-totally bounded and hence, since D is d_0-complete, the closure A^- of A is compact.

This proves the sufficiency of (14.32) and (14.33). If A^- is compact, then it is bounded, and (14.32) follows immediately (note that $\sup_t |x(t)|$ is the distance, in the sense either of d or of d_0, from x to the function that is identically 0). The present theorem differs from the Arzelà–Ascoli theorem in that for no single t do

$$\sup_{x \in A} |x(t)| < \infty \tag{14.38}$$

and (14.33) together imply (14.32). It is not hard, however, to prove that (14.32) follows if (14.33) holds and if (14.38) holds for each individual value of t (or even just for values of t including 1 and dense in $[0, 1]$).

It remains to prove the necessity of (14.33). By (14.8), $w'_x(\delta)$ goes to 0 with δ for each x. If we prove that $w'_x(\delta)$ is upper semicontinuous in x for each δ, then (see p. 218) the convergence will be uniform on compact sets and (14.33) will follow.

Let x, δ, and ε be given. To prove upper semicontinuity, we must find an η such that $d(x, y) < \eta$ implies

$$w'_y(\delta) < w'_x(\delta) + \varepsilon. \tag{14.39}$$

First choose points t_i satisfying (14.7) and

$$w_x[t_{i-1}, t_i) < w'_x(\delta) + \frac{\varepsilon}{2}, \qquad i = 1, 2, \ldots, r. \tag{14.40}$$

Now choose η so small that

$$t_i - t_{i-1} > \delta + 2\eta, \qquad i = 1, 2, \ldots, r,$$

and

$$\eta < \tfrac{1}{4}\varepsilon. \tag{14.41}$$

If $d(x, y) < \eta$, then, for some λ in Λ, we have

$$\sup_t |\lambda t - t| = \sup_t |\lambda^{-1}t - t| < \eta$$

and

$$\sup_t |y(t) - x(\lambda t)| < \eta. \tag{14.42}$$

Let $s_i = \lambda^{-1}t_i$. Then

$$s_i - s_{i-1} > t_i - t_{i-1} - 2\eta > \delta. \tag{14.43}$$

Moreover, if s and t both lie in $[s_{i-1}, s_i)$, then λs and λt both lie in $[t_{i-1}, t_i)$ and therefore, by (14.40), (14.41), and (14.42), $|y(s) - y(t)| < w'_x(\delta) + \varepsilon$. Thus

$$w_y[s_{i-1}, s_i) < w'_x(\delta) + \varepsilon,$$

which, in view of (14.43), implies (14.39). This completes the proof of Theorem 14.3.

A Second Characterization of Compactness

Although the modulus $w'(\delta)$ is natural in that it leads to a complete characterization of compact sets, a different one is sometimes more convenient to work with, namely

$$w''_x(\delta) = \sup \min\{|x(t) - x(t_1)|, |x(t_2) - x(t)|\}, \tag{14.44}$$

where the supremum extends over t_1, t, and t_2 satisfying

$$t_1 \leq t \leq t_2, \qquad t_2 - t_1 \leq \delta. \tag{14.45}$$

Given δ and ε, decompose $[0, 1)$ into subintervals $[s_{i-1}, s_i)$ such that $s_i - s_{i-1} > \delta$ and $w_x[s_{i-1}, s_i) < w'_x(\delta) + \varepsilon$. If (14.45) holds, then either t_1 and t_2 lie in the same subinterval $[s_{i-1}, s_i)$, in which case $|x(t) - x(t_1)| < w'_x(\delta) + \varepsilon$ and $|x(t_2) - x(t)| < w'_x(\delta) + \varepsilon$, or else they lie in abutting intervals

$_{i-1}, s_i)$ and $[s_i, s_{i+1})$, in which case $|x(t) - x(t_1)| < w'_x(\delta) + \varepsilon$ for $\leq t < s_i$ and $|x(t_2) - x(t)| < w'_x(\delta) + \varepsilon$ for $s_i \leq t \leq t_2$. Therefore

$$w''_x(\delta) \leq w'_x(\delta). \tag{14.46}$$

There can be no inequality in the other direction. If, for example,

$$x_n(t) = \begin{cases} 1 & \text{for } 0 \leq t < n^{-1}, \\ 0 & \text{for } n^{-1} < t \leq 1, \end{cases} \tag{14.47}$$

r if

$$x_n(t) = \begin{cases} 0 & \text{for } 0 \leq t \leq 1 - n^{-1}, \\ 1 & \text{for } 1 - n^{-1} \leq t \leq 1, \end{cases} \tag{14.48}$$

hen $w''_{x_n}(\delta) = 0$, whereas

$$w'_{x_n}(\delta) = \begin{cases} 0 & \text{if } \delta < n^{-1}, \\ 1 & \text{if } \delta \geq n^{-1}. \end{cases}$$

ince $\{x_n\}$ has compact closure in neither case, there can be no compactness ondition involving (in addition to (14.32)) a restriction on $w''_x(\delta)$ alone. It is ossible, however, to formulate a compactness condition in terms of $w''_x(\delta)$ nd the behavior of x near 0 and 1.

THEOREM 14.4 *A set A has compact closure in the Skorohod topology f and only if*

$$\sup_{x \in A} \sup_t |x(t)| < \infty \tag{14.49}$$

and

$$\begin{cases} \lim\limits_{\delta \to 0} \sup\limits_{x \in A} w''_x(\delta) = 0, \\ \lim\limits_{\delta \to 0} \sup\limits_{x \in A} w_x[0, \delta) = 0, \\ \lim\limits_{\delta \to 0} \sup\limits_{x \in A} w_x[1 - \delta, 1) = 0. \end{cases} \tag{14.50}$$

Proof. It suffices, in view of Theorem 14.3, to show that (14.50) is equivalent to

$$\lim_{\delta \to 0} \sup_{x \in A} w'(\delta) = 0. \tag{14.51}$$

That (14.51) implies (14.50) follows by (14.46) and the definition of $w'(\delta)$. We prove the reverse implication.

Given a positive ε, choose a positive δ such that, for all x in A,

$$w''_x(\delta) < \varepsilon, \qquad w_x[0, \delta) < \varepsilon, \qquad w_x[1 - \delta, 1) < \varepsilon. \tag{14.52}$$

Assume that x lies in A; we shall show that

$$w'_x(\tfrac{1}{2}\delta) \leq 6\varepsilon, \tag{14.53}$$

which will suffice to prove the theorem.

Let us first show that

$$(14.54)\qquad t_1 \le s \le t \le t_2, \qquad t_2 - t_1 \le \delta$$

implies

$$(14.55)\qquad \min\{|x(s) - x(t_1)|, |x(t_2) - x(t)|\} < 2\varepsilon.$$

Indeed, if $|x(s) - x(t_1)| > \varepsilon$, then, by (14.52), $|x(t) - x(s)| < \varepsilon$ and $|x(t_2) - x(s)| < \varepsilon$, so that $|x(t_2) - x(t)| < 2\varepsilon$.

Suppose x has a jump exceeding 2ε at each of two points τ_1 and τ_2. If $0 < \tau_2 - \tau_1 < \delta$, then there exist points t_1, s, t, t_2 satisfying (14.54), $t_1 < \tau_1 = s$, and $t < \tau_2 = t_2$; if t_1 is close enough to τ_1, and if t is close enough to τ_2, then (14.55) is violated, which we have seen is impossible. Thus [0, 1] cannot contain two points, within δ of each other, at each of which x jumps by more than 2ε. And, by (14.52), neither $[0, \delta)$ nor $[1 - \delta, 1)$ can contain a point at which x jumps by more than 2ε.

Thus there exist points s_i, with $0 = s_0 < s_1 < \cdots < s_r = 1$, such that $s_i - s_{i-1} \ge \delta$ and such that any point at which x jumps by more than 2ε is one of the s_i. If $s_j - s_{j-1} > \delta$ for a pair of adjacent points, enlarge the system $\{s_i\}$ by including their midpoint. Continuing in this way leads to a new, enlarged system $s_0, \ldots, s_r$ (with a new r) that satisfies

$$\tfrac{1}{2}\delta < s_i - s_{i-1} \le \delta, \qquad i = 1, 2, \ldots, r.$$

Now (14.53) will follow if we show that

$$(14.56)\qquad w_x[s_{i-1}, s_i) \le 6\varepsilon$$

for each i. Suppose $s_{i-1} \le t_1 < t_2 < s_i$. Then $t_2 - t_1 < \delta$. Let σ_1 be the supremum of those σ in $[t_1, t_2]$ for which $\sup_{t_1 \le u \le \sigma} |x(u) - x(t_1)| \le 2\varepsilon$; let σ_2 be the infimum of those σ in $[t_1, t_2]$ for which $\sup_{\sigma \le u \le t_2} |x(t_2) - x(u)| \le 2\varepsilon$. If $\sigma_1 < \sigma_2$, there exist points s just to the right of σ_1 with $|x(s) - x(t_1)| > 2\varepsilon$ and there exist points t just to the left of σ_2 with $|x(t_2) - x(t)| > 2\varepsilon$; since we may ensure $s < t$, this contradicts the fact that (14.54) implies (14.55). Therefore $\sigma_2 \le \sigma_1$, and it follows that $|x(\sigma_1 -) - x(t_1)| \le 2\varepsilon$ and $|x(t_2) - x(\sigma_1)| \le 2\varepsilon$. Since $t_1 < \sigma_1 \le t_2$, σ_1 is interior to (s_{i-1}, s_i), so the jump at σ_1 is at most 2ε. Thus $|x(t_2) - x(t_1)| \le 6\varepsilon$. This establishes (14.56), which proves (14.53) and the theorem.†

Finite-Dimensional Sets

Finite-dimensional sets play in D the same role they do in C. For points $t_1, \ldots, t_k$ in [0, 1], define the natural projection $\pi_{t_1 \cdots t_k}$ from D to R^k as

† The points 0 and 1 play a special role in the theory of the space D because each λ in Λ fixes them. Judicious enlargement of Λ might simplify the theory.

usual:

$$\pi_{t_1 \cdots t_k}(x) = \big(x(t_1), \ldots, x(t_k)\big). \tag{14.57}$$

Now π_0 and π_1 are everywhere continuous. Suppose $0 < t < 1$. If points x_n converge to x in the Skorohod topology and x is continuous at t, then (p. 112) $x_n(t) \to x(t)$. Suppose, on the other hand, that x is discontinuous at t. If λ_n is the element of Λ that is linear on $[0, t]$ and on $[t, 1]$ and satisfies $\lambda_n t = t - 1/n$, and if $x_n(s) \equiv x(\lambda_n s)$, then x_n converges to x in the Skorohod topology but $x_n(t)$ does not converge to $x(t)$. Thus: *If* $0 < t < 1$, *then* π_t *is continuous at* x *if and only if* x *is continuous at* t.

We must prove that $\pi_{t_1 \cdots t_k}$ is measurable with respect to the σ-field $\mathscr{D}$ of Borel sets for the Skorohod topology. We need consider only a single time point t (since a mapping into R^k is measurable if each component mapping is), and we may assume $t < 1$ (since π_1 is continuous). If x_n converges to x in the Skorohod topology, then $x_n(s) \to x(s)$ for continuity points s of x and hence for points s outside a set of Lebesgue measure 0. Since the x_n are uniformly bounded,

$$\frac{1}{\varepsilon}\int_t^{t+\varepsilon} x_n(s)\, ds \to \frac{1}{\varepsilon}\int_t^{t+\varepsilon} x(s)\, ds \qquad (n \to \infty) \tag{14.58}$$

for each positive ε. Thus $h_\varepsilon(x) = \varepsilon^{-1}\int_t^{t+\varepsilon} x(s)\, ds$ is continuous in the Skorohod topology. By right-continuity, $h_\varepsilon(x) \to \pi_t(x)$ for each x as $\varepsilon \to 0$. Thus $\pi_t(x)$ is measurable.

Having proved $\pi_{t_1 \cdots t_k}$ measurable, we may, as in C, define the finite-dimensional sets as sets of the form $\pi_{t_1 \cdots t_k}^{-1} H$ with $k \geq 1$ and $H \in \mathscr{R}^k$. Each finite-dimensional set lies in $\mathscr{D}$ by the definition of measurability.

If T_0 is a subset of $[0, 1]$, let $\mathscr{F}_{T_0}$ be the class of sets $\pi_{t_1 \cdots t_k}^{-1} H$, where k is arbitrary, the t_i are arbitrary points of T_0, and $H \in \mathscr{R}^k$. Then $\mathscr{F}_{T_0}$ is a (finitely additive) field. Of course, $\mathscr{F}_{[0,1]}$ is the class of finite-dimensional sets.

THEOREM 14.5 *If* T_0 *contains* 1 *and is dense in* $[0, 1]$, *then* $\mathscr{F}_{T_0}$ *generates* $\mathscr{D}$.

These conditions are also necessary in order that $\mathscr{F}_{T_0}$ generate $\mathscr{D}$, although we shall not need this fact. Taking $T_0 = [0, 1]$ makes the theorem no easier to prove.

Proof. Since D is separable, it suffices to show that each open d_0-sphere $S_{d_0}(x, r)$ lies in the σ-field generated by $\mathscr{F}_{T_0}$. Fix the center x and radius r.

Choose in T_0 a sequence $t_1, t_2, \ldots$ that is dense in $[0, 1]$, with $t_1 = 1$. For $0 < \varepsilon < r$ and $k \geq 1$, let $A_k(\varepsilon)$ consist of those y for which there exists in Λ a λ satisfying

$$\|\lambda\| < r - \varepsilon$$

and

$$\max_{1 \le i \le k} |y(t_i) - x(\lambda t_i)| < r - \varepsilon.$$

It is enough to prove that

$$S_{d_0}(x, r) = \bigcup_{\varepsilon} \bigcap_{k=1}^{\infty} A_k(\varepsilon), \tag{14.59}$$

where the union extends over the rational ε in $(0, r)$ and to prove that each $A_k(\varepsilon)$ lies in $\mathscr{F}_{T_0}$.

We prove the second fact first. For fixed ε and k, let H_1 be the set of points $(x(\lambda t_1), \ldots, x(\lambda t_k))$ in R^k, where λ ranges over those functions in Λ satisfying $\|\lambda\| < r - \varepsilon$. Let H_2 be the set of points $(\alpha_1, \ldots, \alpha_k)$ in R^k such that $|\alpha_i - \beta_i| < r - \varepsilon$, $i = 1, \ldots, k$, for some $(\beta_1, \ldots, \beta_k)$ in H_1. Then H_2 is open and hence lies in $\mathscr{R}^k$, and $A_k(\varepsilon) = \pi^{-1}_{t_1 \cdots t_k} H_2$. Thus $A_k(\varepsilon) \in \mathscr{F}_{T_0}$.

It is easy to see that the left member of (14.59) is contained in the right member. We complete the proof by showing that

$$\bigcap_{k=1}^{\infty} A_k(\varepsilon) \subset S_{d_0}(x, r). \tag{14.60}$$

If y lies in the intersection on the left, choose for each k a λ_k in Λ with

$$\|\lambda_k\| < r - \varepsilon \tag{14.61}$$

and

$$\max_{1 \le i \le k} |y(t_i) - x(\lambda_k t_i)| < r - \varepsilon. \tag{14.62}$$

By Helly's selection theorem (p. 227), there is a subsequence $\{\lambda_{k'}\}$ and a nondecreasing function λ such that

$$\lim_{k'} \lambda_{k'} t = \lambda t \tag{14.63}$$

holds for continuity points t of λ. We shall show that $\lambda \in \Lambda$, that $\|\lambda\| \le r - \varepsilon$, and that $\sup_t |y(t) - x(\lambda t)| \le r - \varepsilon$. This will imply $d_0(x, y) < r$ and prove (14.60).

If s and t are distinct continuity points of λ, then, by (14.61),

$$\left|\log \frac{\lambda t - \lambda s}{t - s}\right| = \lim_{k'} \left|\log \frac{\lambda_{k'} t - \lambda_{k'} s}{t - s}\right| \le r - \varepsilon. \tag{14.64}$$

This relation makes a jump in λ impossible (in particular, (14.63) holds for all t) and it implies that λ is strictly increasing; thus $\lambda \in \Lambda$. The inequality (14.64) also implies $\|\lambda\| \le r - \varepsilon$. By (14.62), $|y(t_1) - x(t_1)| < r - \varepsilon$ (recall that $t_1 = 1$). If $i > 1$, then, by (14.62), $|y(t_i) - x(\lambda_{k'} t_i)| < r - \varepsilon$ for $k' \ge i$. Since $\lambda_{k'} t \to \lambda t$, we have either $|y(t_i) - x(\lambda t_i)| \le r - \varepsilon$ or $|y(t_i) - x((\lambda t_i) -)| \le r - \varepsilon$ and, since $\{t_i\}$ is dense, $\sup_t |y(t) - x(\lambda t)| \le r - \varepsilon$ follows. This proves Theorem 14.5.

Since $\mathscr{F}_{T_0}$ is always a field, this theorem implies that, if T_0 contains 1 and is dense in [0, 1], then $\mathscr{F}_{T_0}$ is a determining class. Not even $\mathscr{F}_{[0,1]}$ is a convergence-determining class—the counterexamples for C apply equally well here.

If P is a probability measure on $(D, \mathscr{D})$, its finite-dimensional distributions are the measures $P\pi_{t_1\cdots t_k}^{-1}$. If T_0 contains 1 and is dense in [0, 1], then P is completely determined by its finite-dimensional distributions for time points in T_0.

We shall use in D the same conventions about coordinate variables we used in C—the conventions set out at the end of Section 8.

Remarks. The topology studied here is the J_1 topology of Skorohod (1956); see also Kolmogorov (1956), Prohorov (1956), and Skorohod (1961).

PROBLEMS

1. Let D^+ be the class of functions on [0, 1] that have only discontinuities of the first kind in (0, 1) and have right-hand limits at 0 and left-hand limits at 1. Convert D^+ into a pseudometric space in such a way that (i) x and y are at distance 0 if and only if they agree at their common continuity points and (ii) D is isometric with the standard quotient space [Kelley (1955, p. 123)].

2. The Skorohod topology is finer than that given by the metric $\int |x(t) - y(t)|\, dt$.

3. Under the Skorohod topology and pointwise addition of functions, D is not a topological group.

4. The set C is nowhere dense in D.

5. Put

$$w_x'''(\delta) = \sup_{t_2 - t_1 < \delta} \sup_{t_1 < t < t_2} \min \{w_x(t_1, t), w_x(t, t_2)\}.$$

Show that

$$w_x'''(\delta) = \sup_{t_2 - t_1 < \delta} \inf_{t_1 < t < t_2} \max \{w_x(t_1, t), w_x(t, t_2)\}$$

and

$$w_x''(\delta) \le w_x'''(\delta) \le 2w_x''(\delta).$$

15. WEAK CONVERGENCE AND TIGHTNESS IN D

Prove weak convergence in function space by proving weak convergence of the finite-dimensional distributions and then proving tightness—this was the technique so useful in C, and we want to adapt it to D. Since D is separable and complete under the metric d_0, a family of probability measures on $(D, \mathscr{D})$ is relatively compact if and only if it is tight, so there is no difficulty on that point. On the other hand, the fact that the natural projections are not continuous complicates matters somewhat.

Finite-Dimensional Distributions

For a probability measure P on $(D, \mathscr{D})$, let T_P consist of those t in [0, 1] for which the projection π_t is continuous except at points forming a set of

P-measure 0. The points 0 and 1 always lie in T_P. If $0 < t < 1$, then $t \in T_P$ if and only if $P(J_t) = 0$, where

$$J_t = \{x : x(t) \neq x(t-)\} \tag{15.1}$$

is the set of x that are discontinuous at t—recall (p. 121) that, for $0 < t < 1$, π_t is continuous at x if and only if x is continuous at t.

An element of D has at most countably many jumps. Let us prove the corresponding fact that $P(J_t) > 0$ is possible for at most countably many t. Let $J_t(\varepsilon)$ be the set of x having at t a jump $|x(t) - x(t-)|$ exceeding ε. For fixed positive ε and δ there can be at most finitely many points t for which $P(J_t(\varepsilon)) \geq \delta$, since if this inequality held for a sequence of distinct points $t_1, t_2, \ldots$, then the set $\limsup_n J_{t_n}(\varepsilon)$ would have measure at least δ and hence would be nonempty, contradicting the fact that for each x the saltus can exceed ε at only finitely many places. For a fixed positive ε, therefore, $P(J_t(\varepsilon))$ can exceed 0 for at most countably many t. Since $P(J_t(\varepsilon)) \uparrow P(J_t)$ as $\varepsilon \downarrow 0$, the result follows.

Thus: *T_P contains 0 and 1 and its complement in $[0, 1]$ is at most countable.* If $t_1, \ldots, t_k$ all lie in T_P, then $\pi_{t_1 \cdots t_k}$ is continuous except on a set of P-measure 0.

Suppose

$$P_n \Rightarrow P, \tag{15.2}$$

where P_n and P are probability measures on $(D, \mathscr{D})$; by Theorem 5.1,

$$P_n \pi_{t_1 \cdots t_k}^{-1} \Rightarrow P \pi_{t_1 \cdots t_k}^{-1} \tag{15.3}$$

holds if all the t_i lie in T_P. In general, (15.3) will not follow from (15.2) if some t_i lies outside T_P: If P is a unit mass at the function $I_{[0,\frac{1}{2})}$ and P_n a unit mass at $I_{[0,\frac{1}{2}+n^{-1})}$, then $P_n \Rightarrow P$ holds while $P_n \pi^{-1} \Rightarrow P\pi_{\frac{1}{2}}^{-1}$ fails.

Let us now prove the analogue of Theorem 8.1.

THEOREM 15.1 *If $\{P_n\}$ is tight and if $P_n \pi_{t_1 \cdots t_k}^{-1} \Rightarrow P \pi_{t_1 \cdots t_k}^{-1}$ holds whenever $t_1, \ldots, t_k$ all lie in T_P, then $P_n \Rightarrow P$.*

Proof. By tightness, each subsequence $\{P_{n'}\}$ contains a further subsequence $\{P_{n''}\}$ converging weakly to some limit Q. It is enough (Theorem 2.3) to show that Q always coincides with P.

If $t_1, \ldots, t_k$ all lie in $T_P \cap T_Q$, we have $P_{n''} \pi_{t_1 \cdots t_k}^{-1} \Rightarrow P \pi_{t_1 \cdots t_k}^{-1}$ by the hypothesis and $P_{n''} \pi_{t_1 \cdots t_k}^{-1} \Rightarrow Q \pi_{t_1 \cdots t_k}^{-1}$ because $P_{n''} \Rightarrow Q$. Thus

$$P \pi_{t_1 \cdots t_k}^{-1} = Q \pi_{t_1 \cdots t_k}^{-1}, \qquad t_1, \ldots, t_k \in T_P \cap T_Q. \tag{15.4}$$

Since T_P and T_Q each contain all but countably many points of $[0, 1]$, the same is true of $T_P \cap T_Q$; in particular, this intersection is dense. Since $T_P \cap T_Q$ contains 1, it follows by Theorem 14.5 that $\mathscr{D}$ is generated by the finite-dimensional sets based on time points in $T_P \cap T_Q$. Thus $P = Q$ follows from (15.4), which completes the proof.

Tightness

The analysis of tightness in C began with a result (Theorem 8.2) which substituted for compactness its Arzelá–Ascoli characterization. Theorem 14.3, which characterizes compactness in D, leads in exactly the same way to the following result. Let $\{P_n\}$ be a sequence of probability measures on $(D, \mathscr{D})$.

THEOREM 15.2 *The sequence $\{P_n\}$ is tight if and only if these two conditions hold:*

(i) *For each positive η, there exists an a such that*

$$P_n\{x: \sup_t |x(t)| > a\} \leq \eta, \qquad n \geq 1. \tag{15.5}$$

(ii) *For each positive ε and η, there exist a δ, $0 < \delta < 1$, and an integer n_0 such that*

$$P_n\{x: w'_x(\delta) \geq \varepsilon\} \leq \eta, \qquad n \geq n_0. \tag{15.6}$$

Using Theorem 14.4 in place of Theorem 14.3 gives a second set of conditions for tightness.

THEOREM 15.3 *The sequence $\{P_n\}$ is tight if and only if these two conditions hold:*

(i) *For each positive η, there exists an a such that*

$$P_n\{x: \sup_t |x(t)| > a\} \leq \eta, \qquad n \geq 1.$$

(ii) *For each positive ε and η, there exist a δ, $0 < \delta < 1$, and an integer n_0 such that*

$$P_n\{x: w''_x(\delta) \geq \varepsilon\} \leq \eta, \qquad n \geq n_0, \tag{15.7}$$

such that

$$P_n\{x: w_x[0, \delta) \geq \varepsilon\} \leq \eta, \qquad n \geq n_0, \tag{15.8}$$

and such that

$$P_n\{x: w_x[1 - \delta, 1) \geq \varepsilon\} \leq \eta, \qquad n \geq n_0. \tag{15.9}$$

In the most interesting cases, (15.8) and (15.9) are automatically satisfied, as the next theorem shows. Let P_n and P be probability measures on $(D, \mathscr{D})$; by the definition (15.1), $J_1 = \{x: x(1) \neq x(1-)\}$.

THEOREM 15.4 *Suppose that*

$$P_n \pi_{t_1 \cdots t_k}^{-1} \Rightarrow P \pi_{t_1 \cdots t_k}^{-1} \tag{15.10}$$

holds whenever $t_1, \ldots, t_k$ *all lie in* T_P. *Suppose further that* $P(J_1) = 0$. *Suppose finally that, for each positive* ε *and* η, *there exist a* δ, $0 < \delta < 1$, *and an integer* n_0 *such that*

$$P_n\{x : w_x''(\delta) \geq \varepsilon\} \leq \eta, \qquad n \geq n_0. \tag{15.11}$$

Then $P_n \Rightarrow P$.

Proof. It will suffice, by Theorem 15.1, to show that $\{P_n\}$ is tight. We first verify condition (i) of the preceding theorem. Given a positive η, choose δ and n_0 satisfying (15.11) with $\varepsilon = 1$ (say). Since T_P is dense, it contains points $t_1, \ldots, t_k$, with $0 = t_1 < \cdots < t_k = 1$, such that $t_i - t_{i-1} < \delta$. From (15.10) it follows that $\{P_n \pi_{t_1 \cdots t_k}^{-1}\}$ is a tight sequence in R^k and hence that

$$P_n\left\{x : \max_{1 \leq i \leq k} |x(t_i)| > a_0\right\} \leq \eta, \qquad n \geq 1, \tag{15.12}$$

for an appropriately chosen a_0. If $|x(t_i)| \leq a_0$, $i = 1, \ldots, k$, and if $w_x''(\delta) < 1$, then $|x(t)| < a_0 + 1$ for all t. If $a = a_0 + 1$, then by (15.12) and (15.11) (with $\varepsilon = 1$) we have $P_n\{x : \sup_t |x(t)| > a\} \leq 2\eta$ for $n \geq n_0$. For the finitely many n preceding n_0, we can ensure that this inequality holds by increasing a, each individual P_n being tight. This establishes condition (i).

As for condition (ii) of the preceding theorem, we must find, given ε and η, a δ and n_0 satisfying (15.8) and (15.9). Consider (15.8). Since each x in D is right-continuous, we have

$$P\{x : |x(\delta) - x(0)| \geq \varepsilon\} < \eta \tag{15.13}$$

for all sufficiently small δ. Choose in the dense set T_P a δ small enough that (15.13) holds and small enough that (15.11) holds for an appropriate n_0. Since 0 and δ both lie in T_P, we have $P_n \pi_{0,\delta}^{-1} \Rightarrow P \pi_{0,\delta}^{-1}$ as a special case of (15.10), and it follows by (15.13) that

$$P_n\{x : |x(\delta) - x(0)| \geq \varepsilon\} < \eta \tag{15.14}$$

for all sufficiently large n. If $w_x''(\delta) < \varepsilon$ and $|x(\delta) - x(0)| < \varepsilon$, then $|x(s) - x(0)| < 2\varepsilon$ for all s in $[0, \delta]$, and hence $w_x[0, \delta] < 4\varepsilon$. Thus (15.11) and (15.14) together imply $P_n\{x : w_x[0, \delta) \geq 4\varepsilon\} \leq 2\eta$. This takes care of (15.8).

With one change, the symmetric argument works for (15.9). In place of (15.13), we need

$$P\{x : |x(1) - x(1 - \delta)| \geq \varepsilon\} < \eta \tag{15.15}$$

for small δ. Since an element of D need not be left-continuous, (15.15) will not in general hold; since we have assumed $P(J_1) = 0$, x is left-continuous at $t = 1$ except for x in a set of P-measure 0, and (15.15) does hold for all sufficiently small δ.

This completes the proof of Theorem 15.4. The condition $P(J_1) = 0$ is essential: If P is a unit mass at the function $I_{\{1\}}$ and P_n is a unit mass at $I_{[1-n^{-1},1]}$, then all the finite-dimensional distributions converge and the condition involving (15.11) holds, but P_n does not converge weakly to P.

Since

$$\lim_{t\to 1} P\{x:|x(1-) - x(t)| > \varepsilon\} = 0, \qquad \varepsilon > 0,$$

we have $P(J_1) = 0$ if and only if

$$\lim_{t\to 1} P\{x:|x(1) - x(t)| > \varepsilon\} = 0, \qquad \varepsilon > 0. \tag{15.16}$$

It is usually easy to check the condition $P(J_1) = 0$ via (15.16).

Our next result shows what happens if in place of $w'_x(\delta)$ or $w''_x(\delta)$ we use the modulus $w_x(\delta)$ appropriate to C.

THEOREM 15.5 *Suppose that, for each positive η, there exists an a such that*

$$P_n\{x:|x(0)| > a\} \leq \eta, \qquad n \geq 1. \tag{15.17}$$

Suppose further that, for each positive ε and η, there exist a δ, $0 < \delta < 1$, and an integer n_0, such that

$$P_n\{x:w_x(\delta) \geq \varepsilon\} \leq \eta, \qquad n \geq n_0. \tag{15.18}$$

Then $\{P_n\}$ is tight, and, if P is the weak limit of a subsequence $\{P_{n'}\}$, then $P(C) = 1$.

Proof. Since $w'_x(\delta) \leq w_x(2\delta)$ for $\delta < \frac{1}{2}$ (see (14.9)), condition (ii) of Theorem 15.2 is satisfied. That condition (i) of that theorem is satisfied follows easily from the hypotheses here and the inequality

$$|x(t)| \leq |x(0)| + \sum_{i=1}^{k} |x(it/k) - x((i-1)t/k)|.$$

Hence $\{P_n\}$ is tight.

If $w_y(\frac{1}{2}\delta) \geq 2\varepsilon$, then y is interior to $\{x:w_x(\delta) \geq \varepsilon\}$; $P_{n'} \Rightarrow P$ therefore implies

$$P\{y:w_y(\tfrac{1}{2}\delta) \geq 2\varepsilon\} \leq \liminf_{n'} P_{n'}\{x:w_x(\delta) \geq \varepsilon\}. \tag{15.19}$$

Given ε and η, choose δ and n_0 so that (15.18) holds; it follows by (15.19) that $P\{y:w_y(\frac{1}{2}\delta) \geq 2\varepsilon\} \leq \eta$. For each k, therefore, there exists a positive δ_k

such that, if $A_k = \{y : w_y(\delta_k) \geq 1/k\}$, then $P(A_k) < 1/k$. Put $A = \liminf_k A_k$; then $P(A) = 0$, and $x \notin A$ implies $\lim_{\delta \to 0} w_x(\delta) = 0$, so that $P(C) = 1$, which proves the theorem.

In applications, the hypothesis (15.18) is generally verified by using the corollary to Theorem 8.3, which carries over to D with no change. Note that, if the n_0 in (15.18) is 1 for every ε and η, then $P_n(C) = 1$ for all n.

Random Elements of D

A random element of D we shall often call a random function. It will always be clear from the context whether a random function under consideration is a random element of C or of D.

If X is a mapping from $(\Omega, \mathscr{B}, \mathsf{P})$ into D, then, for each ω, $X(\omega)$ is an element of D whose value at t we denote by $X(t, \omega)$. For each t, $X(t)$ denotes the real function $\pi_t X$ on Ω—its value at ω is $X(t, \omega)$. Just as in the space C, X is a random element of D $(X^{-1}\mathscr{D} \subset \mathscr{B})$ if and only if, for each t, $X(t)$ is a random variable (use Theorem 14.5).

A sequence $\{X_n\}$ of random elements of D is by definition tight when the sequence of corresponding distributions is tight. Each of Theorems 15.1 through 15.5 can be routinely reformulated in terms of random functions.

A Criterion for Convergence

Let us now combine the results of this section with those of Section 12 to obtain concrete criteria for convergence in distribution.

Let X_n and X be random elements of D. Write T_X for T_P, where P is the distribution of X. Thus T_X contains 0 and 1, and, if $0 < t < 1$, $t \in T_X$ if and only if $\mathsf{P}\{X(t) \neq X(t-)\} = 0$. We shall write $w''(X, \delta)$ in place of $w''_X(\delta)$.

THEOREM 15.6 *Suppose that*

$$(X_n(t_1), \ldots, X_n(t_k)) \xrightarrow{\mathscr{D}} (X(t_1), \ldots, X(t_k)) \tag{15.20}$$

holds whenever $t_1, \ldots, t_k$ *all lie in* T_X; *that* $\mathsf{P}\{X(1) \neq X(1-)\} = 0$; *and that*

$$\mathsf{P}\{|X_n(t) - X_n(t_1)| \geq \lambda, |X_n(t_2) - X_n(t)| \geq \lambda\} \leq \frac{1}{\lambda^{2\gamma}} [F(t_2) - F(t_1)]^{2\alpha} \tag{15.21}$$

for $t_1 \leq t \leq t_2$ *and* $n \geq 1$, *where* $\gamma \geq 0$, $\alpha > \frac{1}{2}$, *and* F *is a nondecreasing, continuous function on* $[0, 1]$. *Then* $X_n \xrightarrow{\mathscr{D}} X$.

There is a more restrictive version of (15.21) involving moments, namely

$$\mathsf{E}\{|X_n(t) - X_n(t_1)|^{\gamma} |X_n(t_2) - X_n(t)|^{\gamma}\} \leq [F(t_2) - F(t_1)]^{2\alpha}.$$

Proof. By Theorem 15.4, it suffices to show that for each positive ε and η there exists a positive δ such that

$$\mathsf{P}\{w''(X_n, \delta) \geq \varepsilon\} \leq \eta \tag{15.22}$$

holds for all n. For fixed τ, δ, and n, and for m a positive integer, consider the random variables

$$\xi_i = X_n\left(\tau + \frac{i}{m}\,\delta\right) - X_n\left(\tau + \frac{i-1}{m}\,\delta\right), \qquad i = 1, 2, \ldots, m. \tag{15.23}$$

Let

$$M''_m = \max \min \left\{\left|X_n\left(\tau + \frac{j}{m}\,\delta\right) - X_n\left(\tau + \frac{i}{m}\,\delta\right)\right|,\right.$$
$$\left.\left|X_n\left(\tau + \frac{k}{m}\,\delta\right) - X_n\left(\tau + \frac{j}{m}\,\delta\right)\right|\right\},$$

where the maximum extends over $0 \leq i \leq j \leq k \leq m$. By (15.21) and Theorem 12.5,

$$\mathsf{P}\{M''_m \geq \varepsilon\} \leq \frac{K}{\varepsilon^{2\gamma}}\,[F(\tau + \delta) - F(\tau)]^{2\alpha}, \tag{15.24}$$

where $K = K''_{\gamma,\alpha}$ depends only on γ and α.

Put

$$w''(X_n, [\tau, \tau + \delta]) = \sup \min \{|X_n(t) - X_n(t_1)|, |X_n(t_2) - X_n(t)|\}, \tag{15.25}$$

where the supremum extends over t_1, t, t_2 satisfying $\tau \leq t_1 \leq t \leq t_2 \leq \tau + \delta$. Since X_n is a right-continuous function on $[0, 1]$, letting $m \to \infty$ in (15.24) yields

$$\mathsf{P}\{w''(X_n, [\tau, \tau + \delta]) > \varepsilon\} \leq \frac{K}{\varepsilon^{2\gamma}}\,[F(\tau + \delta) - F(\tau)]^{2\alpha}. \tag{15.26}$$

Suppose now that $\delta = 1/(2u)$ is the reciprocal of an even integer, and suppose that

$$w''(X_n, [2i\delta, (2i + 2)\delta]) \leq \varepsilon, \qquad 0 \leq i \leq u - 1, \tag{15.27}$$

and

$$w''(X_n, [(2i + 1)\delta, (2i + 3)\delta]) \leq \varepsilon, \qquad 0 \leq i \leq u - 2. \tag{15.28}$$

If $t_1 \leq t \leq t_2$ and $t_2 - t_1 \leq \delta$, then t_1 and t_2 both lie in some one of the $2u - 1$ intervals involved in (15.27) and (15.28), so that

$$\min \{|X_n(t) - X_n(t_1)|, |X_n(t_2) - X_n(t)|\} \leq \varepsilon.$$

Thus (15.27) and (15.28) together imply $w''(X_n, \delta) \leq \varepsilon$. It now follows by

(15.26) that

$$P\{w''(X_n, \delta) \geq \varepsilon\} \leq \frac{K}{\varepsilon^{2\gamma}}(\Sigma' + \Sigma''), \tag{15.29}$$

where each of Σ' and Σ'' is a sum of the form

$$\sum_{k=1}^{r}[F(t_k) - F(t_{k-1})]^{2\alpha}$$

with $0 \leq t_1 \leq \cdots \leq t_r \leq 1$ and $t_k - t_{k-1} \leq 2\delta$. But then

$$P\{w''(X_n, \delta) \geq \varepsilon\} \leq \frac{2K}{\varepsilon^{2\gamma}}[F(1) - F(0)][w_F(2\delta)]^{2\alpha-1}. \tag{15.30}$$

Since $2\alpha > 1$ and F is continuous, the right member of this inequality goes to 0 with δ, which proves (15.22).

Criteria for Existence†

These ideas lead also to a condition for the existence in D of a random element with specified finite-dimensional distributions. As in Theorem 12.4, for each k-tuple $t_1, \ldots, t_k$ of points of $[0, 1]$, let $\mu_{t_1 \cdots t_k}$ be a probability measure on $(R^k, \mathscr{R}^k)$, and assume these measures satisfy the consistency conditions of Kolmogorov's existence theorem.

THEOREM 15.7 *There exists in D a random element with finite-dimensional distributions $\mu_{t_1 \cdots t_k}$, provided these distributions are consistent; provided*

$$\mu_{t_1 t t_2}\{(\beta_1, \beta, \beta_2) \colon |\beta - \beta_1| \geq \lambda, |\beta_2 - \beta| \geq \lambda\} \leq \frac{1}{\lambda^{2\gamma}}[F(t_2) - F(t_1)]^{2\alpha} \tag{15.31}$$

for $t_1 \leq t \leq t_2$, where $\gamma \geq 0$, $\alpha > \frac{1}{2}$, and F is a nondecreasing, continuous function on $[0, 1]$; *and provided*

$$\lim_{h \downarrow 0} \mu_{t,t+h}\{(\beta_1, \beta_2) : |\beta_1 - \beta_2| \geq \varepsilon\} = 0, \qquad 0 \leq t < 1. \tag{15.32}$$

There is a more restrictive version of condition (15.31):

$$\int_{R^3} |\beta - \beta_1|^\gamma |\beta_2 - \beta|^\gamma \, d\mu_{t_1 t t_2}(\beta_1, \beta, \beta_2) \leq [F(t_2) - F(t_1)]^{2\alpha}.$$

Proof. It is easy to construct for each n a random function X_n which is constant over each interval $[(i-1)2^{-n}, i2^{-n})$ and for which the joint distribution of

$$\left(X_n(0), X_n\left(\frac{1}{2^n}\right), \ldots, X_n\left(\frac{2^n - 1}{2^n}\right), X_n(1)\right)$$

† This topic may be omitted.

is $\mu_{t_0\cdots t_k}$ with $k = 2^n$ and $t_i = i2^{-n}$. We shall prove that $\{X_n\}$ is tight. We first show that, for given ε and η, there exists a δ such that

$$\mathsf{P}\{w''(X_n, \delta) \geq \varepsilon\} \leq \eta \tag{15.33}$$

holds for sufficiently large n.

If t_1, t, and t_2 are integral multiples of 2^{-n} and $t_1 \leq t \leq t_2$, then, by (15.31),

$$\mathsf{P}\{|X_n(t) - X_n(t_1)| \geq \lambda, |X_n(t_2) - X_n(t)| \geq \lambda\} \leq \frac{1}{\lambda^{2\gamma}}[F(t_2) - F(t_1)]^{2\alpha}.$$

Consider again the variables (15.23). If the points $\tau + j\delta/m, j = 0, 1, \ldots, m$, are all integral multiples of 2^{-n}, then (15.24) follows as before. If τ and δ are integral multiples of 2^{-n}, then the same is true of the points $\tau + j\delta/m$, provided $m = \delta 2^n$. But, since X_n is constant over each interval $[(i-1)2^{-n}, i2^{-n})$, M''_m for $m = \delta 2^n$ is just the quantity defined by (15.25). Thus (15.26) holds if τ and δ are both integral multiples of 2^{-n}.

Suppose now that $\delta = 1/2^v$ for an integer $v > 1$. If $n \geq v$, then δ is an integral multiple of 2^{-n} and so are the endpoints of all the intervals involved in (15.27) and (15.28). It follows as before that (15.30) holds for $n \geq v$. Thus there does exist a δ such that (15.33) holds for all large n.

The X_n will be tight if their distributions P_n satisfy the hypotheses of Theorem 15.3. Now (15.7) is satisfied because of (15.33). If $2^{-k} < \delta$, then

$$\sup_t |X_n(t)| \leq \max_{i \leq 2^k} \left|X_n\left(\frac{i}{2^k}\right)\right| + w''(X_n, \delta).$$

Since the distributions of the first term on the right all coincide for $n \geq k$, it follows by (15.33) that condition (i) of Theorem 15.3 is satisfied.

To take care of (15.8) and (15.9) we make the temporary assumption that, for some positive δ_0, $h \leq \delta_0$ implies

$$\mu_{0,h}\{(\beta_1, \beta_2) : \beta_1 = \beta_2\} = 1, \qquad \mu_{1,1-h}\{(\beta_1, \beta_2) : \beta_1 = \beta_2\} = 1. \tag{15.34}$$

With this assumption, (15.8) and (15.9) certainly hold, so that $\{X_n\}$ is tight. If X is the limit in distribution of some subsequence, then $(X(t_1), \ldots, X(t_k))$ has distribution $\mu_{t_1\cdots t_k}$ for dyadic rational t_i, and the general case follows via (15.32) by approximation from above.

It remains to remove the restriction (15.34). Take $\delta_0 < \frac{1}{2}$, put

$$f(t) = \begin{cases} 0 & \text{for } 0 \leq t \leq \delta_0, \\ \dfrac{t - \delta_0}{1 - 2\delta_0} & \text{for } \delta_0 \leq t \leq 1 - \delta_0, \\ 1 & \text{for } 1 - \delta_0 \leq t \leq 1, \end{cases}$$

and define $\nu_{t_1\ldots t_k}$ as $\mu_{s_1\ldots s_k}$ with $s_i = f(t_i)$. Then the $\nu_{t_1\ldots t_k}$ satisfy the conditions of the theorem (with a new F), as well as the special condition (15.34), so that there is a random element Y of D with these finite-dimensional distributions. We now need only define X by $X(t) = Y(\delta_0 + t(1 - 2\delta_0))$, $0 \leq t \leq 1$.

As an illustration of Theorem 15.7, let us use it to prove the existence in D of a random element describing a Markov chain in continuous time. Let $p_{ij}(t)$, $i, j = 1, 2, \ldots$ be nonnegative numbers defined for $t > 0$ and satisfying $\Sigma_j p_{ij}(t) = 1$ and $\Sigma_j p_{ij}(s) p_{jk}(t) = p_{ik}(s + t)$. Let p_i, $i = 1, 2, \ldots$, be nonnegative numbers satisfying $\Sigma_i p_i = 1$. Define $p_j(0) = p_j$ and $p_j(t) = \Sigma_i p_i p_{ij}(t)$ for $t > 0$.

Under the regularity assumption that

$$\lim_{t\to 0} p_{ii}(t) = 1 \tag{15.35}$$

holds uniformly in i, we shall prove the existence in D of a random element X with finite-dimensional distributions specified by

$$\mathsf{P}\{X(t_u) = i_u, u = 0, 1, \ldots, m\} = p_{i_0}(t_0) \prod_{u=1}^{m} p_{i_{u-1} i_u}(t_u - t_{u-1}) \tag{15.36}$$

for $0 \leq t_0 < t_1 < \cdots < t_m \leq 1$. The consistency of the finite-dimensional distributions implied by (15.36) is easy to establish.

We first prove the existence of a positive K such that

$$0 \leq 1 - p_{ii}(t) \leq Kt \tag{15.37}$$

for all i and all positive t. Because of the uniformity in (15.35), whatever ε, there is a δ such that $p_{ii}(s) > 1 - \varepsilon$ for all i if $s \leq 2\delta$. If $t \leq \delta$, then $\delta \leq mt \leq 2\delta$ for some positive integer m, and we have

$$\begin{aligned} 1 - \varepsilon \leq p_{ii}(mt) &\leq [p_{ii}(t)]^m + \sum_{l=0}^{m-2} [p_{ii}(t)]^l \sum_{j \neq i} p_{ij}(t) p_{ji}((m - l - 1)t) \\ &\leq [p_{ii}(t)]^m + \sum_{l=0}^{m-2} [p_{ii}(t)]^l \sum_{j \neq i} p_{ij}(t)\varepsilon \\ &\leq [p_{ii}(t)]^m + \varepsilon\{1 - [p_{ii}(t)]^m\}. \end{aligned}$$

Take $\varepsilon = \frac{1}{3}$; it follows (since $\log u \leq u - 1$) that $m(1 - p_{ii}(t)) \leq \log 2 \leq 1$. Take $K = \delta^{-1}$. Since $mt \geq \delta$, it follows that (15.37) holds for $t \leq \delta$. But the inequality is trivial for $t \geq \delta$.

If (15.36) holds, then

$$\mathsf{P}\{X(t) \neq X(t_1), X(t_2) \neq X(t)\} = \sum p_i(t_1)\, p_{ij}(\delta_1)\, p_{jk}(\delta_2), \tag{15.38}$$

where $\delta_1 = t - t_1$ and $\delta_2 = t_2 - t$ and where the summation extends over i and over those j and k for which $i \neq j$ and $j \neq k$. By (15.37), the sum in 5.38) is at most $K^2\delta_1\delta_2 \leq K^2(\delta_1 + \delta_2)^2$. Thus the finite-dimensional stributions implied by (15.36) satisfy (15.31) with $\gamma = 0$, $\alpha = 1$, and $t) = Kt$. Since (15.36) and (15.37) also imply $\mathsf{P}\{X(t+h) \neq X(t)\} \leq Kh$, 5.32) is also satisfied. Thus some random element X of D satisfies (15.36). ince the possible values of $X(t)$ are at mutual distance at least 1 on the line, is, with probability 1, a step function constant over intervals.)

ther Criteria†

ne function F in Theorem 15.6 was assumed nondecreasing and continuous. 'e can replace the assumption of continuity by that of continuity from the ght, provided we strengthen (15.21) to

$$\text{5.39)} \quad \mathsf{P}\{|X_n(t) - X_n(t_1)| \geq \lambda, |X_n(t_2) - X_n(t)| \geq \lambda\} \leq \frac{1}{\lambda^{2\gamma}}[F(t) - F(t_1)]^\alpha[F(t_2) - F(t)]^\alpha.$$

ıdeed, (15.39) and Theorem 12.6 yield, in place of (15.24), the stronger ıequality

$$\text{5.40)} \quad \mathsf{P}\{M''_m \geq \varepsilon\} \leq \frac{K}{\varepsilon^{2\gamma}}[F(\tau+\delta) - F(\tau)]^{2\alpha} \times \min_{1 \leq i \leq m}\left[1 - \frac{F\left(\tau + \frac{i}{m}\delta\right) - F\left(\tau - \frac{i-1}{m}\delta\right)}{F(\tau+\delta) - F(\tau)}\right]^\alpha$$

et $J(\tau, \tau+\delta]$ denote the maximum jump of F in the interval $(\tau, \tau+\delta]$; 15.40) implies

$$\text{(15.41)} \quad \mathsf{P}\{M''_m \geq \varepsilon\} \leq \frac{K}{\varepsilon^{2\gamma}}[F(\tau+\delta) - F(\tau)]^{2\alpha}\left[1 - \frac{J(\tau, \tau+\delta]}{F(\tau+\delta) - F(\tau)}\right]^\alpha.$$

The right member of this inequality is to be interpreted as 0 if $F(\tau) = $ $'(\tau+\delta)$.)

Exactly as before, we obtain (15.29), where now each of Σ' and Σ'' is a um of the form

$$\text{15.42)} \qquad \sum_{k=1}^{r}[F(t_k) - F(t_{k-1})]^{2\alpha}\left[1 - \frac{J(t_{k-1}, t_k]}{F(t_k) - F(t_{k-1})}\right]^\alpha$$

vith $0 \leq t_1 \leq \cdots \leq t_r \leq 1$ and $t_k - t_{k-1} \leq 2\delta$. To prove (15.22) it therefore uffices to show that, for small enough δ, all sums (15.42) are less than $\eta_0 = \eta\varepsilon^{2\gamma}/2K$. Write Δ_k for $F(t_k) - F(t_{k-1})$ and J_k for $J(t_{k-1}, t_k]$. Since $\alpha > \frac{1}{2}$

---

† This topic may be omitted.

and the Δ_k add to at most $F(1) - F(0)$,

$$\sum_k \Delta_k^\alpha(\Delta_k - J_k)^\alpha \le [F(1) - F(0)]^{\alpha+\frac{1}{2}} \max_k (\Delta_k - J_k)^{\alpha-\frac{1}{2}}.$$

But the left member of this inequality is the sum (15.42), and hence it is enough to show that, for small enough δ, $0 \le t - s \le 2\delta$ implies

$$F(t) - F(s) - J(s, t] < \eta_1 = \frac{\eta_0^{1/(\alpha-\frac{1}{2})}}{[F(1) - F(0)]^{(\alpha+\frac{1}{2})/(\alpha-\frac{1}{2})}}. \tag{15.43}$$

Since F is an element of D, there exist (see (14.8)) points s_i, with $0 = s_0 < s_1 < \cdots < s_l = 1$, such that $w_F[s_{i-1}, s_i) < \frac{1}{2}\eta_1$. Take 2δ smaller than the minimum of the $s_i - s_{i-1}$. If $t - s \le 2\delta$, then either s and t lie in a common interval $[s_{i-1}, s_i)$, in which case the left member of (15.43) is less than $\frac{1}{2}\eta_1$, or else they lie in adjacent intervals $[s_{i-1}, s_i)$ and $[s_i, s_{i+1})$, in which case it is at most $F(t) - F(s_i) + F(s_i-) - F(s) \le \eta_1$.

Thus (15.39) for a right-continuous F implies (15.22). The rest of the proof goes through as before, and we can conclude that $X_n \xrightarrow{\mathscr{D}} X$. In the same way, we can in Theorem 15.7 take F to be continuous from the right if we replace (15.31) by

$$\mu_{t_1 t t_2}\{(\beta_1, \beta, \beta_2): |\beta - \beta_1| \ge \lambda, |\beta_2 - \beta| \ge \lambda\} \le \frac{1}{\lambda^{2\gamma}}[F(t) - F(t_1)]^\alpha[F(t_2) - F(t)]^\alpha. \tag{15.44}$$

Separable Stochastic Processes†

According to Theorem 9.2, if the finite-dimensional distributions of a separable stochastic process can be realized as the finite-dimensional distributions of a probability measure on $(C, \mathscr{C})$, then the sample paths of the process are continuous with probability 1. We shall prove a similar result for the space D.

Let $\{\xi_t: 0 \le t \le 1\}$ be a stochastic process and let P be a probability measure on $(D, \mathscr{D})$. Assume that the process and the measure have the same finite-dimensional distributions. Assume further that the process is separable, so that (9.16) holds for every α and β and every open interval I. It is no restriction to assume $0 \in T_0$, in which case (9.16) holds also if I has the form $[0, s)$.

We now construct sets serving the same function as the sets (9.20) did in the arguments in Section 9. This time let V denote the general system

$$V: k; r_1, \ldots, r_k; s_1, \ldots, s_k; \alpha_1, \ldots, \alpha_k, \tag{15.45}$$

† This topic may be omitted.

ıere k is an arbitrary integer, where the r_i, s_i, and α_i are rational, and ıere

$$0 = r_1 < s_1 < r_2 < s_2 < \cdots < r_k < s_k = 1.$$

efine

5.46) $$\Omega_{T_0}(V, \varepsilon) = \bigcap_{i=1}^{k} \{\omega : \alpha_i \le \xi_t(\omega) \le \alpha_i + \varepsilon, t \in (r_i, s_i) \cap T_0\}$$

$$\cap \bigcap_{i=2}^{k} \{\omega : \min\{\alpha_{i-1}, \alpha_i\} \le \xi_t(\omega) \le \max\{\alpha_{i-1}, \alpha_i\} + \varepsilon, t \in (s_{i-1}, r_i) \cap T_0\};$$

the first intersection, we replace (r_1, s_1) by $[r_1, s_1)$ (for $i = 1$). Let $\mathscr{V}_{k\delta}$ ›note the class of systems (15.45) that have a fixed value of k and satisfy $- s_{i-1} < \delta$, $i = 2, \ldots, k$. And now define

5.47) $$\Omega_{T_0} = \bigcap_{\varepsilon} \bigcup_{k} \bigcap_{\delta} \bigcup_{V \in \mathscr{V}_{k\delta}} \Omega_{T_0}(V, \varepsilon),$$

here ε and δ are restricted to positive rationals. Finally, define $\Omega_T(V, \varepsilon)$ ıd Ω_T by substituting $T = [0, 1]$ for T_0 in (15.46) and (15.47).

It is not hard to show that, if $\omega \in \Omega_T$, then the sample path corresponding ω is right-continuous at $t = 0$, has a left-hand limit at $t = 1$, and is ›ntinuous in $(0, 1)$ except possibly for discontinuities of the first kind. And the sample path has this property, then $\omega \in \Omega_T$, as follows by Lemma 1 of ection 14 (applied to the sample path normalized to be right-continuous).

We shall prove that $\Omega_T \in \mathscr{B}$ and $\mathsf{P}(\Omega_T) = 1$. The relations (9.22) and $\Omega_{T_0} \in \mathscr{B}$ follow by the same arguments as in Section 9; hence $\Omega_T \in \mathscr{B}$ and $\mathsf{P}(\Omega_T) = \mathsf{P}(\Omega_{T_0})$, and it suffices to prove

.5.48) $$\mathsf{P}(\Omega_{T_0}) = 1.$$

With $(t_1, t_2, \ldots)$ an enumeration of the points in T_0, define $\varphi : \Omega \to R^\infty$ nd $\psi : D \to R^\infty$ by $\varphi(\omega) = (\xi_{t_1}(\omega), \xi_{t_2}(\omega), \ldots)$ and $\psi(x) = (x(t_1), x(t_2), \ldots)$. ıst as before, (9.25) holds for every H in $\mathscr{R}^\infty$, where now P denotes the ıeasure on D (not C) having the finite-dimensional distributions of $\{\xi_t\}$.

For the system (15.45) and $\varepsilon > 0$, let $H_{T_0}(V, \varepsilon)$ consist of those points $= (z_1, z_2, \ldots)$ in R^∞ such that, first, for each $i = 1, \ldots, k$, the inequality

$$\alpha_i \le z_u \le \alpha_i + \varepsilon$$

olds for every u for which $t_u \in (r_i, s_i)$ (or $t_u \in [r_1, s_1)$ in the case $i = 1$) and, ›cond, for each $i = 2, \ldots k$, the inequality

$$\min\{\alpha_{i-1}, \alpha_i\} \le z_u \le \max\{\alpha_{i-1}, \alpha_i\} + \varepsilon$$

olds for every u for which $t_u \in (s_{i-1}, r_i)$. Now define

$$H_{T_0} = \bigcap_{\varepsilon} \bigcup_{k} \bigcap_{\delta} \bigcup_{V \in \mathscr{V}_{k\delta}} H_{T_0}(V, \varepsilon).$$

Then $H_{T_0}(V, \varepsilon)$ and H_{T_0} lie in $\mathscr{R}^\infty$, $\varphi^{-1}H_{T_0}(V, \varepsilon) = \Omega_{T_0}(V, \varepsilon)$, and $\varphi^{-1}H_{T_0} = \Omega_{T_0}$. By (9.25) we have $\mathsf{P}(\Omega_{T_0}) = P(\psi^{-1}H_{T_0})$, and, since $\psi^{-1}H_{T_0} = D$, (15.48) follows.

We have proved this result:

THEOREM 15.8 *If* $\{\xi_t : 0 \leq t \leq 1\}$ *is a separable stochastic process, and if there exists on* $(D, \mathscr{D})$ *a probability measure having the same finite-dimensional distributions as* $\{\xi_t\}$, *then the sample paths, with probability* 1, *are right-continuous at* $t = 0$, *have left-hand limits at* $t = 1$, *and are continuous on* $(0, 1)$ *except possibly for discontinuities of the first kind.*

An example shows that it is not possible to go on and prove that the paths are right-continuous and hence lie in D (with probability 1): Define

$$\xi_t(\omega) = \begin{cases} 0 & \text{if} \quad 0 \leq t \leq \omega \\ 1 & \text{if} \quad \omega < t \leq 1, \end{cases}$$

where ω is drawn uniformly from $\Omega = (0, 1)$. This points up the fact that in taking the elements of D to be right-continuous we were simply adopting a convention.

PROBLEMS

1. Show that Theorem 15.4 remains true if we only require that (15.10) hold for all $t_1, \ldots, t_k$ in T_0, where T_0 is dense in $[0, 1]$ and contains 0 and 1. Show that it fails if $0 \notin T_0$ or if $1 \notin T_0$.

2. The condition (15.32) in Theorem 15.7 is necessary.

3. If a random element X of D has the property that

$$\lim_{\delta \to 0} \sup_{0 \leqq t \leqq 1-\delta} \frac{1}{\delta} \mathsf{P}\{|X(t+\delta) - X(t)| \geq \varepsilon\} = 0$$

for every positive ε, then $\mathsf{P}\{X \in C\} = 1$.

4. Use Theorem 15.7 and Problem 3 to show that distributions $\mu_{t_1 \cdots t_k}$ can be realized as the finite-dimensional distributions of a random element of C if (15.31) holds (where $\gamma \geq 0$, $\alpha > \frac{1}{2}$, and F is nondecreasing and continuous) and if

$$\lim_{\delta \to 0} \sup_{0 \leqq t \leqq 1-\delta} \frac{1}{\delta} \mu_{t, t+\delta}\{(\beta_1, \beta_2) : |\beta_2 - \beta_1| \geq \varepsilon\} = 0$$

for every positive ε. There is the analogous result with (15.44) in place of (15.31). Use Theorem 9.2 to derive conditions for the continuity of sample paths for a separable process. (Compare Problem 8 in Section 12.)

5. Use Theorem 15.7 to construct a Poisson process in D. Now make a direct construction, starting with independent, exponentially distributed random variables.

16. APPLICATIONS

Donsker's Theorem

In some ways, the space D is more convenient than C for the formulation of Donsker's theorem. Given random variables $\xi_1, \xi_2, \ldots$ on $(\Omega, \mathscr{B}, \mathsf{P})$, with partial sums $S_n = \xi_1 + \cdots + \xi_n$, let $X_n(\omega)$ be the function in D with value

$$X_n(t, \omega) = \frac{1}{\sigma\sqrt{n}} S_{[nt]}(\omega) \tag{16.1}$$

at t. Since each $X_n(t)$ is a random variable, X_n is a random function—a random element of D.

We want to prove, under suitable conditions, that the distribution of X_n converges weakly to Wiener measure W. Now W is defined on $(C, \mathscr{C})$, but it is easily extended to $(D, \mathscr{D})$: Since $C \in \mathscr{D}$, and since the relative Skorohod topology in C coincides with the uniform topology there, $A \in \mathscr{D}$ implies $A \cap C \in \mathscr{C}$. We may therefore extend W to $(D, \mathscr{D})$ by giving it the value $W(A \cap C)$ for A in $\mathscr{D}$. Of course, C supports W. From now on we shall interpret W as this probability measure on $(D, \mathscr{D})$ or as a random element of D with this probability measure as its distribution.

THEOREM 16.1 *Suppose the random variables ξ_n are independent and identically distributed with mean* 0 *and finite, positive variance σ^2:*

$$\mathsf{E}\{\xi_n\} = 0, \qquad \mathsf{E}\{\xi_n^2\} = \sigma^2. \tag{16.2}$$

Then the random functions X_n defined by (16.1) *satisfy*

$$X_n \xrightarrow{\mathscr{D}} W. \tag{16.3}$$

Proof. The proof in Section 10 that the finite-dimensional distributions converge carries over with no difficulty. Thus we need only prove tightness.

There are several ways to prove tightness. One way is to verify the hypotheses of Theorem 15.5. Now condition (ii) of Theorem 8.3, formulated for random elements of D, requires that, for positive ε and η, there exist a δ, $0 < \delta < 1$, and an integer n_0 such that, for $0 \leq t \leq 1$,

$$\frac{1}{\delta} \mathsf{P}\left\{ \sup_{t \leq s \leq t+\delta} |X_n(s) - X_n(t)| \geq \varepsilon \right\} \leq \eta, \qquad n \geq n_0$$

(we replace $t + \delta$ by 1 if it exceeds 1). Furthermore, this condition reduces, by the arguments used in the proof of Theorem 8.4, to the following one:

For each positive ε there exist a $\lambda > 1$ and an integer n_0 such that

$$\mathsf{P}\left\{\max_{i\le n}|S_i| \ge \lambda\sigma\sqrt{n}\right\} \le \frac{\varepsilon}{\lambda^2}, \qquad n \ge n_0.$$

It suffices therefore to verify this last condition, which can be done by using the central limit theorem as in Section 10 (see (10.7)) or by using the bounds in Section 12 (see (12.23)). Thus the methods of the preceding chapter carry over without essential change.

Given the results in Section 15, we can construct a very simple proof. By Theorem 15.6, it is enough to establish the inequality

$$(16.4)\quad \mathsf{E}\{|X_n(t) - X_n(t_1)|^2 \cdot |X_n(t_2) - X_n(t)|^2\} \le 4(t_2 - t_1)^2, \qquad t_1 \le t \le t_2.$$

By (16.1), (16.2), and the independence of the ξ_n, the left side of (16.4) is

$$(16.5)\quad \frac{1}{\sigma^4 n^2}\mathsf{E}\{|S_{[nt]} - S_{[nt_1]}|^2\}\mathsf{E}\{|S_{[nt_2]} - S_{[nt]}|^2\} = \frac{1}{n^2}([nt] - [nt_1])([nt_2] - [nt]) \le \left\{\frac{[nt_2] - [nt_1]}{n}\right\}^2.$$

If $t_2 - t_1 \ge 1/n$, (16.4) follows from this. If $t_2 - t_1 < 1/n$, then either t_1 and t lie in the same subinterval $[(i-1)/n, i/n)$ or else t and t_2 do; in either of these cases the left side of (16.4) vanishes. This establishes (16.4) in general and proves the theorem.

Since the value of $\sup_t X_n(t)$ is the same for the random element of D defined by (16.1) as for the random element of C defined by (10.1), namely $\max_{i\le n} S_i/\sigma\sqrt{n}$, we can obviously use Theorem 16.1 to rederive (10.18). (We must prove that the mapping $h: D \to R^1$ defined by $h(x) = \sup_t x(t)$ is continuous, but this presents no problem.) We may rederive the results in Section 11 in the same way.

Some functions of the partial sums S_i have a simple expression in terms of the X_n in Theorem 16.1 but have no simple expression in terms of the X_n in Theorem 10.1. Such functions are better analyzed in D than in C. For example, for $x \in D$, let $h(x)$ be the Lebesgue measure of the set of t for which $x(t) > 0$. Then (p. 232) h is measurable ($h^{-1}\mathscr{R}^1 \subset \mathscr{D}$) and is continuous except on a set of Wiener measure 0. If X_n is defined by (16.1), then $h(X_n)$ is exactly n^{-1} times the number of positive sums among $S_1, \ldots, S_{n-1}$, whereas this is generally untrue if X_n is defined instead by (10.1). Combining Theorem 5.1, Theorem 16.1, and (11.26) leads to the arc sine law under the assumptions of the Lindeberg–Lévy theorem.†

† See the problems for some further applications.

ominated Measures

he random variables in Theorem 16.1 are defined on a probability space 2, $\mathscr{B}$, P). We shall show that the result remains true if P is replaced by an bitrary probability measure P_0 on $(\Omega, \mathscr{B})$ dominated by (absolutely ontinuous with respect to) P. For example, suppose that $\Omega = [0, 1]$, that consists of the linear Borel subsets of $[0, 1]$, and that $\xi_n(\omega) = 2\omega_n - 1$, here ω_n is the nth digit in the binary expansion of ω. Then Theorem 16.1 oplies to $\{\xi_n\}$ if ω is drawn from $[0, 1]$ according to the uniform distribu- on. According to the theorem we are going to prove, this is true also if ω drawn from $[0, 1]$ according to an arbitrary distribution having a density ith respect to Lebesgue measure.

We shall need the following preliminary result (in which $\sigma(\mathscr{B}_0)$ denotes e σ-field generated by $\mathscr{B}_0$).

HEOREM 16.2 *Let $E_1, E_2, \ldots$ be measurable sets in a probability space* 2, $\mathscr{B}$, P). *Suppose there exist a constant α and a subfield $\mathscr{B}_0$ of $\mathscr{B}$ such that*

$$P(E_n \cap E) \rightarrow \alpha P(E) \tag{6.6}$$

r every E in $\mathscr{B}_0$. Suppose further that all the E_n lie in $\sigma(\mathscr{B}_0)$. If P *dominates* $_0$, *a second probability measure on $\mathscr{B}$, then*

$$P_0(E_n) \rightarrow \alpha. \tag{16.7}$$

roof. From the hypotheses it follows just as in the proof of Theorem 4.5 iat

$$\int_{E_n} g\, dP \rightarrow \alpha \int g\, dP \tag{16.8}$$

or every integrable g. But (16.8) and (16.7) are the same thing if g is the adon–Nikodym derivative of P_0 with respect to P.

HEOREM 16.3 *Theorem* 16.1 *remains valid if* P *is replaced by an rbitrary probability measure* P_0 *dominated by it.*

roof. Define X_n' by

$$X_n'(t, \omega) = \frac{1}{\sigma\sqrt{n}} \sum_{p_n \le i \le nt} \xi_i(\omega), \tag{16.9}$$

here $\{p_n\}$ is a sequence of integers going to infinity slowly enough that $_n/\sqrt{n} \rightarrow 0$ $(X_n'(t, \omega) = 0$ if $t < p_n/n)$. If

$$\delta_n = \sup_t |X_n(t) - X_n'(t)|,$$

hen

$$\delta_n \le \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{p_n} |\xi_i|.$$

By Minkowski's inequality and the fact that $p_n/\sqrt{n} \to 0$,

$$\mathsf{E}^{\frac{1}{2}}\{\delta_n^2\} \leq \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{p_n} \mathsf{E}^{\frac{1}{2}}\{\xi_i^2\} \to 0,$$

so that, by Chebyshev's inequality,

$$\delta_n \xrightarrow{P} 0, \tag{16.10}$$

where (16.10) is interpreted in the sense of P (that is, P governs the distribution of the ξ_n). By Theorem 16.1,

$$X_n \xrightarrow{\mathscr{D}} W, \tag{16.11}$$

where again the relation is interpreted in the sense of P. Since $d(X_n, X_n') \leq \delta_n$, where d is the metric in D (either one), it follows by Theorem 4.1 that

$$X_n' \xrightarrow{\mathscr{D}} W \tag{16.12}$$

in the sense of P.

Let A be a W-continuity set (in D), temporarily fixed; (16.12) implies

$$\mathsf{P}\{X_n' \in A\} \to W(A). \tag{16.13}$$

Let $\mathscr{B}_0$ be the field of cylinders; $\mathscr{B}_0$ consists of sets of the form

$$\{\omega : (\xi_1(\omega), \ldots, \xi_k(\omega)) \in H\}$$

with $k \geq 1$ and $H \in \mathscr{R}^k$. If $E \in \mathscr{B}_0$, then, since $p_n \to \infty$, $\mathsf{P}(\{X_n' \in A\} \cap E) = \mathsf{P}\{X_n' \in A\}\mathsf{P}(E)$ for large n and it follows by (16.13) that

$$\mathsf{P}(\{X_n' \in A\} \cap E) \to W(A)\mathsf{P}(E).$$

Since the sets $\{X_n' \in A\}$ all lie in the σ-field generated by $\mathscr{B}_0$, it follows by Theorem 16.2 with $\alpha = W(A)$ that

$$\mathsf{P}_0\{X_n' \in A\} \to W(A) \tag{16.14}$$

if P_0 is absolutely continuous with respect to P.

Since (16.14) holds for every W-continuity set A, (16.12) holds when interpreted in the sense of P_0. Now (16.10) asserts that $\mathsf{P}\{\delta_n \geq \varepsilon\} \to 0$ for each positive ε; if P_0 is absolutely continuous with respect to P, then $\mathsf{P}_0\{\delta_n \geq \varepsilon\} \to 0$ as well, and (16.10) persists when interpreted in the sense of P_0. Applying Theorem 4.1 once more, we see that (16.11) holds in the sense of P_0, which completes the proof.

Theorem 16.3 concerns convergence in distribution. Notice that passing from P to a dominated measure P_0 trivially preserves such properties of $\{S_n\}$ as the law of the iterated logarithm, which hold with probability 1.

Empirical Distribution Functions

Let $\xi_1, \xi_2, \ldots$ be random variables with

$$0 \leq \xi_n(\omega) \leq 1,$$

and let $F_n(t, \omega)$ be the proportion of the points $\xi_1(\omega), \ldots, \xi_n(\omega)$ not exceeding t, so that $F_n(\cdot, \omega)$ is the empirical distribution function. Let $Y_n(\omega)$ be the element of D with value

$$Y_n(t, \omega) = \sqrt{n}\big(F_n(t, \omega) - F(t)\big) \tag{16.15}$$

at the point $t \in [0, 1]$, where F is the distribution function common to the ξ_n. Since each $Y_n(t)$ is a random variable, Y_n is a random element of D. In Section 13 we studied the related random element (13.7) of C; here we shall investigate Y_n directly, and we shall remove the restriction that the ξ_n be uniformly distributed. The stochastic process $\{Y_n(t): 0 \leq t \leq 1\}$ is sometimes called the empirical process.

THEOREM 16.4 *Suppose the ξ_n are independent and have a common distribution function $F(t)$. If Y_n is defined by* (16.15), *then*

$$Y_n \xrightarrow{\mathscr{D}} Y, \tag{16.16}$$

where Y is the Gaussian random element of D specified by

$$\begin{cases} \mathsf{E}\{Y(t)\} = 0 \\ \mathsf{E}\{Y(s)Y(t)\} = F(s)\big(1 - F(t)\big), \qquad s \leq t. \end{cases} \tag{16.17}$$

Proof. Just as Wiener measure does, the measure W° extends from $(C, \mathscr{C})$ to $(D, \mathscr{D})$. Denote by W° a random element of D with this extended measure as its distribution. We first show, under the assumption that the ξ_n are uniformly distributed over $[0, 1]$, that $Y_n \xrightarrow{\mathscr{D}} W^\circ$.

Let $U_n(t, \omega)$ be the number of points among $\xi_1(\omega), \ldots, \xi_n(\omega)$ not exceeding t. For $0 = t_0 < t_1 < \cdots < t_k = 1$, the random variables $U_n(t_i) - U_n(t_{i-1})$, $i = 1, \ldots, k$, are multinomially distributed with parameters n and $p_i = t_i - t_{i-1}$, and it follows as in Section 13 by the central limit theorem for multinomial trials that the finite-dimensional distributions of the Y_n converge weakly to those of W°. By Theorem 15.6, it suffices to prove that

$$\mathsf{E}\{|Y_n(t) - Y_n(t_1)|^2 \cdot |Y_n(t_2) - Y_n(t)|^2\} \leq 6(t - t_1)(t_2 - t)$$

for $t_1 \leq t \leq t_2$. But this is just the inequality (13.17) already established in Section 13. Hence $Y_n \xrightarrow{\mathscr{D}} W^\circ$.

Suppose now that the ξ_n have an arbitrary distribution function F over $[0, 1]$. Define an "inverse" to F by

$$\varphi(s) = \inf\{t : s \leq F(t)\}.$$

Then $s \leq F(t)$ if and only if $\varphi(s) \leq t$, so that, if η_n is uniformly distributed over $[0, 1]$, then $\mathsf{P}\{\varphi(\eta_n) \leq t\} = F(t)$. Since the theorem involves only the joint distribution of the ξ_n, we may represent them as $\xi_n = \varphi(\eta_n)$, with the η_n independent and uniformly distributed.

If $G_n(\cdot, \omega)$ is the empirical distribution function for $\eta_1(\omega), \ldots, \eta_n(\omega)$ and $Z_n(t, \omega) = \sqrt{n}\big(G_n(t, \omega) - t\big)$, then, by the case of the theorem already established, $Z_n \xrightarrow{\mathscr{D}} W^\circ$. But the empirical distribution of $\xi_1(\omega), \ldots, \xi_n(\omega)$ is $F_n(t, \omega) = G_n\big(F(t), \omega\big)$, so that, if Y_n is defined by (16.15), $Y_n(t, \omega) = Z_n\big(F(t), \omega\big)$. Define $\psi: D \to D$ by $(\psi x)(t) = x\big(F(t)\big)$. If x_n converges to x in the Skorohod topology and $x \in C$, then the convergence is uniform, so that ψx_n converges to ψx uniformly and hence in the Skorohod topology. Hence $Z_n \xrightarrow{\mathscr{D}} W^\circ$ and Theorem 5.1 imply $Y_n = \psi(Z_n) \xrightarrow{\mathscr{D}} \psi(W^\circ)$; since $Y = \psi(W^\circ)$ is Gaussian and satisfies (16.17), the proof is complete.

If we knew the distribution of

$$\sup_t Y(t) = \sup_t W^\circ\big(F(t)\big), \tag{16.18}$$

we could find the limiting distribution of

$$\sqrt{n}\, \sup_t \big(F_n(t, \omega) - F(t)\big). \tag{16.19}$$

Now (16.18) has the same distribution as $\sup_t W_t^\circ$ (namely that given by (11.40)) if F is continuous, but not otherwise. The same remarks apply to

$$\sup_t |Y(t)| = \sup_t |W^\circ\big(F(t)\big)|. \tag{16.20}$$

Remarks. The random function (16.1) has independent increments and so has W; for general results on weak convergence in this circumstance, see Kimme (1957 and 1960) and Skorohod (1957). For convergence of diffusions, see Stone (1963). Theorem 16.2 is due to Rényi (1958).

Concerning Theorem 16.4, see the remarks and references at the end of Section 13. See Schmid (1958) for the distributions of (16.18) and (16.20) for the general F.

PROBLEMS

1. Under the hypotheses of the Lindeberg-Lévy theorem, there are limiting distributions for

(a) $$\frac{1}{n^{\frac{3}{2}}} \sum_{k=1}^{n} |S_k|,$$

(b) $$\frac{1}{n^2} \sum_{k=1}^{n} S_k^2,$$

(c) $$\frac{1}{\sqrt{n}} \min_{\beta n \leq k \leq n} S_k, \qquad 0 < \beta < 1.$$

Construct the three relevant mappings from D to R^1 and prove they are measurable and continuous on a set of W-measure 1. For the forms of the limiting distributions, see Donsker (1951) for (a) and (b) and Mark (1949) for (c).

2. Theorem 16.1 (makes sense and) is true if n goes to infinity in a continuous manner. [See (2.4).]

3. For each n, let $\xi_{n1}, \ldots, \xi_{nn}$ be independent random variables with $\mathsf{P}\{\xi_{nk} = 1\} = p_n$ and $\mathsf{P}\{\xi_{nk} = 0\} = 1 - p_n$, and define Y_n by $Y_n(t) = \Sigma_{k \le nt}\, \xi_{nk}$. Assume $np_n \to \lambda$ and prove that $Y_n \xrightarrow{\mathscr{D}} Y$, where Y is the appropriate Poisson process (with paths in D).

4. Show that Theorem 16.4 still holds if the measure governing the ξ_n is replaced by a probability measure absolutely continuous with respect to it.

5. Prove Theorem 16.4 by an application of Theorem 15.6 with (15.39) in place of (15.21), thus avoiding the representation $\xi_n = \varphi(\eta_n)$.

6. Let P be Lebesgue measure on the unit interval Ω, and define $\xi_n(\omega) = 2\omega_n - 1$, where ω_n is the nth digit in the dyadic expansion of ω. Show that the requirement in Theorem 16.3 that P_0 be dominated by P is essential: The result fails if P_0 is a point mass. It also fails if P_0 is Cantor measure (as defined in Billingsley (1965, p. 36)).

7. Carry Problem 1 of Section 10 over to D. (See also Problem 1 of Section 12.) Generalize the results in Problem 1 of this section.

17. RANDOM CHANGE OF TIME

Sometimes one requires an approximate distribution for a partial sum $S_\nu = \xi_1 + \cdots + \xi_\nu$, where the index ν is itself a random variable. Here we shall prove several functional central limit theorems for such randomly selected partial sums.

Randomly Selected Partial Sums

Suppose the partial sums $S_n = \xi_1 + \cdots + \xi_n$ for a nonrandom index n obey the central limit theorem, say with norming factors $\sigma\sqrt{n}$, so that

$$\frac{1}{\sigma\sqrt{n}}\, S_n \xrightarrow{\mathscr{D}} N \tag{17.1}$$

as $n \to \infty$. If ν is a random integer such that ν is large with high probability, there is some hope that $S_\nu/\sigma\sqrt{\nu}$ will be approximately normally distributed.

To formulate a limit theorem, consider a sequence $\{\nu_n\}$ of random integers. We seek conditions under which

$$\frac{1}{\sigma\sqrt{\nu_n}}\, S_{\nu_n} \xrightarrow{\mathscr{D}} N \tag{17.2}$$

as $n \to \infty$. It is not enough simply to assume that ν_n goes to infinity in probability, in the sense that

$$\mathsf{P}\{\nu_n \le \alpha\} \to 0 \tag{17.3}$$

for each α. For suppose the ξ_n are independent and take the values $+1$ and -1, with probability $\frac{1}{2}$ for each, so that (17.1) holds with $\sigma = 1$. If ν_n is the time of the nth zero crossing (the nth value of k for which $S_k = 0$), then (17.3) holds (because $\nu_n \geq \nu_{n-1} + 1$) but (17.2) does not (because $S_{\nu_n} = 0$).

To derive (17.2) we must assume more than (17.3). Our principal result will be that (17.2) holds if ν_n/n converges in probability to a positive, finite random variable, and if the ξ_n are independent and identically distributed with mean 0 and variance σ^2. We shall derive this result from a functional central limit theorem which gives much more information.

Define a random element X_n of D by

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

and define Y_n by

$$Y_n(t, \omega) = \frac{1}{\sigma\sqrt{n}} S_{\nu_{[nt]}(\omega)}(\omega)$$

(certainly Y_n is a random element of D). Since $Y_n(1) = S_{\nu_n}/\sigma\sqrt{n}$, we shall be in a position to approximate the distribution of S_{ν_n} and derive results like (17.2) if we show that Y_n itself converges in distribution.

The distribution of a random function such as Y_n is most easily investigated by using the fact that Y_n results from subjecting X_n to a random change of time scale. Let $\Phi_n(t, \omega) = \nu_{[nt]}(\omega)/n$. Since

$$Y_n(t, \omega) = X_n(\Phi_n(t, \omega), \omega) \tag{17.4}$$

(we assume for the moment that $\nu_n \leq n$, so that $\Phi_n(t, \omega) \leq 1$), Y_n is X_n with the time scale subjected to a change represented by the random function Φ_n. To clarify the reasoning involved in this section, we first consider such random time changes in a general context. We assume X_n and Φ_n each converge in distribution and look for conditions under which Y_n, as defined by (17.4), converges in distribution.

Random Change of Time

Let D_0 consist of those elements φ of D that are nondecreasing and satisfy $0 \leq \varphi(t) \leq 1$ for all t. Such a φ represents a transformation of the time interval $[0, 1]$. We topologize D_0 by relativizing the Skorohod topology of D. Since $D_0 \in \mathscr{D}$, as is easily shown, the σ-field $\mathscr{D}_0$ of Borel sets in D_0 consists of the subsets of D_0 that lie in $\mathscr{D}$ (see p. 224).

For $x \in D$ and $\varphi \in D_0$, let $x \circ \varphi$ denote the composition of x and φ—the function on $[0, 1]$ whose value at t is

$$(x \circ \varphi)(t) = x(\varphi(t)). \tag{17.5}$$

Now $x \circ \varphi$ lies in D and, if $\psi: D \times D_0 \to D$ is defined by

(17.6) $$\psi(x, \varphi) = x \circ \varphi,$$

then, as is proved on p. 232, ψ is measurable ($\psi^{-1}\mathscr{D} \subset \mathscr{D} \times \mathscr{D}_0$).

Let X be a random element of D and let Φ be a random element of D_0. We assume X and Φ have the same domain, so that (X, Φ) is a random element of $D \times D_0$ with the product topology (see p. 225). If $X \circ \Phi$ has value $X(\omega) \circ \Phi(\omega)$ at ω—that is, if $X \circ \Phi = \psi(X, \Phi)$—then $X \circ \Phi$ is a random element of D; $X \circ \Phi$ results from subjecting X to the time change represented by Φ.

Suppose that, in addition to X and Φ, we have, for each n, random elements X_n and Φ_n of D and D_0, respectively, where X_n and Φ_n have the same domain (which may vary with n). We ask for conditions under which $(X_n, \Phi_n) \xrightarrow{\mathscr{D}} (X, \Phi)$ (convergence in distribution relative to the product topology) implies $X_n \circ \Phi_n \xrightarrow{\mathscr{D}} X \circ \Phi$.

Suppose that

(17.7) $$(X_n, \Phi_n) \xrightarrow{\mathscr{D}} (X, \Phi)$$

and that

(17.8) $$\mathsf{P}\{X \in C\} = \mathsf{P}\{\Phi \in C\} = 1.$$

By Corollary 1 to Theorem 5.1,

(17.9) $$X_n \circ \Phi_n \xrightarrow{\mathscr{D}} X \circ \Phi$$

will follow if we show that ψ (defined by (17.6)) is continuous at (x, φ) for $x \in C$ and $\varphi \in C \cap D_0$. If x_n converges to x and φ_n converges to φ in the Skorohod topology, and if x and φ lie in C, then the convergence in each case is uniform. But

$$\sup_t |x_n(\varphi_n(t)) - x(\varphi(t))| \leq \sup_t |x_n(t) - x(t)| + \sup_t |x(\varphi_n(t)) - x(\varphi(t))|;$$

since x is uniformly continuous, $x_n \circ \varphi_n$ converges to $x \circ \varphi$ uniformly and hence in the Skorohod topology, which proves (17.9).

There remains, of course, the question of when (X_n, Φ_n) converges in distribution, and here Theorems 4.4 and 4.5 can be used.

Applications

We return now to a consideration of sums $S_n = \xi_1 + \cdots + \xi_n$. For each n, let ν_n be a positive-integer-valued random variable defined on the same probability space as the ξ_n. Define X_n by

(17.10) $$X_n(t, \omega) = \frac{1}{\sigma\sqrt{n}} S_{[nt]}(\omega)$$

and Y_n by

$$Y_n(t, \omega) = \frac{1}{\sigma\sqrt{\nu_n(\omega)}} S_{[\nu_n(\omega)t]}(\omega) = X_{\nu_n(\omega)}(t, \omega). \tag{17.11}$$

THEOREM 17.1 *If*

$$\frac{\nu_n}{a_n} \xrightarrow{P} \theta, \tag{17.12}$$

where θ is a positive constant and the a_n are constants going to infinity, then

$$X_n \xrightarrow{\mathscr{D}} W \tag{17.13}$$

implies

$$Y_n \xrightarrow{\mathscr{D}} W. \tag{17.14}$$

Proof. There is no loss of generality in assuming that

$$0 < \theta < 1, \tag{17.15}$$

since this can be arranged by passing to new constants a_n if necessary. Furthermore, there is no loss of generality in assuming that the a_n are integers. Define

$$\Phi_n(t, \omega) = \begin{cases} t\,\dfrac{\nu_n(\omega)}{a_n} & \text{if } \dfrac{\nu_n(\omega)}{a_n} \le 1, \\ t\theta & \text{otherwise.} \end{cases} \tag{17.16}$$

Since

$$\sup_t |\Phi_n(t) - t\theta| \le \left|\frac{\nu_n}{a_n} - \theta\right| \xrightarrow{P} 0, \tag{17.17}$$

Φ_n converges in probability in the sense of the Skorohod topology to the element $\varphi(t) = \theta t$ of D_0. Because of (17.13) and the assumption $a_n \to \infty$, it follows from Theorem 4.4 that $(X_{a_n}, \Phi_n) \xrightarrow{\mathscr{D}} (W, \varphi)$ and hence, since (17.7) and (17.8) imply (17.9), that $X_{a_n} \circ \Phi_n \to W \circ \varphi$.

If

$$Y_n'(t, \omega) = \frac{1}{\sigma\sqrt{a_n}} S_{[\nu_n(\omega)t]}(\omega),$$

then $X_{a_n} \circ \Phi_n$ and Y_n' have the same value at ω if $\nu_n(\omega)/a_n \le 1$, the probability of which goes to 1 by (17.12) and (17.15). Therefore $Y_n' \xrightarrow{\mathscr{D}} W \circ \varphi$. Now

$$\sup_t \left|\frac{1}{\sqrt{\theta}} Y_n'(t, \omega) - Y_n(t, \omega)\right| = \left|\frac{1}{\sqrt{\theta}} - \sqrt{\frac{a_n}{\nu_n(\omega)}}\right| \sup_t |Y_n'(t, \omega)| \xrightarrow{P} 0,$$

and hence Y_n converges in distribution to $\theta^{-\frac{1}{2}}(W \circ \varphi)$. Since $\theta^{-\frac{1}{2}}(W \circ \varphi)$ has the same distribution as W, (17.14) follows.

Of course, (17.14) implies $S_{\nu_n}/\sigma\sqrt{\nu_n} \xrightarrow{\mathscr{D}} N$, as well as an arc sine law and limit laws for the maxima and so on. Note that we have assumed of the ξ_n nothing beyond (17.13), which holds for example in the Lindeberg–Lévy situation. In Chapter 4 we prove (17.13) for various dependent sequences $\{\xi_n\}$, to each of which Theorem 17.1 then applies.

If the limit in (17.12) is not constant, we must make more specific assumptions about the ξ_n.

THEOREM 17.2 *Suppose* $\xi_1, \xi_2, \ldots$ *are independent and identically distributed with* $\mathsf{E}\{\xi_n\} = 0$ *and* $\mathsf{E}\{\xi_n^2\} = \sigma^2$. *If*

$$\frac{\nu_n}{a_n} \xrightarrow{P} \theta, \tag{17.18}$$

where θ *is a positive random variable and the* a_n *are constants going to infinity, then* $Y_n \xrightarrow{\mathscr{D}} W$.

Proof. We assume at first that θ is bounded, so that there exists a constant K such that

$$0 < \theta \leq K$$

with probability 1. We may adjust the a_n so that they are integers and so that $K < 1$.

If we define Φ_n by (17.16), then, as before, $\Phi_n \xrightarrow{P} \Phi$, where Φ is the random element of D_0 defined by

$$\Phi(t) = \theta t.$$

Let $\mathscr{B}_0$ be the field of cylinders and define X'_n by (16.9), where $p_n \to \infty$ and $p_n/\sqrt{n} \to 0$. Then, as in Section 16, we have

$$\sup_t |X_n(t) - X'_n(t)| \xrightarrow{P} 0 \tag{17.19}$$

and

$$\mathsf{P}(\{X'_n \in A\} \cap E) \to W(A)\mathsf{P}(E)$$

for every W-continuity set A and every E in $\mathscr{B}_0$. Now by Theorem 4.4, $(\Phi_n, \nu_n/a_n) \xrightarrow{P} (\Phi, \theta)$ in the sense of the product topology on $D_0 \times R^1$; since every X'_n is measurable $\sigma(\mathscr{B}_0)$, it follows by Theorem 4.5 that

$$\left(X'_{a_n}, \Phi_n, \frac{\nu_n}{a_n}\right) \xrightarrow{\mathscr{D}} (W, \Phi_0, \theta_0)$$

relative to the product topology in $D \times D_0 \times R^1$, where θ_0 is independent of W and $\Phi_0(t) = \theta_0 t$. By (17.19),

$$\left(X_{a_n}, \Phi_n, \frac{\nu_n}{a_n}\right) \xrightarrow{\mathscr{D}} (W, \Phi_0, \theta_0).$$

Now the mapping that carries the point (x, φ, α) to $\alpha^{-\frac{1}{2}}(x \circ \varphi)$ is continuous at that point if $x \in C$, $\varphi \in C \cap D_0$, and $\alpha > 0$. By Corollary 1 to Theorem 5.1, therefore

$$\left(\frac{\nu_n}{a_n}\right)^{-\frac{1}{2}}(X_{a_n} \circ \Phi_n) \xrightarrow{\mathscr{D}} \theta_0^{-\frac{1}{2}}(W \circ \Phi_0).$$

Since θ_0 and W are independent, $\theta_0^{-\frac{1}{2}}(W \circ \Phi_0)$ has the same distribution as W. Moreover $(\nu_n/a_n)^{-\frac{1}{2}}(X_{a_n} \circ \Phi_n)$ coincides with Y_n if $\nu_n/a_n < 1$, the probability of which goes to 1 since $K < 1$. Thus $Y_n \xrightarrow{\mathscr{D}} W$ if θ is bounded.

Suppose θ is not bounded. For $u > 0$, define $\theta_u = \theta$ and $\nu_{un} = \nu_n$ if $\theta \leq u$ and define $\theta_u = u$ and $\nu_{un} = a_n u$ if $\theta > u$. Then, for each u, $\nu_{un}/a_n \xrightarrow{P} \theta_u$ as $n \to \infty$, and, by the case already treated, if

$$Y_{un}(t) = \frac{1}{\sigma\sqrt{\nu_{un}(\omega)}} S_{[\nu_{un}(\omega)t]}(\omega),$$

then $Y_{un} \xrightarrow{\mathscr{D}} W$ as $n \to \infty$. Since $\mathsf{P}\{Y_{un} \neq Y_n\} \leq \mathsf{P}\{\theta > u\}$, $Y_n \xrightarrow{\mathscr{D}} W$ follows by Theorem 4.2.

This proves Theorem 17.2. The proof goes through for many dependent sequences $\{\xi_n\}$ also; see Chapter 4.

Renewal Theory

These ideas can be used to derive a functional central limit theorem connected with renewal theory. Let $\eta_1, \eta_2, \ldots$ be a sequence of positive random variables and define

$$\nu_t = \max\left\{k\colon \sum_{i=1}^{k} \eta_i \leq t\right\}, \qquad t \geq 0, \tag{17.20}$$

with $\nu_t = 0$ if $\eta_1 > t$. If η_k is the time lapse between the occurrences of the $(k - 1)$st and kth events in a series, then ν_t is the number of occurrences up to time t.

We shall assume the existence of positive constants μ and σ such that, if

$$X_n(t, \omega) = \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{[nt]} (\eta_i(\omega) - \mu),$$

then $X_n \xrightarrow{\mathscr{D}} W$. This will be true if, as in the usual renewal theory situation, the η_n are independent and identically distributed with mean μ and variance σ^2. Define Z_n by

$$Z_n(t, \omega) = \frac{\nu_{nt}(\omega) - nt/\mu}{\sigma\mu^{-\frac{3}{2}}\sqrt{n}}.$$

THEOREM 17.3 *If $X_n \xrightarrow{\mathscr{D}} W$, then $Z_n \xrightarrow{\mathscr{D}} W$.*

Proof. We shall assume in the proof that $\mu > 1$; this is only a matter of scale. We first show that

(17.21) $$\sup_{0\le v\le u}\left|\frac{\nu_v}{u}-\frac{1}{\mu}\frac{v}{u}\right|\xrightarrow{P}0$$

as $u\to\infty$. The hypothesis $X_n\xrightarrow{\mathscr{D}}W$ implies that

(17.22) $$\sup_{0\le t\le s}\frac{1}{s}\left|\sum_{i=1}^{[t]}\eta_i-\mu t\right|\xrightarrow{P}0$$

as $s\to\infty$ (because replacing s^{-1} by $s^{-\frac{1}{2}}$ would give convergence in distribution). By the definition (17.20), $\nu_v > t$ implies $\Sigma_{i=1}^{[t]}\eta_i\le v$. Therefore

$$\sup_{0\le v\le u}\left(\frac{\nu_v}{u}-\frac{1}{\mu}\frac{v}{u}\right)>\varepsilon$$

implies

(17.23) $$\sup_{0\le t\le u(\mu^{-1}+\varepsilon)}\left|\sum_{i=1}^{[t]}\eta_i-\mu t\right|\ge\mu u\varepsilon.$$

Similarly,

$$\inf_{0\le v\le u}\left(\frac{\nu_v}{u}-\frac{1}{\mu}\frac{v}{u}\right)<-\varepsilon$$

implies (if $\varepsilon<\mu^{-1}$)

(17.24) $$\sup_{0\le t\le u(\mu^{-1}-\varepsilon)}\left|\sum_{i=1}^{[t]}\eta_i-\mu t\right|\ge\mu u\varepsilon.$$

By (17.22), the probabilities of (17.23) and (17.24) go to 0 as $u\to\infty$, which proves (17.21).

Put

$$\Phi_n(t,\omega)=\begin{cases}\dfrac{\nu_{tn}(\omega)}{n} & \text{if } \dfrac{\nu_n(\omega)}{n}\le 1,\\[2ex] \dfrac{t}{\mu} & \text{otherwise.}\end{cases}$$

By (17.21), $\Phi_n\xrightarrow{P}\varphi$, where $\varphi(t)=t/\mu$, so that, by the usual argument, $X_n\circ\Phi_n\xrightarrow{\mathscr{D}}W\circ\varphi$. Let

$$Y_n(t)=\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{\nu_{tn}(\omega)}(\eta_i(\omega)-\mu);$$

$Y_n=X_n\circ\Phi_n$ if $\nu_n/n\le 1$, the probability of which goes to 1 by (17.21) and the assumption $\mu>1$. Therefore $Y_n\xrightarrow{\mathscr{D}}W\circ\varphi$.

By the definition (17.20),

$$Y_n(t)\le\frac{nt-\mu\nu_{nt}}{\sigma\sqrt{n}}\le Y_n(t)+\frac{\eta_{\nu_{nt}+1}}{\sigma\sqrt{n}}.$$

From $\max_{i\le n}|\eta_i|/\sqrt{n}\xrightarrow{P}0$ it follows that $\sup_{t\le 1}|\eta_{\nu_{nt}}|/\sigma\sqrt{n}$, which in turn implies that, if

$$Z_n'(t) = \frac{nt - \mu v_{nt}}{\sigma\sqrt{n}},$$

then $Z_n' \xrightarrow{\mathscr{D}} W \circ \varphi$. Therefore $\mu^{\frac{1}{2}} Z_n' \xrightarrow{\mathscr{D}} W$ (recall that $\varphi(t) = t/\mu$), from which $Z_n \xrightarrow{\mathscr{D}} W$ follows because of the symmetry of W.

It is possible in Theorem 17.3 to let n tend to infinity in a continuous manner.

Remarks. The results of this section, which are new, extend those in Billingsley (1962). For the central limit theorem for a random number of random variables, see Rényi (1960) and Blum, Hanson, and Rosenblatt (1963). See Lamperti (1962) for another weak-convergence theorem in renewal theory.

18. THE UNIFORM TOPOLOGY

Let us consider briefly the uniform topology on D—the topology given by the uniform metric

$$\rho(x, y) = \sup_t |x(t) - y(t)|. \tag{18.1}$$

With this metric, D is a complete metric space. It is, however, not separable: The uncountably many points x_θ defined for $0 < \theta < 1$ by

$$x_\theta(t) = \begin{cases} 0 & \text{if } 0 \le t < \theta, \\ 1 & \text{if } \theta \le t \le 1, \end{cases} \tag{18.2}$$

are at mutual distance 1 (see p. 216).

A concept prefixed with U will refer to the uniform topology: U-open, U-continuous, etc. In the same way, a prefix S will refer to the Skorohod topology of D. If d is either of the two metrics that give the Skorohod topology (see Section 14), and if ρ is the uniform metric (18.1) then $d(x, y) \le \rho(x, y)$. Therefore the uniform topology is finer than the Skorohod topology: U-convergence implies S-convergence. On the other hand, as observed in Section 14, if x_n S-converges to x and $x \in C$, then x_n U-converges to x. In particular, the uniform and Skorohod topologies coincide when relativized to C.

Let $\mathscr{D}$ denote the class of S-Borel sets as usual, and let $\mathscr{U}$ denote the class of U-Borel sets. Since an S-open set is also U-open,

$$\mathscr{D} \subset \mathscr{U}. \tag{18.3}$$

Let us suppose that P_n and P are probability measures defined on $\mathscr{U}$, the *larger* of $\mathscr{U}$ and $\mathscr{D}$. Let us write $P_n \Rightarrow P[U]$ to indicate weak convergence in the sense of the uniform topology and $P_n \Rightarrow P[S]$ to indicate weak convergence in the sense of the Skorohod topology. Now $P_n \Rightarrow P[U]$ requires

$$\int f\,dP_n \to \int f\,dP \tag{18.4}$$

for all bounded, U-continuous f, whereas $P_n \Rightarrow P[S]$ requires (18.4) only for all bounded, S-continuous f. Hence $P_n \Rightarrow P[U]$ implies $P_n \Rightarrow P[S]$. The converse is false because S-convergence does not imply U-convergence (consider point masses).

Suppose now that P_n and P are defined on $\mathscr{U}$, that $P_n \Rightarrow P[S]$, and that $P(C) = 1$. We shall prove that $P_n \Rightarrow P[U]$. For A in $\mathscr{U}$, let A_U denote its U-closure and let A_S denote its S-closure. Note that $A_U \subset A_S$. Since $P_n \Rightarrow P[S]$ and $P(C) = 1$, $\limsup_n P_n(A) \leq \limsup_n P_n(A_S) \leq P(A_S) = P(A_S \cap C)$. Since a sequence S-converging to a limit in C also U-converges to the limit, $A_S \cap C \subset A_U$. Therefore $\limsup_n P_n(A) \leq P(A_U)$, which implies $P_n \Rightarrow P[U]$ by Theorem 2.1.

We have proved this: *Suppose P_n and P are defined on $\mathscr{U}$.* (i) *If $P_n \Rightarrow P[U]$, then $P_n \Rightarrow P[S]$.* (ii) *If $P_n \Rightarrow P[S]$ and $P(C) = 1$, then $P_n \Rightarrow P[U]$.*† Because of (i), it is better if possible to prove $P_n \Rightarrow P[U]$ than to prove $P_n \Rightarrow P[S]$—for example, one has then the asymptotic distributions for more functions on D. Because of (ii) on the other hand, if $P(C) = 1$, then a proof of $P_n \Rightarrow P[S]$ is automatically a proof of the better result $P_n \Rightarrow P[U]$.

Consider the distributions P_n of the random functions X_n involved in Donsker's theorem as formulated for D (see (16.1)). According to Theorem 16.1, $P_n \Rightarrow W[S]$, where W is Wiener measure. Since $W(C) = 1$, we can draw the stronger conclusion $P_n \Rightarrow W[U]$.

There is a gap in this argument: the measures P_n and W are defined only on $\mathscr{D}$. The question is whether they can be extended to the larger σ-field $\mathscr{U}$. Certainly, W can be extended—the argument used in Section 16 to extend W from $(C, \mathscr{C})$ to $(D, \mathscr{D})$ enables us also to extend it from $(C, \mathscr{C})$ to $(D, \mathscr{U})$. Now P_n is defined as $\mathsf{P}X_n^{-1}$ (the domain of X_n is a probability space $(\Omega, \mathscr{B}, \mathsf{P})$), a definition which is possible because

$$X_n^{-1}\mathscr{D} \subset \mathscr{B}. \tag{18.5}$$

If we are to define P_n on $\mathscr{U}$ instead, we must strengthen (18.5) to

$$X_n^{-1}\mathscr{U} \subset \mathscr{B}. \tag{18.6}$$

If A is a finite-dimensional set in D, then $X_n^{-1}A \in \mathscr{B}$, and it follows that the same is true if A is a U-sphere. Since the uniform topology is not separable, (18.6) does not follow without further argument.

† These statements are true because U-convergence always implies S-convergence and S-convergence to a limit in C implies U-convergence; they involve no further properties of C, D, and the two topologies.

The definition (16.1) of X_n shows that the range $X_n\Omega$ is separable in the uniform topology. If A is U-open, therefore, there exist countably many U-spheres A_i such that $A \cap X_n\Omega = \bigcup_i (A_i \cap X_n\Omega)$. But then $X_n^{-1}A = \bigcup_i X_n^{-1}A_i$, and, since each $X_n^{-1}A_i$ lies in $\mathscr{B}$, so does $X_n^{-1}A$, which proves (18.6).

Thus the distribution P_n is (or can be) defined on $\mathscr{U}$ in the natural way by

$$P_n(A) = \mathsf{P}\{\omega : X_n(\omega) \in A\}, \qquad A \in \mathscr{U},$$

and we can strengthen Theorem 16.1 by interpreting $X_n \xrightarrow{\mathscr{D}} W$ in the uniform topology: $P_n \Rightarrow W[U]$.

Consider next the random functions Y_n defined by

$$Y_n(\omega, t) = \sqrt{n}\big(F_n(t, \omega) - t\big),$$

where $F_n(t, \omega)$ is the empirical distribution function for $\xi_1(\omega), \ldots, \xi_n(\omega)$, assumed independent and uniformly distributed over the unit interval. Now $Y_n \xrightarrow{\mathscr{D}} W^\circ$ by Theorem 16.4. Since W° can be extended to $\mathscr{U}$ just as W was, we can interpret $Y_n \xrightarrow{\mathscr{D}} W^\circ$ in the uniform topology rather than the Skorohod topology if we can extend the distribution of Y_n from $\mathscr{D}$ to $\mathscr{U}$—that is, if we can strengthen

$$Y_n^{-1}\mathscr{D} \subset \mathscr{B} \tag{18.7}$$

to

$$Y_n^{-1}\mathscr{U} \subset \mathscr{B}. \tag{18.8}$$

But here the program collapses: We shall show that (18.8) is false.

To see that (18.8) fails, consider the case $n = 1$. Let h be the mapping from Ω to D that carries ω to the distribution function with a unit jump at $\xi_1(\omega)$; $h(\omega) = x_{\xi_1(\omega)}$ with x_θ defined by (18.2). Clearly, (18.8) is equivalent to

$$h^{-1}\mathscr{U} \subset \mathscr{B}, \tag{18.9}$$

and we shall show that (18.9) is false if $\xi_1(\omega)$ is uniformly distributed.

Let A_θ be the U-sphere with center x_θ and radius $\frac{1}{2}$. Since the x_θ are at mutual distance 1 in the uniform topology, we have

$$h^{-1}\left(\bigcup_{\theta \in H} A_\theta\right) = \{\omega : \xi_1(\omega) \in H\} \tag{18.10}$$

for every subset H of the unit interval. If (18.9) were true, then, since $\bigcup_{\theta \in H} A_\theta$ is U-open, the set (18.10) would lie in $\mathscr{B}$, so that we would have $\{\omega : \xi_1(\omega) \in H\} \in \mathscr{B}$ for *every* subset H of the unit interval. Thus $\mu = \mathsf{P}\xi_1^{-1}$ would be a measure defined for every subset of the unit interval and having the property that $\mu(a, b) = b - a$ for every subinterval (a, b). But no such μ exists.

A small elaboration of these ideas shows that (18.8) fails for every n unless the distribution common to the ξ_i is purely atomic. In treating empirical distribution functions as random elements of D, therefore, we must stay with the Skorohod topology—and this is solely for reasons of measurability.† As a rule, we may interpret $X_n \xrightarrow{\mathcal{D}} X$ in the uniform topology only if (i) X lies in C with probability 1 and (ii) the jumps in X_n occur at fixed time points rather than at time points with random positions.

Remarks. Chibisov (1965) has pointed out that (18.8) is false. Another way around the problems created by the nonseparability of D in the uniform topology is to work within the σ-field $\mathcal{U}_0$ generated by the U-spheres. This requires a different theory of weak convergence; see Dudley (1966 and 1967).

---

† O, wiste a man how manye maladyes
Folwen of excesse and of glotonyes,
He wolde been the moore mesurable . . .
The Pardoner's Tale

CHAPTER 4

Dependent Variables

19. DIFFUSION

This section contains a theorem characterizing Brownian motion, a theorem characterizing more general diffusion processes, and asymptotic forms of these two theorems. In Section 20 we use these results to prove a version of Donsker's theorem for the variables of a stationary sequence satisfying a uniform mixing condition; in Section 21 we extend the result to functions of such sequences; and in Section 22 we prove functional central limit theorems for empirical distribution functions in these two cases. In Sections 23 and 24 the results of the present section are used to derive limit theorems for martingales and for exchangeable random variables.†

Although many of the results of this chapter could be formulated and proved in the space C, we shall consistently use D with the Skorohod topology.

A Characterization of Brownian Motion

The random function W lies in C with probability 1 and has independent increments; moreover, $\mathsf{E}\{W_t\} = 0$ and $\mathsf{E}\{W_t^2\} = t$. These properties characterize W:

THEOREM 19.1 *Let X be a random element of D that has independent increments and satisfies* $\mathsf{P}\{X \in C\} = 1$, $\mathsf{E}\{X(t)\} = 0$, *and* $\mathsf{E}\{X^2(t)\} = t$. *Then X is distributed as W.*

† The succeeding sections all depend upon this one. Sections 20, 21, and 22 require to be read in order, but this sequence and Sections 23 and 24 form three independent units.

Proof. The independence and the moment conditions imply

$$\mathsf{E}\{(X(t) - X(s))^2\} = |t - s|$$

and $\mathsf{P}\{X(0) = 0\} = 1$. We are to show that $X(t) - X(s)$ has distribution $N(0, t - s)$ for $s \leq t$. Because of the independence, it is enough to prove this for $s = 0$, and by continuity we may assume $t < 1$.

Let $\varphi(t, \cdot)$ be the characteristic function of $X(t)$:

$$\varphi(t, u) = \mathsf{E}\{e^{iuX(t)}\}.$$

Since $\mathsf{P}\{X \in C\} = 1$, $\varphi(t, u)$ is continuous in t as well as in u. We shall show that it satisfies the differential equation

$$\frac{\partial}{\partial t}\varphi(t, u) = -\tfrac{1}{2}u^2\varphi(t, u), \qquad 0 \leq t < 1, \quad u \in R^1. \tag{19.1}$$

It will then follow† that

$$\varphi(t, u) = \varphi(0, u)e^{-\frac{1}{2}tu^2};$$

since $X(0)$ vanishes with probability 1, this will entail normality. In proving (19.1), we may take the partial derivative to be a derivative from the right.‡ We are thus to prove

$$\lim_{h \downarrow 0} \frac{1}{h}[\varphi(t + h, u) - \varphi(t, u)] = -\tfrac{1}{2}u^2\varphi(t, u), \qquad 0 \leq t < 1. \tag{19.2}$$

By a standard estimate, §

$$e^{iv} = 1 + iv - \tfrac{1}{2}v^2 + c(v), \tag{19.3}$$

with

$$|c(v)| \leq |v|^3 \tag{19.4}$$

for v real. Write

$$\Delta_{s,t} = \Delta(s, t) = X(t) - X(s). \tag{19.5}$$

From the independence of these increments and the relations $\mathsf{E}\{\Delta_{s,t}\} = 0$

---

† If f satisfies $f'(t) = A(t)f(t)$ with A continuous, then its ratio with the nonvanishing function $f_0(t) = \exp \int_0^t A(\tau)\, d\tau$ has derivative 0, so that $f(t)/f_0(t) = f(0)/f_0(0) = f(0)$.

‡ To prove that a continuous f with continuous right-hand derivative f^+ on $[0, 1)$ necessarily has a two-sided derivative on $(0, 1)$, it suffices to assume f real and prove that $F(t) = f(t) - f(0) - \int_0^t f^+(\tau)\, d\tau$ vanishes identically. If (say) $F(t_0) < 0$, then $G(t) = F(t) - tF(t_0)/t_0$ satisfies $G(0) = G(t_0) = 0$ and (since F^+ vanishes) $G^+(t) > 0$, so that, over $[0, t_0]$, G must have a positive maximum at some interior point s; but then $G^+(s) \leq 0$, a contradiction.

§ Feller (1966, p. 485).

and $\mathsf{E}\{\Delta_{s,t}^2\} = t - s$, we have

$$\begin{aligned}\varphi(t+h, u) - \varphi(t, u) &= \mathsf{E}\{e^{iuX(t)}[e^{iu\Delta(t,t+h)} - 1]\} \\ &= \mathsf{E}\{e^{iuX(t)}[iu\Delta_{t,t+h} - \tfrac{1}{2}u^2\Delta_{t,t+h}^2 + c(u\Delta_{t,t+h})\} \\ &= -\tfrac{1}{2}u^2h\varphi(u, t) + \mathsf{E}\{e^{iuX(t)}c(u\Delta_{t,t+h})\}.\end{aligned}$$

Because of (19.4), (19.2) will follow if we prove that

$$\lim_{h \downarrow 0} \frac{1}{h} \mathsf{E}\{|\Delta_{t,t+h}|^3\} = 0. \tag{19.6}$$

Using again the independence of the increments, we obtain

$$\mathsf{E}\{\Delta_{t_1,t}^2\Delta_{t,t_2}^2\} = (t - t_1)(t_2 - t) \le (t_2 - t_1)^2, \qquad t_1 \le t \le t_2. \tag{19.7}$$

Fix t and $t + h$, with $h > 0$. If

$$M'_m = \max_{0 \le i \le m} \min\left\{\left|\Delta\left(t, t + \frac{i}{m}h\right)\right|, \left|\Delta\left(t + \frac{i}{m}h, t + h\right)\right|\right\}, \tag{19.8}$$

then, by (19.7) and Theorem 12.1,

$$\mathsf{P}\{M'_m \ge \lambda\} \le K_{2,1}\frac{h^2}{\lambda^4}. \tag{19.9}$$

And, by (12.6),

$$|\Delta_{t,t+h}| \le 3M'_m + \max_{1 \le i \le m}\left|\Delta\left(t + \frac{i-1}{m}h, t + \frac{i}{m}h\right)\right|. \tag{19.10}$$

Since X lies in C with probability 1, the maximum here goes to 0 with probability 1 as $m \to \infty$, and it follows by (19.9) that

$$\mathsf{P}\{|\Delta_{t,t+h}| \ge \lambda\} \le K\frac{h^2}{\lambda^4} \tag{19.11}$$

with $K = 4^4K_{2,1}$.

From (19.11) we obtain (see (3) on p. 223)

$$\mathsf{E}\{|\Delta_{t,t+h}|^3\} \le a + \int_a^\infty \frac{Kh^2}{\alpha^{\frac{4}{3}}}\,d\alpha = a + 3Kh^2\frac{1}{a^{\frac{1}{3}}},$$

valid for $a > 0$. Taking $a = K^{\frac{3}{4}}h^{\frac{3}{2}}$ (this minimizes the function on the right), we obtain

$$\mathsf{E}\{|\Delta_{t,t+h}|^3\} \le 4K^{\frac{3}{4}}h^{\frac{3}{2}},$$

which implies (19.6) and completes the proof.

The condition that X lie in C with probability 1 is essential here, since the other hypotheses of the theorem are satisfied if

$$X(t) = Y(t) - t \tag{19.12}$$

and Y is distributed as a Poisson process with parameter 1.

We now derive a limit theorem from Theorem 19.1. Let X_n be random elements of D. We say X_n has *asymptotically independent increments* if

$$0 \le s_1 \le t_1 < s_2 \le t_2 < \cdots < s_r \le t_r \le 1 \tag{19.13}$$

implies, for all linear Borel sets $H_1, \ldots, H_r$, that the difference

$$\mathsf{P}\{X_n(t_i) - X_n(s_i) \in H_i,\ i = 1, \ldots, r\} - \prod_{i=1}^{r} \mathsf{P}\{X_n(t_i) - X_n(s_i) \in H_i\} \tag{19.14}$$

converges to 0 as $n \to \infty$. Note that we require (19.14) to hold only when the intervals $[s_i, t_i]$ are strictly separated—when $t_{i-1} < s_i$.

Recall that the modulus of continuity $w_x(\delta) = w(x, \delta)$, defined by (8.1), is well defined if x lies in D. It enters into the next theorem because the limiting distribution on D is actually supported by C.

THEOREM 19.2 *Suppose that X_n has asymptotically independent increments, that $\{X_n^2(t): n \ge 1\}$ is uniformly integrable for each t, and that $\mathsf{E}\{X_n(t)\} \to 0$ and $\mathsf{E}\{X_n^2(t)\} \to t$ as $n \to \infty$. Suppose finally that, for each positive ε and η, there exists a positive δ such that*

$$\mathsf{P}\{w(X_n, \delta) \ge \varepsilon\} \le \eta \tag{19.15}$$

for all sufficiently large n. Then $X_n \xrightarrow{\mathscr{D}} W$.

Proof. Since $\mathsf{E}\{X_n^2(0)\} \to 0$, $\{X_n(0)\}$ is tight. It follows from (19.15) and Theorem 15.5 that $\{X_n\}$ is tight and that, if X is the limit in distribution of a subsequence, then $\mathsf{P}\{X \in C\} = 1$. It is enough to show that any such X must be distributed as W. But $\mathsf{E}\{X_n(t)\} \to 0$, $\mathsf{E}\{X_n^2(t)\} \to t$, and the uniform integrability of $\{X_n^2(t): n \ge 1\}$ imply $\mathsf{E}\{X(t)\} = 0$ and $\mathsf{E}\{X^2(t)\} = t$ (Theorem 5.4). Now the increments $X(t_i) - X(s_i)$, $i = 1, \ldots, r$, are independent when (19.13) holds, and from $\mathsf{P}\{X \in C\} = 1$ it follows that the same is true if we allow $t_i = s_{i+1}$ in (19.13). Thus the result follows by Theorem 19.1.

It is not possible to replace the w in (19.15) by w' as defined by (14.6) (define $X_n \equiv X$ by (19.12)).

The condition that $\{X_n^2(t): n \ge 1\}$ is uniformly integrable for each t cannot be dispensed with in this theorem; indeed the condition follows via Theorem 5.4 from $\mathsf{E}\{X_n^2(t)\} \to t$ and $X_n \xrightarrow{\mathscr{D}} W$. For a specific example in which the condition fails, take $X_n(t) = \sqrt{t}\,\xi_n$, where ξ_n assumes the values $\sqrt{n}$, $-\sqrt{n}$, and 0 with respective probabilities $1/2n$, $1/2n$, and $1 - 1/n$; although X_n does not converge in distribution to W, the hypotheses of the theorem are satisfied except for the requirement of uniform integrability.

Theorem 19.2, together with the results of Section 12, leads to another proof of Donsker's theorem. If X_n is the random function (10.1) involved in that theorem, then, by (12.23) and Theorem 8.4 (carried over to D), the condition involving (19.15) is satisfied. Further, by (12.20),

$$\int_{\{S_n^2/n\sigma^2 \geq a\}} S_n^2/n \, d\mathsf{P} \leq K'\left[\frac{1}{a} + \frac{1}{\sigma^2}\int_{\{|\xi_1| \geq \frac{1}{4}a\sigma\sqrt{n}\}} \xi_1^2 \, d\mathsf{P}\right]$$

for a universal constant K', which implies $\{X_n^2(t): n \geq 1\}$ uniformly integrable. The remaining conditions in Theorem 19.2 are easily checked, and $X_n \xrightarrow{\mathscr{D}} W$ follows. Notice that this proof does not presuppose the central limit theorem; in particular, we have an independent proof of the Lindeberg–Lévy theorem.† As we shall see in Section 20, this method applies also to certain sequences of dependent random variables.

Other Diffusion Processes‡

If X satisfies the hypotheses of Theorem 19.1, then, for $0 \leq t_1 \leq \cdots \leq t_k < 1$ and h positive,

$$\text{(19.16)} \qquad \begin{cases} \mathsf{E}\{X(t_k + h) - X(t_k) \,\|\, X(t_1), \ldots, X(t_k)\} = 0, \\ \mathsf{E}\{(X(t_k + h) - X(t_k))^2 \,\|\, X(t_1), \ldots, X(t_k)\} = h. \end{cases}$$

If $X \in C$ and $X(0) = 0$ with probability 1, that X is Brownian motion follows, as we shall see, from (19.16), and we can dispense with the assumption of independent increments. The same is true if, in addition to certain regularity assumptions, we require only that, for small h, the equations in (19.16) hold *approximately*, in a sense presently to be made precise.

We can in a similar way characterize diffusion processes other than Brownian motion by replacing the right sides of the equations in (19.16) by functions of h, t_k, and $X(t_k)$. We shall assume that

$$\text{(19.17)} \qquad \mathsf{E}\{X(t_k + h) - X(t_k) \,\|\, X(t_1), \ldots, X(t_k)\} \approx h\rho(t_k)X(t_k)$$

and

$$\text{(19.18)} \qquad \mathsf{E}\{(X(t_k + h) - X(t_k))^2 \,\|\, X(t_1), \ldots, X(t_k)\} \approx h\sigma^2(t_k)$$

for small h (the meaning to be made exact), where $\rho(t)$ and $\sigma^2(t)$ are given functions, and prove (Theorem 19.3 below) that X is a Gaussian random function satisfying

$$\text{(19.19)} \qquad \mathsf{E}\{X(t)\} = 0, \qquad 0 \leq t \leq 1,$$

† The argument is easily adapted to the Lindeberg case; see Problem 1 of Section 10 and Problem 1 of Section 12.

‡ The remainder of this section is required only for Sections 23 and 24.

and

$$\text{(19.20)}\quad \mathsf{E}\{X(s)\,X(t)\} = \int_0^s \sigma^2(r)\exp\left[2\int_r^s \rho(\tau)\,d\tau + \int_s^t \rho(\tau)\,d\tau\right] dr,$$

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

The choice $\rho(t) = 0$ and $\sigma^2(t) = 1$ leads back to Brownian motion, and the choice $\rho(t) = -1/(1 - t)$ and $\sigma^2(t) = 1$ leads to the Brownian bridge.

Of the functions $\sigma^2(t)$ and $\rho(t)$ we shall assume that $\sigma^2(t)$ is nonnegative, that both are continuous on $[0, 1)$, and that, if

$$g(t) = \exp\left[\int_0^t \rho(\tau)\,d\tau\right]$$

and

$$G(t) = \int_0^t \sigma^2(r) g^{-2}(r)\,dr,$$

then the limit

$$\text{(19.21)}\qquad \lim_{t\to 1} g(t)G(t)$$

exists and is finite. As $t \to 1$, the nondecreasing function $G(t)$ has a limit (finite or infinite), and hence $g(t)$ has a finite limit (which vanishes if $G \to \infty$). Thus the right side of (19.20), which is well defined for $0 \le s \le t < 1$, has as $t \to 1$ a limit which we shall take as its value for $t = 1$.

If we define

$$X(t) = g(t)(1 + G(t))W^\circ(G(t)/(1 + G(t)))$$

for $0 \le t < 1$, where W° is the Brownian bridge, then (19.20) holds for $0 \le s \le t < 1$. Since $g(t)$, $g(t)G(t)$, and $G(t)/(1 + G(t))$ all have limits as $t \to 1$, we can define $X(1)$ by passing to the limit. Thus, there does exist a continuous, Gaussian random function satisfying (19.19) and (19.20). Clearly $X(0) = 0$ with probability 1. We shall regard X as a random element of D lying with probability 1 in C.

We shall give three conditions on X which characterize it as this random function. Each condition will be given in two versions, first in a form convenient for most applications and second in a distinctly weaker form that suffices for the characterization. The first condition defines the sense in which the approximations (19.17) and (19.18) are to hold.

Condition 1. *If* $0 \le t_1 \le \cdots \le t_k < 1$, *then*

$$\text{(19.22)}\quad \lim_{h\downarrow 0}\frac{1}{h}\mathsf{E}\{|\mathsf{E}\{X(t_k + h) - X(t_k) \,\|\, X(t_1), \ldots, X(t_k)\} - h\rho(t_k)X(t_k)|\} = 0$$

and

$$\text{(19.23)}\quad \lim_{h\downarrow 0}\frac{1}{h}\mathsf{E}\{|\mathsf{E}\{(X(t_k + h) - X(t_k))^2 \,\|\, X(t_1), \ldots, X(t_k)\} - h\sigma^2(t_k)|\} = 0.$$

Condition 1a. *If* $0 \le t_1 \le \cdots \le t_k < 1$, *then for all real* $u_1, \ldots, u_k$,

$$(19.24)\quad \lim_{h \downarrow 0} \frac{1}{h} \mathsf{E}\left\{\left[\exp \sum_{j=1}^{k} iu_j X(t_j)\right][X(t_k + h) - X(t_k) - h\rho(t_k)X(t_k)]\right\} = 0$$

and

$$(19.25)\quad \lim_{h \downarrow 0} \frac{1}{h} \mathsf{E}\left\{\left[\exp \sum_{j=1}^{k} iu_j X(t_j)\right][(X(t_k + h) - X(t_k))^2 - h\sigma^2(t_k)]\right\} = 0.$$

Condition 2. *We have*

$$(19.26)\qquad \sup_t \mathsf{E}\{X^2(t)\} < \infty.$$

Condition 2a. *The variables* $X(t)$, $0 \le t \le 1$, *are uniformly integrable.*
Condition 3. *There is a constant* K *such that*

$$(19.27)\quad \mathsf{E}\{(X(t) - X(t_1))^2(X(t_2) - X(t))^2\} \le K(t_2 - t_1)^2, \qquad t_1 \le t \le t_2.$$

Condition 3a. *For* $t < 1$,

$$(19.28)\quad \lim_{\alpha \to \infty} \limsup_{h \downarrow 0} \frac{1}{h} \int_{\{(X(t+h)-X(t))^2 \ge \alpha h\}} (X(t + h) - X(t))^2 \, d\mathsf{P} = 0.$$

Clearly, Condition 1 implies 1a and 2 implies 2a. To see that 3 implies 3a (which is essentially a uniform integrability condition), observe that, with the notation (19.5), the inequality (19.27) is just (19.7) with an extra factor of K on the right, and, if X lies in C with probability 1, we may deduce (19.11) just as before (with a new K—the value is irrelevant). An application of (3) on p. 223 shows that the integral in (19.28) cannot exceed $2Kh/\alpha$, so that 3a does follow from 3. (For conditions intermediate between 2 and 2a and between 3 and 3a, replace the exponent 2 on the left in (19.26) and (19.27) by $1 + \varepsilon$.)

THEOREM 19.3 *Let* X *be a random element of* D *with* $\mathsf{P}\{X \in C\} = 1$ *and* $\mathsf{P}\{X(0) = 0\} = 1$. *Suppose that* $\sigma^2(t)$ *and* $\rho(t)$ *are continuous on* $[0,1)$ *and there exists the finite limit* (19.21). *If* X *satisfies Conditions* 1 (*or* 1a), 2 (*or* 2a), *and* 3 (*or* 3a), *then* X *is the continuous, Gaussian random function specified by* (19.19) *and* (19.20).

Before proving the theorem, let us connect it with our two familiar examples. If

$$(19.29)\qquad \rho(t) = 0, \qquad \sigma^2(t) = 1,$$

then (19.20) becomes $\mathsf{E}\{X(s)\,X(t)\} = s$, $s \le t$, so that X is Brownian motion. If

$$(19.30)\qquad \mathsf{E}\{X(t_k + h) - X(t_k) \,\|\, X(t_1), \ldots, X(t_k)\} = 0$$

and

$$\mathsf{E}\{(X(t_k+h)-X(t_k))^2 \,\|\, X(t_1),\ldots,X(t_k)\} = h \tag{19.31}$$

with probability 1, then (19.22) and (19.23) are certainly satisfied, (19.26) holds, and (19.27) holds (even with $(t-t_1)(t_2-t)$ on the right). Hence (if $X \in C$ and $X(0)=0$ with probability 1) X must be Brownian motion. Since the hypotheses of Theorem 19.1 imply (19.30) and (19.31), that theorem is a corollary of the present one.

If

$$\rho(t) = -\frac{1}{1-t}, \qquad \sigma^2(t) = 1, \tag{19.32}$$

then (19.20) becomes $\mathsf{E}\{X(s)X(t)\} = s(1-t)$, $s \le t$, and X must be distributed as the Brownian bridge W°.

Proof of the theorem. Fix points t_i with $0 \le t_1 \le \cdots \le t_k < 1$, fix real numbers $u_1,\ldots,u_k$, and let t and u vary over the strip

$$t_k \le t < 1, \qquad -\infty < u < \infty. \tag{19.33}$$

Put

$$\psi(t,u) = \mathsf{E}\{\exp i(u_1 X(t_1) + \cdots + u_k X(t_k) + u\,X(t))\}; \tag{19.34}$$

we shall derive a differential equation for $\psi(t,u)$.

For notational convenience, write $Z = u_1 X(t_1) + \cdots + u_k X(t_k)$, so that (19.34) becomes

$$\psi(t,u) = \mathsf{E}\{e^{iZ}e^{iuX(t)}\}. \tag{19.35}$$

Since $X \in C$ with probability 1, ψ is continuous (jointly in t and u) in the strip (19.33). Since $\mathsf{E}\{X(t)\}$ exists, we have

$$\frac{\partial}{\partial u}\psi(t,u) = \mathsf{E}\{iX(t)e^{iZ}e^{iuX(t)}\}, \tag{19.36}$$

and Condition 2a implies that this function also is continuous in the strip (19.33).

With the notations (19.3) and (19.5), we have

$$\begin{aligned}\psi(t+h,u)-\psi(t,u) &= \mathsf{E}\{e^{iZ}e^{iuX(t)}[e^{iu\Delta(t,t+h)}-1]\}\\ &= \mathsf{E}\{e^{iZ}e^{iuX(t)}[iu\Delta_{t,t+h} - \tfrac12 u^2\Delta_{t,t+h}^2 + c(u\Delta_{t,t+h})]\}.\end{aligned}$$

By (19.35) and (19.36),

$$\begin{aligned}&\left|\frac{1}{h}[\psi(t+h,u)-\psi(t,u)] - u\rho(t)\frac{\partial}{\partial u}\psi(t,u) + \tfrac12 u^2\sigma^2(t)\psi(t,u)\right| \\ &\quad\le \frac{|u|}{h}\left|\mathsf{E}\{e^{iZ}e^{iuX(t)}[\Delta_{t,t+h} - h\rho(t)X(t)]\}\right| \\ &\qquad + \frac{u^2}{2h}\left|\mathsf{E}\{e^{iZ}e^{iuX(t)}[\Delta_{t,t+h}^2 - h\sigma^2(t)]\}\right| + \frac{1}{h}\mathsf{E}\{|c(u\Delta_{t,t+h})|\}.\end{aligned} \tag{19.37}$$

Since $|e^{iv} - 1 - iv| \leq \frac{1}{2}v^2$, we have, in addition to (19.4), the inequality $|c(v)| \leq v^2$. Therefore

$$\frac{1}{h}\,\mathsf{E}\{|c(u\Delta_{t,t+h})|\} \leq |u|^3\alpha^{\frac{3}{2}}h^{\frac{1}{2}} + \frac{u^2}{h}\int_{\{\Delta_{t}^{2},_{t+h} > \alpha h\}} \Delta_{t,t+h}^2\, d\mathsf{P},$$

and it follows by Condition 3a that the third term on the right in (19.37) tends to 0 with h. And by Condition 1a (with $(t_1, \ldots, t_k, t)$ replacing $(t_1, \ldots, t_k)$ and $(u_1, \ldots, u_k, u)$ replacing $(u_1, \ldots, u_k)$), the other two terms also tend to 0. Therefore

$$\text{(19.38)} \qquad \frac{\partial}{\partial t}\,\psi(t, u) = u\rho(t)\frac{\partial}{\partial u}\,\psi(t, u) - \tfrac{1}{2}u^2\sigma^2(t)\psi(t, u);$$

since the right side of this equation is continuous in the strip (19.33), so is the left.†

We now solve the differential equation (19.38). For v arbitrary, define

$$\text{(19.39)} \qquad \lambda_v(s) = v\exp\left[-\int_{t_k}^{s}\rho(\tau)\,d\tau\right], \qquad t_k \leq s < 1,$$

and put

$$\text{(19.40)} \qquad g_v(s) = \psi(s, \lambda_v(s)), \qquad t_k \leq s < 1.$$

By the chain rule (ψ_1 and ψ_2 denote the partial derivatives with respect to the first and second arguments of ψ),

$$g_v'(s) = \psi_1(s, \lambda_v(s)) + \psi_2(s, \lambda_v(s))\lambda_v'(s),$$

which, together with (19.38) and the definitions (19.39) and (19.40), gives

$$g_v'(s) = -\tfrac{1}{2}\lambda_v^2(s)\sigma^2(s)g_v(s).$$

Therefore‡

$$\text{(19.41)} \quad \psi(s, \lambda_v(s)) = \psi(t_k, v)\exp\left[-\frac{1}{2}\int_{t_k}^{s}\sigma^2(r)\lambda_v^2(r)\,dr\right], \qquad t_k \leq s < 1.$$

Given an arbitrary (t, u) in the strip (19.33), take $v = u\exp\int_{t_k}^{t}\rho(\tau)\,d\tau$. Then $\lambda_v(t) = u$ and

$$\lambda_v^2(r) = u^2\exp\left[2\int_{r}^{t}\rho(\tau)\,d\tau\right], \qquad t_k \leq r \leq t,$$

so that, by (19.41) with $s = t$,

$$\text{(19.42)} \qquad \psi(t, u) = \psi(t_k, ua)e^{-\frac{1}{2}u^2b^2},$$

† The derivative with respect to t here is two-sided, even though h tends to 0 through positive values; see the second footnote on p. 155.

‡ See the first footnote on p. 155.

where

$$a = \exp \int_{t_k}^{t} \rho(\tau)\, d\tau$$

and

$$b^2 = \int_{t_k}^{t} \sigma^2(r) \exp\left[2\int_{r}^{t} \rho(\tau)\, d\tau\right] dr.$$

Going back to the definition of ψ, we see by (19.42) that

$$\begin{aligned}\mathsf{E}\{\exp i(u_1 X(t_1) + \cdots + u_k X(t_k) + u X(t))\} \\ = \mathsf{E}\{\exp i(u_1 X(t_1) + \cdots + u_k X(t_k) + uaX(t_k))\}e^{-\frac{1}{2}u^2b^2}.\end{aligned}$$

We have proved this for $0 \le t_1 \le \cdots \le t_k \le t < 1$; by continuity, it is true for $t = 1$ as well. It implies that, for arbitrary v,

$$\begin{aligned}\mathsf{E}\{\exp i(u_1 X(t_1) + \cdots + u_k X(t_k) + v(X(t) - a X(t_k)))\} \\ = \mathsf{E}\{\exp i(u_1 X(t_1) + \cdots + u_k X(t_k))\}e^{-\frac{1}{2}v^2b^2}.\end{aligned}$$

Thus the random vector $(X(t_1), \ldots, X(t_k))$ and the random variable $X(t) - a X(t_k)$ are independent of each other, and the latter has distribution $N(0, b^2)$.†

Taking $k = 1$ and $t_1 = 0$ and using the fact that $\mathsf{P}\{X(0) = 0\} = 1$, we see that $X(t)$ is normally distributed with mean 0. Taking $k = 1$ and t_1 general, we see that (19.20) holds and that $X(t_1)$ and $X(t) - a X(t_1)$ are jointly normal (for an appropriate a) and hence that the same is true of $X(t_1)$ and $X(t)$. Taking $k = 2$, we see that $X(t_1)$, $X(t_2)$, and $X(t) - a X(t_2)$ are jointly normal and hence that the same is true of $X(t_1)$, $X(t_2)$, and $X(t)$. Continuing in this way, we conclude that all the finite-dimensional distributions are normal, which completes the proof.

For the asymptotic form of Theorem 19.3, we shall need three conditions which parallel the Conditions 1, 2, and 3. As before, each condition will be given two forms. Let X_n be random elements of D.

Condition 1°. *If* $0 \le t_1 \le \cdots \le t_k < 1$, *then*

$$(19.43)\quad \lim_{h\downarrow 0} \limsup_{n\to\infty} \frac{1}{h} \mathsf{E}\{|\mathsf{E}\{X_n(t_k + h) - X_n(t_k) \,\|\, X_n(t_1), \ldots, X_n(t_k)\} - h\rho(t_k)X_n(t_k)|\} = 0$$

and

$$(19.44)\quad \lim_{h\downarrow 0} \limsup_{n\to\infty} \frac{1}{h} \mathsf{E}\{|\mathsf{E}\{(X_n(t_k + h) - X_n(t_k))^2 \,\|\, X_n(t_1), \ldots, X_n(t_k)\} - h\,\sigma^2(t_k)|\} = 0.$$

† This implies that $\{X(t)\}$ is a Markov process.

Condition 1°a. *If* $0 \le t_1 \le \cdots \le t_k < 1$, *then, for all real* $u_1, \ldots, u_k$,

$$\lim_{h \downarrow 0} \limsup_{n \to \infty} \frac{1}{h} \left| \mathsf{E}\left\{ \left[\exp \sum_{j=1}^{k} i u_j X_n(t_j) \right] \times \left[X_n(t_k + h) - X_n(t_k) - h\rho(t_k) X_n(t_k) \right] \right\} \right| = 0 \tag{19.45}$$

and

$$\lim_{h \downarrow 0} \limsup_{n \to \infty} \frac{1}{h} \left| \mathsf{E}\left\{ \left[\exp \sum_{j=1}^{k} i u_j X_n(t_j) \right] \times \left[(X_n(t_k + h) - X_n(t_k))^2 - \sigma^2(t_k) \right] \right\} \right| = 0. \tag{19.46}$$

Condition 2°. *We have*

$$\sup_t \limsup_{n \to \infty} \mathsf{E}\{X_n^2(t)\} < \infty. \tag{19.47}$$

Condition 2°a. *We have*

$$\lim_{\alpha \to \infty} \sup_t \limsup_{n \to \infty} \int_{\{|X_n(t)| \ge \alpha\}} |X_n(t)| \, d\mathsf{P} = 0. \tag{19.48}$$

Condition 3°. *There is a constant* K *such that*

$$\mathsf{E}\{(X_n(t) - X_n(t_1))^2 (X_n(t_2) - X_n(t))^2\} \le K(t_2 - t_1)^2, \quad t_1 \le t \le t_2, \tag{19.49}$$

for all n.

Condition 3°a. *For* $t < 1$,

$$\lim_{\alpha \to \infty} \limsup_{h \downarrow 0} \limsup_{n \to \infty} \frac{1}{h} \int_{\{(X_n(t+h) - X_n(t))^2 \ge \alpha h\}} (X_n(t+h) - X_n(t))^2 \, d\mathsf{P} = 0. \tag{19.50}$$

It is easy to show that Condition 1° implies 1°a and that 2° implies 2°a. Although 3° does not imply 3°a, each of them does suffice, as we shall see, for the theorem.

THEOREM 19.4 *Suppose that* $\{X_n^2(t) : n \ge 1\}$ *is uniformly integrable for each* t, *that* $X_n(0) \xrightarrow{P} 0$, *and that for each positive* ε *and* η *there exists a* δ *such that*

$$\mathsf{P}\{w(X_n, \delta) \ge \varepsilon\} \le \eta \tag{19.51}$$

for all sufficiently large n. *Suppose* $\sigma^2(t)$ *and* $\rho(t)$ *are continuous on* $[0, 1)$, *and there exists the finite limit* (19.21). *If* $\{X_n\}$ *satisfies Conditions* 1° (*or* 1°a), 2° (*or* 2°a), *and* 3° (*or* 3°a), *then* $X_n \xrightarrow{\mathcal{D}} X$, *where* X *is the continuous, Gaussian random function specified by* (19.19) *and* (19.20).

Proof. By Theorem 15.5, each subsequence of $\{X_n\}$ contains a further subsequence converging in distribution to some random element X of D with $\mathsf{P}\{X(0) = 0\} = 1$ and $\mathsf{P}\{X \in C\} = 1$. It is enough to show that any such X must satisfy the remaining hypotheses of Theorem 19.3. By Theorem 5.4 and the assumed uniform integrability of $\{X_n^2(t): n \geq 1\}$, Condition 1°a implies Condition 1a; moreover, by Theorem 5.3, 2°a implies 2a, 3°a implies 3a, and 3° implies 3 (which, as we saw before, implies 3a). Thus X satisfies Conditions 1a, 2a, and 3a, which completes the proof.

Remark. Suppose that Condition 3° holds, that

$$X_n(0) = 0 \tag{19.52}$$

holds with probability 1, and that the maximum jump in X_n satisfies

$$\max_t |X_n(t) - X_n(t-)| \leq \varepsilon_n \tag{19.53}$$

with probability 1, where

$$\varepsilon_n \to 0. \tag{19.54}$$

By Theorem 12.1 and the inequality (12.6) we then have

$$\mathsf{P}\left\{\sup_{t \leq s \leq t+\delta} |X_n(s) - X_n(t)| \geq \lambda\right\} \leq K' \frac{\delta^2}{\lambda^4} \tag{19.55}$$

with $K' = 4^4 K K_{2,1}$, provided $\varepsilon_n < \frac{1}{4}\lambda$, from which (19.51) follows by the corollary to Theorem 8.3. Now (19.52) and (19.55) imply

$$\mathsf{P}\{|X_n(t)| \geq \lambda\} \leq \frac{K'}{\lambda^4}$$

if $\varepsilon_n < \frac{1}{4}\lambda$; therefore (by (3) on p. 223)

$$\int_{\{X_n^2(t) \geq \alpha\}} X_n^2(t)\, d\mathsf{P} \leq 2 \frac{K'}{\alpha} \tag{19.56}$$

if $\varepsilon_n^2 < \frac{1}{16}\alpha$, which implies that $\{X_n(t): n \geq 1\}$ is uniformly integrable. Finally, (19.56) implies (19.48). Therefore, if (19.52), (19.53), and (19.54) hold, the conclusion of Theorem 19.4 will follow if we verify only Condition 1° (or 1°a) and Condition 3°.

Remarks. The results here are those of Rosén (1967a), with the conditions altered so as to bring Theorem 12.1 to bear. The idea of deducing the central limit theorem from a differential equation goes back to Khinchine (1933); Theorem 19.3 under the special conditions (19.30) and (19.31) is due to Lévy (1948) and Doob (1953).

20. MIXING PROCESSES

φ-*Mixing*

Let

$$\ldots, \xi_{-1}, \xi_0, \xi_1, \ldots \tag{20.1}$$

be a strictly stationary sequence of random variables defined on a probability space $(\Omega, \mathscr{B}, \mathsf{P})$. For $a \le b$, define $\mathscr{M}_a^b$ as the σ-field generated by the random variables $\xi_a, \ldots, \xi_b$; define $\mathscr{M}_{-\infty}^a$ as the σ-field generated by $\ldots, \xi_{a-1}, \xi_a$; and define $\mathscr{M}_a^\infty$ as the σ-field generated by $\xi_a, \xi_{a+1}, \ldots$.

Consider a nonnegative function φ of positive integers. We shall say that the sequence $\{\xi_n\}$ is *φ-mixing* if, for each k $(-\infty < k < \infty)$ and for each n $(n \ge 1)$, $E_1 \in \mathscr{M}_{-\infty}^k$ and $E_2 \in \mathscr{M}_{k+n}^\infty$ together imply

$$|\mathsf{P}(E_1 \cap E_2) - \mathsf{P}(E_1)\mathsf{P}(E_2)| \le \varphi(n)\mathsf{P}(E_1). \tag{20.2}$$

This is a joint property of $\{\xi_n\}$ and φ. We consider only functions φ satisfying

$$\lim_{n\to\infty} \varphi(n) = 0, \tag{20.3}$$

and usually we require that $\varphi(n)$ go to 0 at some specified minimum rate. If we say that $\{\xi_n\}$ is φ-mixing without specifying φ, we mean that (20.2) holds for some φ satisfying (20.3).

If $\mathsf{P}(E_1) > 0$, then (20.2) is equivalent to

$$|\mathsf{P}(E_2 \mid E_1) - \mathsf{P}(E_2)| \le \varphi(n), \tag{20.4}$$

whereas (20.2) holds trivially if $\mathsf{P}(E_1) = 0$. We shall regard the left side of (20.4) as vanishing if $\mathsf{P}(E_1) = 0$, so that the condition for φ-mixing is

$$\sup\{|\mathsf{P}(E_2 \mid E_1) - \mathsf{P}(E_2)| : E_1 \in \mathscr{M}_{-\infty}^k, E_2 \in \mathscr{M}_{k+n}^\infty\} \le \varphi(n). \tag{20.5}$$

If $\varphi(n)$ is small, (20.2) and (20.4) say that E_2 is virtually independent of E_1. In a φ-mixing process, the distant future is virtually independent of the past and present.† This property enables us to prove various central limit theorems and functional central limit theorems.

We shall often write φ_n in place of $\varphi(n)$. If each φ_n has the minimal value consistent with (20.5), then

$$1 \ge \varphi_1 \ge \varphi_2 \ge \cdots. \tag{20.6}$$

† Since (20.2) and (20.4) are not symmetric in E_1 and E_2, past and future cannot be interchanged; for a φ-mixing process that ceases to be φ-mixing when time is reversed, take $\xi_n = \sum_{k=1}^\infty \eta_{n-k}/2^k$, where η_k are independent and assume the values ± 1 with probability $\frac{1}{2}$ each.

If we replace φ_n by

(20.7) $$\min\{1, \varphi_1, \ldots, \varphi_n\},$$

then $\{\xi_n\}$ is still φ-mixing and (20.6) holds. Thus there is no loss of generality in assuming (20.6). It will be convenient to define $\varphi_0 = 1$, so that (20.5) holds trivially if $n = 0$.

For fixed E_2,

(20.8) $$\{E_1 : |\mathsf{P}(E_2 \mid E_1) - \mathsf{P}(E_2)| \leq \varepsilon\}$$

is a monotone class† (is closed under the formation of countable increasing unions and countable decreasing intersections); for fixed E_1, the same is true of the class

(20.9) $$\{E_2 : |\mathsf{P}(E_2 \mid E_1) - \mathsf{P}(E_2)| \leq \varepsilon\}.$$

It follows that, if (20.4) holds for E_1 in $\mathscr{A}$ and E_2 in $\mathscr{B}$, where $\mathscr{A}$ and $\mathscr{B}$ are fields generating $\mathscr{M}_{-\infty}^{k}$ and $\mathscr{M}_{k+n}^{\infty}$, respectively, then (20.5) holds, an observation useful in checking whether a given process is φ-mixing.

Example 1. The sequence (20.1) is said to be m-dependent if the random vectors $(\xi_i, \ldots, \xi_k)$ and $(\xi_{k+n}, \ldots, \xi_j)$ are independent whenever $n > m$. (In this terminology, an independent process is 0-dependent.) Such a process is φ-mixing with $\varphi_n = 0$ for $n > m$. For a specific example, take

(20.10) $$\xi_n = a_0\zeta_n + a_1\zeta_{n-1} + \cdots + a_m\zeta_{n-m},$$

where the a_i are constants and the ζ_n are independent and identically distributed.

Example 2. Let $\{\zeta_n\}$ be a stationary Markov process with finite state space, and let $\xi_n = f(\zeta_n)$, where f is some real function on the state space. If $\mathscr{M}_a{}^b$ is defined as before and if $\mathscr{N}_a{}^b$ is the σ-field generated by $\zeta_a, \ldots, \zeta_b$, then $\mathscr{M}_a{}^b \subset \mathscr{N}_a{}^b$.

Let p_u be the stationary probabilities for the Markov process, and let p_{uv} be the transition probabilities, so that

$$\mathsf{P}\{\zeta_i = u_0, \ldots, \zeta_{i+l} = u_l\} = p_{u_0}p_{u_0u_1}\cdots p_{u_{l-1}u_l}$$

for each finite sequence $u_0, u_1, \ldots, u_l$ of states. Assume that the p_u are all positive, so that

(20.11) $$\varphi_n = \max_{u,v}\left|\frac{p_{uv}^{(n)}}{p_v} - 1\right|$$

is finite, where $p_{uv}^{(n)}$ are the nth order transition probabilities. Let H_1 be a set

† See Halmos (1950, p. 26).

of $(i+1)$-tuples of states, and let H_2 be a set of $(j+1)$-tuples of states. For the special element

$$E_1 = \{(\zeta_{k-i}, \ldots, \zeta_k) \in H_1\} \tag{20.12}$$

of $\mathscr{N}_{-\infty}^{k}$ and the special element

$$E_2 = \{(\zeta_{k+n}, \ldots, \zeta_{k+n+j}) \in H_2\} \tag{20.13}$$

of $\mathscr{N}_{k+n}^{\infty}$, we have

$$|\mathsf{P}(E_1 \cap E_2) - \mathsf{P}(E_1)\mathsf{P}(E_2)| \le \sum p_{u_0} p_{u_0 u_1} \cdots p_{u_{i-1} u_i} |p_{u_i v_0}^{(n)} - p_{v_0}| \, p_{v_0 v_1} \cdots p_{v_{j-1} v_j},$$

where the sum extends over $(u_0, \ldots, u_i)$ in H_1 and over $(v_0, \ldots, v_j)$ in H_2. From the relation $\Sigma_v p_{uv} = 1$ and the definition (20.11), we obtain (20.2). And now from the fact that (20.8) and (20.9) are monotone classes, it follows that (20.2) holds for all E_1 in $\mathscr{N}_{-\infty}^{k}$ and E_2 in $\mathscr{N}_{k+n}^{\infty}$ and hence for all E_1 in $\mathscr{M}_{-\infty}^{k}$ and E_2 in $\mathscr{M}_{k+n}^{\infty}$.

If the transition matrix (p_{uv}) is irreducible and aperiodic, then†

$$p_{uv}^{(n)} \to p_v \tag{20.14}$$

for all u and v. Since the state space is assumed finite, the convergence must be uniform in u and v, so that φ_n, defined by (20.11), converges to 0: $\{\xi_n\}$ is φ-mixing. In these circumstances more is known, namely that the rate of convergence in (20.14) is exponential: There exist positive constants a and ρ, $\rho < 1$, such that, if

$$\varphi_n = a\rho^n, \tag{20.15}$$

then $\{\xi_n\}$ is φ-mixing. This is also true of certain process $\xi_n = f(\zeta_n)$ for which $\{\zeta_n\}$ is a Markov process with infinite state space—it is true, for example, if $\{\zeta_n\}$ satisfies Doeblin's condition, has one ergodic class, and is aperiodic.‡

In some applications, we are presented only with a one-sided (strictly stationary) sequence

$$\xi_1, \xi_2, \ldots. \tag{20.16}$$

In such cases we shall say the sequence is φ-mixing if

$$\sup\{|\mathsf{P}(E_2 \mid E_1) - \mathsf{P}(E_2)| : E_1 \in \mathscr{M}_1^{k}, E_2 \in \mathscr{M}_{k+n}^{\infty}\} \le \varphi_n \tag{20.17}$$

for positive integers k and n. Given a one-sided sequence (20.16), we can

† Doob (1953, pp. 172 ff.).

‡ Doob (1953, pp. 190 ff.).

always construct a two-sided sequence (20.1) with the same finite-dimensional distributions—the new sequence will in general be defined on a new sample space (which we can always take to be the product of a doubly infinite sequence of copies of the real line). We shall then call (20.1) the doubly infinite extension of (20.16).

If (20.16) is φ-mixing, then, by stationarity, the doubly infinite extension satisfies (20.4) for $E_1 \in \mathscr{M}_{k-i}^{k}$ and $E_2 \in \mathscr{M}_{k+n}^{\infty}$. Since (20.8) is a monotone class, it follows that the doubly infinite extension is φ-mixing, with the same φ as before. It is easy to verify the converse: If the extension (20.1) is φ-mixing, then the original sequence (20.16) is φ-mixing with the same function φ.

If we prove, for example, that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i \xrightarrow{\mathscr{D}} N,$$

where the ξ_i are interpreted as elements of (20.1), then the result is also true if they are interpreted as elements of (20.16), since only finite-dimensional distributions are involved. All our theorems will be of this kind, and it is indifferent to the results whether we work with (20.16) or with (20.1). In proofs we shall consistently work with (20.1), which is more convenient because then the past is an infinite sequence instead of a finite sequence of changing length.

Example 3. Let Ω be the unit interval, let $\mathscr{B}$ consist of the linear Borel subsets of Ω, and let P be Lebesgue measure on $\mathscr{B}$. Suppose

(20.18) $$\xi_n(\omega) = g(\omega_n, \omega_{n+1}, \ldots, \omega_{n+m}), \qquad n = 1, 2, \ldots,$$

where ω has dyadic expansion $\omega = \cdot\, \omega_1\omega_2 \cdots$ and g is some function of $(m + 1)$-long sequences of 0's and 1's. The sequence $\{\xi_n\}$ is stationary and m-dependent. This sequence is of the form (20.16) and comes under the theory as developed here because it is possible to construct, as outlined above, a doubly infinite extension (on some new space).

Example 4. For another example in which the sequence $\{\xi_n\}$ is one-sided, take Ω and $\mathscr{B}$ as in Example 3, let P be Gauss's measure

(20.19) $$\mathsf{P}(A) = \frac{1}{\log 2} \int_A \frac{dx}{1 + x}, \qquad A \in \mathscr{B},$$

and take $\xi_n(\omega)$ to be $a_n(\omega)$, the nth coefficient in the continued-fraction expansion of ω. The sequence $\{a_1(\omega), a_2(\omega), \ldots\}$ is stationary and φ-mixing with φ of the form

(20.20) $$\varphi_n = a\rho^n, \qquad a > 0, \qquad 0 < \rho < 1.†$$

† Lévy (1937, Chapter 9).

Inequalities for Moments

Suppose ξ is measurable $\mathscr{M}_{-\infty}^{k}$ and η is measurable $\mathscr{M}_{k+n}^{\infty}$. If ξ and η are the indicators of sets, then, by (20.2),

$$|\mathsf{E}\{\xi\eta\} - \mathsf{E}\{\xi\}\mathsf{E}\{\eta\}| \tag{20.21}$$

is at most $\varphi_n\mathsf{E}\{\xi\}$. We shall require bounds on (20.21) for random variables ξ and η more general than indicators. In all that follows, $\{\xi_n\}$ is assumed stationary and φ-mixing unless the contrary is explicitly stated.

LEMMA 1 *If ξ is measurable $\mathscr{M}_{-\infty}^{k}$ and η is measurable $\mathscr{M}_{k+n}^{\infty}$ $(n \geq 0)$, then*

$$\mathsf{E}\{|\xi|^r\} < \infty, \qquad \mathsf{E}\{|\eta|^s\} < \infty, \qquad r, s > 1, \qquad \frac{1}{r} + \frac{1}{s} = 1, \tag{20.22}$$

implies

$$|\mathsf{E}\{\xi\eta\} - \mathsf{E}\{\xi\}\mathsf{E}\{\eta\}| \leq 2\varphi_n^{1/r}\mathsf{E}^{1/r}\{|\xi|^r\}\mathsf{E}^{1/s}\{|\eta|^s\}. \tag{20.23}$$

Proof. For an understanding of the lemma, consider first two extreme cases. If $\varphi_n = 0$, so that ξ and η are independent, then the inequality (20.23) holds because each of its members vanishes. If $\varphi_n = 1$, which imposes no restriction at all, then, by the inequalities of Hölder and Minkowski,

$$\begin{aligned} |\mathsf{E}\{\xi\eta\} - \mathsf{E}\{\xi\}\mathsf{E}\{\eta\}| &= |\mathsf{E}\{\xi(\eta - \mathsf{E}\{\eta\})\}| \\ &\leq \mathsf{E}^{1/r}\{|\xi|^r\}\mathsf{E}^{1/s}\{|\eta - \mathsf{E}\{\eta\}|^s\} \leq 2\mathsf{E}^{1/r}\{|\xi|^r\}\mathsf{E}^{1/s}\{|\eta|^s\}. \end{aligned}$$

Since we can approximate to ξ and η by simple random variables, in treating the general case we may suppose that

$$\xi = \Sigma_i u_i I_{A_i} \tag{20.24}$$

and

$$\eta = \Sigma_j v_j I_{B_j}, \tag{20.25}$$

where $\{A_i\}[\{B_j\}]$ is a finite decomposition of the sample space Ω into elements of $\mathscr{M}_{-\infty}^{k}$ $[\mathscr{M}_{k+n}^{\infty}]$. For such random variables, we have, by Hölder's inequality,

$$\begin{aligned} |\mathsf{E}\{\xi\eta\} - \mathsf{E}\{\xi\}\mathsf{E}\{\eta\}| &= \left|\sum_i u_i \mathsf{P}(A_i)^{1/r}\left[\mathsf{P}(A_i)^{1/s}\sum_j v_j\big(\mathsf{P}(B_j \mid A_i) - \mathsf{P}(B_j)\big)\right]\right| \\ &\leq \mathsf{E}^{1/r}\{|\xi|^r\}\left\{\sum_i \mathsf{P}(A_i)\left|\sum_j v_j\big(\mathsf{P}(B_j \mid A_i) - \mathsf{P}(B_j)\big)\right|^s\right\}^{1/s}, \end{aligned}$$

and hence it suffices to prove

$$\sum_i \mathsf{P}(A_i)\left|\sum_j v_j\big(\mathsf{P}(B_j \mid A_i) - \mathsf{P}(B_j)\big)\right|^s \leq 2^s\varphi_n^{s/r}\mathsf{E}\{|\eta|^s\}. \tag{20.26}$$

For each i, Hölder's inequality gives

$$\left|\sum_j v_j(\mathsf{P}(B_j \mid A_i) - \mathsf{P}(B_j))\right|$$

$$\leq \left\{\sum_j |v_j|^s |\mathsf{P}(B_j \mid A_i) - \mathsf{P}(B_j)|\right\}^{1/s} \left\{\sum_j |\mathsf{P}(B_j \mid A_i) - \mathsf{P}(B_j)|\right\}^{1/r}.$$

Since

$$\sum_i \mathsf{P}(A_i) \sum_j |v_j|^s |\mathsf{P}(B_j \mid A_i) - \mathsf{P}(B_j)| \leq 2\mathsf{E}\{|\eta|^s\},$$

(20.26) will follow if we show that

(20.27) $$\sum_j |\mathsf{P}(B_j \mid A_i) - \mathsf{P}(B_j)| \leq 2\varphi_n$$

holds for each i. If C_i^+ $[C_i^-]$ is the union of those B_j for which $\mathsf{P}(B_j \mid A_i) - \mathsf{P}(B_j)$ is positive [nonpositive], then C_i^+ and C_i^- lie in $\mathscr{M}_{k+n}^{\infty}$, and hence

$$\sum_j |\mathsf{P}(B_j \mid A_i) - \mathsf{P}(B_j)| = [\mathsf{P}(C_i^+ \mid A_i) - \mathsf{P}(C_i^+)] + [\mathsf{P}(C_i^-) - \mathsf{P}(C_i^- \mid A_i)] \leq 2\varphi_n.$$

Thus (20.27) holds, which completes the proof.

Taking $r = 1$ and $s = \infty$ in Lemma 1 leads formally to this result: If $\mathsf{E}\{|\xi|\} < \infty$ and $\mathsf{P}\{|\eta| > C\} = 0$ (ξ measurable $\mathscr{M}_{-\infty}^k$ and η measurable $\mathscr{M}_{k+n}^{\infty}$), then

(20.28) $$|\mathsf{E}\{\xi\eta\} - \mathsf{E}\{\xi\}\mathsf{E}\{\eta\}| \leq 2\varphi_n \mathsf{E}\{|\xi|\}C.$$

The inequality is easily proved directly: It suffices to consider simple random variables given by (20.24) and (20.25), where $|v_j| \leq C$. Since

$$|\mathsf{E}\{\xi\eta\} - \mathsf{E}\{\xi\}\mathsf{E}\{\eta\}| \leq C \sum_i |u_i| \, \mathsf{P}(A_i) \sum_j |\mathsf{P}(B_j \mid A_i) - \mathsf{P}(B_j)|,$$

(20.28) follows by (20.27).

We shall only need the special case of (20.28) in which both random variables are bounded.

LEMMA 2 *If ξ is measurable $\mathscr{M}_{-\infty}^k$ and $|\xi| \leq C_1$, and if η is measurable $\mathscr{M}_{k+n}^{\infty}$ $(n \geq 0)$ and $|\eta| \leq C_2$, then*

(20.29) $$|\mathsf{E}\{\xi\eta\} - \mathsf{E}\{\xi\}\mathsf{E}\{\eta\}| \leq 2C_1C_2\varphi_n.$$

Let

(20.30) $$S_n = \xi_1 + \cdots + \xi_n$$

and $S_0 = 0$. Let us temporarily abandon the hypothesis that $\{\xi_n\}$ is φ-mixing (retaining the hypothesis of stationarity, however).

LEMMA 3 *Suppose that* $\mathsf{E}\{\xi_0\} = 0$ *and*

$$\sum_{k=0}^{\infty} |\mathsf{E}\{\xi_0\xi_k\}| < \infty. \tag{20.31}$$

Then

$$\frac{1}{n}\mathsf{E}\{S_n^2\} \le 2\sum_{k=0}^{\infty} |\mathsf{E}\{\xi_0\xi_k\}| \tag{20.32}$$

for all n, *and*

$$\frac{1}{n}\mathsf{E}\{S_n^2\} \to \sigma^2, \tag{20.33}$$

where

$$\sigma^2 = \mathsf{E}\{\xi_0^2\} + 2\sum_{k=1}^{\infty} \mathsf{E}\{\xi_0\xi_k\}. \tag{20.34}$$

Proof. If $\rho_k = \mathsf{E}\{\xi_0\xi_k\}$, then, by stationarity,

$$\mathsf{E}\{S_n^2\} = n\rho_0 + 2\sum_{k=1}^{n-1}(n-k)\rho_k;$$

(20.32) follows immediately and (20.33) follows from

$$\left|\sigma^2 - \frac{1}{n}\mathsf{E}\{S_n^2\}\right| \le 2\sum_{k=n}^{\infty}|\rho_k| + \frac{2}{n}\sum_{i=1}^{n-1}\sum_{k=i}^{\infty}|\rho_k|.$$

Suppose once more that $\{\xi_n\}$ is φ-mixing. If ξ_0 has mean 0 and finite variance, then, by Lemma 1 with $r = s = 2$,

$$|\mathsf{E}\{\xi_0\xi_k\}| \le 2\varphi_k^{\frac{1}{2}}\mathsf{E}\{\xi_0^2\}. \tag{20.35}$$

If $\Sigma\varphi_n^{\frac{1}{2}} < \infty$, then (20.31) holds, so that the variance of S_n is asymptotically $n\sigma^2$ with σ^2 defined by (20.34). The quantity σ^2 may vanish even though ξ_0 has positive variance. This is true, for example, if $\xi_n = \zeta_n - \zeta_{n-1}$, where the ζ_n are independent and identically distributed (see (20.10)). The case $\sigma = 0$ is a degenerate one, to be excluded in most of our theorems.

LEMMA 4 *If* ξ_0 *is bounded by* C *and* $\mathsf{E}\{\xi_0\} = 0$, *and if* $\Sigma\varphi_n^{\frac{1}{2}} < \infty$, *then*

$$\mathsf{E}\{S_n^4\} \le K_\varphi C^4 n^2,$$

where K_φ *depends on* φ *alone.*

Proof. We shall show that

$$\mathsf{E}\{S_n^4\} \le 768C^4\left[\sum_{i=0}^{\infty}\varphi_i^{\frac{1}{2}}\right]^2 n^2. \tag{20.36}$$

If we replace φ_i by (20.7), the series in (20.36) does not increase; hence we may assume

$$1 = \varphi_0 \ge \varphi_1 \ge \varphi_2 \ge \cdots. \tag{20.37}$$

By stationarity,

$$\mathsf{E}\{S_n{}^4\} \leq 4!\, n \sum |\mathsf{E}\{\xi_0 \xi_i \xi_{i+j} \xi_{i+j+k}\}|, \tag{20.38}$$

where the indices in the sum are constrained by

$$i, j, k \geq 0, \qquad i + j + k \leq n. \tag{20.39}$$

By Lemma 2,

$$|\mathsf{E}\{\xi_0(\xi_i \xi_{i+j} \xi_{i+j+k})\}| \leq 2C^4 \varphi_i \tag{20.40}$$

and

$$|\mathsf{E}\{(\xi_0 \xi_i \xi_{i+j})\xi_{i+j+k}\}| \leq 2C^4 \varphi_k. \tag{20.41}$$

By Lemma 2 again (and stationarity),

$$|\mathsf{E}\{(\xi_0\xi_i)(\xi_{i+j}\xi_{i+j+k})\}| \leq |\mathsf{E}\{\xi_0\xi_i\}\mathsf{E}\{\xi_0\xi_k\}| + 2C^4\varphi_j.$$

Two further applications of Lemma 2 yield

$$|\mathsf{E}\{\xi_0\xi_i\}| \leq 2C^2\varphi_i, \qquad |\mathsf{E}\{\xi_0\xi_k\}| \leq 2C^2\varphi_k;$$

inserting these two inequalities into the preceding one, we obtain

$$|\mathsf{E}\{(\xi_0\xi_i)(\xi_{i+j}\xi_{i+j+k})\}| \leq 4C^4\varphi_i\varphi_k + 2C^4\varphi_j. \tag{20.42}$$

By (20.40), (20.41), and (20.42), the summand in (20.38) does not exceed $4C^4$ times the minimum of the three quantities

$$\varphi_i, \qquad \varphi_k, \qquad \varphi_i\varphi_k + \varphi_j,$$

and hence the sum itself does not exceed $4C^4$ times

$$\sum_{j,k\leq i} \varphi_i + \sum_{i,j\leq k} \varphi_k + \sum_{i,k\leq j} (\varphi_i\varphi_k + \varphi_j) = \sum_{i,k\leq j} \varphi_i\varphi_k + 3\sum_{j,k\leq i} \varphi_i, \tag{20.43}$$

where in each sum the indices obey (20.39) as well as the restrictions indicated.

Since $\varphi_i \leq 1$,

$$\sum_{i,k\leq j} \varphi_i\varphi_k \leq \sum_{j=0}^{n} \sum_{i,k=0}^{\infty} \varphi_i\varphi_k \leq 2n\left[\sum_{i=0}^{\infty} \varphi_i^{\frac{1}{2}}\right]^2.$$

Furthermore, by (20.37),

$$\begin{aligned}\sum_{j,k\leq i} \varphi_i \leq \sum_{i=0}^{n} \sum_{j,k=0}^{i} \varphi_i &\leq 2n\sum_{i=0}^{n}(i+1)\varphi_i \\ &= 2n\sum_{u=0}^{n}\sum_{v=u}^{n}\varphi_v \leq 2n\sum_{u=0}^{\infty}\varphi_u^{\frac{1}{2}}\sum_{v=u}^{\infty}\varphi_v^{\frac{1}{2}} \leq 2n\left[\sum_{i=0}^{\infty}\varphi_i^{\frac{1}{2}}\right]^2.\end{aligned}$$

From these two inequalities it follows that (20.43) is at most

$$8n\left[\sum_{i=0}^{\infty} \varphi_i^{\frac{1}{2}}\right]^2 \tag{20.44}$$

and hence that the sum in (20.38) is at most $4C^4$ times this quantity. Multiplying (20.44) by $4!\,n \cdot 4C^4$, we arrive at (20.36).

Functional Central Limit Theorem

Define a random element X_n of D by

$$X_n(t, \omega) = \frac{1}{\sigma\sqrt{n}} S_{[nt]}(\omega), \qquad 0 \le t \le 1. \tag{20.45}$$

THEOREM 20.1 *Suppose that* $\{\xi_n\}$ *is* φ*-mixing with* $\sum_n \varphi_n^{\frac{1}{2}} < \infty$ *and that* ξ_0 *has mean* 0 *and finite variance. Then the series*

$$\sigma^2 = \mathsf{E}\{\xi_0^2\} + 2\sum_{k=1}^{\infty} \mathsf{E}\{\xi_0\xi_k\} \tag{20.46}$$

converges absolutely; if $\sigma^2 > 0$ *and* X_n *is defined by* (20.45), *then*

$$X_n \xrightarrow{\mathscr{D}} W. \tag{20.47}$$

Proof. That the series (20.46) converges absolutely follows from (20.35). Suppose $\sigma^2 > 0$. We shall prove (20.47) by an application of Theorem 19.2, thus avoiding a separate argument for the central limit theorem.

We shall prove that $\{S_n^2/n\}$ is uniformly integrable. If we write $\mathsf{E}_\alpha\{U\}$ as shorthand for the integral of U over the set $\{U \ge \alpha\}$, the condition for uniform integrability becomes

$$\lim_{\alpha\to\infty} \sup_n \mathsf{E}_\alpha\left\{\frac{1}{n} S_n^2\right\} = 0. \tag{20.48}$$

Assuming this condition for the moment, let us see how the rest of the proof goes through. We must first show that X_n has asymptotically independent increments, in the sense that (19.13) implies that the difference (19.14) converges to 0 as $n \to 0$.

Suppose u_i and v_j are integers with $u_1 \le v_1 < u_2 \le v_2 < \cdots < u_r \le v_r$ and $u_i - v_{i-1} \ge b$, $i = 2, \ldots, r$, and suppose that, for $i = 1, \ldots, r$, E_i lies in $\mathscr{M}_{u_i}^{v_i}$. It follows from the definition of φ-mixing by induction on r that

$$\left| \mathsf{P}\left(\bigcap_{i=1}^{r} E_i\right) - \prod_{i=1}^{r} \mathsf{P}(E_i) \right| \le r\varphi(b). \tag{20.49}$$

Suppose that (19.13) holds and let E_i be the event $\{X_n(t_i) - X_n(s_i) \in H_i\}$, where $H_i \in \mathscr{R}^1$. Then E_i lies in $\mathscr{M}_{[ns_i]+1}^{[nt_i]}$ and, if δ is the smallest difference $s_i - t_{i-1}$, then $[ns_i] + 1 - [nt_{i-1}] \ge [n\delta]$, so that, by (20.49), the difference (19.14) has modulus at most $r\varphi([n\delta])$. Since δ is positive, X_n does have asymptotically independent increments.

Since $X_n^2(t) \le S_{[nt]}^2/\sigma^2[nt]$, (20.48) implies that $\{X_n^2(t)\}$ is uniformly integrable for each t. Certainly, $\mathsf{E}\{X_n(t)\} = 0$, and Lemma 3 implies that $\mathsf{E}\{X_n^2(t)\} \to t$. By Theorem 8.4 (adapted to D) the tightness condition involving (19.15) will follow if we prove that, for each positive ε, there exist a λ, $\lambda > \sigma$, and an integer n_0 such that $n \ge n_0$ implies

$$\mathsf{P}\left\{\max_{i \le n} |S_i| \ge \lambda\sqrt{n}\right\} \le \frac{\varepsilon}{\lambda^2}. \tag{20.50}$$

Let us set

$$S_n^* = \sum_{j=1}^{n} |\xi_j|.$$

Since ξ_0 has a finite second moment, there exists an increasing sequence of integers m_i such that $n \ge m_i$ implies $\mathsf{P}\{|\xi_0| \ge \sqrt{n}/i^2\} \le 1/ni^2$. If we define $p_n = i$ for $m_i \le n < m_{i+1}$ (and $p_n = 1$ for $n < m_1$), then p_n goes to infinity, but so slowly that

$$\lim_{n\to\infty} n\mathsf{P}\{S_{p_n}^* \ge \lambda\sqrt{n}\} = 0 \tag{20.51}$$

for each positive λ. We may at the same time choose p_n in such a way that $p_n \le n$.

Given ε, choose λ so that $\lambda > \sigma$ and so that

$$\mathsf{P}\{|S_i| > \lambda\sqrt{i}\} < \frac{\varepsilon}{\lambda^2} \tag{20.52}$$

for all i, which is possible because of (20.48). If

$$E_i = \left\{\max_{j<i} |S_j| < 3\lambda\sqrt{n} \le |S_i|\right\},$$

then

$$\mathsf{P}\left\{\max_{i\le n} |S_i| \ge 3\lambda\sqrt{n}\right\} \le \mathsf{P}\{|S_n| \ge \lambda\sqrt{n}\} + \sum_{i=1}^{n-1} \mathsf{P}(E_i \cap \{|S_n - S_i| \ge 2\lambda\sqrt{n}\}).$$

With $p = p_n$, the sum here is at most

$$\sum_{i=1}^{n-p-1} \mathsf{P}\{|S_i - S_{i+p}| \ge \lambda\sqrt{n}\} + \sum_{i=1}^{n-p-1} \mathsf{P}(E_i \cap \{|S_n - S_{i+p}| \ge \lambda\sqrt{n}\}) + \sum_{i=n-p}^{n-1} \mathsf{P}\{|S_n - S_i| \ge \lambda\sqrt{n}\}.$$

Each term in the first and third of these sums is at most $\mathsf{P}\{S_p^* \ge \lambda\sqrt{n}\}$, and we can estimate the second sum by using the fact that $E_i \in \mathscr{M}_{-\infty}^{i}$:

$$\mathsf{P}\left\{\max_{i\le n} |S_i| \ge 3\lambda\sqrt{n}\right\} \le \mathsf{P}\{|S_n| \ge \lambda\sqrt{n}\} + (n+p)\mathsf{P}\{S_p^* \ge \lambda\sqrt{n}\} + \sum_{i=1}^{n-p-1} \mathsf{P}(E_i)[\mathsf{P}\{|S_n - S_{i+p}| \ge \lambda\sqrt{n}\} + \varphi_p].$$

And now (20.52) and the fact that the E_i are disjoint yields

$$\mathsf{P}\left\{\max_{i\leq n}|S_i| \geq 3\lambda\sqrt{n}\right\} \leq \frac{\varepsilon}{\lambda^2} + 2n\mathsf{P}\{S_p^* \geq \lambda\sqrt{n}\} + \frac{\varepsilon}{\lambda^2} + \varphi_p.$$

From (20.51) and the fact that $p_n \to \infty$, we conclude that

$$\mathsf{P}\left\{\max_{i\leq n}|S_i| \geq 3\lambda\sqrt{n}\right\} < \frac{3\varepsilon}{\lambda^2}$$

for large n, and this is (20.50) except for the two irrelevant factors of 3.

It now remains to prove (20.48). If $|\xi_0|$ is bounded by C, then, by Lemma 4,

$$\mathsf{E}_\alpha\left\{\frac{1}{n}S_n^{\ 2}\right\} \leq \frac{1}{\alpha}\mathsf{E}\left\{\frac{1}{n^2}S_n^{\ 4}\right\} \leq \frac{K_\varphi C^4}{\alpha}, \tag{20.53}$$

where K_φ depends on φ alone. Thus (20.48) holds if ξ_0 is bounded. If ξ_0 is not bounded, define, for real x and positive u,

$$f_u(x) = \begin{cases} x & \text{if } |x| \leq u, \\ 0 & \text{if } |x| > u, \end{cases} \qquad g_u(x) = \begin{cases} 0 & \text{if } |x| \leq u, \\ x & \text{if } |x| > u, \end{cases}$$

and put

$$\bar{f}_u(x) = f_u(x) - \mathsf{E}\{f_u(\xi_0)\},\ \bar{g}_u(x) = g_u(x) - \mathsf{E}\{g_u(\xi_0)\}.$$

Then $x = f_u(x) + g_u(x) = \bar{f}_u(x) + \bar{g}_u(x)$, so that, if $S_{nu} = \Sigma_{j=1}^n \bar{f}_u(\xi_j)$ and $D_{nu} = \Sigma_{j=1}^n \bar{g}_u(\xi_j)$, then $S_n = S_{nu} + D_{nu}$ and hence

$$S_n^{\ 2} \leq 2S_{nu}^2 + 2D_{nu}^2. \tag{20.54}$$

Since $\bar{f}_u(\xi_0)$ is bounded by $2u$, it follows by (20.53) that

$$\mathsf{E}_\alpha\left\{\frac{1}{n}S_{nu}^2\right\} \leq \frac{K_\varphi(2u)^4}{\alpha}. \tag{20.55}$$

By Lemma 1,

$$|\mathsf{E}\{\bar{g}_u(\xi_0)\bar{g}_u(\xi_k)\}| \leq 2\varphi_k^{\frac{1}{2}}\mathsf{E}\{(\bar{g}_u(\xi_0))^2\} \leq 2\varphi_k^{\frac{1}{2}}\mathsf{E}\{(g_u(\xi_0))^2\} = 2\varphi_k^{\frac{1}{2}}\mathsf{E}_{u^2}\{\xi_0^{\ 2}\},$$

and it follows by (20.32) that

$$\mathsf{E}\left\{\frac{1}{n}D_{nu}^2\right\} \leq 4\left[\sum_{k=0}^{\infty}\varphi_k^{\frac{1}{2}}\right]\mathsf{E}_{u^2}\{\xi_0^{\ 2}\}. \tag{20.56}$$

From (20.54) and the relation $\mathsf{E}_\alpha\{U + V\} \leq 2\mathsf{E}_{\frac{1}{2}\alpha}\{U\} + 2\mathsf{E}\{V\}$, we now obtain

$$\mathsf{E}_\alpha\left\{\frac{1}{n}S_n^{\ 2}\right\} \leq 4\mathsf{E}_{\frac{1}{4}\alpha}\left\{\frac{1}{n}S_{nu}^2\right\} + 4\mathsf{E}\left\{\frac{1}{n}D_{nu}^2\right\},$$

and (20.55) and (20.56) lead to

$$\mathsf{E}_\alpha\left\{\frac{1}{n}S_n^{\,2}\right\} \le K'_\varphi\left[\frac{u^4}{\alpha} + \mathsf{E}_{u^2}\{\xi_0^{\,2}\}\right]$$

for an appropriate K'_φ depending only on φ. We can achieve (20.48) by choosing u so that $K'_\varphi \mathsf{E}_{u^2}\{\xi_0^2\} < \frac{1}{2}\varepsilon$ and then choosing α so that $K'_\varphi u^4/\alpha < \frac{1}{2}\varepsilon$. This completes the proof of Theorem 20.1.

Under the hypotheses of the theorem we of course have

$$\frac{1}{\sqrt{n}} S_n \xrightarrow{\mathscr{D}} N(0, \sigma^2) \tag{20.57}$$

if $\sigma^2 > 0$. And if $\sigma^2 = 0$ (50.57) is equivalent to $S_n/\sqrt{n} \xrightarrow{P} 0$ and follows easily by Chebyshev's inequality and Lemma 3.†

It is easy to make up a multivariate version of (20.57). If ξ_n is the random vector

$$\xi_n = (\xi_n^{(1)}, \ldots, \xi_n^{(r)}), \tag{20.58}$$

we can define the notion of φ-mixing just as before. If $\Sigma\varphi_n^{\frac{1}{2}} < \infty$ and if the $\xi_n^{(i)}$ have mean 0 and finite variances, it follows by an application of the Cramér–Wold technique (Theorem 7.7) to (20.57) that $n^{-\frac{1}{2}} \Sigma_{k=1}^n \xi_k$ has asymptotically a normal distribution, centered at the origin, with covariances

$$\sigma_{ij} = \mathsf{E}\{\xi_0^{(i)}\xi_0^{(j)}\} + \sum_{k=1}^{\infty}\mathsf{E}\{\xi_0^{(i)}\xi_k^{(j)}\} + \sum_{k=1}^{\infty}\mathsf{E}\{\xi_k^{(i)}\xi_0^{(j)}\}, \tag{20.59}$$

where the series converge absolutely. The matrix (σ_{ij}) may be singular or nonsingular.

These results apply to each of the four examples given above. Note that in Example 1 (p. 167) and in Example 3 (p. 169) the series (20.46) are finite. In Example 2, with $\xi_n = f(\zeta_n)$, where the transition matrix for $\{\zeta_n\}$ is irreducible and aperiodic, the asymptotic variance σ^2 can be written

$$\sigma^2 = \sum_{uv} \gamma_{uv} f(u) f(v),$$

where

$$\gamma_{uv} = \delta_{uv}p_u - p_u p_v + p_u\sum_{k=1}^{\infty}(p_{uv}^{(k)} - p_v) + p_v\sum_{k=1}^{\infty}(p_{vu}^{(k)} - p_u).$$

If the matrix (γ_{uv}) annihilates the vector $(f(u))$, then $\sigma^2 = 0$.

---

† If $\sigma^2 = 0$, it is even possible, by adapting the proof of (20.50), to show that

$$\max_{i \le n} |S_i|/\sqrt{n} \xrightarrow{P} 0.$$

Integrals in Place of Sums

Theorem 20.1 has a natural formulation with $\{\xi_n\}$ replaced by a process in continuous time. Let

$$\{v_t(\omega): -\infty < t < \infty\} \tag{20.60}$$

be a stationary stochastic process satisfying

$$\mathsf{E}\{v_0\} = 0, \qquad \mathsf{E}\{v_0^2\} < \infty. \tag{20.61}$$

Suppose that the process (20.60) is measurable,† so that the integrals

$$\int_s^t v_u(\omega)\, du \tag{20.62}$$

are well defined and finite with probability 1. Let Y_n be the random element of D defined by

$$Y_n(t, \omega) = \frac{1}{\sigma\sqrt{n}} \int_0^{nt} v_s(\omega)\, ds,$$

where σ will be determined later (Y_n, of course, lies in C). We should like to prove

$$Y_n \xrightarrow{\mathscr{D}} W. \tag{20.63}$$

We define $\{v_t\}$ to be φ-mixing if

$$|\mathsf{P}(E_1 \cap E_2) - \mathsf{P}(E_1)\, \mathsf{P}(E_2)| \leq \varphi(t)\, \mathsf{P}(E_1) \tag{20.64}$$

holds whenever E_1 lies in the σ-field generated by $\{v_u: u \leq s\}$ and E_2 lies in the σ-field generated by $\{v_u: u \geq s + t\}$, where now $\varphi(t)$ is defined for all positive t. We shall suppose that $\{v_t\}$ is φ-mixing with

$$\int_0^\infty \varphi^{\frac{1}{2}}(t)\, dt < \infty. \tag{20.65}$$

Under this assumption and the assumption (20.61), we shall prove (20.63) with σ (assumed positive) defined by

$$\sigma^2 = 2\int_0^\infty \mathsf{E}\{v_0 v_t\}\, dt. \tag{20.66}$$

Although we could imitate the arguments used before for discrete time, it is simpler to reduce the present case to the previous one. If

$$\xi_i(\omega) = \int_{i-1}^{i} v_s(\omega)\, ds,$$

---

† Doob (1953, p. 62).

then, by two applications of Fubini's theorem, $\mathsf{E}\{\xi_0\} = 0$ and

$$(20.67)\qquad \mathsf{E}\{\xi_0^2\} \le \mathsf{E}\left\{\left[\int_0^1 |v_s|\, ds\right]^2\right\} \le \mathsf{E}\left\{\int_0^1 v_s^2\, ds\right\} = \mathsf{E}\{v_0^2\} < \infty.$$

It is not hard to check that (20.66) and (20.46) define the same σ. Since (20.64) implies that $\{\xi_n\}$ satisfies (20.2), and since (20.65) implies $\Sigma\varphi^{\frac{1}{2}}(n) < \infty$ (provided φ has the minimal value consistent with (20.64)), it follows by Theorem 20.1 that, if X_n is defined by (20.45), then

$$(20.68)\qquad X_n \xrightarrow{\mathscr{D}} W.$$

Now

$$\delta_n = \sup_t |Y_n(t) - X_n(t)| \le \frac{1}{\sigma\sqrt{n}} \max_{1\le i\le n} \int_{i-1}^{i} |v_t|\, dt,$$

so that

$$\mathsf{P}\{\delta_n \ge \varepsilon\} \le \frac{1}{\varepsilon^2\sigma^2} \int_{\{|\zeta|\ge\varepsilon\sigma\sqrt{n}\}} \zeta^2\, d\mathsf{P},$$

where $\zeta = \int_0^1 |v_s|\, ds$. Since $\mathsf{E}\{\zeta^2\} < \infty$ by (20.67), $\delta_n \xrightarrow{P} 0$, and (20.63) follows by (20.68) and Theorem 4.1.

The relation (20.63) persists if n goes to infinity in a continuous manner. If v_t is interpreted as the velocity of a particle at time t, the integral (20.62) is the displacement it undergoes during the time interval (s, t); (20.63) asserts that the particle is approximately in Brownian motion.

Nonstationarity

Returning to the case of discrete time, let us generalize Theorem 16.3 by showing that in Theorem 20.1 we may replace P, the probability measure governing the ξ_n, by an arbitrary probability measure P_0 absolutely continuous with respect to it. Under P_0 the process need not be stationary.

THEOREM 20.2 *Theorem* 20.1 *remains valid if* P *is replaced by an arbitrary probability measure* P_0 *dominated by it.*

Proof. Define X_n by (20.45) and X_n' by

$$X_n'(t, \omega) = \frac{1}{\sigma\sqrt{n}} \sum_{p_n \le i \le nt} \xi_i(\omega),$$

where $p_n \to \infty$ but $p_n/\sqrt{n} \to 0$. As in Section 16 (see (16.10)), we have

$$\sup_t |X_n(t) - X_n'(t)| \xrightarrow{P} 0.$$

If E lies in $\mathscr{M}_{-\infty}^k$, then, for each A in $\mathscr{D}$,

$$|\mathsf{P}(\{X_n' \in A\} \cap E) - \mathsf{P}\{X_n' \in A\}\mathsf{P}(E)| \le \varphi(p_n - k) \to 0.$$

Since this is true for every k, the proof now goes through precisely as the proof of Theorem 16.3.

In Example 4 (p. 169) we may thus substitute Lebesgue measure for Gauss's measure.

The proof of Theorem 17.2 carries over in exactly the same way:

THEOREM 20.3 *If the hypotheses of Theorem* 20.1 *are satisfied, if* ν_n *are positive-integer-valued random variables such that* $\nu_n/a_n \xrightarrow{P} \theta$, *where* θ *is a positive random variable and* $a_n \to \infty$, *then* (*if* $\sigma > 0$) $Y_n \xrightarrow{\mathscr{D}} W$, *where*

$$Y_n(t, \omega) = \frac{1}{\sigma\sqrt{\nu_n(\omega)}} S_{[\nu_n(\omega)t]}(\omega).$$

If $\mathsf{P}\{\xi_0 \in H\} > 0$, then, by Theorem 20.2, all our limit theorems persist if we compute the probabilities conditionally on the event $\{\xi_0 \in H\}$. It is possible to go further and work with probabilities conditional on an event $\{\xi_0 = \alpha\}$. To carry this through, we need another lemma. As usual, we assume $\{\xi_n\}$ to be φ-mixing.

LEMMA 5 *If* $E_2 \in \mathscr{M}_{k+n}^{\infty}$, *then*

$$|\mathsf{P}\{E_2 \,\|\, \mathscr{M}_{-\infty}^{k}\} - \mathsf{P}(E_2)| \leq \varphi_n \tag{20.69}$$

with probability 1.

Proof. Suppose $\mathsf{P}\{E_2 \,\|\, \mathscr{M}_{-\infty}^{k}\} - \mathsf{P}(E_2) > \varphi_n$ holds on some set E_1 that has positive measure and lies in $\mathscr{M}_{-\infty}^{k}$. Integrating over E_1 produces a contradiction to (20.2). Thus $\mathsf{P}\{E_2 \,\|\, \mathscr{M}_{-\infty}^{k}\} - \mathsf{P}(E_2) \leq \varphi_n$ with probability 1. The opposite inequality is treated in the same way.

Let us now suppose that $\{\xi_n\}$ is regular in the following sense. In the first place, we suppose that the ξ_n generate the σ-field $\mathscr{B}$ (this is really no restriction). In the second place, we assume that there exist conditional probability measures on $\mathscr{B}$ relative to $\mathscr{M}_{-\infty}^{0}$. We assume, that is, that there exists a function $\mathsf{P}_\omega(M)$ such that, for each ω in Ω, $\mathsf{P}_\omega(\cdot)$ is a probability measure on $\mathscr{B}$, and such that, for each E in $\mathscr{B}$, $\mathsf{P}_{.}(E)$ is a version of the conditional probability $\mathsf{P}\{E \,\|\, \mathscr{M}_{-\infty}^{0}\}$ (and hence is measurable $\mathscr{M}_{-\infty}^{0}$). This regularity condition holds, for example, if $(\Omega, \mathscr{B})$ is the product of a doubly infinite sequence of copies of $(R^1, \mathscr{R}^1)$ and the ξ_n are the coordinate variables.†

THEOREM 20.4 *Suppose* $\{\xi_n\}$ *is regular in the sense described. Under the hypotheses of Theorem* 20.1, *there exists in* $\mathscr{B}$ *a set* E^*, *with* $\mathsf{P}(E^*) = 0$, *such*

† See Loeve (1960, p. 361).

hat, if $\omega_0 \notin E^*$, then the conclusion of this theorem remains valid when P *s replaced by* P_{ω_0}.

Proof. Let $S_n = \xi_1 + \cdots + \xi_n$ and define X_n by (20.45). For positive ntegers u and n let Z_{un} be the random element of D defined by

$$Z_{un}(t, \omega) = \frac{1}{\sigma\sqrt{n}} \sum_{u \le i \le nt} \xi_i(\omega).$$

f

$$\delta_{un}(\omega) = \sup_t |X_n(t, \omega) - Z_{un}(t, \omega)|,$$

hen

$$|\delta_{un}(\omega)| \le \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{u} |\xi_i(\omega)|.$$

Let E_0 be the set of ω for which $\xi_i(\omega)$ is $\pm\infty$ for some i; then $\mathsf{P}(E_0) = 0$ and, for each u,

(20.70) $$\lim_{n\to\infty} \delta_{un}(\omega) = 0, \qquad \omega \notin E_0.$$

Now $X_n \xrightarrow{\mathscr{D}} W$ by Theorem 20.1, and it follows by (20.70) and Theorem 4.1 that, for each u,

(20.71) $$Z_{un} \xrightarrow{\mathscr{D}} W$$

as $n \to \infty$. In (20.71) P is the measure governing the ξ_n.

Since $\mathsf{P}(E_0) = 0$, we have $\mathsf{P}_{\omega_0}(E_0) = 0$ for ω_0 outside some exceptional set E_1 of P-measure 0. If F is a closed set in D and $F_\varepsilon = \{x : d(x, F) \le \varepsilon\}$ with d either of the metrics for the Skorohod topology, it follows by (20.70) that

(20.72) $$\limsup_{n\to\infty} \mathsf{P}_{\omega_0}\{\omega : X_n(\omega) \in F\} \le \limsup_{n\to\infty} \mathsf{P}_{\omega_0}\{\omega : Z_{un}(\omega) \in F_\varepsilon\}$$

for ω_0 outside E_1.

According to Lemma 5, if $E \in \mathscr{M}_u^\infty$, then

(20.73) $$|\mathsf{P}_{\omega_0}(E) - \mathsf{P}(E)| \le \varphi(u)$$

for ω_0 outside an exceptional set (of P-measure 0) depending on E. Since $\mathscr{M}_u^\infty$ is generated by $\{\xi_u, \xi_{u+1}, \ldots\}$, it is generated by a countable subfield. Since $\mathsf{P}_{\omega_0}(\cdot)$ is actually a measure, it follows that (20.73) holds for all $u \ge 1$ and all $E \in \mathscr{M}_u^\infty$, provided ω_0 lies outside some grand exceptional set E_2. Now $\{\omega : Z_{un}(\omega) \in F_\varepsilon\}$ lies in $\mathscr{M}_u^\infty$; hence

(20.74) $$\mathsf{P}_{\omega_0}\{\omega : Z_{un}(\omega) \in F_\varepsilon\} \le \mathsf{P}\{\omega : Z_{un}(\omega) \in F_\varepsilon\} + \varphi(u)$$

for $\omega_0 \notin E_2$.

If $\omega_0 \notin E^* = E_1 \cup E_2$, then (20.72) and (20.74) both apply, so that

$$\limsup_{n\to\infty} \mathsf{P}_{\omega_0}\{\omega: X_n(\omega) \in F\} \leq \limsup_{n\to\infty} \mathsf{P}\{\omega: Z_{un}(\omega) \in F_\varepsilon\} + \varphi(u)$$

for each F and each ε. Since (20.71) holds in the sense of P and F_ε is closed, the limit superior on the right here is at most $\mathsf{P}\{W \in F_\varepsilon\}$. Letting $\varepsilon \to 0$ and $u \to \infty$, we conclude that

$$\limsup_{n\to\infty} \mathsf{P}_{\omega_0}\{\omega: X_n(\omega) \in F\} \leq \mathsf{P}\{W \in F\}$$

for all closed F if $\omega_0 \notin E^*$, which, by Theorem 2.1, proves the theorem.

If we have a one-sided sequence $\xi_1, \xi_2, \ldots$, then Theorem 20.4 still holds if we compute probabilities conditionally on $\{\xi_1 = \alpha\}$.

21. FUNCTIONS OF MIXING PROCESSES

In this section we shall generalize the results of the preceding one by analyzing stationary processes $\{\eta_n\}$ for which each η_n is a function of the entire process $\{\xi_n\}$, which we assume to be stationary and φ-mixing, as before.

Preliminaries

Let f be a measurable mapping from the space of doubly infinite sequences $(\ldots, \alpha_{-1}, \alpha_0, \alpha_1, \ldots)$ of real numbers into the real line:

$$f(\ldots, \alpha_{-1}, \alpha_0, \alpha_1, \ldots) \in R^1.$$

Define random variables

$$(21.1)\quad \eta_n = f(\ldots, \xi_{n-1}, \xi_n, \xi_{n+1}, \ldots), \qquad n = 0, \pm 1, \pm 2, \ldots,$$

where ξ_n occupies the 0th place in the argument of f. We shall derive limit theorems for the process $\{\eta_n\}$. Although the η_n must be real valued, the ξ_n may now take values in R^k or even in a general measurable space.

If the value of f depends on only finitely many of the coordinates of its argument, then $\{\eta_n\}$ is again φ-mixing (with a new function φ), and hence the theorems of Section 20 apply. (Note that $\{\eta_n\}$ is in any case stationary.) If f effectively involves all the coordinates of its argument, $\{\eta_n\}$ need not be φ-mixing. We shall obtain limit theorems for $\{\eta_n\}$ under the assumption that η_n can be closely approximated by functions depending only on finitely many of the ξ_n.

Let f_l be a measurable mapping from R^{2l+1} into R^1. We write the general point of R^{2l+1} as $(\alpha_{-l}, \ldots, \alpha_0, \ldots, \alpha_l)$, so that the value of f_l is in the form

$$f_l(\alpha_{-l}, \ldots, \alpha_0, \ldots, \alpha_l).$$

Put

(21.2) $$\eta_{ln} = f_l(\xi_{n-l}, \ldots, \xi_n, \ldots, \xi_{n+l})$$

(the subscript ln is a pair, not a product). The process $\{\eta_{ln}\}$ is stationary for each l. We shall assume of the function f and the process $\{\xi_n\}$ that

$$\mathsf{E}\{\eta_0\} = 0, \qquad \mathsf{E}\{\eta_0^2\} < \infty.$$

We shall assume further that there exist functions f_l such that η_{ln} (defined by (21.2)) has a finite second moment and such that, if

(21.3) $$\nu_l = \nu(l) = \mathsf{E}\{|\eta_0 - \eta_{l0}|^2\},$$

then

$$\sum_{l=1}^{\infty} \nu_l^{\frac{1}{2}} < \infty.$$

This is the sense in which η_n is assumed well approximable by functions of only finitely many of the ξ_n.

If, in place of the doubly infinite sequence (20.1), we are presented with a one-sided sequence (20.16), then we must assume in place of (21.1) that η_n is given by

(21.4) $$\eta_n = f(\xi_n, \xi_{n+1}, \ldots),$$

where now f is a real, measurable function of one-sided sequences $(\alpha_1, \alpha_2, \ldots)$. In this case, we take f_l to be a measurable mapping from R^l into R^1, we define

(21.5) $$\eta_{ln} = f_l(\xi_n, \ldots, \xi_{n+l-1}),$$

and we assume that, if

(21.6) $$\nu_l = \mathsf{E}\{|\eta_1 - \eta_{l1}|^2\},$$

then $\Sigma \nu_l^{\frac{1}{2}}$ converges as before. Let us replace $\{\xi_n\}$ by its doubly infinite extension (see the discussion following (20.16)). If we rewrite $f(\alpha_1, \alpha_2, \ldots)$ as a function $f(\ldots, \alpha_{-1}, \alpha_0, \alpha_1, \ldots)$ effectively involving only the coordinates with positive indices, and similarly for f_l, then (21.1) and (21.2) have the same finite-dimensional distributions as (21.4) and (21.5). It follows that the square root of (21.3) is summable in l and hence that to each limit theorem for functions of (20.1) there corresponds a limit theorem for functions of (20.16). As before, therefore, we may confine our attention to the doubly infinite case.

Since η_{l0}, as defined by (21.2), is measurable $\mathscr{M}_{-l}^{l}$, it follows by Jensen's inequality for conditional expected values† that

$$\mathsf{E}\{|\eta_{l0} - \mathsf{E}\{\eta_0 \,\|\, \mathscr{M}_{-l}^{l}\}|^2\} \leq \mathsf{E}\{\mathsf{E}\{|\eta_{l0} - \eta_0|^2 \,\|\, \mathscr{M}_{-l}^{l}\}\} = \nu_l.$$

† Doob (1953) or Billingsley (1965). In all that follows, we take $\mathsf{E}\{\xi \,\|\, \mathscr{F}\}$ to be measurable $\mathscr{F}$ (rather than equal almost everywhere to a function that is measurable $\mathscr{F}$).

Therefore, by Minkowski's inequality,

$$\mathsf{E}^{\frac{1}{2}}\{|\eta_0 - \mathsf{E}\{\eta_0 \parallel \mathscr{M}_{-l}^{l}\}|^2\} \leq 2\nu_l^{\frac{1}{2}}.$$

Since $\Sigma 2\nu_l^{\frac{1}{2}}$ converges if $\Sigma\nu_l^{\frac{1}{2}}$ does, there is no loss of generality in assuming that

(21.7) $$\eta_{ln} = \mathsf{E}\{\eta_n \parallel \mathscr{M}_{n-l}^{n+l}\}.\dagger$$

The following lemma implies that, if η_{ln} is given by (21.7), then ν_l is nonincreasing.

LEMMA 1 *Let $\mathscr{F}$ and $\mathscr{G}$ be σ-fields with $\mathscr{F} \subset \mathscr{G}$. If $\mathsf{E}\{\xi^2\} < \infty$, then*

(21.8) $$\mathsf{E}\{|\xi - \mathsf{E}\{\xi \parallel \mathscr{G}\}|^2\} \leq \mathsf{E}\{|\xi - \mathsf{E}\{\xi \parallel \mathscr{F}\}|^2\}.$$

Proof. The following equalities and inequalities all hold with probability 1. If $\eta = \xi - \mathsf{E}\{\xi \parallel \mathscr{F}\}$, then $\xi - \mathsf{E}\{\xi \parallel \mathscr{G}\} = \eta - \mathsf{E}\{\eta \parallel \mathscr{G}\}$. Hence it suffices to prove $\mathsf{E}\{|\eta - \mathsf{E}\{\eta \parallel \mathscr{G}\}|^2\} \leq \mathsf{E}\{\eta^2\}$. But

$$\mathsf{E}\{|\eta - \mathsf{E}\{\eta \parallel \mathscr{G}\}|^2 \parallel \mathscr{G}\} = \mathsf{E}\{\eta^2 \parallel \mathscr{G}\} - \mathsf{E}^2\{\eta \parallel \mathscr{G}\} \leq \mathsf{E}\{\eta^2 \parallel \mathscr{G}\}.$$

Take expected values.

Functional Central Limit Theorem

Write

(21.9) $$S_n = \eta_1 + \cdots + \eta_n$$

and define X_n by

(21.10) $$X_n(t, \omega) = \frac{1}{\sigma\sqrt{n}} S_{[nt]}(\omega).$$

THEOREM 21.1 *Suppose that $\{\xi_n\}$ is φ-mixing with $\Sigma\, \varphi_n^{\frac{1}{2}} < \infty$ and that the η_n defined by* (21.1) *have mean* 0 *and finite variance. Suppose further that there exist random variables of the form* (21.2) *such that $\Sigma\nu_l^{\frac{1}{2}} < \infty$, where ν_l is defined by* (21.3). *Then the series*

(21.11) $$\sigma^2 = \mathsf{E}\{\eta_0\} + 2\sum_{k=1}^{\infty}\mathsf{E}\{\eta_0\eta_k\}$$

converges absolutely; if $\sigma^2 > 0$ and X_n is defined by (21.10), *then*

(21.12) $$X_n \xrightarrow{\mathscr{D}} W.$$

† According to Doob (1953, p. 603), $\mathsf{E}\{\eta_n \parallel \mathscr{M}_{n-l}^{n+l}\}$ can be represented in the form (21.2) because it is measurable $\mathscr{M}_{n-l}^{n+l}$. This does not really matter, however: In the proofs we use only the fact that $\{\eta_{ln}\}$ is stationary for each l and the fact that η_{ln} is measurable $\mathscr{M}_{n-l}^{n+l}$.

Proof. We first show that (21.11) converges absolutely. For each k and i,

$$|\mathsf{E}\{\eta_0\eta_k\}| \le |\mathsf{E}\{\mathsf{E}\{\eta_0 \,\|\, \mathscr{M}^i_{-i}\}\mathsf{E}\{\eta_k \,\|\, \mathscr{M}^{k+i}_{k-i}\}\}| + 2\mathsf{E}^{\frac{1}{2}}\{|\mathsf{E}\{\eta_0 \,\|\, \mathscr{M}^i_{-i}\}|^2\}\mathsf{E}^{\frac{1}{2}}\{|\eta_0 - \mathsf{E}\{\eta_0 \,\|\, \mathscr{M}^i_{-i}\}|^2\} + \mathsf{E}\{|\eta_0 - \mathsf{E}\{\eta_0 \,\|\, \mathscr{M}^i_{-i}\}|^2\}.$$

If $i = [\frac{1}{3}k]$, then, by Lemma 1 of the preceding section,

$$|\mathsf{E}\{\mathsf{E}\{\eta_0 \,\|\, \mathscr{M}^i_{-i}\}\mathsf{E}\{\eta_k \,\|\, \mathscr{M}^{k+i}_{k-i}\}\}| \le 2\varphi_i^{\frac{1}{2}}\mathsf{E}\{\mathsf{E}^2\{\eta_0 \,\|\, \mathscr{M}^i_{-i}\}\} \le 2\varphi_i^{\frac{1}{2}}\mathsf{E}\{\eta_0^2\},$$

and therefore

$$(21.13)\quad |\mathsf{E}\{\eta_0\eta_k\}| \le 2\varphi^{\frac{1}{2}}\left(\left[\frac{k}{3}\right]\right)\mathsf{E}\{\eta_0^2\} + 2\mathsf{E}^{\frac{1}{2}}\{\eta_0^2\}\nu^{\frac{1}{2}}\left(\left[\frac{k}{3}\right]\right) + \nu\left(\left[\frac{k}{3}\right]\right).$$

Because of the assumed convergence of $\Sigma\varphi_n^{\frac{1}{2}}$ and $\Sigma\nu_l^{\frac{1}{2}}$, (21.11) does converge absolutely.

Although it is possible to derive (21.12) from Theorem 19.2, it is more convenient to use the results of the preceding section. We first prove

$$(21.14)\qquad \frac{1}{\sqrt{n}} S_n \xrightarrow{\mathscr{D}} N(0, \sigma^2).$$

We shall assume, as we may, that η_{ln} is given by (21.7). Put

$$(21.15)\qquad \delta_{ln} = \eta_n - \eta_{ln},$$

so that $\nu(l) = \mathsf{E}\{\delta^2_{ln}\}$ and $\mathsf{E}\{\delta_{ln}\} = \mathsf{E}\{\eta_{ln}\} = 0$. We have

$$(21.16)\qquad \frac{1}{\sqrt{n}} S_n = \frac{1}{\sqrt{n}}\sum_{j=1}^n \eta_{lj} + \frac{1}{\sqrt{n}}\sum_{j=1}^n \delta_{lj},$$

and the idea in proving (21.14) is to show, by using Theorem 20.1, that the first sum on the right in (21.16) is approximately normal for large n and then to show that the second sum is small for large l.

First of all

$$\mathsf{E}\{|\eta_{l0} - \mathsf{E}\{\eta_{l0} \,\|\, \mathscr{M}^i_{-i}\}|^2\} \le \nu_i,$$

since $i \ge l$ implies that the left side vanishes, whereas $i \le l$ implies that it coincides with $\mathsf{E}\{\mathsf{E}^2\{\eta_0 - \mathsf{E}\{\eta_0 \,\|\, \mathscr{M}^i_{-i}\} \,\|\, \mathscr{M}^l_{-l}\}\}$, which by Jensen's theorem does not exceed ν_i. In the argument leading to (21.13) we may therefore replace η_0 by η_{l0} and η_k by η_{lk} and, using the inequality $\mathsf{E}\{\eta_{l0}^2\} \le \mathsf{E}\{\eta_0^2\}$, conclude that $|\mathsf{E}\{\eta_{l0}\eta_{lk}\}|$ does not exceed the right side of (21.13). Therefore the

$$\sigma_l^2 = \mathsf{E}\{\eta_{l0}^2\} + 2\sum_{k=1}^{\infty}\mathsf{E}\{\eta_{l0}\eta_{lk}\}$$

series converges absolutely and uniformly in l; since it converges termwise to (21.11), we have

$$\lim_{l\to\infty} \sigma_l^2 = \sigma^2. \tag{21.17}$$

It is easy to check that, for each l, $\{\eta_{ln}\}$ is $\varphi^{(l)}$-mixing, where

$$\varphi^{(l)}(n) = \begin{cases} 1 & \text{if } \ n \le 2l, \\ \varphi(n-2l) & \text{if } \ n > 2l. \end{cases}$$

The convergence of $\Sigma(\varphi(n))^{\frac{1}{2}}$ implies that of $\Sigma(\varphi^{(l)}(n))^{\frac{1}{2}}$, and it follows by Theorem 20.1 that, for each fixed l,

$$\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \eta_{lj} \xrightarrow{\mathscr{D}} N(0, \sigma_l^2) \tag{21.18}$$

as $n \to \infty$.

Now $N(0, \sigma_l^2) \xrightarrow{\mathscr{D}} N(0, \sigma^2)$ by (21.17); because of (21.16) and (21.18), the relation (21.14) will follow by Theorem 4.2 if we show that

$$\lim_{l\to\infty} \limsup_{n\to\infty} \mathsf{P}\left\{\left|\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \delta_{lj}\right| \ge \varepsilon\right\} = 0 \tag{21.19}$$

for each positive ε. To this end, we estimate the variance of $\Sigma_{j=1}^{n} \delta_{lj}$.

If m denotes the maximum of l and i, then, by the definition (21.15), $\delta_{l0} - \mathsf{E}\{\delta_{l0} \,\|\, \mathscr{M}_{-i}^{i}\} = \delta_{m0}$. Since $\nu(m) \le \nu(i)$, we obtain

$$\mathsf{E}\{|\delta_{l0} - \mathsf{E}\{\delta_{l0} \,\|\, \mathscr{M}_{-i}^{i}\}|^2\} \le \nu_i.$$

In the argument leading to (21.13) we may therefore replace η_0 by δ_{l0} and η_k by δ_{lk}; since $\mathsf{E}\{\delta_{l0}^2\} \le \mathsf{E}\{\eta_0^2\}$ by Lemma 1, $|\mathsf{E}\{\delta_{l0}\delta_{lk}\}|$ cannot exceed the right side of (21.13). Therefore the series

$$\tau_l^2 = \mathsf{E}\{\delta_{l0}^2\} + 2\sum_{k=1}^{\infty} \mathsf{E}\{\delta_{l0}\delta_{lk}\}$$

converges absolutely and uniformly in l. Since each term goes to 0 as $l \to \infty$,

$$\lim_{l\to\infty} \tau_l^2 = 0. \tag{21.20}$$

By Lemma 3 of the preceding section applied to $\{\delta_{ln}\}$ (the sequence $\{\delta_{ln}\}$ is not φ-mixing, but the lemma does not require this),

$$\lim_{n\to\infty} \frac{1}{n} \mathsf{E}\left\{\left|\sum_{j=1}^{n} \delta_{lj}\right|^2\right\} = \tau_l^2. \tag{21.21}$$

Chebyshev's inequality and (21.20) now yield (21.19), which completes the proof of (21.14).

From the absolute convergence of the series (21.11), it follows via Lemma 3 of the preceding section (recall once again that this lemma requires only stationarity—φ-mixing is not needed) that

$$\frac{1}{n}\,\mathsf{E}\{S_n^{\,2}\} \to \sigma^2. \tag{21.22}$$

If $\sigma^2 = 0$, (21.14) is the same as $S_n/\sqrt{n} \xrightarrow{P} 0$. From now on we assume that $\sigma^2 > 0$. To prove (21.12), we establish the convergence of the finite-dimensional distributions and then tightness. Let p_n be positive integers going to infinity at a rate to be specified later and define

$$U_{ni} = \mathsf{E}\{S_{i-2p_n} \,\|\, \mathscr{M}_{-\infty}^{i-p_n}\}, \tag{21.23}$$

and

$$V_{ni} = \mathsf{E}\{S_n - S_{i+2p_n} \,\|\, \mathscr{M}_{i+p_n}^{\infty}\}. \tag{21.24}$$

In these definitions we adopt the conventions that $S_{i-2p_n} = 0$ if $i < 2p_n$ and $S_n - S_{i+2p_n} = 0$ if $i + 2p_n > n$. We shall often write p in place of p_n.

By Minkowski's inequality and Lemma 1,

$$\begin{aligned}\mathsf{E}^{\frac{1}{2}}\{|S_k - \mathsf{E}\{S_k \,\|\, \mathscr{M}_{-\infty}^{k+p}\}|^2\} &\le \sum_{j=1}^{k} \mathsf{E}^{\frac{1}{2}}\{|\eta_j - \mathsf{E}\{\eta_j \,\|\, \mathscr{M}_{-\infty}^{k+p}\}|^2\} \\ &\le \sum_{j=1}^{k} \nu^{\frac{1}{2}}(k + p - j).\end{aligned}$$

If

$$\mu(p) = \sum_{i=p}^{\infty} \nu^{\frac{1}{2}}(i), \tag{21.25}$$

then

$$\mathsf{E}\{|S_k - \mathsf{E}\{S_k \,\|\, \mathscr{M}_{-\infty}^{k+p}\}|^2\} \le \mu^2(p), \tag{21.26}$$

and it follows that

$$\mathsf{E}\{|U_{ni} - S_i|^2\} \le 2\mathsf{E}\{S_{2p_n}^2\} + 2\mu^2(p_n) \tag{21.27}$$

for all i. In the same way we obtain

$$\mathsf{E}\{|V_{ni} - (S_n - S_i)|^2\} \le 2\mathsf{E}\{S_{2p_n}^2\} + 2\mu^2(p_n). \tag{21.28}$$

Since U_{ni} and V_{ni} are measurable $\mathscr{M}_{-\infty}^{i-p}$ and $\mathscr{M}_{i+p}^{\infty}$, respectively,

$$|\mathsf{P}\{U_{ni} \in H_1, V_{ni} \in H_2\} - \mathsf{P}\{U_{ni} \in H_1\}\mathsf{P}\{V_{ni} \in H_2\}| \le \varphi(2p_n) \tag{21.29}$$

for all linear Borel sets H_1 and H_2.

Consider now the finite-dimensional distributions. From (21.14) it follows that

$$X_n(t) - X_n(s) \xrightarrow{\mathscr{D}} W_t - W_s. \tag{21.30}$$

We shall prove that

$$(21.31) \qquad (X_n(t), X_n(1) - X_n(t)) \xrightarrow{\mathscr{D}} (W_t, W_1 - W_t);$$

the argument is easily adapted to cases of dimension exceeding 2. Let p_n go to infinity slowly enough that $p_n/n \to 0$. Since $\mathsf{E}\{S_k^2\} \sim k\sigma^2$ by (21.22), it follows by the inequality (21.27) and Chebyshev's inequality that we have the convergence relation

$$(21.32) \qquad \frac{1}{\sigma\sqrt{n}} U_{n,[nt]} - X_n(t) \xrightarrow{P} 0.$$

Similarly,

$$(21.33) \qquad \frac{1}{\sigma\sqrt{n}} V_{n,[nt]} - (X_n(1) - X_n(t)) \xrightarrow{P} 0,$$

and (21.31) will follow if

$$\left(\frac{1}{\sigma\sqrt{n}} U_{n,[nt]}, \frac{1}{\sigma\sqrt{n}} V_{n,[nt]}\right) \xrightarrow{\mathscr{D}} (W_t, W_1 - W_t).$$

Because of (21.29), it is enough (Theorem 3.1) to show that there is convergence in distribution in each of the two coordinates here; but this follows from (21.30) by (21.32) and (21.33).

We turn now to the question of tightness and prove that, for each positive ε, there exists a $\lambda > \sigma$ such that

$$(21.34) \qquad \mathsf{P}\left\{\max_{i \le n} |S_i| \ge \lambda\sqrt{n}\right\} \le \frac{\varepsilon}{\lambda^2}$$

for all sufficiently large n. Put

$$S_n^* = \sum_{j=1}^{n} |\eta_j|.$$

By the argument leading to (20.51), there is a sequence p_n going to infinity so slowly that, if

$$(21.35) \qquad \beta_n(\lambda) = \mathsf{P}\{S_{2p_n}^* \ge \lambda\sqrt{n}\},$$

then

$$(21.36) \qquad \lim_{n \to \infty} n\beta_n(\lambda) = 0$$

for each positive λ.

Consider the variables (21.23) and (21.24) for this choice of p_n. With the definition (21.35), we have

$$\mathsf{P}\{|S_i - U_{ni}| \ge \lambda\sqrt{n}\} \le \mathsf{P}\{|S_{i-2p_n} - \mathsf{E}\{S_{i-2p_n} \| \mathscr{M}_{-\infty}^{i-p_n}\}| \ge \tfrac{1}{2}\lambda\sqrt{n}\} + \beta_n(\tfrac{1}{2}\lambda),$$

so that, by (21.26),

$$(21.37) \qquad \mathsf{P}\{|U_{ni} - S_i| \ge \lambda\sqrt{n}\} \le \frac{4\mu^2(p_n)}{\lambda^2 n} + \beta_n(\tfrac{1}{2}\lambda)$$

for all i. Similarly,

$$\mathsf{P}\{|V_{ni} - (S_n - S_i)| \geq \lambda\sqrt{n}\} \leq \frac{4\mu^2(p_n)}{\lambda^2 n} + \beta_n(\tfrac{1}{2}\lambda). \tag{21.38}$$

By (21.14) and (21.22) it follows via Theorem 5.4 that the variables S_n^2/n are uniformly integrable. Therefore there exists a $\lambda > \sigma$ such that

$$\mathsf{P}\{|S_k| \geq \lambda\sqrt{k}\} \leq \frac{\varepsilon}{\lambda^2} \tag{21.39}$$

for all k. Fix such a λ. By (21.37),

$$\mathsf{P}\left\{\max_{i \leq n} |S_i| \geq 6\lambda\sqrt{n}\right\} \leq \mathsf{P}\left\{\max_{i \leq n} |U_{ni}| \geq 5\lambda\sqrt{n}\right\} + 4\frac{\mu^2(p_n)}{\lambda^2} + n\beta_n(\tfrac{1}{2}\lambda). \tag{21.40}$$

Consider the sets

$$E_i = \left\{\max_{j \leq i} |U_{nj}| < 5\lambda\sqrt{n} \leq |U_{ni}|\right\}.$$

We have

$$\mathsf{P}\left\{\max_{i \leq n} |U_{ni}| \geq 5\lambda\sqrt{n}\right\} \leq \mathsf{P}\{|S_n| \geq \lambda\sqrt{n}\} + \sum_{i=1}^{n-1} \mathsf{P}(E_i \cap \{|S_n - U_{ni}| \geq 4\lambda\sqrt{n}\}). \tag{21.41}$$

Since

$$|S_n - U_{ni}| \leq |S_n - S_i - V_{ni}| + |V_{ni}| + |S_i - U_{ni}|,$$

the ith summand in (21.41) is at most

$$\mathsf{P}\{|S_n - S_i - V_{ni}| \geq \lambda\sqrt{n}\} + \mathsf{P}(E_i \cap \{|V_{ni}| \geq 2\lambda\sqrt{n}\}) + \mathsf{P}\{|S_i - U_{ni}| \geq \lambda\sqrt{n}\}. \tag{21.42}$$

Since E_i lies in $\mathscr{M}_{-\infty}^{i-p}$ and V_{ni} is measurable $\mathscr{M}_{i+p}^{\infty}$,

$$\mathsf{P}(E_i \cap \{|V_{ni}| \geq 2\lambda\sqrt{n}\}) \leq \mathsf{P}(E_i)\mathsf{P}\{|V_{ni}| \geq 2\lambda\sqrt{n}\} + \mathsf{P}(E_i)\,\varphi(2p_n).$$

By (21.38), $\mathsf{P}\{|V_{ni}| \geq 2\lambda\sqrt{n}\} \leq \mathsf{P}\{|S_{n-i}| \geq \lambda\sqrt{n-i}\} + 4\mu^2(p_n)/\lambda^2 n + \beta_n(\tfrac{1}{2}\lambda)$, so that

$$\mathsf{P}(E_i \cap \{|V_{ni}| \geq 2\lambda\sqrt{n}\}) \leq \frac{4\mu^2(p_n)}{\lambda^2 n} + \beta_n(\tfrac{1}{2}\lambda) + \mathsf{P}(E_i)\mathsf{P}\{|S_{n-i}| \geq \lambda\sqrt{n-i}\} + \mathsf{P}(E_i)\,\varphi(2p_n).$$

Using this estimate for the middle term in (21.42) and the estimates (21.37) and (21.38) for the other two, we see that the ith summand in (21.41) is at most

$$\mathsf{P}(E_i)\mathsf{P}\{|S_{n-i}| \geq \lambda\sqrt{n-i}\} + \mathsf{P}(E_i)\varphi(2p_n) + \frac{12\mu^2(p_n)}{\lambda^2 n} + 3\beta_n(\tfrac{1}{2}\lambda).$$

Therefore, by (21.39) and the disjointness of the E_i,

$$\mathsf{P}\Big\{\max_{i\leq n}|U_{ni}| \geq 5\lambda\sqrt{n}\Big\} \leq \frac{2\varepsilon}{\lambda^2} + \varphi(2p_n) + \frac{12\mu^2(p_n)}{\lambda^2} + 3n\beta_n(\tfrac{1}{2}\lambda).$$

The last three terms on the right here go to 0 (see (21.25) and (21.36)), and it follows by (21.40) that

$$\mathsf{P}\Big\{\max_{i\leq n}|S_i| \geq 6\lambda\sqrt{n}\Big\} \leq \frac{3\varepsilon}{\lambda^2}$$

for all sufficiently large n, which proves (21.34).

This completes the proof of Theorem 21.1. In particular, we have the central limit theorem (21.14), valid even if σ^2 vanishes, and this result has a multivariate version. Suppose

$$\eta_n = (\eta_n^{(1)}, \ldots, \eta_n^{(r)}), \tag{21.43}$$

where each $\eta_n^{(i)}$ is a function

$$\eta_n^{(i)} = f^{(i)}(\ldots, \xi_{n-1}, \xi_n, \xi_{n+1}, \ldots)$$

with mean 0 and finite variance. Suppose these random variables have approximations

$$\eta_{ln}^{(i)} = f_l^{(i)}(\xi_{n-l}, \ldots, \xi_n, \ldots, \xi_{n+l}).$$

If $\sum_l \nu_l^{\frac{1}{2}}(i) < \infty$ for each i, where

$$\nu_l(i) = \mathsf{E}\{|\eta_0^{(i)} - \eta_{l0}^{(i)}|^2\},$$

then, since

$$\mathsf{E}^{\frac{1}{2}}\left\{\left|\sum_{i=1}^r \alpha_i\eta_0^{(i)} - \sum_{i=1}^r \alpha_i\eta_{l0}^{(i)}\right|^2\right\} \leq \sum_{i=1}^r |\alpha_i|\, \nu_l^{\frac{1}{2}}(i),$$

any linear combination of the components of η_n satisfies the hypotheses of Theorem 21.1, so that the Cramér–Wold method applies. It follows that $n^{-\frac{1}{2}}\sum_{k=1}^n \eta_k$ is asymptotically normal; the covariances for the limit are

$$\mathsf{E}\{\eta_0^{(i)}\eta_0^{(j)}\} + \sum_{k=1}^\infty \mathsf{E}\{\eta_0^{(i)}\eta_k^{(j)}\} + \sum_{k=1}^\infty \mathsf{E}\{\eta_k^{(i)}\eta_0^{(j)}\}, \tag{21.44}$$

where the series converge absolutely.

Applications

Example 1. Let the ξ_n be independent and identically distributed with mean 0 and variance 1 and define

$$\eta_n = \sum_{i=-\infty}^{\infty} a_i \xi_{n+i},$$

where we assume $\Sigma_i a_i^2 < \infty$. The random variable (21.7) is

$$\eta_{ln} = \sum_{i=-l}^{l} a_i \xi_{n+i}$$

in this case, and the requirement $\Sigma \nu_l^{\frac{1}{2}} < \infty$ becomes

$$\sum_{l=1}^{\infty} \left(\sum_{|i|>l} a_i^2 \right)^{\frac{1}{2}} < \infty.$$

Example 2. Let $(\Omega, \mathscr{B}, \mathsf{P})$ be the unit interval with Lebesgue measure, as in Example 3 of the preceding section (p. 169), and let $\xi_n(\omega)$ be the nth digit in the dyadic expansion of ω $(n \geq 1)$. Then $\{\xi_1, \xi_2, \ldots\}$ is stationary and independent. A random variable η_n of the form (21.4) can be regarded as a function $\eta_n(\omega) = f(T^{n-1}\omega)$, where f is a function on the unit interval and T is the transformation $T\omega = 2\omega \pmod 1$. Suppose that $\int_0^1 f(\omega)^2 \, d\omega$ is finite and that $\Sigma \nu_l^{\frac{1}{2}}$ converges, where

$$\nu_l = \int_0^1 |f(\omega) - f_l(\omega)|^2 \, d\omega$$

and each f_l depends on ω only through the first l digits $\xi_1(\omega), \ldots, \xi_l(\omega)$ of its dyadic expansion. It then follows by Theorem 21.1 that, if $\mu = \int_0^1 f(\omega) \, d\omega$,

$$\frac{1}{\sqrt{n}} \left(\sum_{i=1}^{n} f(T^i \omega) - \mu n \right) \xrightarrow{\mathscr{D}} N(0, \sigma^2) \tag{21.45}$$

for an appropriate asymptotic variance σ^2. And there is the corresponding functional central limit theorem. If there exist such functions f_l at all, they may be taken to be

$$f_l(\omega) = 2^l \int_{(i-1)/2^l}^{i/2^l} f(s) \, ds, \qquad \omega \in \left(\frac{i-1}{2^l}, \frac{i}{2^l} \right] \tag{21.46}$$

(see (21.7)).

If $f(\omega) = I_{(0,t)}(\omega)$ and f_l is defined by (21.46), then $\nu_l \leq 2^{-l}$, so that (21.45) holds $(\mu = t)$. In this case $\Sigma_{i=1}^n f(T^i\omega)$ is the number of i, $1 \leq i \leq n$, for which $T^{i-1}\omega < t$. It t is a dyadic rational, f coincides with f_l for some l, and (21.45) follows by the central limit theorem for $(l-1)$-dependent variables. If t is not a dyadic rational, $f(\omega)$ involves the entire expansion of ω.

If f is continuous and f_l is defined by (21.46), then $\nu_l \leq 4w_f^2(2^{-l})$, where $w_f(\delta)$ is the modulus of continuity of f. Therefore (21.45) is true if $\Sigma_l\, w_f(2^{-l})$ converges, a condition which holds, for example, if f satisfies a uniform Hölder condition of some positive order:

$$|f(\omega) - f(\omega')| \leq K\,|\omega - \omega'|^\theta, \qquad K, \theta > 0. \tag{21.47}$$

For example, we may take $f(\omega) = \omega$, in which case $\eta_n(\omega) = T^n(\omega)$. (It can be verified that this last sequence $\{\eta_1, \eta_2, \ldots\}$ is not φ-mixing for any φ going to 0—to know $T^n\omega$ is to know $T^{n+1}\omega$, $T^{n+2}\omega, \ldots$ exactly.)

Example 3. Consider the $(\Omega, \mathscr{B}, \mathrm{P})$ and $\{a_1(\omega), a_2(\omega), \ldots\}$ of Example 4 of the preceding section. If T_1 denotes the continued-fraction transformation

$$T_1\omega = \frac{1}{\omega} - \left[\frac{1}{\omega}\right], \tag{21.48}$$

and if $a(\omega) = [1/\omega]$, then $a_n(\omega) = a(T_1^{n-1}\omega)$, $n \geq 1$.† As in the preceding example, (21.4) can be regarded as a function $\eta_n(\omega) = f(T_1^{n-1}\omega)$ with f defined over the unit interval. And

$$\frac{1}{\sqrt{n}}\left(\sum_{i=1}^{n} f(T_1^{\,i}\omega) - \mu n\right) \xrightarrow{\mathscr{D}} N(0, \sigma^2)$$

will hold if

$$\mu = \frac{1}{\log 2}\int_0^1 \frac{f(\omega)}{1+\omega}\, d\omega,$$

if

$$\frac{1}{\log 2}\int_0^1 \frac{f(\omega)^2}{1+\omega}\, d\omega < \infty, \tag{21.49}$$

and if $\Sigma \nu_l^{\frac{1}{2}}$ converges, where

$$\nu_l = \frac{1}{\log 2}\int_0^1 \frac{\left(f(\omega) - f_l(\omega)\right)^2}{1+\omega}\, d\omega \tag{21.50}$$

and where each f_l depends on ω only through its first l partial quotients $a_1(\omega), \ldots, a_l(\omega)$. Since $1 \leq 1 + \omega \leq 2$ (21.49) is equivalent to

$$\int_0^1 f(\omega)^2\, d\omega < \infty$$

and the requirement $\Sigma \nu_l^{\frac{1}{2}} < \infty$ is unaffected if we define ν_l by

$$\nu_l = \int_0^1 |f(\omega) - f_l(\omega)|^2\, d\omega$$

instead of by (21.50).

† See Section 4 of Billingsley (1965) for this and the other facts about T_1 used here.

Since a set $\{\omega : a_i(\omega) = a_i,\ 1 \le i \le l\}$ is an interval of length at most 2^{-l+1}, we may, for the same reasons as in Example 2 preceding, take $f(\omega)$ to be $I_{(0,t)}(\omega)$ or to be any function satisfying a Hölder condition (21.47). For another example, note that, since†

$$\left|\log \omega - \log \frac{p_l(\omega)}{q_l(\omega)}\right| \le \frac{1}{2^{l-2}},$$

where $p_l(\omega)/q_l(\omega)$ is the lth convergent, we may also take $f(\omega) = -\log \omega$; for this f, $\mu = \pi^2/(12 \log 2)$. Moreover,

$$\log q_n(\omega) = -\sum_{k=0}^{n-1} \log T_1^k\omega + 4\theta, \qquad |\theta| \le 1, \tag{21.51}$$

and it follows that

$$\frac{1}{\sqrt{n}}\left(\log q_n(\omega) - \frac{n\pi^2}{12 \log 2}\right) \xrightarrow{\mathscr{D}} N(0, \sigma^2) \tag{21.52}$$

for an appropriate positive σ^2.

The functional versions of these limit theorems are easily written out.

Diophantine Approximation

These results lead to limit theorems connected with Diophantine approximation. A fraction p/q is a best approximation to ω if it minimizes the form

$$|q' \cdot \omega - p'| \tag{21.53}$$

over fractions p'/q' with denominators q' not exceeding q. The successive best approximations to ω are‡ just the convergents $p_n(\omega)/q_n(\omega)$, $n = 1, 2, \ldots$, and so the value of the form (21.53) for the nth in the series of best approximations is

$$d_n(\omega) = |q_n(\omega) \cdot \omega - p_n(\omega)|. \tag{21.54}$$

Since§

$$|-\log d_n(\omega) - \log q_{n+1}(\omega)| \le \log 2, \tag{21.55}$$

(21.52) implies

$$\frac{1}{\sqrt{n}}\left(-\log d_n(\omega) - \frac{n\pi^2}{12 \log 2}\right) \xrightarrow{\mathscr{D}} N(0, \sigma^2). \tag{21.56}$$

---

† Billingsley (1965, p. 42).

‡ See Khinchine (1961).

§ See (4.6) on p. 42 of Billingsley (1965).

This limit theorem has a functional form. From (21.51) and Theorem 4.1 it follows that, if Y_n is the random element of D defined by

$$Y_n(t, \omega) = \frac{1}{\sigma\sqrt{n}}\left(\log q_{[nt]}(\omega) - \frac{[nt]\pi^2}{12 \log 2}\right)$$

(with the same σ as in (20.52)), then $Y_n \xrightarrow{\mathscr{D}} W$. Let

$$Z_n(t, \omega) = \frac{1}{\sigma\sqrt{n}}\left(-\log d_{[nt]}(\omega) - \frac{[nt]\pi^2}{12 \log 2}\right).$$

Now (21.55) implies a functional central limit theorem corresponding to (21.56):

$$Z_n \xrightarrow{\mathscr{D}} W. \tag{21.57}$$

From (21.57) we can derive an arc sine law for best approximations. By (21.56), $k^{-1} \log d_k \xrightarrow{P} \pi^2/(12 \log 2)$ (it can be shown that there is convergence with probability 1), so that the measure of discrepancy $d_k(\omega)$ has normal order $e^{-k\pi^2/(12 \log 2)}$. Let us say that the kth best approximation $p_k(\omega)/q_k(\omega)$ is "superior" if

$$d_k(\omega) < e^{-k\pi^2/(12 \log 2)}$$

and "inferior" otherwise. If $\theta_n(\omega)$ is the fraction of superior ones among the convergents

$$\frac{p_1(\omega)}{q_1(\omega)}, \ldots, \frac{p_n(\omega)}{q_n(\omega)}, \tag{21.58}$$

then, by (21.57) and (11.26),

$$\mathsf{P}\{\theta_n \le \alpha\} \to \frac{2}{\pi} \arcsin \sqrt{\alpha}, \qquad 0 < \alpha < 1.$$

For example, if n is large and if $p_n(\omega)/q_n(\omega)$ is superior, then the odds are approximately even that more than 85 per cent of the convergents (21.58) are also superior.

Nonstationarity

If X_n' is defined by

$$X_n'(t, \omega) = \frac{1}{\sigma\sqrt{n}} \mathsf{E}\{S_{[nt]} \,\|\, \mathscr{M}_{p_n}^{\infty}\},$$

where $p_n \to \infty$ but $p_n/\sqrt{n} \to 0$, if X_n is defined by (21.10), and if

$$\delta_n = \sup_t |X_n(t) - X_n'(t)|,$$

then

$$\delta_n \le \frac{1}{\sigma\sqrt{n}} \sum_{j=1}^{n} |\eta_j - \mathsf{E}\{\eta_j \,\|\, \mathscr{M}_{p_n}^{\infty}\}|.$$

By Minkowski's inequality and stationarity,

$$\mathsf{E}^{\frac{1}{2}}\{\delta_n^{\,2}\} \le \frac{1}{\sigma\sqrt{n}} \sum_{j=1}^{p_n} \mathsf{E}^{\frac{1}{2}}\{|\eta_j - \mathsf{E}\{\eta_j \,\|\, \mathscr{M}_{p_n}^{\infty}\}|^2\}$$
$$+ \frac{1}{\sigma\sqrt{n}} \sum_{l=1}^{\infty} \mathsf{E}^{\frac{1}{2}}\{|\eta_0 - \mathsf{E}\{\eta_0 \,\|\, \mathscr{M}_{-l}^{\infty}\}|^2\}.$$

If $\Sigma \nu_l^{\frac{1}{2}} < \infty$, then by Lemma 1 and the choice of p_n, $\mathsf{E}\{\delta_n^{\,2}\} \to 0$. By the same arguments as before, therefore, Theorems 20.2 and 20.3 carry over to processes $\{\eta_n\}$. It follows, for example, that we may replace Gauss's measure by Lebesgue measure in the applications to Diophantine approximations.

Theorem 20.4, however, does not carry over. To see this, consider the process $\{\xi_n\}$ of Example 2 of this section (p. 191) and put $\eta_n(\omega) = \Sigma_{k=n}^{\infty} \xi_k(\omega)/2^{l-n+1}$. Although $\Sigma_{i=1}^{n} \eta_i(\omega)$ is asymptotically normal (when properly normalized), this cannot be true if the probabilities are computed conditionally on $\eta_1(\omega)$: Since $\eta_1(\omega) = \omega$ it follows that, if $\eta_1(\omega) = \alpha$ $(0 < \alpha < 1)$, then the conditional distribution of $\Sigma_{i=1}^{n} \eta_i(\omega)$ is a unit mass at $\Sigma_{i=1}^{n} \eta_i(\alpha)$.

Remarks. The results of this and the preceding section, which are new, extend those in Billingsley (1956 and 1962). Central limit theorems under the hypotheses of Theorems 20.1 and 21.1 have been proved by Ibragimov (1962); see his paper for references to the earlier literature (see also Doeblin (1940), Fortet (1940), and Kac (1946)).

22. EMPIRICAL DISTRIBUTION FUNCTIONS

We shall analyze the empirical distribution functions for sequences $\{\xi_n\}$ and $\{\eta_n\}$ of the kind discussed in the preceding two sections.

φ-Mixing Processes

Throughout the section, $\{\xi_n\}$ is stationary and φ-mixing. We shall need the following estimate, related to Lemma 4 of Section 20.

LEMMA 1 *Suppose that* $|\xi_0| \le 1$ *with probability* 1, *that* $\mathsf{E}\{\xi_0\} = 0$, *and that* $\Sigma k^2 \varphi_k^{\frac{1}{2}} < \infty$. *Then*

$$\mathsf{E}\{S_n^{\,4}\} \le 288[n^2\mathsf{E}^2\{\xi_0^{\,2}\} + n\mathsf{E}\{\xi_0^{\,2}\}]\left[\sum_{k=0}^{\infty}(k+1)^2\varphi_k^{\frac{1}{2}}\right]^2. \tag{22.1}$$

Proof. We have

(22.2) $$\mathsf{E}\{S_n^4\} \leq 4!\, n \sum |\mathsf{E}\{\xi_0 \xi_i \xi_{i+j} \xi_{i+j+k}\}|,$$

with the indices constrained by

(22.3) $$i, j, k \geq 0, \qquad i + j + k \leq n.$$

Write p in place of $\mathsf{E}\{\xi_0^2\}$. We first prove these three inequalities:

(22.4) $$|\mathsf{E}\{\xi_0(\xi_i \xi_{i+j} \xi_{i+j+k})\}| \leq 2\varphi_i^{\frac{1}{2}} p,$$

(22.5) $$|\mathsf{E}\{(\xi_0 \xi_i \xi_{i+j}) \xi_{i+j+k}\}| \leq 2\varphi_k^{\frac{1}{2}} p,$$

(22.6) $$|\mathsf{E}\{(\xi_0 \xi_i)(\xi_{i+j} \xi_{i+j+k})\}| \leq 4\varphi_i^{\frac{1}{2}} \varphi_k^{\frac{1}{2}} p^2 + 2\varphi_j^{\frac{1}{2}} p.$$

By Lemma 1 of Section 20, the left member of (22.4) is at most

$$2\varphi_i^{\frac{1}{2}} \mathsf{E}^{\frac{1}{2}}\{\xi_0^2\} \mathsf{E}^{\frac{1}{2}}\{\xi_i^2 \xi_{i+j}^2 \xi_{i+j+k}^2\};$$

since $|\xi_n| \leq 1$, (22.4) follows. A similar argument leads to (22.5).

By another application of Lemma 1 of Section 20, the left member of (22.6) is at most

(22.7) $$|\mathsf{E}\{\xi_0 \xi_i\} \mathsf{E}\{\xi_0 \xi_k\}| + 2\varphi_j^{\frac{1}{2}} \mathsf{E}^{\frac{1}{2}}\{\xi_0^2 \xi_i^2\} \mathsf{E}^{\frac{1}{2}}\{\xi_0^2 \xi_k^2\}.$$

Two further applications of the same lemma give

$$|\mathsf{E}\{\xi_0 \xi_i\}| \leq 2\varphi_i^{\frac{1}{2}} p,\ |\mathsf{E}\{\xi_0 \xi_k\}| \leq 2\varphi_k^{\frac{1}{2}} p;$$

from $|\xi_n| \leq 1$ it follows that (22.7) is at most $4\varphi_i^{\frac{1}{2}} \varphi_k^{\frac{1}{2}} p^2 + 2\varphi_j^{\frac{1}{2}} p$, which proves (22.6).

Now (22.2) and the three inequalities just proved imply

$$\mathsf{E}\{S_n^4\} \leq 4!\, 4n\left(p^2 \sum_{i,k \leq j} \varphi_i^{\frac{1}{2}} \varphi_k^{\frac{1}{2}} + 3p \sum_{j,k \leq i} \varphi_i^{\frac{1}{2}}\right),$$

where the indices also obey (22.3). Since

$$\sum_{i,k \leq j} \varphi_i^{\frac{1}{2}} \varphi_k^{\frac{1}{2}} \leq \sum_{j=0}^{n} \sum_{i,k=0}^{\infty} \varphi_i^{\frac{1}{2}} \varphi_k^{\frac{1}{2}} \leq 2n\left[\sum_{k=0}^{\infty} \varphi_k^{\frac{1}{2}}\right]^2$$

and

$$\sum_{j,k \leq i} \varphi_i^{\frac{1}{2}} \leq \sum_{i=0}^{n} \sum_{j,k=0}^{i} \varphi_i^{\frac{1}{2}} \leq \sum_{k=0}^{\infty} (k+1)^2 \varphi_k^{\frac{1}{2}},$$

(22.1) is proved.

Suppose now that

$$0 \leq \xi_n(\omega) \leq 1;$$

let $F_n(t, \omega)$ be the empirical distribution function for $\xi_1(\omega), \ldots, \xi_n(\omega)$, and define Y_n by

$$Y_n(t, \omega) = \sqrt{n}\big(F_n(t, \omega) - F(t)\big), \tag{22.8}$$

where F is the distribution function for ξ_0, Let

$$g_t(\alpha) = I_{[0,t]}(\alpha) - F(t). \tag{22.9}$$

THEOREM 22.1 *Suppose $\{\xi_n\}$ is φ-mixing with $\Sigma n^2 \varphi_n^{\frac{1}{2}} < \infty$, and suppose ξ_0 has a continuous distribution function F on* $[0, 1]$. *Then*

$$Y_n \xrightarrow{\mathscr{D}} Y, \tag{22.10}$$

where Y_n is defined by (22.8) *and Y is the Gaussian random function specified by*

$$\mathsf{E}\{Y(t)\} = 0 \tag{22.11}$$

and

$$\mathsf{E}\{Y(s)Y(t)\} = \mathsf{E}\{g_s(\xi_0)g_t(\xi_0)\} + \sum_{k=1}^{\infty} \mathsf{E}\{g_s(\xi_0)g_t(\xi_k)\} + \sum_{k=1}^{\infty} \mathsf{E}\{g_s(\xi_k)g_t(\xi_0)\}. \tag{22.12}$$

These series converge absolutely and $\mathsf{P}\{Y \in C\} = 1$.

Proof. We first show that we may confine our attention to the case in which ξ_0 is uniformly distributed over $[0, 1]$. Now $\{F(\xi_n)\}$ is φ-mixing, and, since F is continuous, $F(\xi_0)$ is uniformly distributed. If $F_n'(t, \omega)$ is the empirical distribution function for $F(\xi_1(\omega)), \ldots, F(\xi_n(\omega))$, and if

$$Y_n'(t) = \sqrt{n}\big(F_n'(t, \omega) - t\big),$$

then, with probability 1,

$$Y_n'\big(F(t)\big) = Y_n(t)$$

for all t. If the theorem is true in the uniform case, then $Y_n' \xrightarrow{\mathscr{D}} Y'$, where Y' is a Gaussian random function with $\mathsf{E}\{Y'(t)\} = 0$ and (write $g_t'(\alpha) = I_{[0,t]}(\alpha) - t$)

$$\mathsf{E}\{Y'(s)Y'(t)\} = \mathsf{E}\{g_s'(F(\xi_0))g_t'(F(\xi_0))\} + \sum_{k=1}^{\infty} \mathsf{E}\{g_s'(F(\xi_0))g_t'(F(\xi_k))\} + \sum_{k=1}^{\infty} \mathsf{E}\{g_s'(F(\xi_k))g_t'(F(\xi_0))\},$$

and where $\mathsf{P}\{Y' \in C\} = 1$. Define $h: D \to D$ by $(hx)(t) = x\big(F(t)\big)$ and let $Y = h(Y')$. Since $C \subset h^{-1}C$, $\mathsf{P}\{Y \in C\} = 1$; furthermore, Y is Gaussian and satisfies (22.11) and (22.12). Since h is continuous on C, it follows by Corollary 1 to Theorem 5.1 that $Y_n' \xrightarrow{\mathscr{D}} Y'$ implies $Y_n \xrightarrow{\mathscr{D}} Y$. Thus we need only treat the uniform case.

Assume then that ξ_0 is uniformly distributed, and take $F(t) = t$ in (22.9). Since $\{g_t(\xi_n)\}$ is φ-mixing and

$$Y_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g_t(\xi_i),$$

it follows by Theorem 20.1 that $Y_n(t)$ is asymptotically normal. From the multivariate version of this result, it follows that, for each k-tuple $t_1, \ldots, t_k$, the distribution of $\bigl(Y_n(t_1), \ldots, Y_n(t_k)\bigr)$ approaches a normal distribution centered at the origin; by (20.59), the covariances of these limit distributions are those specified by (22.12) for Y. It follows also from Theorem 20.1 that the series in (22.12) are absolutely convergent.

We shall show that for each positive ε and η there exists a δ, $0 < \delta < 1$, such that

$$\mathsf{P}\{w(Y_n, \delta) \geq \varepsilon\} \leq \eta \tag{22.13}$$

for all sufficiently large n. It will follow by Theorem 15.5 that $\{Y_n\}$ is tight and that, if Y is taken as the limit in distribution of some subsequence $\{Y_{n'}\}$, then $\mathsf{P}\{Y \in C\} = 1$, $Y_n \xrightarrow{\mathscr{D}} Y$, and Y is Gaussian and satisfies (22.11) and (22.12). This will complete the proof.

Fix ε and η. Since ξ_0 is uniformly distributed,

$$\mathsf{E}\{|g_t(\xi_0) - g_s(\xi_0)|^2\} \leq |t - s|.$$

By Lemma 1 applied to $\{g_t(\xi_n) - g_s(\xi_n)\}$,

$$\mathsf{E}\left\{\left|\sum_{i=1}^{n} \bigl(g_t(\xi_i) - g_s(\xi_i)\bigr)\right|^4\right\} \leq K_1(n^2\,|t - s|^2 + n\,|t - s|),$$

where K_1 depends on φ alone. Therefore, if

$$\frac{\varepsilon}{n} \leq t - s, \tag{22.14}$$

we have (assume $\varepsilon < 1$)

$$\mathsf{E}\{|Y_n(t) - Y_n(s)|^4\} \leq \frac{2K_1}{\varepsilon}(t - s)^2. \tag{22.15}$$

Assume now that p is a number satisfying $\varepsilon/n \leq p$, and consider the random variables

$$Y_n(s + ip) - Y_n\bigl(s + (i - 1)p\bigr), \qquad i = 1, 2, \ldots, m,$$

where m is a positive integer. By (22.15) and Theorem 12.2,

$$\mathsf{P}\left\{\max_{i \leq m} |Y_n(s + ip) - Y_n(s)| \geq \lambda\right\} \leq \frac{K_2}{\varepsilon\lambda^4} m^2 p^2, \tag{22.16}$$

where $K_2 = 2K'_{4,2}K_1$ depends on φ alone.

We next show that

(22.17)
$$|Y_n(t) - Y_n(s)| \leq |Y_n(s+p) - Y_n(s)| + p\sqrt{n}, \qquad s \leq t \leq s+p.$$

Taking $s = 0$ for notational simplicity, we see that (22.17) is equivalent to
$$|U_n(t) - nt| \leq |U_n(p) - np| + np, \qquad 0 \leq t \leq p,$$
where $U_n(t)$ is the number among $\xi_1, \ldots, \xi_n$ that satisfy $\xi_i \leq t$. But $U_n(t) - nt \leq U_n(p) - nt = U_n(p) - np + n(p-t) \leq |U_n(p) - np| + np$ and $nt - U_n(t) \leq nt \leq |U_n(p) - np| + np$.

Now (22.17) implies

(22.18)
$$\sup_{s \leq t \leq s+mp} |Y_n(t) - Y_n(s)| \leq 3 \max_{i \leq m} |Y_n(s+ip) - Y_n(s)| + p\sqrt{n}.$$

If

(22.19)
$$\frac{\varepsilon}{n} \leq p < \frac{\varepsilon}{\sqrt{n}},$$

then (22.16) applies, and it follows by (22.18) that

(22.20)
$$\mathsf{P}\left\{\sup_{s \leq t \leq s+mp} |Y_n(t) - Y_n(s)| \geq 4\varepsilon\right\} \leq \frac{K_2}{\varepsilon^5} m^2 p^2.$$

Choose δ so that $K_2\delta/\varepsilon^5 < \eta$. From (22.20) it will follow that

(22.21)
$$\mathsf{P}\left\{\sup_{s \leq t \leq s+\delta} |Y_n(t) - Y_n(s)| \geq 4\varepsilon\right\} < \eta\delta,$$

provided there exist a p and an integer m such that (22.19) holds and $mp = \delta$. But this is equivalent to the existence of an integer m with $(\delta/\varepsilon)\sqrt{n} < m \leq (\delta/\varepsilon)n$, which is true for all sufficiently large n.

For given ε and η, therefore, there exists a δ such that (22.21) holds (for all $s \leq 1$ and with $s + \delta$ replaced by 1 if it exceeds 1) when n is large enough. But this implies (see the corollary to Theorem 8.3) that
$$\mathsf{P}\{w(Y_n, \delta) \geq 12\varepsilon\} \leq \eta$$
for large n, which is (22.13) except for the factor 12.

Thus $Y_n \xrightarrow{\mathscr{D}} Y$. The assumption that F is continuous serves only to simplify the proof. It can be avoided by using Theorem 12.5 where we have used Theorem 12.2. Of course, Y will not lie in C with probability 1 if F has discontinuities.

Functions of φ-Mixing Processes

Consider now a function

(22.22)
$$\eta_n = f(\ldots, \xi_{n-1}, \xi_n, \xi_{n+1}, \ldots)$$

of a φ-mixing process, as in the preceding section. We shall assume as before that η_n has close approximations

$$\eta_{ln} = f_l(\xi_{n-l}, \ldots, \xi_n, \ldots, \xi_{n+l}) \tag{22.23}$$

depending on only finitely many of the ξ_i. Again we may work with one-sided sequences just as well.

Suppose that

$$0 \leq \eta_n(\omega) \leq 1,$$

and define Y_n by (22.8), where now F denotes the distribution function of η_0 and $F_n(t, \omega)$ denotes the empirical distribution of $\eta_1(\omega), \ldots, \eta_n(\omega)$. In addition to a restriction on the magnitude of φ_n, we shall need, in analyzing the asymptotic distribution of Y_n, an assumption the effect of which is to ensure that the empirical distribution $F_{ln}(t, \omega)$ of $\eta_{l1}(\omega), \ldots, \eta_{ln}(\omega)$ agrees with $F_n(t, \omega)$ for points t in a set H_l which rapidly becomes dense in $[0, 1]$ as $l \to \infty$.

We shall suppose in the first place that

$$0 \leq \eta_{ln}(\omega) \leq 1.$$

We shall suppose in the second place that there exist sets H_l in $[0, 1]$ with these three properties:

(i) If $t \in H_l$, then

$$I_{[0,t]}(\eta_0) = I_{[0,t]}(\eta_{l0}) \tag{22.24}$$

with probability 1.

(ii) If

$$J_l = \{F(t) : t \in H_l\}, \tag{22.25}$$

then J_l is a ρ_l-net in $[0, 1]$, where ρ_l goes to 0 exponentially: There exist positive A and ρ, $\rho < 1$, such that

$$\rho_l \leq A\rho^l, \qquad l = 1, 2, \ldots. \tag{22.26}$$

(iii) We have $H_l \subset H_{l+1}$.

From (i) it follows that, if $t \in H_l$, then $F_{ln}(t, \omega) = F_n(t, \omega)$ with probability 1. Define g_t by (22.9), as before.

THEOREM 22.2 *Suppose that* $\{\xi_n\}$ *is* φ*-mixing with* $\Sigma n^2 \varphi_n^{\frac{1}{2}} < \infty$, *that* η_0 *has a continuous distribution function* F *on* $[0, 1]$, *and that there exist sets* H_l *with the three properties just described. Then*

$$Y_n \xrightarrow{\mathscr{D}} Y,$$

where Y *is the Gaussian random function specified by*

$$\mathsf{E}\{Y(t)\} = 0 \tag{22.27}$$

and

$$(22.28)\quad \mathsf{E}\{Y(s)Y(t)\} = \mathsf{E}\{g_s(\eta_0)g_t(\eta_0)\} + \sum_{k=1}^{\infty} \mathsf{E}\{g_s(\eta_0)g_t(\eta_k)\} + \sum_{k=1}^{\infty} \mathsf{E}\{g_s(\eta_k)g_t(\eta_0)\}.$$

The series converge absolutely and $\mathsf{P}\{Y \in C\} = 1$.

Before proving the theorem, we give two examples of its application.

Example 1. Let $(\Omega, \mathscr{B}, \mathsf{P})$ be the unit interval with Lebesgue measure and let $\xi_n(\omega)$ be the nth digit in the dyadic expansion of ω, as in Example 2 of the preceding section (p. 191). If $T\omega = 2\omega \pmod 1$ and $\eta_n(\omega) = T^{n-1}\omega$, then η_n has the form (22.22) (one-sided version). The η_n are uniformly distributed: $F(t) \equiv t$.

Define η_{ln} by

$$\eta_{ln}(\omega) = \frac{i-1}{2^l} \quad \text{if} \quad \frac{i-1}{2^l} < \eta_n(\omega) \le \frac{i}{2^l}, \qquad 1 \le i \le 2^l.$$

Then $\eta_{l0}(\omega)$ is a function of the first l digits of ω and hence η_{ln} has the form (22.23) (one-sided version). If t has the form $i/2^l$, then (22.24) holds. Since F is the identity, we may take

$$H_l = J_l = \left\{0, \frac{1}{2^l}, \ldots, \frac{2^l - 1}{2^l}, 1\right\},$$

which is a 2^{-l}-net in $[0, 1]$. Thus $\{\eta_n\}$ satisfies the hypotheses of the theorem.

Example 2. Let $(\Omega, \mathscr{B}, \mathsf{P})$ be the unit interval with Gauss's measure (20.19), as in Example 3 of the preceding section (p. 192). Let $\xi_n(\omega) = a_n(\omega)$ be the nth partial quotient in the continued-fraction expansion of ω, and let $\eta_n(\omega) = T_1^{n-1}\omega$, where T_1 is the continued-fraction transformation (21.48). Then η_n has the form (22.22) (one-sided version), and the distribution function of the η_n is

$$F(t) = \frac{1}{\log 2}\int_0^t \frac{dx}{1+x} = \frac{\log(1+t)}{\log 2}.$$

Define $\eta_{ln}(\omega)$ as the lth convergent of $T_1^{n-1}\omega$:

$$\eta_{ln}(\omega) = \frac{p_l(T_1^{n-1}\omega)}{q_l(T_1^{n-1}\omega)}.$$

Then η_{ln} has the form (22.23) (one-sided version). Let H_l be the set of all lth order convergents (for all ω). Now the unit interval splits into countably many subintervals† having the elements of H_l as endpoints; $\eta_1(\omega) = \omega$

† The fundamental intervals of rank l—see Billingsley (1965, p. 42).

and $\eta_{l1}(\omega) = p_l(\omega)/q_l(\omega)$ lie in the same subinterval, so that (22.24) holds for t in H_l. Now the subintervals have length at most 2^{-l+1}. Since the derivative of F is everywhere less than or equal to $1/\log 2$, the set (22.25) is a ρ_l-net in $[0, 1]$ with

$$\rho_l \leq \frac{2}{\log 2} \frac{1}{2^l}.$$

Thus $\{\eta_n\}$ satisfies the hypotheses of the theorem.

Proof of Theorem 22.2. As in the proof of Theorem 22.1, we first show that it suffices to consider the case in which η_0 is uniformly distributed on $[0, 1]$. Let $\eta'_n = F(\eta_n)$ and $\eta'_{ln} = F(\eta_{ln})$. Then η'_0 is uniformly distributed and, if t lies in the set J_l defined by (22.25), then

$$I_{[0,t]}(\eta'_0) = I_{[0,t]}(\eta'_{l0})$$

with probability 1. If $H'_l = J'_l = J_l$, then $\{\eta'_n\}$ and $\{\eta'_{ln}\}$ satisfy the hypotheses of the theorem relative to H'_l and J'_l. If the result is true in the uniform case, then, by the opening arguments of the preceding proof, it is true in general.

Assume then that η_0 is uniformly distributed; take $F(t) = t$ in (22.9); assume that, if $t \in H_l$, then (22.24) holds with probability 1; assume $H_l \subset H_{l+1}$; and, finally, assume that H_l is a ρ_l-net in $[0, 1]$, where ρ_l satisfies (22.26). To include 0 and 1 in H_l involves no loss of generality.

From these assumptions, it follows that $|\eta_0 - \eta_{l0}| \leq 2\rho_l$ with probability 1. Therefore, whatever t may be,

$$\begin{aligned} \mathsf{P}\{\eta_0 \leq t < \eta_{l0}\} &+ \mathsf{P}\{\eta_{l0} \leq t < \eta_0\} \\ &\leq \mathsf{P}\{t - 2\rho_l < \eta_0 \leq t\} + \mathsf{P}\{t < \eta_0 \leq t + 2\rho_l\}, \end{aligned}$$

so that, since η_0 is uniformly distributed,

$$\mathsf{E}\{|g_t(\eta_0) - g_t(\eta_{l0})|^2\} \leq 4\rho_l.$$

Since $g_t(\eta_{l0})$ is a function of $\xi_{-l}, \ldots, \xi_l$ alone, and since (22.26) implies $\Sigma\rho_l^{\frac{1}{2}} < \infty$, it follows by Theorem 21.1 that

$$Y_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g_t(\eta_i)$$

is asymptotically normal. The multivariate version of this result (see (21.43) and (21.44)) shows, just as the multivariate version of Theorem 20.1 did in the preceding proof, that the finite-dimensional distributions of Y_n converge to those specified for Y and that the series in (22.28) converge absolutely.

Again as in the preceding proof, it now suffices to produce, given positive ε and η, a δ, $0 < \delta < 1$, such that

$$\mathsf{P}\{w(Y_n, \delta) \geq \varepsilon\} \leq \eta \tag{22.29}$$

for all large n.

If s and t both lie in H_l, then the process

$$g_t(\eta_n) - g_s(\eta_n) = g_t(\eta_{ln}) - g_s(\eta_{ln}), \qquad n = 0, \pm 1, \pm 2, \ldots$$

s $\varphi^{(l)}$-mixing with

$$\varphi^{(l)}(n) = \begin{cases} 1 & \text{if} \quad n \le 2l, \\ \varphi(n - 2l) & \text{if} \quad n > 2l. \end{cases}$$

Since η_0 is uniformly distributed, and since

$$\sum_{k=0}^{\infty} (k+1)^2 [\varphi_k^{(l)}]^{\frac{1}{2}} = O(l^3),$$

Lemma 1 implies

$$\mathsf{E}\left\{ \left| \sum_{i=1}^{n} \left(g_t(\eta_i) - g_s(\eta_i) \right) \right|^4 \right\} \le K_3 l^6 (n^2 |t - s|^2 + n |t - s|),$$

where K_3 depends on φ alone. Therefore

$$s, t \in H_l, \qquad \frac{\varepsilon}{n} \le t - s \tag{22.30}$$

mply (assume $\varepsilon < 1$)

$$\mathsf{P}\{|Y_n(t) - Y_n(s)| \ge \lambda\} \le \frac{2K_3}{\varepsilon} \frac{l^6}{\lambda^4} (t - s)^2. \tag{22.31}$$

Choose and fix an integer m such that

$$\rho_{mv} \le \frac{1}{2^{2v+2}}, \qquad v = 1, 2, \ldots, \tag{22.32}$$

which is possible because of (22.26). We shall inductively define sets

$$L_v\colon t_0^{(v)}, t_1^{(v)}, \ldots, t_{2^v}^{(v)}$$

for $v \ge 0$. The sets L_v will be such that

$$0 = t_0^{(v)} < t_1^{(v)} < \cdots < t_{2^v}^{(v)} = 1;$$

such that

$$t_i^{(v)} \in H_{mv}$$

and

$$t_{2i}^{(v)} = t_i^{(v-1)}, \qquad v \ge 1;$$

and such that

$$\frac{1}{2^{v+1}}\left(1 + \frac{1}{2^v}\right) \le \alpha_v \le \beta_v \le \frac{1}{2^{v+1}}\left(3 - \frac{1}{2^v}\right), \tag{22.33}$$

where

$$\alpha_v = \min_{1 \le i \le 2^v} (t_i^{(v)} - t_{i-1}^{(v)}) \le \max_{1 \le i \le 2^v} (t_i^{(v)} - t_{i-1}^{(v)}) = \beta_v.$$

If $t_0^{(0)} = 0$ and $t_1^{(0)} = 1$, then L_0 has these properties. Suppose $L_0, \ldots, L_v$ already constructed so as to have these properties. Define $t_{2i}^{(v+1)} = t_i^{(v)}$, and take as $t_{2i+1}^{(v+1)}$ any point of $H_{m(v+1)}$ satisfying

$$|t_{2i+1}^{(v+1)} - \tfrac{1}{2}(t_i^{(v)} + t_{i+1}^{(v)})| < \rho_{m(v+1)};$$

such a point exists because $H_{m(v+1)}$ is a $\rho_{m(v+1)}$-net in $[0, 1]$. From (22.32) it follows that (22.33) holds also with v replaced by $v + 1$. Thus L_{v+1} satisfies the requirements.

Fix a number θ such that

$$\tfrac{1}{2} < \theta_0 = \theta^4 < 1. \tag{22.34}$$

Suppose ε, n, and v satisfy

$$\frac{\varepsilon}{n} \le \frac{1}{2^{v+1}} \le \frac{1}{2^{v-1}} \le \frac{\varepsilon}{\sqrt{n}}, \tag{22.35}$$

which, by (22.33) implies

$$\frac{\varepsilon}{n} \le \alpha_v \le \beta_v \le \frac{\varepsilon}{\sqrt{n}}. \tag{22.36}$$

Suppose further that $0 < u < v$ and $0 \le h < 2^{v-u}$. We shall show that

$$\mathsf{P}\left\{\max_{0 \le i \le 2^u} |Y_n(t_{h2^u+i}^{(v)}) - Y_n(t_{h2^u}^{(v)})| \ge \varepsilon\right\} \le \frac{K_5}{\varepsilon^5} \theta_0^{v-u} [t_{(h+1)2^u}^{(v)} - t_{h2^u}^{(v)}] \tag{22.37}$$

for a K_5 depending only on φ and the fixed numbers m and θ involved in (22.32) and (22.34). For notational convenience, we carry through the proof with $h = 0$.

If $0 \le i \le 2^u$, then

$$|Y_n(t_i^{(v)})| \le \sum_{k=0}^{u} \max_{1 \le j \le 2^{u-k}} |Y_n(t_{j2^k}^{(v)}) - Y_n(t_{(j-1)2^k}^{(v)})|$$

(consider the base-2 representation of the integer i), and therefore

$$\mathsf{P}\left\{\max_{0 \le i \le 2^u} |Y_n(t_i^{(v)})| \ge \varepsilon\right\} \le \sum_{k=0}^{u} \sum_{j=1}^{2^{u-k}} \mathsf{P}\{|Y_n(t_{j2^k}^{(v)}) - Y_n(t_{(j-1)2^k}^{(v)})| \ge \varepsilon(1-\theta)\theta^{u-k}\}.$$

Now $t_{j2^k}^{(v)} = t_j^{(v-k)} \in H_{m(v-k)}$. Since (22.30) implies (22.31), it follows by (22.36) that

$$\begin{aligned}\mathsf{P}\left\{\max_{0 \le i \le 2^u} |Y_n(t_i^{(v)})| \ge \varepsilon\right\} &\le \sum_{k=0}^{u} \sum_{j=1}^{2^{u-k}} \frac{2K_3}{\varepsilon} \frac{(m(v-k))^6}{(\varepsilon(1-\theta)\theta^{u-k})^4} (t_j^{(v-k)} - t_{j-1}^{(v-k)})^2 \\ &\le \frac{K_4}{\varepsilon^5} \sum_{k=0}^{u} \frac{(v-k)^6 \beta_{v-k}}{\theta_0^{u-k}} t_{2^u}^{(v)},\end{aligned}$$

where $K_4 = 2K_3m^6/(1-\theta)^4$. Since $\beta_{v-k} \le 1/2^{v-k-1}$, (22.37) follows with

$$K_5 = 2K_4 \sum_{i=0}^{\infty} \frac{i^6}{(2\theta_0)^i}$$

(recall that $2\theta_0 > 1$ and that the K_3 in (22.31) depends on φ alone).

The purely geometric inequality (22.17) applies as before and yields

$$\sup_{t_{h2^u}^{(v)} \le t \le t_{(h+1)2^u}^{(v)}} |Y_n(t) - Y_n(t_{h2^u}^{(v)})| \le 3 \max_{0 \le i \le 2^u} |Y_n(t_{h2^u+i}^{(v)}) - Y_n(t_{h2^u}^{(v)})| + \sqrt{n}\,\beta_v.$$

It now follows by (22.36) and (22.37) that

$$\sum_{h=0}^{2^{v-u}-1} \mathsf{P}\left\{\sup_{t_{h2^u}^{(v)} \le t \le t_{(h+1)2^u}^{(v)}} |Y_n(t) - Y_n(t_{h2^u}^{(v)})| > 4\varepsilon\right\} \le \frac{K_5}{\varepsilon^5}\,\theta_0^{v-u}.$$

Since $t_{h2^u}^{(v)} = t_h^{(v-u)} \in L_{v-u}$ and $\alpha_{v-u} \ge 1/2^{v-u+1}$, we have, by the corollary to Theorem 8.3,

$$\mathsf{P}\left\{w\left(Y_n, \frac{1}{2^{v-u+1}}\right) > 12\varepsilon\right\} \le \frac{K_5}{\varepsilon^5}\,\theta_0^{v-u}. \tag{22.38}$$

This inequality holds under the assumption of (22.35) and $0 < u < v$.

If ε and η are given, choose an integer r such that $K_5\theta_0{}^r/\varepsilon^5 < \eta$ (recall that $\theta_0 < 1$) and define $\delta = 1/2^{r+1}$. For all n exceeding some n_0, there exists a v satisfying (22.35). If we take $u = v - r$, then (22.38) holds, which implies (22.29) (with 12ε in place of ε). This proves Theorem 22.2.

It is not hard to show that the theorem persists if the probability measure P governing the ξ_n and η_n is replaced by a probability measure it dominates. Thus we can replace Gauss's measure by Lebesgue measure in Example 2.

Remarks. Theorems 22.1 and 22.2 are new; the case presented here as Example 1 is due to Ciesielski and Kesten (1962).

23. MARTINGALES

In Section 20 we proved a functional central limit theorem for stationary processes $\{\xi_n\}$ satisfying a uniform mixing condition. This mixing condition can be weakened to the assumption of ergodicity if we assume that the partial sums

$$S_n = \xi_1 + \cdots + \xi_n$$

form a martingale.†

† See Doob (1953) for the ergodic theory and martingale theory required here.

THEOREM 23.1 *Let $\{\xi_1, \xi_2, \ldots\}$ be a stationary, ergodic process for which*

$$\mathsf{E}\{\xi_n \,\|\, \xi_1, \ldots, \xi_{n-1}\} = 0 \tag{23.1}$$

with probability 1 and for which $\mathsf{E}\{\xi_n^2\} = \sigma^2$ is positive and finite. If $X_n(t, \omega) = S_{[nt]}(\omega)/\sigma\sqrt{n}$, then $X_n \xrightarrow{\mathscr{D}} W$.

Proof. There is no loss of generality in working with a doubly infinite sequence $\ldots, \xi_{-1}, \xi_0, \xi_1, \ldots,$ stationary and ergodic.† If $\mathscr{F}_k$ denotes the σ-field generated by $\ldots, \xi_{k-1}, \xi_k$, then, by (23.1) and stationarity,

$$\mathsf{E}\{\xi_k \,\|\, \mathscr{F}_{k-1}\} = 0 \tag{23.2}$$

with probability 1.

We shall verify the hypotheses of Theorem 19.4 with the functions $\rho(t) = 0$ and $\sigma^2(t) = 1$ appropriate to Brownian motion (see (19.29)); $X_n \xrightarrow{\mathscr{D}} W$ will follow. For Condition 1⁰ of that theorem, it is enough to show that, for $t < 1$,

$$\lim_{h \downarrow 0} \limsup_{n \to \infty} \frac{1}{h} \mathsf{E}\{|\mathsf{E}\{X_n(t+h) - X_n(t) \,\|\, \mathscr{F}_{[nt]}\}|\} = 0 \tag{23.3}$$

and

$$\lim_{h \downarrow 0} \limsup_{n \to \infty} \frac{1}{h} \mathsf{E}\{|\mathsf{E}\{(X_n(t+h) - X_n(t))^2 \,\|\, \mathscr{F}_{[nt]}\} - h|\} = 0. \tag{23.4}$$

Now (23.3) is an immediate consequence of (23.2). As for (23.4), the relation (23.2) implies

$$\mathsf{E}\{(S_{k+l} - S_k)^2 \,\|\, \mathscr{F}_k\} = \sum_{i=1}^{l} \mathsf{E}\{\xi_{k+i}^2 \,\|\, \mathscr{F}_k\},$$

so that, by stationarity, the expected value in (23.4) is

$$\frac{h}{\sigma^2} \mathsf{E}\left\{\left| \frac{1}{nh} \sum_{i=1}^{k_n} \mathsf{E}\{\xi_i^2 \,\|\, \mathscr{F}_0\} - \sigma^2 \right|\right\},$$

where $k_n = [n(t+h)] - [nt]$. Since, by Jensen's inequality, the above expression does not exceed

$$\frac{h}{\sigma^2} \mathsf{E}\left\{\left| \frac{1}{nh} \sum_{i=1}^{k_n} \xi_i^2 - \sigma^2 \right|\right\},$$

(23.4) follows by the mean ergodic theorem.

† Take the ξ_n as the coordinate variables on the product of a doubly infinite sequence of copies of the real line, with the finite-dimensional distributions prescribed by the original process; ergodicity is preserved because it depends only on the finite-dimensional distributions.

By (23.2), the variables ξ_k are orthogonal, and hence

$$\mathsf{E}\{S_n^2\} = n\sigma^2, \tag{23.5}$$

which implies Condition 2°.

Suppose we succeed in proving that $\{\max_{k\le n} S_k^2/n\}$ is uniformly integrable, which, if we define $\mathsf{E}_\alpha\{U\}$ as the integral of U over $\{U \ge \alpha\}$, means that

$$\lim_{\alpha\to\infty} \sup_n \mathsf{E}_\alpha\left\{\frac{1}{n} \max_{k\le n} S_k^2\right\} = 0. \tag{23.6}$$

It will certainly follow that $\{X_n^2(t): n \ge 1\}$ is uniformly integrable for each t, which is one of the hypotheses of Theorem 19.4, and Condition 3°a will follow easily by stationarity. Finally, since

$$\mathsf{P}\left\{\max_{k\le n} |S_k| \ge \lambda\sqrt{n}\right\} \le \frac{1}{\lambda^2} \mathsf{E}_{\lambda^2}\left\{\frac{1}{n} \max_{k\le n} S_k^2\right\},$$

it will follow by Theorem 8.4 that the tightness condition (19.51) is satisfied. It suffices therefore to prove (23.6).

If ξ_0 has a fourth moment, then $\mathsf{E}\{S_n^4\} = \Sigma \mathsf{E}\{\xi_i\xi_j\xi_k\xi_l\}$, with the indices running independently from 1 to n. If the largest index is not matched by any other, then, by (23.2), the term vanishes; hence

$$\mathsf{E}\{S_n^4\} = \sum_k \mathsf{E}\{\xi_k^4\} + 4\sum_{i<k} \mathsf{E}\{\xi_i\xi_k^3\} + 6\sum_{i,j<k} \mathsf{E}\{\xi_i\xi_j\xi_k^2\}.$$

If $|\xi_0| \le C$ with probability 1, then the first two sums on the right contribute at most $3n^2C^4$ in all, and the last sum is $6\sum_{k=2}^n \mathsf{E}\{S_{k-1}^2\xi_k^2\}$, which cannot, by (23.5), exceed $3n^2C^4$. Thus

$$\mathsf{E}\{S_n^4\} \le 6n^2C^4 \tag{23.7}$$

if $|\xi_0| \le C$ with probability 1.

By a martingale inequality†

$$\mathsf{E}\left\{\max_{k\le n} |S_k|^\gamma\right\} \le \left(\frac{\gamma}{\gamma-1}\right)^\gamma \mathsf{E}\{|S_n|^\gamma\} \tag{23.8}$$

for $\gamma > 1$. Therefore (23.5) implies

$$\mathsf{E}\left\{\max_{k\le n} S_k^2\right\} \le 4n\mathsf{E}\{\xi_0^2\}, \tag{23.9}$$

and, if $|\xi_0|$ is bounded by C, (23.7) implies

$$\mathsf{E}\left\{\max_{k\le n} S_k^4\right\} \le (\tfrac{4}{3})^4 \cdot 6n^2C^4. \tag{23.10}$$

---

† Doob (1953, p. 317).

For $u > 0$, define

$$\xi_{iu} = \begin{cases} \xi_i & \text{if } |\xi_i| \leq u \\ 0 & \text{if } |\xi_i| > u, \end{cases}$$

and put

$$\eta_{iu} = \xi_{iu} - \mathsf{E}\{\xi_{iu} \,\|\, \mathscr{F}_{i-1}\}$$

and

$$\delta_{iu} = \xi_i - \eta_{iu} = \xi_i - \xi_{iu} - \mathsf{E}\{\xi_i - \xi_{iu} \,\|\, \mathscr{F}_{i-1}\}.$$

If $S_{ku} = \sum_{i=1}^k \eta_{iu}$ and $D_{ku} = \sum_{i=1}^k \delta_{iu}$, then

$$\frac{1}{n} \max_{k \leq n} S_k^2 \leq \frac{2}{n} \max_{k \leq n} S_{ku}^2 + \frac{2}{n} \max_{k \leq n} D_{ku}^2. \tag{23.11}$$

Now $\{\eta_{nu} : n \geq 1\}$ has the martingale property (23.1) and $|\eta_{nu}| \leq 2u$; it follows by (23.10) that

$$\mathsf{E}_\alpha\left\{\frac{1}{n} \max_{k \leq n} S_{ku}^2\right\} \leq \frac{1}{\alpha} \mathsf{E}\left\{\frac{1}{n^2} \max_{k \leq n} S_{ku}^4\right\} \leq \frac{1}{\alpha}\left(\frac{4}{3}\right)^4 \cdot 6(2u)^4.$$

And $\{\delta_{nu} : n \geq 1\}$ also has the martingale property, so that, by (23.9) and Lemma 1 of Section 21 (see p. 184),

$$\mathsf{E}\left\{\frac{1}{n} \max_{k \leq n} D_{ku}^2\right\} \leq 4\mathsf{E}\{\delta_{0u}^2\} \leq 4\mathsf{E}\{(\xi_0 - \xi_{0u})^2\}$$

$$\leq 4\mathsf{E}_{u^2}\{\xi_0^2\}.$$

From the relation $\mathsf{E}_\alpha\{U + V\} \leq 2\mathsf{E}_{\frac{1}{2}\alpha}\{U\} + 2\mathsf{E}_{\frac{1}{2}\alpha}\{V\}$, together with (23.11), it now follows that

$$\mathsf{E}_\alpha\left\{\frac{1}{n} \max_{k \leq n} S_k^2\right\} \leq K\left[\frac{u^4}{\alpha} + \mathsf{E}_{u^2}\{\xi_0^2\}\right]$$

for a universal constant K. Since ξ_0 has a finite second moment, (23.6) follows.

Remarks. The central limit theorem under the hypotheses of Theorem 23.1 was proved independently by Billingsley (1961) and Ibragimov (1963); the proof of Theorem 23.1 itself (for bounded ξ_n) is due to Rosén (1967a).

24. EXCHANGEABLE RANDOM VARIABLES

Sampling

For each n, let

$$x_{n1}, x_{n2}, \ldots, x_{nk_n} \tag{24.1}$$

be a sequence of real numbers (not necessarily distinct), and let $\xi_{n1}(\omega), \ldots, \xi_{nk_n}(\omega)$ be a random permutation of these numbers, each of the $k_n!$ permutations having probability $1/k_n!$ Define a random element X_n of D by

$$X_n(t, \omega) = \sum_{i=1}^{[k_n t]} \xi_{ni}(\omega), \tag{24.2}$$

with $X_n(t, \omega) = 0$ for $0 \leq t < 1/k_n$. If we sample until the finite population (24.1) is exhausted, X_n describes the course the sampling takes.

THEOREM 24.1 *If*

$$\sum_{i=1}^{k_n} x_{ni} = 0, \qquad \sum_{i=1}^{k_n} x_{ni}^2 = 1, \tag{24.3}$$

and

$$\max_{1 \leq i \leq k_n} |x_{ni}| \rightarrow 0, \tag{24.4}$$

and if X_n is defined by (24.2), *then*

$$X_n \xrightarrow{\mathscr{D}} W^\circ. \tag{24.5}$$

Proof. It will be enough to verify the hypotheses of Theorem 19.4 with the functions $\rho(t) = -1/(1-t)$ and $\sigma^2(t) = 1$ that characterize the Brownian bridge W° (see (19.32)).

We shall need a preliminary computation. Let $y_1, \ldots, y_u$ be real numbers, suppose $1 \leq m \leq u$, and let $\eta_1, \ldots, \eta_m$ be an ordered sample of size m made without replacement (there are $\binom{u}{m}$ possible samples, all equally likely). By symmetry, the sum $\Sigma_{i=1}^m \eta_i$ has mean $m\mathsf{E}\{\eta_1\}$, and hence

$$\mathsf{E}\left\{\sum_{i=1}^m \eta_i\right\} = \frac{m}{u} \sum_{i=1}^u y_i. \tag{24.6}$$

The second moment of this sum is, again by symmetry,

$$m\mathsf{E}\{\eta_1{}^2\} + m(m-1)\mathsf{E}\{\eta_1\eta_2\},$$

which computation reduces to

$$\mathsf{E}\left\{\left[\sum_{i=1}^m \eta_i\right]^2\right\} = \frac{m(u-m)}{u(u-1)} \sum_{i=1}^u y_i{}^2 + \frac{m(m-1)}{u(u-1)} \left[\sum_{i=1}^u y_i\right]^2. \tag{24.7}$$

Applying these results to $\Sigma_{i=1}^m \xi_{ni}$ and using (24.3), we see that

$$\mathsf{E}\{X_n(t)\} = 0, \qquad \mathsf{E}\{X_n{}^2(t)\} = \frac{[k_n t](k_n - [k_n t])}{k_n(k_n - 1)}.$$

The second moment here converges to $t(1-t)$ (note that (24.3) and (24.4) imply $k_n \rightarrow \infty$).

Suppose now that $1 \le m_1 < m_1 + m_2 \le k_n$ and that we know the values of $\xi_{n1}, \ldots, \xi_{nm_1}$. Then $\xi_{n,m_1+1}, \ldots, \xi_{n,m_1+m_2}$ is conditionally distributed as a sample of size m_2 from the population (24.1) with the values $\xi_{n1}, \ldots \xi_{nm_1}$ removed from it. Applying (24.6) and (24.7) with $m = m_2$ and $u = k_n - m_1$ and using (24.3), we obtain

$$\mathsf{E}\left\{\sum_{i=m_1+1}^{m_1+m_2} \xi_{ni} \,\middle\|\, \xi_{n1}, \ldots, \xi_{nm_1}\right\} = -\frac{m_2}{k_n - m_1} \sum_{i=1}^{m_1} \xi_{ni} \tag{24.8}$$

and

$$\mathsf{E}\left\{\left[\sum_{i=m_1+1}^{m_1+m_2} \xi_{ni}\right]^2 \,\middle\|\, \xi_{ni}, \ldots, \xi_{nm_1}\right\} \tag{24.9}$$

$$= \frac{m_2(k_n - m_1 - m_2)}{(k_n - m_1)(k_n - m_1 - 1)}\left[1 - \sum_{i=1}^{m_1} \xi_{ni}^2\right] + \frac{m_2(m_2 - 1)}{(k_n - m_1)(k_n - m_1 - 1)}\left[\sum_{i=1}^{m_1} \xi_{ni}\right]^2.$$

Fix $t < 1$; taking $m_1 = [k_n t]$ and $m_2 = [k_n(t + h)] - [k_n t]$ in (24.8) gives

$$\mathsf{E}\{X_n(t + h) - X_n(t) \,\|\, X_n(t)\} = -A_n(h) X_n(t)$$

with

$$A_n(h) = \frac{[k_n(t + h)] - [k_n t]}{k_n - [k_n t]}.$$

Therefore

$$\frac{1}{h}\mathsf{E}\left\{\left|\, \mathsf{E}\{X_n(t + h) - X_n(t) \,\|\, X_n(t)\} + \frac{h}{1 - t} X_n(t)\right|\right\}$$

$$= \frac{1}{h}\left| A_n(h) - \frac{h}{1 - t}\right| \mathsf{E}\{|X_n(t)|\}.$$

Since $\mathsf{E}\{|X_n(t)|\} \le \mathsf{E}^{\frac{1}{2}}\{X_n^2(t)\}$ is bounded and $\lim_n A_n(h) = h/(1 - t)$, (19.43) follows. (Although we have for notational convenience conditioned with respect to but one $X(t)$, (24.8) easily yields the general case as well. The same remark applies to the next argument.)

It follows from (24.9) that

$$\mathsf{E}\{(X_n(t + h) - X_n(t))^2 \,\|\, X_n(t)\} = B_n(h)\left[1 - \sum_{i=1}^{[k_n t]} \xi_{ni}^2\right] + C_n(h) X_n^2(t)$$

with

$$B_n(h) = \frac{([k_n(t + h)] - [k_n t])(k_n - [k_n(t + h)])}{(k_n - [k_n t])(k_n - [k_n t] - 1)}$$

and

$$C_n(h) = \frac{([k_n(t + h)] - [k_n t])([k_n(t + h)] - [k_n t] - 1)}{(k_n - [k_n t])(k_n - [k_n t] - 1)}.$$

Therefore

(24.10) $$\frac{1}{h}\mathsf{E}\{|\mathsf{E}\{(X_n(t+h)-X_n(t))^2 \,\|\, X_n(t)\} - h|\}$$

$$\le \frac{|B_n(h)|}{h}\mathsf{E}\left\{\left|\sum_{i=1}^{[k_nt]}\xi_{ni}^2 - \frac{[k_nt]}{k_n}\right|\right\}$$

$$+ \left|\frac{B_n(h)}{h}\left(1-\frac{[k_nt]}{k_n}\right) - 1\right| + \frac{|C_n(h)|}{h}\mathsf{E}\{X_n^{\,2}(t)\}.$$

Now $\Sigma_{i=1}^{[k_nt]}\,\xi_{ni}^2$ has mean $[k_nt]/k_n$ and second moment (use (24.7))

$$\frac{[k_nt](k_n-[k_nt])}{k_n(k_n-1)}\sum_{i=1}^{k_n} x_{ni}^4 + \frac{[k_nt]([k_nt]-1)}{k_n(k_n-1)} \to t^2;$$

hence the variance of this sum tends to 0. Since

$$\lim_n B_n(h) = h(1-t-h)/(1-t)^2,$$

the first term on the right in (24.10) tends to 0 as $n\to\infty$. Since $\mathsf{E}\{X_n^{\,2}(t)\}$ is bounded and $\lim_n C_n(h) = h^2/(1-t)^2$, (19.44) follows.

We have verified Condition 1^0 of Theorem 19.4. Since the maximum jump $\max_i |x_{ni}|$ in X_n tends to 0, it follows by the remark following the proof of Theorem 19.4 that we need only verify $X_n(0) = 0$ (which is obvious) and Condition 3^0. Hence it remains only to show that

(24.11)

$$\mathsf{E}\{|X_n(t)-X_n(t_1)|^2\,|X_n(t_2)-X_n(t)|^2\} \le K(t_2-t_1)^2, \qquad t_1\le t\le t_2,$$

for some K independent of n, t_1, t, and t_2. Now the left member of (24.11) is

(24.12) $$\sum \mathsf{E}\{\xi_{ni}\xi_{nj}\xi_{nk}\xi_{nl}\},$$

where i and j range over $[k_nt_1] < i,\ j \le [k_nt]$ and k and l range over $[k_nt] < k,\ l \le [k_nt_2]$. Put $[k_nt]-[k_nt_1] = m_1$ and $[k_nt_2]-[k_nt] = m_2$; by symmetry, (24.12) reduces to

$$m_1m_2\mathsf{E}\{\xi_{n1}^2\xi_{n2}^2\} + m_1m_2(m_1+m_2-2)\mathsf{E}\{\xi_{n1}^2\xi_{n2}\xi_{n3}\}$$
$$+ m_1(m_1-1)m_2(m_2-1)\mathsf{E}\{\xi_{n1}\xi_{n2}\xi_{n3}\xi_{n4}\}.$$

Write $\tau_n = \Sigma_{i=1}^{k_n}\,x_{ni}^4$. Calculations (routine if tedious) further reduce (24.12) to

$$m_1m_2\frac{1-\tau_n}{k_n(k_n-1)} + m_1m_2(m_1+m_2-2)\frac{2\tau_n-1}{k_n(k_n-1)(k_n-2)}$$
$$+ m_1(m_1-1)m_2(m_2-1)\frac{3(1-2\tau_n)}{k_n(k_n-1)(k_n-2)(k_n-3)}.$$

Since $0 \leq \tau_n \leq 1$ by (24.3), $k_n \geq 6$ implies that this last expression is at most

$$\frac{2m_1m_2}{k_n^{\ 2}} + \frac{8m_1m_2(m_1+m_2)}{k_n^{\ 3}} + \frac{24m_1^{\ 2}m_2^{\ 2}}{k_n^{\ 4}} \leq 34\left(\frac{m_1+m_2}{k_n}\right)^2.$$

If $t_2 - t_1 \geq 1/k_n$, then $(m_1 + m_2)/k_n \leq 3(t_2 - t_1)$, so that (24.11) holds with $K = 306$. If $t_2 - t_1 < 1/k_n$, then (24.11) holds because one or the other of the two factors inside the expected value vanishes. This proves Theorem 24.1.

We shall need the following consequence of Theorem 24.1. For each W°-continuity set A and for each positive ε, there exists a positive $\delta(\varepsilon, A)$ such that, if $x_{n1}, \ldots, x_{nk_n}$ satisfy (24.3) and

$$\max_{1\leq i\leq k_n} |x_{ni}| \leq \delta(\varepsilon, A), \tag{24.13}$$

then

$$|\mathsf{P}\{X_n \in A\} - W^\circ(A)| < \varepsilon \tag{24.14}$$

$\big(X_n$ being defined by (24.2)$\big)$. Indeed, if for some ε and A there were no such $\delta(\varepsilon, A)$, we could construct a sequence of sets $\{x_{ni}\}$ satisfying (24.3) and (24.4) but violating (24.5).

Exchangeable Variables

We turn now to a more general problem. For each n, let

$$\xi_{n1}, \xi_{n2}, \ldots, \xi_{nk_n} \tag{24.15}$$

be random variables that are exchangeable in the sense that each permutation of the set (24.15) has the same joint distribution as the set itself. Define X_n, as before, by (24.2). We shall assume the ξ_{ni} satisfy the three conditions,

$$\sum_{i=1}^{k_n} \xi_{ni} \xrightarrow{P} 0, \qquad \sum_{i=1}^{k_n} \xi_{ni}^2 \xrightarrow{P} 1, \qquad \max_{1\leq i\leq k_n} |\xi_{ni}| \xrightarrow{P} 0, \tag{24.16}$$

as $n \to \infty$.

If (24.15) is a random permutation of (24.1), then it is exchangeable and (24.3) and (24.4) together imply (24.16). The following result thus generalizes Theorem 24.1.

THEOREM 24.2 *If the variables* (24.15) *are exchangeable and satisfy* (24.16), *then* $X_n \xrightarrow{\mathscr{D}} W^\circ$.

Proof. If $\alpha_n = k_n^{-1} \sum_{i=1}^{k_n} \xi_{ni}$ and $\beta_n^{\ 2} = \sum_{i=1}^{k_n} (\xi_{ni} - \alpha_n)^2$, then, by (24.16),

$$k_n\alpha_n \xrightarrow{P} 0, \qquad \beta_n \xrightarrow{P} 1. \tag{24.17}$$

If $\eta_{ni} = (\xi_{ni} - \alpha_n)/\beta_n$, then the variables $\eta_{n1}, \eta_{n2}, \ldots, \eta_{nk_n}$ are exchangeable

and

$$\text{(24.18)} \qquad \sum_{i=1}^{k_n} \eta_{ni} = 0, \qquad \sum_{i=1}^{k_n} \eta_{ni}^2 = 1, \qquad \max_{1 \le i \le k_n} |\eta_{ni}| \xrightarrow{P} 0.$$

Define a random element Y_n of D by

$$Y_n(t, \omega) = \sum_{i=1}^{[k_n t]} \eta_{ni}(\omega).$$

Since

$$X_n(t) = \beta_n Y_n(t) + [k_n t]\alpha_n,$$

$X_n \xrightarrow{\mathscr{D}} W^\circ$ and $Y_n \xrightarrow{\mathscr{D}} W^\circ$ are equivalent by (24.17) and Theorem 4.1. We shall prove $Y_n \xrightarrow{\mathscr{D}} W^\circ$.

Let $\zeta_{n1}, \ldots, \zeta_{nk_n}$ be a random permutation of $\eta_{n1}, \ldots, \eta_{nk_n}$. (The η_{ni} are subjected to a random permutation that is independent of them. This is a probative device; to support the random permutation, the probability space on which the η_{ni} are defined may require to be enlarged.) The ζ_{ni} have the same joint distribution as the η_{ni}, since the latter are exchangeable, and hence Y_n has the same distribution as the random function defined by

$$Z_n(t, \omega) = \sum_{i=1}^{[k_n t]} \zeta_{ni}(\omega).$$

To prove $Y_n \xrightarrow{\mathscr{D}} W^\circ$ it therefore suffices to show that

$$\text{(24.19)} \qquad \mathsf{P}\{Z_n \in A\} \to W^\circ(A)$$

for each W°-continuity set A.

Given a W°-continuity set A and a positive ε, choose $\delta(\varepsilon, A)$ in such a way that (24.13) implies (24.14) (where, in (24.14), X_n is defined via (24.2) from a random permutation of numbers (24.1) satisfying (24.3)). If E_n is the event

$$E_n = \left\{ \max_{1 \le i \le k_n} |\eta_{ni}| < \delta(A, \varepsilon) \right\},$$

then $\mathsf{P}(E_n) \to 1$ by (24.18). By the definition of conditional probability,

$$\mathsf{P}\{Z_n \in A\} = \int_{E_n} \mathsf{P}\{Z_n \in A \parallel \eta_{n1}, \ldots, \eta_{nk_n}\}\, d\mathsf{P} + \int_{E_n{}^c} \mathsf{P}\{Z_n \in A \parallel \eta_{n1}, \ldots, \eta_{nk_n}\}\, d\mathsf{P}.$$

Because of (24.18),

$$|\mathsf{P}\{Z_n \in A \parallel \eta_{n1}, \ldots, \eta_{nk_n}\} - W^\circ(A)| < \varepsilon$$

holds on the set E_n, and therefore

$$|\mathsf{P}\{Z_n \in A\} - W^\circ(A)| \le 2\varepsilon + 2\mathsf{P}(E_n{}^c).$$

Since $\mathsf{P}(E_n) \to 1$, (24.19) follows, which completes the proof of the theorem.

Combining Theorem 24.2 with the computations in Section 11 (see (11.39), (11.40), and (11.42)) gives the limiting distributions for $\max_{k \le k_n} \sum_{i=1}^{k} \xi_{ni}$, for $\max_{k \le k_n} |\sum_{i=1}^{k} \xi_{ni}|$, and for the fraction of k, $1 \le k \le k_n$, for which $\sum_{i=1}^{k} \xi_{ni} > 0$.

Remarks. For Theorem 24.1 and extensions to other sampling procedures, see Rosén (1964, 1967a, and 1967c). Theorem 24.2 is new. Chernoff and Teicher (1958) proved $X_n(t) \xrightarrow{\mathscr{D}} N(0, t(1-t))$ (t fixed) under the hypotheses of this theorem; see their paper for examples of variables satisfying these hypotheses. Related limit theorems concern probabilities computed conditionally on $X_n(1) = 0$, where X_n is some random function; see Dwass and Karlin (1963)—and the references there—and Trumbo (1965), as well as forthcoming papers by T. Liggett and M. Wichura.

APPENDIX I

Metric Spaces

We review here a few facts about metric spaces, taking as known the first definitions (open, closed, dense, limit point, continuity, etc.) and properties.†

We denote the metric space by S and the metric itself by $\rho(x, y)$. We denote the closure of a subset A of S by A^-, its interior by A°, and its boundary by ∂A $(= A^- - A^\circ)$. We define the distance from x to A as

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

it is easy to check that $\rho(x, A)$ is uniformly continuous in x. Denote by $S(x, \varepsilon)$ the open sphere with center x and radius ε: $S(x, \varepsilon) = \{y : \rho(x, y) < \varepsilon\}$. By "sphere" we shall mean "open sphere"; we shall call $S(x, \varepsilon)$ the ε-sphere about x. (Topologists use "ball" in place of "sphere.")

Two metrics ρ_1 and ρ_2 on S are said to be equivalent if (write $S_i(x, \varepsilon) = \{y : \rho_i(x, y) < \varepsilon\}$) for each x and ε there is a δ with $S_1(x, \delta) \subset S_2(x, \varepsilon)$ and $S_2(x, \delta) \subset S_1(x, \varepsilon)$, so that S with ρ_1 is homeomorphic to S with ρ_2.

Separability

The space S is by definition separable if it contains a countable, dense subset. A base for S is a class of open sets such that each open subset of S is the union of some of the members of the class. An open cover of A is a class of open sets whose union contains A. A set A is discrete if about each point of

† For full accounts, see Dieudonné (1960), Royden (1963), or Simmons (1963), for example.

A there is a sphere containing no other points of A—in other words, if each point of A is isolated in the relative topology. If S itself is discrete, then taking the distance between distinct points to be 1 defines a metric equivalent to the original one.

These three conditions are equivalent:

(i) *S is separable.*

(ii) *S has a countable base.*

(iii) *Each open cover of each subset of S has a countable subcover.*

Moreover, separability implies

(iv) *S contains no uncountable discrete set,*

and this in turn implies

(v) *S contains no uncountable set A with*

$$\inf\{\rho(x, y): x, y \in A, x \neq y\} > 0. \tag{1}$$

Proof of (i) → (ii). Assuming D is a countable set dense in S, let $\mathscr{V}$ be the countable class of spheres with rational radii and with centers in D. Let G be open; to prove that $\mathscr{V}$ is a base, we must show that, if G_1 is the union of those elements of $\mathscr{V}$ that are contained in G, then $G = G_1$. Clearly, $G_1 \subset G$, and to prove $G \subset G_1$, it suffices to find, for a given x in G, a d in D and a rational r such that $x \in S(d, r) \subset G$. But if $x \in G$, then $S(x, \varepsilon) \subset G$ for some positive ε. Since D is dense, there is a d in D with $\rho(x, d) < \frac{1}{2}\varepsilon$. Take a rational r with $\rho(x, d) < r < \frac{1}{2}\varepsilon$.

Proof of (ii) → (iii). Let $\{V_1, V_2, \ldots\}$ be a countable base, and suppose $\{G_\alpha\}$ is an open cover of A (α ranges over an arbitrary index set). For each V_k for which there exists a G_α with $V_k \subset G_\alpha$, let G_{α_k} be some one of these G_α containing it. Then $A \subset \bigcup_k G_{\alpha_k}$.

Proof of (iii) → (i). For each n, $\{S(x, n^{-1}): x \in S\}$ is an open cover of S. If (iii) holds, there is a countable subcover $\{S(x_{n,k}, n^{-1}): k = 1, 2, \ldots\}$. The countable set $\{x_{n,k}: n, k = 1, 2, \ldots\}$ is dense in S.

Proof of (iii) → (iv). If A is discrete, then for each x in A there is a positive ε_x such that $S(x, \varepsilon_x)$ contains no other point of A. Since $\{S(x, \varepsilon_x): x \in A\}$ is an open cover of A without any subcover, (iii) cannot hold if A is uncountable.

Since a set satisfying (1) is discrete, certainly (iv) implies (v). Although we shall not need the fact, (v) implies separability, so that (i) through (v) are all equivalent. (For each positive ε, use Zorn's lemma to find a maximal set A_ε of points distant at least ε from one another; the union of the A_ε for rational ε is dense and, if (v) holds, countable.)

Compactness

A set A in S is by definition compact if each open cover of A contains a finite subcover. An ε-net for A is a set of points $\{x_k\}$ with the property that for each x in A there is an x_k such that $\rho(x, x_k) < \varepsilon$ (the x_k are not required to lie in A). A set is totally bounded if, for every positive ε, it has a finite ε-net. A set A is complete if each fundamental sequence in A converges to some point of A.

For an arbitrary set A in S, these four conditions are equivalent:

(i) *A^- is compact.*
(ii) *Each countable open cover of A^- has a finite subcover.*
(iii) *Each sequence in A has a limit point (has a subsequence converging to a limit, which necessarily lies in A^-).*
(iv) *A is totally bounded and A^- is complete.*

It is easy to show that (iii) holds if and only if each sequence in A^- has a limit point (necessarily in A^-) and that A is totally bounded if and only if A^- is totally bounded. Therefore we may assume in the proof that $A = A^-$ is closed. The implication (i) $\rightarrow$ (ii) is trivial.

Proof of (ii) $\rightarrow$ (iii). Given a sequence $\{x_n\}$ in A, define F_n to be the closure of the set $\{x_k : k \geq n\}$. If $\bigcap_n F_n = 0$, then the open sets $F_n{}^c$ cover A, and hence, if (ii) holds, $A \subset F_1{}^c \cup \cdots \cup F_n{}^c$ for some n, which implies the impossible relation $F_n \cap A = 0$. Thus $\bigcap_n F_n$ contains some x, which must be a limit point of $\{x_n\}$.

Proof of (iii) $\rightarrow$ (iv). If A is not totally bounded, there exist some positive ε and some sequence $\{x_n\}$ such that $\rho(x_m, x_n) \geq \varepsilon$ for $m \neq n$; $\{x_n\}$ can have no limit point. Hence (iii) implies total boundedness. And (iii) implies completeness because, if $\{x_n\}$ is fundamental and has a limit point x, then $\{x_n\}$ converges to x.

Proof of (iv) $\rightarrow$ (i). Assume (iv) and suppose $\{G_\alpha\}$, where α ranges over an arbitrary index set, is an open cover of A having no finite subcover. We shall derive a contradiction.

Since A is totally bounded, it can, for each n, be covered by finitely many open spheres $B_{n1}, \ldots, B_{nk_n}$ of radius 2^{-n}. At least one of the B_{ni} must have the property that no finite subfamily of $\{G_\alpha\}$ covers $A \cap B_{ni}$ (which must therefore be nonempty); let C_n be one such B_{ni}. Since the $B_{ni} \cap C_{n-1} \cap A$ cover $C_{n-1} \cap A$ (if $n > 1$), we may also insist that no finite subfamily of $\{G_\alpha\}$ cover $C_n \cap C_{n-1} \cap A$, so that, in particular, $C_n \cap C_{n-1} \neq 0$.

Let x_n be one of the points that C_n shares with A. Since $C_n \cap C_{n-1} \neq 0$ and C_n has radius 2^{-n}, we have $\rho(x_n, x_{n-1}) < 6 \cdot 2^{-n}$, which implies

$\rho(x_n, x_{n+k}) < 6 \cdot 2^{-n}$. Thus $\{x_n\}$ is fundamental and has, by the completeness assumption, a limit x, which must lie in A.

Now $x \in G_\alpha$ for some α, and $S(x, \varepsilon) \subset G_\alpha$ for some positive ε. But then $\rho(x, x_n) < \varepsilon/3$ for some n satisfying $2^{-n} < \varepsilon/3$, which implies $C_n \subset G_\alpha$, a contradiction.†

A subset of k-dimensional Euclidean space R^k (with the usual metric) has compact closure if and only if it is bounded.

If h is a continuous mapping of S into another metric space S', and if K is a compact subset of S, then hK is a compact subset of S'.

Proof. If $\{G'_\alpha\}$ is an open cover of hK, then $\{h^{-1}G'_\alpha\}$ is an open cover of K and hence has a finite subcover $\{h^{-1}G'_{\alpha_i}: i = 1, \ldots, n\}$. Clearly, $\{G'_{\alpha_i}: i = 1, \ldots, n\}$ covers hK.

Upper Semicontinuity

A function f is upper semicontinuous at x if for every positive ε there exists a positive δ such that $\rho(x, y) < \delta$ implies $f(y) < f(x) + \varepsilon$. It is easy to see that f is everywhere upper semicontinuous if and only if, for each real α, the set $\{x: f(x) < \alpha\}$ is open.

Let f_n be real functions on S such that, for each x, $f_n(x)$ is nonincreasing and

$$\lim_{n \to \infty} f_n(x) = 0. \tag{2}$$

If the f_n are upper semicontinuous, then the convergence in (2) *is uniform on each compact set.*

Proof. For each positive ε, the open sets $G_n = \{x: f_n(x) < \varepsilon\}$ cover S. If K is compact, then $K \subset G_n$ for some n and uniformity follows.

The Space R^∞

Let R^∞ be the space of sequences $x = (x_1, x_2, \ldots)$ of real numbers. If $\rho_0(\alpha, \beta) = |\alpha - \beta|/(1 + |\alpha - \beta|)$, then ρ_0 is a metric on the line R^1 equivalent to the ordinary metric $|\alpha - \beta|$. The line is complete under ρ_0. It follows that, if $\rho(x, y) = \sum_{k=1}^\infty \rho_0(x_k, y_k) 2^{-k}$, then ρ is a metric on R^∞. If

$$N_{k,\varepsilon}(x) = \{y: |y_i - x_i| < \varepsilon, i = 1, \ldots, k\}, \tag{3}$$

then $N_{k,\varepsilon}(x)$ is open in the sense of the metric ρ. Moreover, since $\rho(x, y) < \varepsilon 2^{-k}/(1 + \varepsilon)$ implies $y \in N_{k,\varepsilon}(x)$, which in turn implies $\rho(x, y) < \varepsilon + 2^{-k}$, the sets (3) form a base for the topology given by ρ.

† Proof from Dieudonné (1960).

We shall always take R^∞ with this topology. It is the product topology, or the topology of coordinatewise convergence: $\lim_n x(n) = x$ if and only if $\lim_n x_k(n) = x_k$ for each k. The space R^∞ is separable; one countable, dense set consists of those points with coordinates that are all rational and that, with only finitely many exceptions, vanish.

Suppose $\{x(n)\}$ is a fundamental sequence in R^∞. Since

$$\rho_0\big(x_k(m), x_k(n)\big) \leq 2^k \rho\big(x(m), x(n)\big),$$

it follows easily that, for each k, $\{x_k(1), x_k(2), \ldots\}$ is a fundamental sequence on the line with the usual metric, so that the limit $x_k = \lim_n x_k(n)$ exists. If $x = (x_1, x_2, \ldots)$, then $x(n)$ converges to x in the sense of R^∞. Thus R^∞ is complete.

A subset A of R^∞ has compact closure if and only if the set $\{x_k : x \in A\}$ is, for each k, a bounded set on the line.

Proof. It is easy to show that the stated condition is necessary for compactness. We may prove sufficiency by the classical diagonal method. Given a sequence $\{x(n)\}$ in A, we may choose a sequence of subsequences

$$(4) \qquad \begin{cases} x(n_{11}), & x(n_{12}), & x(n_{13}), & \ldots \\ x(n_{21}), & x(n_{22}), & x(n_{23}), & \ldots \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{cases}$$

in the following way. The first row of (4) is a subsequence of $\{x(n)\}$, so chosen that $x_1 = \lim_i x_1(n_{1i})$ exists; there is such a subsequence because $\{x_1 : x \in A\}$ is a bounded set of real numbers. The second row of (4) is a subsequence of the first row, so chosen that $x_2 = \lim_i x_2(n_{2i})$ exists; there is such a subsequence because $\{x_2 : x \in A\}$ is bounded.

We continue in this way; row k is a subsequence of row $k - 1$, and $x_k = \lim_i x_k(n_{ki})$ exists. Let x be the point of R^∞ with coordinates x_k. If $n_i = n_{ii}$, then $\{x(n_i)\}$ is a subsequence of $\{x(n)\}$. For each k, moreover, $x(n_k), x(n_{k+1}), \ldots$ all lie in the kth row of (4), so that $\lim_i x_k(n_i) = x_k$. Thus $\lim_i x(n_i) = x$, and it follows that A^- is compact.

Our final result about R^∞ is a special case of Urysohn's embedding theorem.

Every separable metric space S is homeomorphic to a subset of R^∞.

Proof. Let $\{d_1, d_2, \ldots\}$ be a sequence of points dense in S, and define a mapping h from S into R^∞ by

$$h(x) = \big(\rho(x, d_1), \rho(x, d_2), \ldots\big), \qquad x \in S,$$

where ρ denotes the metric on S. If points x_n of S converge to a limit x, then $\lim_n \rho(x_n, d_k) = \rho(x, d_k)$ for each k, so that, since the topology of R^∞ is that of coordinatewise convergence, $h(x_n)$ converges to $h(x)$. Thus h is continuous.

Suppose x_n does not converge to x. Then $\rho(x_{n'}, x) > \varepsilon$ for some positive ε and for each element of some subsequence $\{x_{n'}\}$. If $\rho(x, d_k) < \frac{1}{2}\varepsilon$, which must be true for some element of the dense sequence $\{d_k\}$, then $\rho(x_{n'}, d_k) > \frac{1}{2}\varepsilon$ for all n'; thus $\rho(x_n, d_k)$ cannot converge to $\rho(x, d_k)$ for this value of k and hence $h(x_n)$ cannot converge to $h(x)$.

Thus $h(x_n) \to h(x)$ implies $x_n \to x$. The same argument shows that $h(x) = h(y)$ implies $x = y$ (take $x_n \equiv y$). Therefore h is a one-to-one, bicontinuous mapping of S onto the subset hS of R^∞.

The Space C

Let $C = C[0, 1]$ be the space of continuous, real-valued functions on the unit interval $[0, 1]$ with the uniform metric. The distance between two elements $x = x(t)$ and $y = y(t)$ of C is

$$\rho(x, y) = \sup_t |x(t) - y(t)|;$$

it is easy to check that ρ is a metric. Convergence in the topology is uniform pointwise convergence of (continuous) functions.

The space C is separable; one countable, dense set consists of the (polygonal) functions that are linear on each subinterval $[(i-1)/k, i/k]$, $i = 1, \ldots, k$, for some integer k, and assume rational values at the points i/k, $i = 0, 1, \ldots, k$.

If $\{x_n\}$ is a fundamental sequence in C, then, for each value of t, $\{x_n(t)\}$ is a fundamental sequence on the line and hence has a limit $x(t)$. It is easy to show that the convergence $x_n(t) \to x(t)$ is uniform in t, so that x lies in C and is the limit in C of $\{x_n\}$. Thus C is complete.

We define the modulus of continuity of an element x of C by

$$w_x(\delta) = w(x, \delta) = \sup_{|s-t|<\delta} |x(s) - x(t)|, \qquad 0 < \delta < 1. \tag{5}$$

Since

$$|w_x(\delta) - w_y(\delta)| \leq 2\rho(x, y), \tag{6}$$

$w_x(\delta)$ is, for fixed positive δ, continuous in x. Note also that, since an element of C is uniformly continuous, we have

$$\lim_{\delta \to 0} w_x(\delta) = 0, \qquad x \in C. \tag{7}$$

(The Arzelà–Ascoli Theorem.) A subset A of C has compact closure if and only if

(8) $$\sup_{x \in A} |x(0)| < \infty$$

and

(9) $$\lim_{\delta \to 0} \sup_{x \in A} w_x(\delta) = 0.$$

Proof. If A^- is compact, (8) follows easily. Since $w_x(1/n)$ is continuous in x and nonincreasing in n, (7) holds uniformly on A if A^- is compact (see p. 218) and (9) follows.

Suppose now that (8) and (9) hold. Choose k large enough that $\sup_{x \in A} w_x(1/k)$ is finite. Since

$$|x(t)| \leq |x(0)| + \sum_{i=1}^{k} \left| x\left(\frac{i}{k}\, t\right) - x\left(\frac{i-1}{k}\, t\right) \right|,$$

it follows that

(10) $$\sup_t \sup_{x \in A} |x(t)| < \infty.$$

We shall deduce from (9) and (10) that A is totally bounded; since S is complete, this will imply that A^- is compact. Given ε, we must construct a finite ε-net for A. Let α denote the finite quantity in (10), and let H denote the finite set of points

$$\frac{u}{v}\,\alpha, \qquad u = 0, \pm 1, \ldots, \pm v,$$

where v is an integer such that $\alpha/v < \varepsilon$ (H is an ε-net for the linear interval $[-\alpha, \alpha]$). Now choose k large enough that $w_x(1/k) < \varepsilon$ for all x in A, and take B to consist of those elements of C that are linear on each subinterval $[(i-1)/k, i/k]$, $i = 1, \ldots, k$, and assume values in H at the points i/k, $i = 0, 1, \ldots, k$. The set B is finite (it contains $(2v+1)^{k+1}$ points); we shall show that it is a 2ε-net for A.

If $x \in A$, then $|x(i/k)| \leq \alpha$. Therefore there exists a point y of B such that

(11) $$\left| x\left(\frac{i}{k}\right) - y\left(\frac{i}{k}\right) \right| < \varepsilon, \qquad i = 0, 1, \ldots, k.$$

Since $w_x(1/k) < \varepsilon$, and since y is linear on each subinterval $[(i-1)/k, i/k]$, it follows from (11) that $\rho(x, y) < 2\varepsilon$. This proves the theorem.

We have seen that (8) and (10) are equivalent in the presence of (9). When (9) holds, A is said to be uniformly equicontinuous. Thus the Arzelà–Ascoli theorem asserts that A^- is compact if and only if A is uniformly bounded and uniformly equicontinuous.

APPENDIX II

Miscellany

Measurability

Let $(\Omega, \mathscr{B})$ and $(\Omega', \mathscr{B}')$ be measurable spaces; $\mathscr{B}[\mathscr{B}']$ is a σ-field of subsets of $\Omega[\Omega']$. A mapping $h:\Omega \to \Omega'$ from Ω into Ω' is said to be measurable $(\mathscr{B}, \mathscr{B}')$ if the inverse image $h^{-1}M'$ lies in $\mathscr{B}$ for each M' in $\mathscr{B}'$. If $h^{-1}\mathscr{B}'$ denotes the class $\{h^{-1}M' : M' \in \mathscr{B}'\}$, this condition may be succinctly stated as $h^{-1}\mathscr{B}' \subset \mathscr{B}$. Since $\{M' : h^{-1}M' \in \mathscr{B}\}$ is a σ-field, *if* $\mathscr{B}_0'$ *is contained in* $\mathscr{B}'$ *and generates it, then* $h^{-1}\mathscr{B}_0' \subset \mathscr{B}$ implies $h^{-1}\mathscr{B}' \subset \mathscr{B}$.

Let $(\Omega'', \mathscr{B}'')$ be a third measurable space, let $j:\Omega' \to \Omega''$ map Ω' into Ω'', and denote by jh the composition of h and j: $(jh)(\omega) = j(h(\omega))$. It is easy to show that, *if* $h^{-1}\mathscr{B}' \subset \mathscr{B}$ *and* $j^{-1}\mathscr{B}'' \subset \mathscr{B}'$, *then* $(jh)^{-1}\mathscr{B}'' \subset \mathscr{B}$.

If $\Omega = S$ and $\Omega' = S'$ are metric spaces, h is continuous when $h^{-1}G'$ is open in S for each open G' in S'. Let $\mathscr{S}$ and $\mathscr{S}'$ be the σ-fields of Borel sets in S and S'. If h is continuous, then, since $h^{-1}G' \in \mathscr{S}$ for G' open in S', and since the open sets in S' generate $\mathscr{S}'$, $h^{-1}\mathscr{S}' \subset \mathscr{S}$, so that h is measurable.

If Ω' is k-dimensional Euclidean space R^k, we always take $\mathscr{B}'$ to be the class $\mathscr{R}^k$ of k-dimensional Borel sets (the σ-field generated by the open subsets of R^k with the Euclidean metric), and we say h is measurable $\mathscr{B}$ if $h^{-1}\mathscr{R}^k \subset \mathscr{B}$; h is measurable $\mathscr{B}$ if $h^{-1}\{\alpha \in R^k : \alpha_i < a\} \in \mathscr{B}$ for each $i = 1, \ldots, k$ and each real a. In particular, if $\Omega = S$ is a metric space, if $k = 1$, and if h is upper semicontinuous (see p. 218), then h is measurable $\mathscr{S}$.

Change of Variable

If P is a probability measure on $(\Omega, \mathscr{B})$, and if $h^{-1}\mathscr{B}' \subset \mathscr{B}$, $\mathsf{P}h^{-1}$ denotes the probability measure on $(\Omega', \mathscr{B}')$ defined by $(\mathsf{P}h^{-1})(M') = \mathsf{P}(h^{-1}M')$ for

$M' \in \mathscr{B}'$. If f is a real function on Ω', measurable $\mathscr{B}'$, then the real function fh on Ω is measurable $\mathscr{B}$.

Under these circumstances,† *f is integrable with respect to* $\mathsf{P}h^{-1}$ *if and only if fh is integrable with respect to* P, *in which case we have*

$$\int_{h^{-1}M'} f(h(\omega))\,\mathsf{P}(d\omega) = \int_{M'} f(\omega')\,\mathsf{P}h^{-1}(d\omega') \tag{1}$$

for each M' in $\mathscr{B}'$.

If X takes Ω into a metric space S and $X^{-1}\mathscr{S} \subset \mathscr{B}$, so that X is a random element of S (see Section 4), we generally write $\mathsf{E}\{f(X)\}$ in place of $\int f(X(\omega))\mathsf{P}(d\omega)$. If $P = \mathsf{P}X^{-1}$ is the distribution of X in S, (1) implies

$$\mathsf{E}\{f(X)\} = \int_S f(x)\,\mathsf{P}(dx). \tag{2}$$

Tail Probabilities

Let X be a random variable on a probability space $(\Omega, \mathscr{B}, \mathsf{P})$. If X is nonnegative and integrable, and if $\alpha \geq 0$, then

$$\int_{\{X \geq \alpha\}} X\,d\mathsf{P} = \alpha\,\mathsf{P}\{X \geq \alpha\} + \int_\alpha^\infty \mathsf{P}\{X \geq t\}\,dt. \tag{3}$$

This will follow if we prove

$$\mathsf{E}\{X\} = \int_0^\infty \mathsf{P}\{X \geq t\}\,dt, \tag{4}$$

since we may replace X by its product with the indicator of $\{X \geq \alpha\}$. If X has finite range, (4) follows by summation by parts. The general result follows from the fact that any nonnegative X can be represented as the limit of a nondecreasing sequence of random variables X_n with finite range. (The equation also holds in the extended sense that if one side of (4) is infinite so is the other.)

Still assuming X nonnegative and integrable, and assuming $\alpha > 0$, we have

$$\mathsf{P}\{X \geq \alpha\} \leq \frac{1}{\alpha}\int_{\{X \geq \alpha\}} X\,d\mathsf{P} \leq \frac{1}{\alpha}\,\mathsf{E}\{X\}. \tag{5}$$

Scheffé's Theorem

Let λ be a measure (not necessarily finite) on a space $(\Omega, \mathscr{B})$, and let $p(\omega)$ and $p_n(\omega)$ be probability densities with respect to λ; p and p_n are nonnegative functions on Ω, measurable $\mathscr{B}$, such that $\int p\,d\lambda = \int p_n\,d\lambda = 1$.

† For a proof, see Halmos (1950, p. 163), for example.

(Scheffé's Theorem.)† *If $p_n(\omega) \to p(\omega)$ except for ω in a set of λ-measure* 0, *then*

$$\sup_{E\in\mathscr{B}} \left| \int_E p\, d\lambda - \int_E p_n\, d\lambda \right| = \frac{1}{2}\int |p - p_n|\, d\lambda \to 0 \tag{6}$$

Proof. If $\delta_n = p - p_n$, then $\int \delta_n\, d\lambda = 0$. For E in $\mathscr{B}$, therefore,

$$2\left|\int_E \delta_n\, d\lambda\right| = \left|\int_E \delta_n\, d\lambda\right| + \left|\int_{E^c} \delta_n\, d\lambda\right| \le \int |\delta_n|\, d\lambda;$$

and, if $E = \{\delta_n \ge 0\}$, there is equality here. This proves the equality in (6).

If δ_n^+ is the positive part of δ_n, then $\delta_n^+ \to 0$ except on a set of λ-measure 0. Since $0 \le \delta_n^+ \le p$, $\int |\delta_n|\, d\lambda = 2\int \delta_n^+\, d\lambda \to 0$ follows by Lebesgue's dominated convergence theorem.

A special case: If $\Sigma_i\, p_n(i) = \Sigma_i\, p(i) = 1$, the terms being nonnegative, and if $\lim_n p_n(i) = p(i)$ for each i, then $\Sigma_i\, |p(i) - p_n(i)| \to 0$. If $a(i)$ is bounded, it follows that $\Sigma_i\, a(i)p_n(i) \to \Sigma_i\, a(i)p(i)$.

Subspaces

Let S be a metric space, and let $\mathscr{S}$ be its σ-field of Borel sets. A subset S_0 (not necessarily in $\mathscr{S}$) is a metric space in its own right in the relative topology. Let $\mathscr{S}_0$ be the σ-field of Borel sets in S_0. We shall prove

$$\mathscr{S}_0 = \{S_0 \cap A : A \in \mathscr{S}\}. \tag{7}$$

If $h(x) = x$ for $x \in S_0$, then h is a continuous mapping from S_0 to S and hence $h^{-1}\mathscr{S} \subset \mathscr{S}_0$, so that $S_0 \cap A \in \mathscr{S}_0$ if $A \in \mathscr{S}$. Since $\{S_0 \cap A : A \in \mathscr{S}\}$ is a σ-field in S_0 and contains all sets $S_0 \cap G$ with G open in S, that is, all open sets in S_0, (7) follows.

If S_0 lies in $\mathscr{S}$, (7) becomes

$$\mathscr{S}_0 = \{A : A \subset S_0,\ A \in \mathscr{S}\}. \tag{8}$$

Product Spaces

Let S' and S'' be metric spaces with metrics ρ' and ρ'' and σ-fields $\mathscr{S}'$ and $\mathscr{S}''$ of Borel sets. The rectangles

$$A' \times A'' \tag{9}$$

with A' open in S' and A'' open in S'' are a basis for the product topology in $S = S' \times S''$. This topology may also be described as the one under which

† Scheffé (1947).

$(x'_n, x''_n) \to (x', x'')$ if and only if $x'_n \to x'$ and $x''_n \to x''$. Finally, the topology may be specified by various metrics, for example,

$$\rho((x', x''), (y', y'')) = \sqrt{[\rho'(x', y')]^2 + [\rho''(x'', y'')]^2} \tag{10}$$

and

$$\rho((x', x''), (y', y'')) = \max\{\rho'(x', y'), \rho''(x'', y'')\}. \tag{11}$$

Let $\mathscr{S}' \times \mathscr{S}''$ be the σ-field generated by the measurable rectangles (sets (9) with $A' \in \mathscr{S}'$ and $A'' \in \mathscr{S}''$), and let $\mathscr{S}$ be the σ-field of Borel sets in S for the product topology. We first show that

$$\mathscr{S}' \times \mathscr{S}'' \subset \mathscr{S}. \tag{12}$$

If $\pi'(x', x'') = x'$ and $\pi''(x', x'') = x''$, then $\pi': S \to S'$ and $\pi'': S \to S''$ are continuous and hence measurable $((\pi')^{-1}\mathscr{S}' \subset \mathscr{S}$ and $(\pi'')^{-1}\mathscr{S}'' \subset \mathscr{S})$. It follows that, if $A' \in \mathscr{S}'$ and $A'' \in \mathscr{S}''$, then $A' \times A'' = \pi'^{-1}A' \cap \pi''^{-1}A''$ lies in $\mathscr{S}$. Since $\mathscr{S}$ contains all the measurable rectangles, (12) follows.

Now S is separable if and only if S' and S'' are both separable. Let us show that, if S is separable, then

$$\mathscr{S}' \times \mathscr{S}'' = \mathscr{S}. \tag{13}$$

In view of (12), it suffices to show that, if G is open in S, then $G \in \mathscr{S}' \times \mathscr{S}''$. But G is a union of rectangles (9) with A' open in S' and A'' open in S'' (so that $A' \times A'' \in \mathscr{S}' \times \mathscr{S}''$), and, if S is separable, G is a countable such union. This proves (13).†

Suppose that X' and X'' are random elements of S' and S'', respectively, and have a common domain $(\Omega, \mathscr{B})$. Now $(X'(\omega), X''(\omega))$ defines a mapping (X', X'') from Ω to $S = S' \times S''$, and clearly $(X', X'')^{-1}(\mathscr{S}' \times \mathscr{S}'') \subset \mathscr{B}$. If S is separable, then, by (13), (X', X'') is a random element of S.

If $S' = S''$, then $\rho'(x', y')$ defines a continuous mapping from $S' \times S'$ to R^1; if S' is separable, then the composition $\rho(X', X'')$ is measurable and hence is a random variable.‡

Measurability of D_h

Let D_h be the set of discontinuities of a mapping h from a metric space S to another metric space S'. Denote the metrics in S and S' by ρ and ρ'. Let $A_{\varepsilon,\delta}$ be the set of x in S for which there exist points y and z in S satisfying

† Without separability, (13) may fail: If $S' = S''$ is a discrete space whose cardinality exceeds that of the continuum, then the diagonal $\{(x, y): x = y\}$ lies in $\mathscr{S}$ but not in $\mathscr{S}' \times \mathscr{S}''$ (see Problem 2 on p. 261 of Halmos (1950)).

‡ The counterexample in the preceding footnote shows that this may be false if S' is not separable.

$\rho(x, y) < \delta$, $\rho(x, z) < \delta$, and $\rho'(hy, hz) \geq \varepsilon$. Then $A_{\varepsilon,\delta}$ is an open set. Since

$$D_h = \bigcup_{\varepsilon} \bigcap_{\delta} A_{\varepsilon,\delta},$$

where ε and δ are restricted to positive rationals, D_h is a Borel set in S. This is true even if h is not measurable (for example, if h is the indicator of a set A, then $D_h = \partial A$ is closed no matter what A may be).

Consider now the set E involved in Theorem 5.5. We shall assume that the mapping h is measurable and that S' is separable and prove $E \in \mathscr{S}$. If $B_{\varepsilon,\delta,i}$ is the set of x such that $\rho'(hx, h_i y) \geq \varepsilon$ for some y with $\rho(x, y) < \delta$, then

$$E = \bigcup_{\varepsilon} \bigcap_{\delta} \bigcap_{k \geq 1} \bigcup_{i \geq k} B_{\varepsilon,\delta,i},$$

where ε and δ range over the positive rationals, and it suffices to find sets $C_{\varepsilon,\delta,i}$ that lie in $\mathscr{S}$ and satisfy

$$B_{\varepsilon,\delta,i} \subset C_{\varepsilon,\delta,i} \subset B_{\frac{1}{2}\varepsilon,\delta,i}.$$

If the sequence u_m is dense in S' and if $H_{\varepsilon,m} = \{x : \rho'(hx, u_m) < \frac{1}{4}\varepsilon\}$, then $H_{\varepsilon,m} \in \mathscr{S}$ and $S = \bigcup_m H_{\varepsilon,m}$. The requirements on $C_{\varepsilon,\delta,i}$ are satisfied by

$$C_{\varepsilon,\delta,i} = \bigcup_{m} (H_{\varepsilon,m} \cap J_{\varepsilon,\delta,i,m}),$$

where $J_{\varepsilon,\delta,i,m}$ is the set of x such that $\rho'(h_i y, hz) \geq \varepsilon$ for some pair of points y and z with $\rho(x, y) < \delta$, $\rho(x, z) < \delta$, and $z \in H_{\varepsilon,m}$ ($J_{\varepsilon,\delta,i,m}$ is open). This proves† $E \in \mathscr{S}$. (The proof of Theorem 5.5 goes through even if E lies outside $\mathscr{S}$, provided it has outer P-measure 0.)

Helly's Theorem

A distribution function is a function $F(x) = F(x_1, \ldots, x_k)$ on R^k with these three properties (we use the terminology and notation of Section 3):

(i) F is everywhere continuous from above;

(ii) $0 \leq F(x) \leq 1$ for all x, F is nondecreasing in each variable, and, for each k-dimensional rectangle $(a, b]$,

$$\sum \pm F(a_1 + \theta_1 d_1, \ldots, a_k + \theta_k d_k) \geq 0, \tag{14}$$

where $d_i = b_i - a_i$, where the sum ranges over all 2^k sequences $(\theta_1, \ldots, \theta_k)$ of 0's and 1's, and where the sign is + or − according as the number of 0's in the sequence is even or odd;

(iii) $F(x) \to 0$ as any one coordinate of x goes to $-\infty$, and $F(x) \to 1$ as all coordinates of x go to ∞.

---

† This proof is due to F. Topsøe.

If P is a probability measure on $(R^k, \mathscr{R}^k)$ and

$$F(x) = P\{y : y \leq x\}, \qquad x \in R^k, \tag{15}$$

then F is a distribution function. It is a standard fact of real variable theory that, if F is a distribution function, then there is exactly one P satisfying (15). If F has properties (i) and (ii) above (but perhaps not (iii)), then there is a finite measure μ on $(R^k, \mathscr{R}^k)$ such that

$$F(x) = \mu\{y : y \leq x\}, \qquad x \in R^k; \tag{16}$$

μ will satisfy $\mu(R^k) \leq 1$ but need not be a probability measure.

(*Helly's Selection Theorem.*) *If $\{F_n\}$ is a sequence of distribution functions on R^k, then there exist a subsequence $\{F_{n'}\}$ and a function F satisfying conditions* (i) *and* (ii) *above (but perhaps not* (iii)*) such that*

$$\lim_{n'} F_{n'}(x) = F(x) \tag{17}$$

for all continuity points x of F.

Proof. Let R_0^k denote the set of rational points in R^k—the set of points whose coordinates are rational—and let $\{r(1), r(2), \ldots\}$ be an enumeration of the points of R_0^k. For each F_n in the given sequence of distribution functions,

$$\big(F_n(r(1)), F_n(r(2)), \ldots\big) \tag{18}$$

is a point of R^∞. Since $0 \leq F_n(x) \leq 1$, it follows by the compactness criterion for R^∞ (see p. 219) that some subsequence of the points (18) converges in the sense of R^∞ to some point $(z_1, z_2, \ldots)$ of R^∞. Define F_0 on R_0^k by $F_0(r(k)) = z_k$. We see in this way that there exists a function F_0 on R_0^k and a subsequence $\{F_{n'}\}$ of $\{F_n\}$ such that

$$\lim_{n'} F_{n'}(r) = F_0(r), \qquad r \in R_0^k. \tag{19}$$

Since each F_n is a distribution function, it follows from (19) that F_0 satisfies condition (ii) on R_0^k: $0 \leq F_0(r) \leq 1$, F_0 is nondecreasing as each coordinate varies over the rationals, and (14) holds if a and b lie in R_0^k. If we define F on R^k by

$$F(x) = \inf\{F_0(r) : x < r,\ r \in R_0^k\}, \qquad x \in R^k,$$

then F has properties (i) and (ii) (but perhaps not (iii)).

If F is continuous at x, then, given $\varepsilon > 0$, we can find points r' and r'' of R_0^k such that $r' < x < r''$ and

$$F(x) - \varepsilon < F_0(r') \leq F_0(r'') < F(x) + \varepsilon.$$

For each n',

$$F_{n'}(r') \leq F_{n'}(x) \leq F_{n'}(r'').$$

From these relations and (19) it follows that

$$F(x) - \varepsilon \leq \liminf_{n'} F_{n'}(x) \leq \limsup_{n'} F_{n'}(x) \leq F(x) + \varepsilon.$$

Since ε was arbitrary, (17) follows. This proves Helly's theorem.

Kolmogorov's Theorem

We shall prove the Kolmogorov (or Daniell–Kolmogorov) existence theorem. Let π_k be the projection from R^∞ to R^k defined by $\pi_k(x) = (x_1, \ldots, x_k)$, as described in Section 3. For $k > 1$, let ψ_k be the projection from R^k to R^{k-1} defined by $\psi_k(x_1, \ldots, x_k) = (x_1, \ldots, x_{k-1})$. Since π_k and ψ_k are continuous, they are measurable ($\pi_k^{-1}\mathscr{R}^k \subset \mathscr{R}^\infty$ and $\psi_k^{-1}\mathscr{R}^{k-1} \subset \mathscr{R}^k$).

If P is a probability measure on $(R^\infty, \mathscr{R}^\infty)$, define probability measures μ_k on $(R^k, \mathscr{R}^k)$ by

(20) $$\mu_k = P\pi_k^{-1}.$$

Since

(21) $$\pi_{k-1} = \psi_k \pi_k, \qquad k > 1,$$

the measures μ_k satisfy

(22) $$\mu_{k-1} = \mu_k \psi_k^{-1}, \qquad k > 1.$$

The problem is to go the other way and construct, from given μ_k satisfying the consistency conditions (22), a P satisfying (20).

(*Kolmogorov's Existence Theorem.*) *If probability measures* μ_k *on* $(R^k, \mathscr{R}^k)$, $k \geq 1$, *satisfy* (22), *then there exists on* $(R^\infty, \mathscr{R}^\infty)$ *a unique probability measure P satisfying* (20).

Proof. For $k > i \geq 1$, define a continuous mapping $\psi_{k,i}$ from R^k to R^i by $\psi_{k,i}(x_1, \ldots, x_k) = (x_1, \ldots, x_i)$. Note that $\psi_{k,k-1} = \psi_k$ and that

(23) $$\psi_{k,i} = \psi_{i+1} \cdots \psi_{k+1}\psi_k, \qquad \pi_i = \psi_{k,i}\pi_k$$

From this and (22) it follows that

(24) $$\mu_i = \mu_k \psi_{k,i}^{-1}, \qquad k > i \geq 1.$$

Let $\mathscr{F}$ be the collection of finite-dimensional sets, as defined in Section 3; $\mathscr{F}$, which consists of the sets of the form

(25) $$A = \pi_i^{-1}H, \qquad H \in \mathscr{R}^i,$$

is a field generating the σ-field $\mathscr{R}^\infty$. A set of the form (25) can also be cast in the form

(26) $$A = \pi_k^{-1}H', \qquad H' \in \mathscr{R}^k,$$

if $k > i$; we need only take $H' = \psi_{k,i}^{-1}H$ and use (23).

Suppose now that A is given both by (25) and by (26), where $i < k$. Then $\alpha = (\alpha_1, \ldots, \alpha_k)$ lies in H' if and only if $(\alpha_1, \ldots, \alpha_k, 0, 0, \ldots)$ lies in $\pi_k^{-1}H' = \pi_i^{-1}H$, which is true if and only if $(\alpha_1, \ldots, \alpha_i)$ lies in H. Thus $H' = \psi_{k,i}^{-1}H$, and (24) implies

$$\mu_i H = \mu_k H'. \tag{27}$$

Since (25) and (26) together imply (27), we may consistently define a function P on $\mathscr{F}$ by setting $P(A) = \mu_i(H)$ if A is given by (25). Clearly, $P(0) = 0$, $P(R^\infty) = 1$, and $0 \le P(A) \le 1$. Suppose A is given by (25) and B by

$$B = \pi_k^{-1}J, \qquad J \in \mathscr{R}^k,$$

where we assume $k \ge i$. Then A also has the form (26), and if $A \cap B = 0$, we must have $H' \cap J = 0$, so that

$$P(A \cup B) = \mu_k(H' \cup J) = \mu_k(H') + \mu_k(J) = P(A) + P(B).$$

Thus P is a finitely additive probability measure on $\mathscr{F}$.

Let us prove that P is completely additive on $\mathscr{F}$ by showing that, if $A_k \in \mathscr{F}$,

$$A_1 \supset A_2 \supset \cdots, \tag{28}$$

and $\bigcap_i A_i = 0$, then $\lim_i P(A_i) = 0$. Since A_i lies in $\mathscr{F}$, it has the form

$$A_i = \pi_{n_i}^{-1}H_i, \qquad H_i \in \mathscr{R}^{n_i}. \tag{29}$$

Since a set of the form (25) can be cast in the form (26) for each k exceeding i, there is no loss of generality in assuming that

$$n_1 < n_2 < n_3 < \cdots. \tag{30}$$

We shall show that, if sets (29) satisfy (28) and (30), and if there exists a positive ε such that

$$P(A_i) > \varepsilon, \qquad i = 1, 2, \ldots, \tag{31}$$

then $\bigcap_i A_i \ne 0$.

From (31) and (29), it follows by the definition of P that $\mu_{n_i}(H_i) > \varepsilon$. By Theorems 1.1 and 1.4, there exists in R^{n_i} a compact subset K_i of H_i with $\mu_{n_i}(H_i - K_i) < \varepsilon/2^{i+1}$. If $B_i = \pi_{n_i}^{-1}K_i$, then $B_i \in \mathscr{F}$, $B_i \subset A_i$, and

$$P(A_i - B_i) < \frac{\varepsilon}{2^{i+1}}.$$

From this and (28) it follows that $C_i = B_1 \cap \cdots \cap B_i$ is a subset of A_i satisfying $P(A_i - C_i) \le \sum_{j=1}^{i} P(A_j - B_j) < \varepsilon/2$. By (31), therefore, $P(C_i) > \varepsilon/2$, so that C_i is nonempty.

We have constructed nonempty sets $C_i \subset \pi_{n_i}^{-1} K_i$ with K_i compact and

(32) $$C_1 \supset C_2 \supset \cdots .$$

Since $C_i \subset A_i$, $\bigcap_i A_i \neq 0$ will follow if we find a point common to all the C_i. Let $x(j)$ be an arbitrary element of C_j. If $j \geq i$, then $x(j) \in C_i$ by (32), and hence $\pi_{n_i} x(j) \in K_i$. Since K_i is compact, we have $\sup_{j \geq i} |x_i(j)| < \infty$ and hence $\sup_j |x_i(j)| < \infty$. For each i, therefore, $\{x_i(1), x_i(2), \ldots\}$ is a bounded set of real numbers. Therefore (see p. 219) some subsequence $\{x(j')\}$ of $\{x(j)\}$ converges in the sense of R^∞ to some limit $x \in R^\infty$. Since $x(j') \in C_{j'}$, and each C_i is closed, (32) implies that x lies in $\bigcap_i C_i$.

We have shown that P is a completely additive probability measure on the finitely additive field $\mathscr{F}$. Since $\mathscr{F}$ generates $\mathscr{R}^\infty$, P can be extended† to a probability measure on R^∞. Clearly, P satisfies (20). Since $\mathscr{F}$ is a determining class, there can be at most one P satisfying (20), which proves Kolmogorov's theorem.

It is not difficult to go on from here and prove a more general version of the theorem. Let R^T be the space of all real functions $x = x(t)$ on an arbitrary set T, which we may as well take to be infinite. (If T consists of the integers, then $R^T = R^\infty$.) For a finite ordered set $\sigma = (s_1, \ldots, s_{k_\sigma})$ of distinct elements of T, define $\pi_\sigma : R^T \to R^{k_\sigma}$ by $\pi_\sigma(x) = \big(x(s_1), \ldots, x(s_{k_\sigma})\big)$; let $\mathscr{R}^T$ be the σ-field generated by the sets $\pi_\sigma^{-1} H$ with $H \in \mathscr{R}^{k_\sigma}$ (for all σ and H). Then $(R^T, \mathscr{R}^T)$ is a measurable space; no topology is involved (or need be involved).

A collection of probability measures μ_σ on $(R^{k_\sigma}, \mathscr{R}^{k_\sigma})$ (one measure for each σ) is consistent if $\mu_{\sigma_2} = \mu_{\sigma_1} \psi^{-1}$ whenever $\sigma_2 = (s_{i_1}, \ldots, s_{i_j})$ is a permutation of j of the elements of $\sigma_1 = (s_1, \ldots, s_k)$ (here $j \leq k$) and $\psi : R^k \to R^j$ is defined by $\psi(x_1, \ldots, x_k) = (x_{i_1}, \ldots, x_{i_j})$.

If the μ_σ are consistent, there exists a unique probability measure P on $(R^T, \mathscr{R}^T)$ such that $P\pi_\sigma^{-1} = \mu_\sigma$ for all σ.

To prove this, define, for a sequence $\tau = (t_1, t_2, \ldots)$ of elements of T, a mapping $\pi_\tau : R^T \to R^\infty$ by $\pi_\tau(x) = \big(x(t_1), x(t_2), \ldots\big)$. From the existence theorem for R^∞, it follows that there exists a probability measure P_τ on $(R^\infty, \mathscr{R}^\infty)$ such that $P_\tau \pi_k^{-1} = \mu_{t_1 \cdots t_k}$ for every k.

Now the class of sets $\pi_\tau^{-1} H$ with $H \in \mathscr{R}^\infty$ (for all τ and H) is exactly $\mathscr{R}^T$ (and does not merely generate it), and it follows easily that $P(\pi_\tau^{-1} H) = P_\tau(H)$ consistently defines the desired P.

Measurability of Some Mappings

Let T denote the unit interval $[0, 1]$; let $\mathscr{T}$ denote the class of linear Borel subsets of T. For each t, the projection π_t from C to R^1 is measurable

† Halmos (1950, Chapter 3).

$\mathscr{C}$. Since the mapping

$$(x, t) \to x(t) \tag{33}$$

from $C \times T$ to R^1 is continuous in the product topology, and since $\mathscr{C} \times \mathscr{T}$ is the σ-field of Borel sets for this topology (see p. 225), (33) is measurable $\mathscr{C} \times \mathscr{T}$.

For $x \in C$, let $h(x)$ be the Lebesgue measure of the set of t in T for which $x(t) > 0$. We want to prove that h is measurable $\mathscr{C}$, and we shall derive this from a more general result. If we define a real function v of a real variable by

$$v(\alpha) = \begin{cases} 1 & \text{if } \alpha > 0 \\ 0 & \text{if } \alpha \le 0, \end{cases} \tag{34}$$

then

$$h(x) = \int_0^1 v(x(t))\, dt. \tag{35}$$

Now v is (i) Borel measurable, (ii) bounded, and (iii) continuous except on a set of Lebesgue measure 0. Using these three properties of v, we shall show that the function h defined by (35) is measurable $\mathscr{C}$ and is continuous except on a set of Wiener measure 0.

Since, for each x in C, $v(x(t))$ is a bounded, measurable function of t, the integral in (35) is well defined. Since the mapping (33) is measurable $\mathscr{C} \times \mathscr{T}$ and since v is measurable, the mapping $\psi: C \times T \to R^1$ defined by $\psi(x, t) = v(x(t))$ is also measurable. Since ψ is bounded, $h(x) = \int_0^1 \psi(x, t)\, dt$ is measurable in x.† Hence h is measurable $\mathscr{C}$.

If D_v is the set of discontinuities of v, then by (iii), $\lambda(D_v) = 0$, where λ denotes Lebesgue measure on the line. Let E denote the set of (x, t) for which $x(t) \in D_v$. For each positive t,

$$W\{x:(x, t) \in E\} = \frac{1}{\sqrt{2\pi t}} \int_{D_v} e^{-\frac{1}{2}u^2}\, du = 0.$$

It follows by Fubini's theorem applied to the measure $W \times \lambda$ on $\mathscr{C} \times \mathscr{T}$ that

$$\lambda\{t:(x, t) \in E\} = 0$$

if $x \notin A$, where A is an element of $\mathscr{C}$ with $W(A) = 0$. Now if x_n converges to x pointwise, then $v(x_n(t)) \to v(x(t))$ for each t such that $x(t) \notin D_v$. If $x \notin A$, this is true for almost all t, and hence

$$\int_0^1 v(x_n(t))\, dt \to \int_0^1 v(x(t))\, dt$$

by the bounded convergence theorem. This proves that h is continuous except at points forming a set of W-measure 0.

† See Halmos (1950, Chapter 7).

The argument goes through if W is replaced by an arbitrary P with the property that $P\pi_t^{-1}$ is absolutely continuous with respect to Lebesgue measure for almost all t. This is true of W°, for example.

Except for the proof that (33) is measurable, this argument goes through word for word if C is replaced by D. And the measurability of (33) can be proved by adapting the proof in Section 14 of the measurability of π_t.

One more function on C remains to be analyzed, namely the supremum $h(x)$ of those t in $[0, 1]$ for which $x(t) = 0$. Since $\{x : h(x) < \alpha\}$ is open, certainly h is measurable. If h is discontinuous at x, then $x(t)$ must keep to one side of 0 in $(h(x), 1)$ and keep to the same side of 0 in $(h(x) - \varepsilon, h(x))$ for some ε. That h is continuous except on a set of Wiener measure 0 will therefore follow if we show that, for each t_0, the supremum and infimum of W over $[t_0, 1]$ have continuous distributions. Since $W_t - W_{t_0}$ with t ranging over $[t_0, 1]$ is distributed as a Wiener path with a linearly transformed time scale, $-W_{t_0} + \sup_{t \geq t_0} W_t$ has a continuous distribution (see (10.17)). This last random variable and W_{t_0} are independent, and hence their sum also has a continuous distribution. The infimum is treated the same way. (This function h extended to have domain D is *not* continuous except on a set of Wiener measure 0.)

More Measurability

In Section 17 we defined the mapping $\psi : D \times D_0 \to D$ by $\psi(x, \varphi) = x \circ \varphi$ (see (17.5) and (17.6)), and we are to prove it measurable ($\psi^{-1}\mathscr{D} \subset \mathscr{D} \times \mathscr{D}_0$). Since the finite-dimensional sets in D generate $\mathscr{D}$, it is enough to prove that, for each t, the mapping

$$(x, \varphi) \to \pi_t(x \circ \varphi) = x(\varphi(t)) \tag{36}$$

is measurable $\mathscr{D} \times \mathscr{D}_0$. If $\varphi_k(t)$ is the smallest ratio i/k not smaller than $\varphi(t)$, then $\varphi_k(t) \downarrow \varphi(t)$ for each t. Hence the mapping

$$(x, \varphi) \to x(\varphi_k(t)) \tag{37}$$

converges pointwise to (36), and it suffices to prove this latter mapping measurable $\mathscr{D} \times \mathscr{D}_0$. Now $\{(x, \varphi) : x(\varphi_k(t)) \leq \alpha\}$ is the union of

$$\{(x, \varphi) : \varphi(t) = 0\} \cap \{(x, \varphi) : x(0) \leq \alpha\} \tag{38}$$

with the sets

$$\left\{(x, \varphi) : \frac{i-1}{k} < \varphi(t) \leq \frac{i}{k}\right\} \cap \left\{(x, \varphi) : x\left(\frac{i}{k}\right) \leq \alpha\right\}, \qquad i = 1, \ldots, k. \tag{39}$$

If $H \in \mathscr{R}^1$, then $\{\varphi \in D_0 : \varphi(t) \in H\}$ is the intersection with D_0 of the subset $\pi_t^{-1}H$ of D and hence lies in $\mathscr{D}_0$. Therefore $\{(x, \varphi) : \varphi(t) \in H\} \in \mathscr{D} \times \mathscr{D}_0$. Similarly, $\{(x, \varphi) : x(t) \in H\} \in \mathscr{D} \times \mathscr{D}_0$. Thus the sets (38) and (39) all lie in $\mathscr{D} \times \mathscr{D}_0$, which proves the measurability of (37).

APPENDIX III

Theoretical Complements

This appendix, most of which requires general (nonmetric) topology, treats questions that arise naturally out of the theory of weak convergence but are irrelevant to its application.

The Problem of Measure

Does there exist on the class of all subsets of a given set U a probability measure that has no atoms (that assigns measure 0 to each individual point)? The answer certainly depends only on the cardinality of U. It is clear that no such measure can exist if U is countable, and it can be shown† under the assumption of the continuum hypothesis that no such measure can exist if U has the power of the continuum.

If there does exist an atomless probability measure on the class of all subsets of U, then the cardinal of U is called *measurable*; otherwise this cardinal is called *nonmeasurable*. Thus $\aleph_0$ is nonmeasurable, and the continuum hypothesis implies that the power of the continuum is also nonmeasurable. (Among sets on the line, it is the nonmeasurable ones that are aberrant. Here the terminology is reversed: Among cardinals, it is the measurable ones that are aberrant.)

Whether there exist measurable cardinals is a famous unsolved problem of set theory—the so-called problem of measure.‡ If measurable cardinals

†See Birkhoff (1961, p. 187).

‡ See Keisler and Tarski (1964) for an account of the problem of measure and a large bibliography.

exist at all, they must be so large as never to arise in a natural way in mathematics. We shall see that certain irregular phenomena in weak convergence are equally remote from ordinary mathematical activity because they can arise only if there exist measuable cardinals.

Separable Measures

Consider now a probability measure P on the class $\mathscr{S}$ of Borel sets in a metric space S. According to Theorem 1.4, P is tight if S is separable and complete. Certainly, completeness can be weakened to topological completeness (the condition that there exists for S an equivalent metric under which it is complete), and an examination of the proof shows that the assumption of separability can be weakened to the assumption that P has separable support. Let us define P itself to be *separable* if it has a separable support.† We then have the following result.

THEOREM 1 *If P is separable and if S is topologically complete, then P is tight.*

Remark 1. The hypothesis here that P is separable cannot be suppressed: If P is tight, then it has a σ-compact support, and hence is separable. But, of course, S itself need not be separable.

Remark 2. The hypothesis of topological completeness cannot be suppressed either: Let S be a thick subset of $[0, 1]$—a set whose inner Lebesgue measure $\lambda_*(S)$ is 0 and whose outer Lebesgue measure $\lambda^*(S)$ is 1—with the relative topology of the line. Here $\mathscr{S}$ consists of the sets $S \cap A$ with A an ordinary linear Borel set (see (7) on p. 224). If P is the restriction of λ^* to $\mathscr{S}$, then P is completely additive; it is a probability measure because $\lambda^*(S) = 1$. If K is compact in S, it is an ordinary Borel set and hence $P(K) = 0$ because $\lambda_*(S) = 0$. Hence P is not tight. Since P is separable (S itself being separable), Theorem 1 becomes false without the completeness hypothesis.

Remark 3. If S consists of the rationals with the relative topology of the line, then, since S is σ-compact, each P on S is tight. On the other hand, by Baire's category theorem,‡ S is not topologically complete. It is an open problem to characterize topologically those metric spaces that support tight probability measures only.

The question now arises, do nonseparable probability measures exist?

---

† This is the specialization to the metric case of the more general notion of a τ-*smooth* measure; see Varadarajan (1958a and 1961a).

‡ See Kelley (1955, p. 200).

THEOREM 2 *A necessary and sufficient condition that each probability measure on $\mathscr{S}$ be separable is that each discrete*† *subset of S have nonmeasurable cardinal.*

Proof. We first prove the necessity. Suppose S contains a discrete subset A_0 with measurable cardinal. We shall construct a nonseparable P on $\mathscr{S}$. Let Q be an atomless probability measure on the class of all subsets of A_0, and define an atomless probability measure P on $\mathscr{S}$ by $P(A) = Q(A \cap A_0)$. If A is separable, then $A \cap A_0$ is both discrete and separable and hence (p. 216) countable, so that $Q(A \cap A_0) = 0$. Thus P can have no separable support.

The proof of sufficiency is much deeper. Suppose that each discrete subset of S has nonmeasurable cardinal and consider an arbitrary probability measure P on $\mathscr{S}$. We are to show that P is separable. Note first that it suffices to show that each open cover $\mathscr{G}$ of S contains a countable subclass $\{G_1, G_2, \ldots\}$ with

(1) $$P(\textstyle\bigcup_n G_n) = 1.$$

Indeed, for each k there is then a sequence $A_{k1}, A_{k2}, \ldots$ of open $1/k$-spheres with $P(\bigcup_n A_{kn}) = 1$, and $\bigcap_k \bigcup_n A_{kn}$ is a separable support for P.

Consider then an open cover $\mathscr{G}$ of S. By the paracompactness theorem,‡ there is a class $\mathscr{H}$ with these properties: (i) $\mathscr{H}$ is an open cover of S. (ii) Each element of $\mathscr{H}$ is contained in some element of $\mathscr{G}$. (iii) $\mathscr{H}$ can be represented as a countable union

(2) $$\mathscr{H} = \textstyle\bigcup_n \mathscr{H}_n,$$

where, for each n,

(3) $$\inf\{\rho(A, B): A, B \in \mathscr{H}_n, A \neq B\} > 0,$$

ρ being the metric on S.

Fix n for the moment. From each element of $\mathscr{H}_n$ choose a single point, forming in this way a set S_n. Because of (3), S_n is discrete. For $A \subset S_n$, let $\lambda_n(A)$ be the P-measure of the union of all those elements of $\mathscr{H}_n$ that meet A (this union is open). Since the elements of $\mathscr{H}_n$ are disjoint, λ_n is a finite measure on the class of all subsets of S_n. Since S_n has nonmeasurable cardinal by hypothesis, λ_n has a countable support. (Otherwise we could subtract away the atomic part of λ_n to obtain on the class of all subsets of S_n an atomless measure ν_n with $0 < \nu_n(S_n) < \infty$, and ν_n could be normalized to a probability measure.) Let A_n be a countable subset of S_n with

(4) $$\lambda_n(S_n - A_n) = 0,$$

† See p. 215.

‡ See Kelley (1955, p. 129).

let $\mathscr{J}_n$ be the class of elements of $\mathscr{H}_n$ that meet A_n, and define $\mathscr{J} = \bigcup_n \mathscr{J}_n$. Since each class $\mathscr{J}_n$ is countable, so is $\mathscr{J}$.

If H_n is the union of the sets in $\mathscr{H}_n$ and J_n is the union of the sets in $\mathscr{J}_n$, then, by (4) and the definition of λ_n, $P(H_n - J_n) = 0$. Since $\mathscr{H}$ covers S, it follows by (2) that $\bigcup_n H_n = S$ and hence, if $J = \bigcup_n J_n$, that $P(J) = 1$. Now J is just the union of the sets in $\mathscr{J}$, which we can enumerate as G_1', G_2', ... Each G_n' lies in $\mathscr{H}$ and hence is contained in some element G_n of $\mathscr{G}$. Since $P(J) = 1$, it follows that the G_n satisfy (1), as required.

That each P on a separable S is itself separable, an obvious fact, follows because each discrete subset must have the nonmeasurable cardinal $\aleph_0$. Theorem 2 also implies that each P on S is separable if S itself has nonmeasurable cardinal, which is true if the power of S does not exceed that of the continuum and the continuum hypothesis is true. In any case, the search for a nonseparable probability measure belongs properly not to probability but to set theory.

The Topology of Weak Convergence

Consider now the space $Z = Z(S)$ of probability measures on $(S, \mathscr{S})$. Make Z into a Hausdorff space by taking as the basic neighborhoods of P the sets of the form

$$\left\{Q : \left| \int f_i \, dQ - \int f_i \, dP \right| < \varepsilon, \quad i = 1, \ldots, k\right\}, \tag{5}$$

where ε is positive and $f_1, \ldots, f_k$ are elements of $C(S)$. The resulting topology we shall call the *topology of weak convergence* and denote by $\mathscr{W}$. We have $P_n \Rightarrow P$ if and only if the sequence $\{P_n\}$ $\mathscr{W}$-converges to P.

We shall show that there are three other bases for $\mathscr{W}$, namely, the sets

$$\{Q : Q(F_i) < P(F_i) + \varepsilon, i = 1, \ldots, k\} \tag{6}$$

with F_i closed, the sets

$$\{Q : Q(G_i) > P(G_i) - \varepsilon, i = 1, \ldots, k\} \tag{7}$$

with G_i open, and the sets

$$\{Q : |Q(A_i) - P(A_i)| < \varepsilon, i = 1, \ldots, k\} \tag{8}$$

with A_i a P-continuity set. Certainly, each of these three classes of sets is a basis (for some topology). Theorem 2.1 is the sequential version of the following result.

THEOREM 3 *The three bases just described all generate* $\mathscr{W}$.

Proof. Since (6) and (7) coincide if $G_i = F_i^c$, the two corresponding bases are identical. We shall show that the basis (6) generates the same topology as does (8) and then that it generates the same topology as (5), that is, $\mathscr{W}$.

Fix P. If A is a P-continuity set, then for Q in a set of the form (6) we have $Q(A) \leq Q(A^-) < P(A^-) + \varepsilon = P(A) + \varepsilon$, and for Q in a set of the form (6) (or (7)) we have $Q(A) \geq Q(A^\circ) > P(A^\circ) - \varepsilon = P(A) - \varepsilon$. Since the sets (6) do have the properties of a basis, each set (8) contains a set (6) (with the same P). On the other hand, if F is closed, then there is some δ for which

$$F^\delta = \{x: \rho(x, F) < \delta\} \tag{9}$$

is a P-continuity set and

$$P(F^\delta) < P(F) + \tfrac{1}{2}\varepsilon. \tag{10}$$

If $|Q(F^\delta) - P(F^\delta)| < \frac{1}{2}\varepsilon$, then $Q(F) < P(F) + \varepsilon$. Thus each set (6) contains a set (8).

We now compare (6) with (5). Given a closed F, choose δ so that (9) satisfies (10), and then choose in $C(S)$ an f with value 1 on F, value 0 outside F^δ, and value everywhere contained between 0 and 1 (Theorem 1.2). If $|\int f\,dQ - \int f\,dP| < \frac{1}{2}\varepsilon$, then $Q(F) < P(F) + \varepsilon$. Thus each (6) contains a (5).

It remains only to find within (5) a set of the form (6). We need consider only a single f in $C(S)$ and we may assume that $0 < f(x) < 1$ for all x. Choose k so that $1/k < \varepsilon$ and let $F_i = \{x: i/k \leq f(x)\}$. By (2.2) we have $\int f\,dQ < \varepsilon + k^{-1}\sum_i Q(F_i)$ and $k^{-1}\sum_i P(F_i) \leq \int f\,dP$. For Q in (6) we thus have $\int f\,dQ < f\,dP + 2\varepsilon$. The same argument applied to $1 - f$ completes the proof.

By $N(P)$ we shall mean a $\mathscr{W}$-neighborhood of P having any of the forms (5) through (8). We are free to work with whatever base is most convenient.

THEOREM 4 *The probability measures with finite support are $\mathscr{W}$-dense in Z.*

Proof. Consider a neighborhood $N(P)$ of the form (6). From each nonempty set B in the finite partition generated by $F_1, \ldots, F_k$ choose a single point and place there a mass $P(B)$. The resulting measure has finite support and lies in $N(P)$ because it agrees with P for each F_i.

Theorem 4 applies even to the approximation of a nonseparable P (if there are any). Therefore, if a subset Π of Z consists exclusively of separable measures, one cannot conclude that each element of the $\mathscr{W}$-closure of Π is separable. If, however, the elements of Π have a *common* separable support and P lies in the $\mathscr{W}$-closure of Π, then P is separable (because, if A is the closure of a common separable support, then A is separable and it follows by Theorem 3 that $P(A) = 1$). Since a countable collection of separable measures has a common separable support, P is separable if $P_n \Rightarrow P$ and each P_n is separable.

It is natural to ask when $\mathscr{W}$ is metrizable. For P and Q in Z, let $p(P, Q)$ be the infinum of those positive ε for which the inequalities $Q(A) \leq P(A^\varepsilon) + \varepsilon$

and $P(A) \leq Q(A^\varepsilon) + \varepsilon$ hold for all A in $\mathscr{S}$, where $A^\varepsilon = \{x: \rho(x, A) < \varepsilon\}$. From $p(P, Q) = 0$ it follows that P and Q agree on closed sets and hence (Theorem 1.1) are identical. The remaining postulates being easy to check, p is a metric on Z—the *Prohorov metric*. We shall see that, if $\mathscr{W}$ can be metrized at all, then p does it.

Fix P. If for each $N(P)$ there is about P a p-sphere contained in $N(P)$, we say p is at least as fine as $\mathscr{W}$ at P. If, in addition, for each p-sphere about P there is an $N(P)$ contained in it, we say p and $\mathscr{W}$ are equivalent at P.

THEOREM 5 *For arbitrary P, p is at least as fine as $\mathscr{W}$ at P. Moreover, p and $\mathscr{W}$ are equivalent at P if and only if $\mathscr{W}$ has a countable basis at P, which in turn holds if and only if P is separable.*

Proof. Given P, F, and ε, choose δ so that $\delta < \varepsilon$ and (9) satisfies (10). If $p(P, Q) < \frac{1}{2}\delta$, then $Q(F) < P(F^\delta) + \delta < P(F) + \varepsilon$. Thus each $N(P)$ of the form (6) contains a p-sphere about P, which proves the first part of the theorem.

If p and $\mathscr{W}$ are equivalent at P, then certainly $\mathscr{W}$ has a countable basis at P. It remains then to prove that this last condition implies the separability of P and that the separability of P implies the equivalence of p and $\mathscr{W}$ at P.

Suppose $\mathscr{W}$ has at P a countable basis $N_n(P)$, $n = 1, 2, \ldots$. By Theorem 4, each $\bigcap_{k=1}^n N_k(P)$ contains a P_n with separable (even finite) support. But then $P_n \Rightarrow P$, which, as remarked after the proof of Theorem 4, implies that P is separable.

Finally, suppose that P is separable. We are to prove that p and $\mathscr{W}$ are equivalent at P, for which, in view of the first part of the theorem, it suffices to show that a p-sphere with radius ε and center P must contain some $N(P)$.

Choose δ so that $3\delta < \varepsilon$. Cover the separable support of P by P-continuity spheres of diameter less than δ, pass to a countable subcover (p. 216), and by the usual procedure construct disjoint P-continuity sets $A_1, A_2, \ldots$ that cover the support. Now choose k so that

$$P\left(\bigcup_{i=1}^{k} A_i\right) > 1 - \delta. \tag{11}$$

Each of $A_1, \ldots, A_k$ has diameter less than δ and, if $\mathscr{A}$ is the (finite) class of unions of these sets, each element of $\mathscr{A}$ is a P-continuity set.

By Theorem 3, there is an $N(P)$ such that $Q \in N(P)$ implies

$$|Q(A) - P(A)| < \delta, \qquad A \in \mathscr{A}. \tag{12}$$

This inequality and (11) imply

$$Q\left(\bigcup_{i=1}^{k} A_i\right) > 1 - 2\delta. \tag{13}$$

We shall show that $p(P, Q) < \varepsilon$ if $Q \in N(P)$.

Given the general B in $\mathscr{S}$, consider the union of those sets among $A_1, \ldots,$ A_k that meet B. If A denotes this union, then it satisfies (12). The relation $B \subset A \cup (\bigcup_{i=1}^k A_i)^c$, the relation $A \subset B^\delta$ (true because diam $A_i < \delta$), and the inequalities (11), (12), and (13) combine to yield $P(B) \le Q(B^\delta) + 2\delta$ and $Q(B) \le P(B^\delta) + 3\delta$. Since $3\delta < \varepsilon$, $p(P, Q) < \varepsilon$ follows.

It follows by Theorems 2 and 5 that $\mathscr{W}$ is metrizable if and only if each separable subset of S has nonmeasurable cardinal, in which case it is metrizable by p.

One can ask after the separability of Z. If S is separable, then the p and Z topologies coincide on Z, and it follows by Theorem 4 that the probability measures with finite support are dense. It is not difficult to go on to show that, if A_0 is a countable, dense set in S and Z_0 consists of those P having finite support contained in A_0 and rational masses, then Z_0 is countable and dense relative to p. Thus Z is metrizable and separable if S is separable.

On the other hand, if $\mathscr{W}$ satisfies the second axiom of countability, it can be metrized by p and is separable. Since S with its metric is homeomorphic with the set of unit masses in Z, it follows that S is separable.

Prohorov's Theorem

Let us reexamine Prohorov's theorem in the light of the preceding results. If Π is a subset of Z, let Π_p and $\Pi_{\mathscr{W}}$ denote its closures with respect to p and $\mathscr{W}$. Theorem 5 implies that $\Pi_p \subset \Pi_{\mathscr{W}}$ (with strict inclusion for some Π if $\mathscr{W}$ is not metrizable).

A set in a topological space is compact if each open cover of it has a finite subcover; it is sequentially compact if each sequence in it contains a subsequence converging to a limit that is also in the set. Consider these three statements:

1^0. Each sequence in Π contains a $\mathscr{W}$-convergent subsequence.
2^0. $\Pi_{\mathscr{W}}$ is sequentially compact.
3^0. $\Pi_{\mathscr{W}}$ is compact.

The limit of the convergent subsequence in 1^0 is not required to lie in Π, but, of course, it will lie in $\Pi_{\mathscr{W}}$. Clearly, 2^0 implies 1^0. From the three statements, one can make up five other implications; whether any of them are true is unknown.† If each separable subspace of S has nonmeasurable cardinal, however, the $\mathscr{W}$ topology is metrizable, so that all three concepts coincide, and we shall see that this is also true if Π is tight. In this book (see the beginning of Section 6), we have *defined* Π to be relatively compact if it satisfies 1^0.

---

† To me.

Consider the direct half of Prohorov's theorem (Theorem 6.1).

THEOREM 6 *If Π is tight, then $\Pi_p = \Pi_{\mathscr{W}}$, $\Pi_{\mathscr{W}}$ is tight, the p and $\mathscr{W}$ topologies coincide on $\Pi_{\mathscr{W}}$, and $\Pi_{\mathscr{W}}$ is both compact and sequentially compact (in either topology).*

Proof. Since Π is tight, there is to each ε a compact K such that $Q(K) > 1 - \frac{1}{2}\varepsilon$ for all Q in Π. Suppose $P \in \Pi_{\mathscr{W}}$. By Theorem 3, there is an $N(P)$ such that $Q \in N(P)$ implies $Q(K) < P(K) + \frac{1}{2}\varepsilon$. Choosing Q in $N(P) \cap \Pi$ leads to $P(K) > 1 - \varepsilon$. Thus $\Pi_{\mathscr{W}}$ is tight.

In particular, each element of $\Pi_{\mathscr{W}}$ is separable. Therefore (Theorem 5) the p and $\mathscr{W}$ topologies coincide on $\Pi_{\mathscr{W}}$; since $\Pi_p \subset \Pi_{\mathscr{W}}$ in any case, $\Pi_p = \Pi_{\mathscr{W}}$ follows. From the tightness of $\Pi_{\mathscr{W}}$, it follows by Theorem 6.1 that $\Pi_{\mathscr{W}}$ is sequentially compact in the sense of $\mathscr{W}$. Since compactness and sequential compactness are the same in the metric case, the proof is complete.

The direct Prohorov theorem can also be used to investigate the completeness of Z. Suppose that S is complete and that each element of Z is separable (and hence tight). Then p and $\mathscr{W}$ coincide on Z, and we shall show that Z is p-complete. For this it is enough to show that a p-fundamental sequence $\{P_n\}$ necessarily contains a weakly convergent subsequence, which will follow by the direct Prohorov theorem if we prove that $\{P_n\}$ is tight. Finally, as we saw in the proof of Theorem 6.2 (see p. 40), since S is complete $\{P_n\}$ is tight if, for each positive ε and δ, there exists a finite collection $A_1, \ldots, A_k$ of δ-spheres such that

$$P_n(A_1 \cup \cdots \cup A_k) > 1 - \varepsilon \tag{14}$$

for all n.

Choose η so that $2\eta < \min\{\varepsilon, \delta\}$, and then choose n_0 so that $n \geq n_0$ implies $p(P_n, P_{n_0}) < \eta$. Since P_{n_0} is separable, there exist finitely many η-spheres $B_1, \ldots, B_k$ such that $P_{n_0}(\bigcup_{i=1}^k B_i) > 1 - \eta$. Let $A_1, \ldots, A_k$ be the 2η-spheres with the same centers. Since $(\bigcup_{i=1}^k B_i)^\eta \subset \bigcup_{i=1}^k A_i$ and $p(P_n, P_{n_0}) < \eta$ for $n \geq n_0$, it follows that (14) holds for $n \geq n_0$. Since each P_n is separable, we can ensure that (14) holds for the finitely many n preceding n_0 by enlarging the system $\{A_i\}$.

Thus Z is p-complete if S is complete and each probability measure on it is separable. Suppose, on the other hand, that Z is p-complete. It is not difficult to show that the set Z_0 of unit masses is p-closed and hence also p-complete and then to show that S is complete.

We turn now to the converse half of Prohorov's theorem.

THEOREM 7 *If Π is $\mathscr{W}$-compact and each measure in it is separable, and if S is topologically complete, then Π is tight.*

Proof. By Theorem 1, the elements of Π are individually tight. To prove Π itself tight, it suffices as usual to produce, given ε and δ, finitely many δ-spheres $A_1, \ldots, A_k$ such that $P(A_1 \cup \cdots \cup A_k) > 1 - \varepsilon$ for all P in Π.

If P lies in Π, then, since P is tight, there is a compact set K such that $P(K) > 1 - \frac{1}{2}\varepsilon$, and K can be covered by finitely many open δ-spheres $B_{P,i}$, $i = 1, \ldots, k_P$. By Theorem 3, there is a $\mathscr{W}$-neighborhood $N(P)$ of P such that $Q \in N(P)$ implies that the union G_P of the $B_{P,i}$ satisfies $Q(G_P) > P(G_P) - \frac{1}{2}\varepsilon$ and hence $Q(G_P) > 1 - \varepsilon$. Since Π is $\mathscr{W}$-compact, it can be covered by a finite selection of these $N(P)$. Take $A_1, \ldots, A_k$ to consist of the $B_{P,i}$ corresponding to the neighborhoods $N(P)$ selected.

In view of Theorem 1, Theorem 7 is the same as the assertion that, if Π is $\mathscr{W}$-compact and each of its elements is tight, and if S is topologically complete, then Π is tight. The hypothesis that each P in Π is tight cannot be suppressed, because a single nontight P forms a compact set. On the other hand, it seems reasonable to conjecture that the completeness hypothesis can be suppressed—to conjecture that Π is tight if it is $\mathscr{W}$-compact and if each of its elements is tight.† Although the problem is open, there is an interesting result in this direction.

THEOREM 8 *Suppose that $P_n \Rightarrow P$, where P and each of the P_n are tight. Then $\{P, P_1, P_2, \ldots\}$ is tight.*

Proof. Since the measures involved are individually tight, it is enough to show that for each ε there is a compact K such that

$$P_n(K) > 1 - \varepsilon \tag{15}$$

for all sufficiently large n.

Given ε, use the tightness of P to choose a compact K' with $P(K') > 1 - \frac{1}{3}\varepsilon$, and put $G_k = \{x: \rho(x, K') < 1/k\}$. Since $P_n \Rightarrow P$, there is an increasing sequence $\{n_k\}$ such that $n \geq n_k$ implies

$$P_n(G_k) > P(G_k) - \tfrac{1}{3}\varepsilon > 1 - \tfrac{2}{3}\varepsilon. \tag{16}$$

For $n_k \leq n \leq n_{k+1}$, use the tightness of P_n to choose a compact K_n' satisfying $K_n' \subset G_k$ and $P_n(G_k - K_n') < \frac{1}{3}\varepsilon$. If $K_k = K' \cup \bigcup_{n=n_k}^{n_{k+1}} K_n'$, then K_k is compact, $K' \subset K_k \subset G_k$, and $n_k \leq n \leq n_{k+1}$ implies $P_n(G_k - K_k) < \frac{1}{3}\varepsilon$. From (16) it now follows that $n_k \leq n \leq n_{k+1}$ implies

$$P_n(K_k) > 1 - \varepsilon. \tag{17}$$

Put $K = \bigcup_k K_k$. Suppose $\{x_u\}$ is a sequence in K; if there is some one K_k containing all the x_u, then $\{x_u\}$ contains a convergent subsequence; otherwise

† Varadarajan (1958a and 1961a) has the result in this form, but his proof contains a lacuna.

there is a subsequence $\{x_{u_m}\}$ and a sequence $\{v_m\}$ of integers such that $x_{u_m} \in K_{v_m} \subset G_{v_m}$ and $v_m \geq m$, so that $\rho(x_{u_m}, K') < 1/m$; $\{x_{u_m}\}$ must contain a further subsequence that converges. Thus K is compact. If $n \geq n_1$, then $n_k \leq n \leq n_{k+1}$ for some k, and (15) follows from (17) and the relation $K \supset K_k$.

It is essential in this theorem to assume that P is tight: Take a nontight P on a separable S, as in the second remark following Theorem 1, and apply Theorem 4.

Remarks. Theorem 8 and the example in Remark 2 following Theorem 1 are due to LeCam (1957); the remaining results are for the most part due to Varadarajan (1958a and 1961a).

Bibliography

Articles in the Soviet journal *Teoriya Veroyatnosti i ee Primeneniya* are listed as they appear in the English translation *Theory of Probability and its Applications*, published by the Society for Industrial and Applied Mathematics.

A. D. Alexandroff (1940–43), "Additive set functions in abstract spaces," *Mat. Sb.* **8,** 307–348; **9,** 563–628; **13,** 169–238.

T. W. Anderson (1963), "Asymptotic theory for principal component analysis," *Ann. Math. Statist.* **34,** 122–148.

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

Stefan Banach (1932), *Théorie des Opérations Linéaires.* Warsaw: Monografje Matematyczne.

Robert Bartoszyński (1961), "A characterization of the weak convergence of measures," *Ann. Math. Statist.* **32,** 561–576.

P. J. Bickel (1968a), "Some contributions to the theory of order statistics," to appear.

(1968b), "A distribution free version of the Smirnov two sample test in the p-variate case," to appear.

Patrick Billingsley (1956), "The invariance principle for dependent random variables," *Trans. Amer. Math. Soc.* **83,** 250–268.

(1961), "The Lindeberg-Lévy theorem for martingales," *Proc. Amer. Math. Soc.* **12,** 788–792.

(1962), "Limit theorems for randomly selected partial sums," *Ann. Math. Statist.* **33,** 85–92.

(1964), "An application of Prohorov's theorem to probabilistic number theory," *Ill. J. Math.* **8,** 697–704.

(1965), *Ergodic Theory and Information.* New York: John Wiley and Sons.

Patrick Billingsley and Flemming Topsøe (1967), "Uniformity in weak convergence," *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **7,** 1–16.

Garret Birkhoff (1961), *Lattice Theory.* Providence: The American Mathematical Society.

Z. W. Birnbaum and Ronald Pyke (1958), "On some distributions related to the statistic D_n^+," *Ann. Math. Statist.* **29**, 179–187.

J. R. Blum, D. L. Hanson, and J. I. Rosenblatt (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.

H. Bohr (1932), *Fastperiodische Funktionen*, Erg. Math. **1**, no. 5. Berlin-Göttingen-Heidelberg: Springer-Verlag. (English translation 1947. New York: Chelsea.)

N. N. Chentsov (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.

H. Chernoff (1956), "Large sample theory: parametric case," *Ann. Math. Statist.* **27**, 1–22.

H. Chernoff and H. Teicher (1958), "A central limit theorem for sequences of exchangeable random variables," *Ann. Math. Statist.* **29**, 118–130.

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

Z. Ciesielski and H. Kesten (1962), "A limit theorem for the fractional parts of the sequence $2^k t$," *Proc. Amer. Math. Soc.* **13**, 596–600.

Harald Cramér (1946), *Mathematical Methods of Statistics*. Princeton: Princeton University Press.

H. Cramér and H. Wold (1936), "Some theorems on distribution functions," *J. London Math. Soc.* **11**, 290–295.

D. A. Darling (1957), "The Kolmogorov-Smirnov, Cramér-von Mises tests," *Ann. Math. Statist.* **28**, 823–838.

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

Jean Dieudonné (1960), *Foundations of Modern Analysis*. New York: Academic Press.

W. Doeblin (1940), "Remarques sur la théorie metrique des fractions continues," *Compositio Math.* **7**, 353–371.

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

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

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

(1953), *Stochastic Processes*. New York: John Wiley and Sons.

R. M. Dudley (1966), "Weak convergence of probabilities on nonseparable metric spaces and empirical measures on Euclidean spaces," *Ill. J. Math.* **10**, 109–126.

(1967), "Measures on non-separable metric spaces," *Ill. J. Math.* **11**, 449–453.

Nelson Dunford and Jacob T. Schwartz (1958), *Linear Operators. Part I: General Theory*. New York: Interscience Publishers.

Meyer Dwass and Samuel Karlin (1963), 'Conditional limit theorems," *Ann. Math. Statist.* **34**, 1147–1167.

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

(1947), "On the number of positive sums of independent random variables," *Bull. Amer. Math. Soc.* **53**, 1011–1020.

William Feller (1957), *An Introduction to Probability Theory and Its Applications*, Volume I, second edition. New York: John Wiley and Sons.

(1966), *An Introduction to Probability Theory and Its Applications*, Volume II. New York: John Wiley and Sons.

R. Fortet (1940), "Sur une suite également répartie," *Studia Math.* **9,** 54–69.

I. I. Gikhman and A. V. Skorohod (1965), *Introduction to the Theory of Random Processes.* Moscow.

Leonard Gross (1963), "Harmonic analysis on Hilbert space," *Mem. Amer. Math. Soc.* **46.**

Paul R. Halmos (1950), *Measure Theory.* New York: D. Van Nostrand.

G. H. Hardy and E. M. Wright (1960), *An Introduction to the Theory of Numbers*, fourth edition. Oxford: Clarendon Press.

P. L. Hennequin and A. Tortrat (1965), *Théorie des Probabilitiés et Quelques Applications.* Paris: Masson et Cie.

I. A. Ibragimov (1962), "Some limit theorems for stationary processes, " *Theor. Probability Appl.* **7,** 349–382.

(1963), "A central limit theorem for a class of dependent random variables," *Theor. Probability Appl.* **8,** 83–89.

Kiyosi Itô and Henry P. McKean, Jr. (1965), *Diffusion Processes and Their Sample Paths.* Berlin-Heidelberg-New York: Springer-Verlag.

M. Kac (1946), "On the distributions of values of sums of type $\Sigma f(2^k t)$," *Ann. of Math.* **47,** 33–49.

(1949), "On deviations between theoretical and empirical distributions," *Proc. Nat. Acad. Sci. U.S.A.* **35,** 252–257.

Gopinath Kallianpur (1961), "The topology of weak convergence of probability measures," *J. Math. Mech.* **10,** 947–969.

Samuel Karlin (1966), *A First Course in Stochastic Processes.* New York: Academic Press.

John L. Kelley (1955), *General Topology.* New York: D. Van Nostrand.

A. Ya. Khinchine (1961), *Continued Fractions*, third edition. Moscow: State Publishing House of Physical-Mathematical Literature. (English translation 1964. Chicago: University of Chicago Press)

(1933), *Asymptotische Gesetze der Wahrscheinlichkeitsrechnung*. Erg. Math. **2,** no. 4. Berlin-Göttingen-Heidelberg: Springer-Verlag.

H. J. Keisler and A. Tarski (1964), "From accessible to inaccessible cardinals," *Fund. Math.* **53,** 225–308.

Ernest G. Kimme (1957), "On the convergence of sequences of stochastic processes," *Trans. Amer. Math. Soc.* **84,** 208–229.

(1960), "Some equivalence conditions for the convergence in distribution of sequences of stochastic processes," *Trans. Amer. Math. Soc.* **95,** 495–515.

Frank B. Knight (1962), "On the random walk and Brownian motion," *Trans. Amer. Math. Soc.* **103,** 218–228.

A. N. Kolmogorov (1931), "Eine Verallgemeinierung des Laplace-Liapounoffschen Satzes," *Izv. Akad. Nauk SSSR Ser. Fiz-Mat.*, 959–962.

(1933), *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Erg. Math. **2,** no. 3. Berlin-Göttingen-Heidelberg: Springer-Verlag.

(1956), "On Skorohod convergence," *Theor. Probability Appl.* **1,** 213–222.

A. N. Kolmogorov and Yu. V. Prohorov (1954), "Zufällige Funktionen und Grenzverteilungssätze," *Berich über die Tagung Wahrscheinlichkeitsrechnung und Mathematische Statistik*, 113–126. Berlin: Deutscher Verlag der Wissenschaften.

John Lamperti (1962), "An invariance principle in renewal theory," *Ann. Math. Statist.* **33,** 685–696.

(1962a), "A new class of probability limit theorems," *J. Math. Mech.* **11,** 749—772.

(1962b), "On convergence of stochastic processes," *Trans. Amer. Math. Soc.* **104,** 430–435.

L. LeCam (1957), "Convergence in distribution of stochastic processes," *Univ. California Publ. Statist.* **2,** no. 11, 207–236.

Paul Lévy (1925), *Calcul des Probabilités*. Paris: Gauthier-Villars.

(1937), *Théorie de l'Addition des Variables Aléatoires*. Paris: Gauthier-Villars.

(1948), *Processus Stochastiques et Mouvement Brownien* Paris: Gauthier-Villars.

Michel Loève (1960), *Probability Theory*, second edition. Princeton: D. Van Nostrand.

H. B. Mann and A. Wald (1943), "On stochastic limit and order relations," *Ann. Math. Statist.* **14,** 217–226.

A. M. Mark (1949), "Some probability theorems," *Bull. Amer. Math. Soc.* **55,** 885–900.

J. C. Oxtoby and S. Ulam (1939), "On the existence of a measure invariant under a transformation," *Ann. of Math.* (2) **40,** 560–566.

K. R. Parthasarathy (1967), *Probability Measures on Metric Spaces*. New York: Academic Press.

Yu. V. Prohorov (1953), "Probability distributions in functional spaces, " *Uspehi Matem. Nauk* (N.S.) **55,** 167.

(1956), "Convergence of random processes and limit theorems in probability theory," *Theor. Probability Appl.* **1,** 157–214.

(1960), "The method of characteristic functionals," *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Volume 2, 403–419.

Ronald Pyke (1965), "Spacings," *J. Roy. Stat. Soc. Ser. B* **27,** 395–449.

(1968), "The weak convergence of the empirical process of random sample size," to appear.

Ronald Pyke and Galen Shorack (1968), "Weak convergence of the two-sample empirical process and the Chernoff-Savage theorem," to appear.

R. Ranga Rao (1962), "Relations between weak and uniform convergence of measures with applications," *Ann. Math. Statist.* **33,** 659–680.

R. Ranga Rao and V. S. Varadarajan (1958), "On a theorem in metric spaces," *Ann. Math. Statist.* **29,** 612–613.

A. Rényi (1958), "On mixing sequences of sets," *Acta Math. Acad. Sci. Hung.* **9,** 215–228.

(1960), "On the central limit theorem for the sum of a random number of independent random variables," *Acta. Math. Acad. Sci. Hung.* **11,** 97–102.

Bengt Rosén (1964), "Limit theorems for sampling from a finite population," *Arkiv för Matematik*, **5,** 383–424.

(1967a), "On the central limit theorem for sums of dependent random variables," *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **7,** 48–82.

(1967b), "On asymptotic normality of sums of dependent random variables," *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **7,** 95–102.

(1967c), "On the central limit theorem for a class of sampling procedures," *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **7,** 103–115.

H. Royden (1963), *Real Analysis*. New York: Macmillan.

H. Scheffé (1947), "A useful convergence theorem for probability distributions," *Ann. Math. Statist.* **18,** 434–438.

Paul Schmid (1958), "On the Kolmogorov and Smirnov limit theorems for discontinuous distribution functions," *Ann. Math. Statist.* **29,** 1011–1027.

J. Sethuraman (1964), "On the probability of large deviations of families of sample means," *Ann. Math. Statist.* **35,** 1304–1316.

(1965), "On the probability of large deviations of the mean for random variables in $D[0, 1]$," *Ann. Math. Statist.* **36,** 280–285.

George F. Simmons (1963), *Introduction to Topology and Modern Analysis*. New York: McGraw-Hill.

A. V. Skorohod (1956), "Limit theorems for stochastic processes," *Theor. Probability Appl.* **1**, 261–290.
(1957), "Limit theorems for stochastic processes with independent increments," *Theor. Probability Appl.* **2**, 138–171.
(1961), *Studies in the Theory of Random Processes.* Kiev: Kiev University Press. (English translation 1965. Reading: Addison-Wesley.)

E. E. Slutsky (1925), "Über Stochastische Asymptoten und Grenzwerte," *Metron* **5**, 1–90.
(1937), "Qualche proposizione relativa alla teoria delle funzioni aleatorie," *Giorn. Ist. Ital. Attuari* **8**, 183–199.

Charles Stone (1963), "Weak convergence of stochastic processes defined on semi-infinite time intervals," *Proc. Amer. Math. Soc.* **14**, 694–696.
(1963), "Limit theorems for random walks, birth and death processes, and diffusion processes," *Ill. J. Math.* **7**, 638–660.

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

E. C. Titchmarsh (1939), *The Theory of Functions*, second edition. Oxford: Oxford University Press.

Flemming Topsøe (1967a), "On the connection between *P*-continuity and *P*-uniformity in weak convergence," *Theor. Probability Appl.* **12**, 281–290.
(1967b), "Preservation of weak convergence under mappings," *Ann. Math. Statist.* **38**, 1661–1665.

Bruce E. Trumbo (1965), "Sufficient conditions for the weak convergence of conditional probability distributions in a metric space." Thesis, University of Chicago.

V. S. Varadarajan (1958a), "Convergence in distribution of stochastic processes." Thesis, Calcutta University.
(1958b), "Weak convergence of measures on separable metric spaces," *Sankyā* **19**, 15–22.
(1958c), "On the convergence of probability distributions,"*Sankyā*, **19**, 23–26.
(1961a), "Measures on topological spaces," *Mat. Sb.* **55** (97), 35–100. (English translation 1965. Providence: American Mathematical Society Translations, series 2, **48**, 161–228.)
(1961b), "Convergence of stochastic processes," *Bull. Amer. Math. Soc.* **67**, 276–280.

Summary of Notation

This list includes only symbols used systematically throughout the book in some special way. Although the generic element of S is asserted to be x, it can actually be any letter in that general region of the lower case roman alphabet, possibly with subscripts and so on; similar remarks apply to other generic elements.

Set Theory

E^c is the complement of E; $E_1 - E_2 = E_1 \cap E_2{}^c$ is the difference of E_1 and E_2; $E_1 + E_2 = (E_1 - E_2) \cup (E_2 - E_1)$ is their symmetric difference; I_E is the indicator or characteristic function of E.

Probability Spaces

$(\Omega, \mathscr{B})$ is a measurable space; ω and E are the generic elements of Ω and $\mathscr{B}$; P is a probability measure on $(\Omega, \mathscr{B})$; E denotes expected value; ξ is a random variable, and $\{\xi_n\}$ and $\{\xi_t\}$ are stochastic processes.

If h maps $(\Omega, \mathscr{B})$ into $(\Omega', \mathscr{B}')$,

$$h^{-1}\mathscr{B}' \subset \mathscr{B}$$

means that $h^{-1}E' \in \mathscr{B}$ for each $E' \in \mathscr{B}'$ — that is, that h is measurable — and $\mathsf{P}h^{-1}$ has value $\mathsf{P}(h^{-1}E')$ at E' (p. 222).

The General Metric Space

S is a metric space with generic point x, metric $\rho(x, y)$ (and $\rho(x, A)$ — see p. 215), and σ-field $\mathscr{S}$ of Borel sets; A°, A^-, and ∂A are the interior, closure, and boundary of the generic subset A (which lies in $\mathscr{S}$ unless the contrary is stated); $S(x, \varepsilon)$ is the open ε-sphere (ball) about x; the general closed set is F (but this can also be a distribution function), the general open set is G, and the general compact set is K (but this can also be a constant or bound).

$C(S)$, with generic element f, is the class of bounded, continuous functions on C; if h maps S into another metric space, D_h is the set of its discontinuities.

P and Q are probability measures on $(S, \mathscr{S})$; $P_n \Rightarrow P$ denotes weak convergence; Π is a family of P's.

Special Metric Spaces

The σ-field of Borel sets in a space is denoted by the script version (with any attendant superscripts) of the upper case roman letter denoting the space itself.

R^k is Euclidean k-space; H is the generic element of $\mathscr{R}^k$; F is a distribution function; $F_n \Rightarrow F$ denotes weak convergence (p. 18); for the notations $x \leq y$, $x < y$, $|x - y|$, and $(a, b]$, see p. 17.

R^∞ is the topological product of a countable sequence of copies of the real line (pp. 19 and 218), with generic point $x = (x_1, x_2, \ldots)$; $\pi_k(x) = (x_1, \ldots, x_k)$ is the natural projection into R^k.

C is $C[0, 1]$ (pp. 19, 54, and 220), with metric ρ and generic point $x = x(t)$; $\pi_{t_1 \ldots t_k}(x) = (x(t_1), \ldots, x(t_k))$ is the natural projection into R^k; for x_t as a coordinate variable and for the phrase "the distribution of x_t under P", see p. 60.

D is $D[0, 1]$ (p. 109), with two equivalent metrics d (p. 111) and d_0 (p. 113), and generic point $x = x(t)$; projections $\pi_{t_1 \ldots t_k}$ and coordinate variables x_t are as for C; special symbols: λ and Λ (p. 111), $\|\lambda\|$ (p. 112), T_P (p. 123), J_t (p. 124), and T_X (p. 128).

Random Elements

X is a random element (p. 22) with value $X(\omega)$ at ω; if its range is a function space, $X(t) = \pi_t X$ and $X(t, \omega) = \pi_t(X(\omega))$ (p. 57 for C; p. 128 for D); $X_n \xrightarrow{\mathscr{D}} X$ and $X_n \xrightarrow{\mathscr{D}} P$ denote convergence in distribution (pp. 23 and 24); $X_n \xrightarrow{P} a$ and $X_n \xrightarrow{P} X$ denote convergence in probability (pp. 24 and 26); for (X, Y) and $\rho(X, Y)$, see pp. 25 and 225.

N and $N(\mu, \sigma^2)$ are normal distributions or variables (p. 24); W is Wiener measure on C (p. 61), a random function in C with that distribution (p. 64), or the corresponding measure or random function in D (p. 137); W° is the Brownian bridge, as a measure or a random function in C or in D (pp. 64, 65, and 141).

Moduli

The modulus of continuity for C and related moduli for D:

$w_x(\delta) = w(x, \delta)$,	pp. 54 and 220,
$w_X(\delta) = w(X, \delta)$,	p. 58,
$w_x(T_0)$,	p. 109,
$w'_x(\delta)$,	p. 110,
$w''_x(\delta)$,	p. 118,
$w''(X, \delta)$,	p. 128.

Index