WEAK CONVERGENCE OF STOCHASTIC PROCESSES WITH SEVERAL PARAMETERS

MIRON L. STRAF
UNIVERSITY OF CALIFORNIA, BERKELEY

1. Stochastic processes, random functions, and weak convergence

Looking at a stochastic process as a function plucked at random may be interesting in its own right, but there is a statistical reason for doing so: a large class of limit theorems for functions of partial sums of a sequence of random variables or for statistics which are functions of empirical processes may be derived.

To achieve such practicality, we are led to ask a number of requirements of a theory of probability measures on a function space. We enumerate a few.

(i) To begin with, our function space S must contain the realizations of processes of interest to us and should contain those which are defined in their simplest forms with empirical observations.

(ii) The σ-field $\mathscr{B}$ of subsets of S over which our probability measures are defined must contain events whose probability we wish to know. Some of these events may be defined by topological or analytical qualifications; for example, the classes of continuous functions in S or integer valued functions in S.

(iii) Processes under investigation must be measurable functions from the underlying probability space over which they are defined; that is, they must induce probability measures on $\mathscr{B}$. Moreover, we should have convenient rules in order to determine whether a process is measurable.

(iv) Each real valued and measurable function h on S transforms a stochastic process or random function X into a statistic or random variable $h(X)$. Statistics of interest to us should be so induced by measurable functions of stochastic processes. In addition, we must have criteria, preferably easy to apply, to verify the measurability of real functions h on S and to determine if $h(X)$ is a random variable.

(v) For a sequence of stochastic processes $\{X_n\}$ a mode of convergence should be defined so that we may infer, for a large class of functions $\{h\}$, that the statistic $h(X_n)$ converges weakly to $h(X_0)$; that is, the distribution function of $h(X_n)$ converges to that of $h(X_0)$ at each continuity point of the latter. In addition, we need facile standards to determine for a function h when such convergence holds.

This work was prepared with the partial support of National Science Foundation Grant GP-15283.

(vi) Finally, and most important, the analytical nature of our function space should lead to adaptable criteria which specify whether or not a sequence of stochastic processes has a limit with respect to our mode of convergence.

Having investigated many formulations of convergence of probability measures on a function space, Prohorov concluded that in the more interesting cases the function spaces were separable and complete metric spaces, each provided with the Borel σ-field generated by the topology of their metric. The treatment of stochastic processes as random elements of such spaces satisfies, by and large, our six requirements. Prohorov [22] formulated and further developed this theory in its now classical setting. It is called weak convergence of probability measures on a separable and complete metric space. A consummate exposition of this theory is presented in the treatise of Billingsley [3]. Briefly, we describe what weak convergence means, and how the theory may be used.

Let S be a metric space, and $\mathscr{B}$ be the collection of its Borel sets (the σ-field generated by the open sets). Let $C(S)$ be the class of those real valued, continuous, and bounded functions on S. For probability measures P_n, $n = 0, 1, \cdots$, on $\mathscr{B}$, the *weak convergence* of P_n to P_0, written $P_n \Rightarrow P_0$, is defined by the requirement that $\lim_n \int_S f\, dP_n = \int_S f\, dP_0$ hold for each f in $C(S)$.

From this definition, it is clear that if h is a real and continuous function on S, then the induced measures $P_n h^{-1}$ on the real line converge weakly to the induced limit measure $P_0 h^{-1}$. What is more general: we may allow h to have discontinuities as long as it is measurable and its set of discontinuities has probability zero under P_0. In practice, we establish weak convergence in order to prove weak convergence of random variables induced by various real functions h.

When does a sequence of probability measures on $(S, \mathscr{B})$ converge weakly to a limit? Concepts useful for the answer to this question are relative compactness and tightness. A family Π of probability measures on $(S, \mathscr{B})$ is *relatively compact* (relatively, sequentially, weakly compact would be precise but verbose) if every sequence of probability measures in Π contains a weakly convergent subsequence; that is, a subsequence which converges weakly to some probability measure not necessarily in Π. The family Π is *tight* (the term is due to Le Cam [16]) if, for each $\varepsilon > 0$, there is a compact set K for which the relation $P(K) > 1 - \varepsilon$ is satisfied at once for all probability measures P in Π.

The Prohorov theorem states that for a metric space S, if Π is tight, then it is also relatively compact. In separable and complete metric spaces (for which Prohorov [22] proved this theorem), tightness is necessary and sufficient for relative weak compactness.

Thus, if our function space is a separable and complete metric space, the problem of characterizing relative compactness of a family of probability measures on the space becomes one of characterizing tightness of the family, which, in turn, reduces to the problem of characterizing compactness in our function space.

Stochastic processes with jump discontinuities—empirical distribution functions, for example—may be handled by treating them as random elements of the space $D[0, 1]$ of all functions on the unit interval with discontinuities of at most the first kind. (For the convenience of discriminating between functions, we require that they be right continuous.) The required measurability of important empirical processes with respect to the Borel sets of $D[0, 1]$, however, precludes the use of uniform distance as a metric on this space. For example, the elementary transformation which assigns to each point p on the unit interval the indicator function of $[p, 1]$ is not a measurable transformation from the Borel sets of the unit interval to the Borel sets of $D[0, 1]$ with the uniform norm. Thus, empirical distribution functions may not be random elements of this space.

An ingenious topology invented by Skorohod [27] satisfies our measurability requirements and makes $D[0, 1]$ a separable and topologically complete metric space with the result that Prohorov's theorem may be applied.

Ironically, it was not Skorohod's intention to apply the Prohorov theory to $D[0, 1]$; his strategy was to establish criteria by which a sequence of stochastic processes $\{X_n\}$ might be replaced by an almost surely convergent sequence $\{\tilde{X}_n\}$ each of whose elements has the same finite dimensional distributions as the corresponding element of $\{X_n\}$.

It was Kolmogorov [14] who exhibited a metric for Skorohod's topology and proved that the metric space was homeomorphic to some complete one. The question which he raised, whether one might exhibit a complete metric, was answered by Prohorov [22]. To the Hausdorff metric between the closure of the graphs of functions in $D[0, 1]$, Prohorov appended the Lévy distance for monotone functions on the real line between those obtained when each function's modulus, which serves to characterize elements of $D[0, 1]$, is considered as a monotone function of its real argument exploded by an exponential transformation. This eclectic metric is equivalent to Skorohod's topology and makes $D[0, 1]$ a separable and complete metric space.

A relatively simple formulation of an equivalent metric which preserves the intuitiveness of Skorohod's convergence is presented by Billingsley [3]. Let Λ be the class of all strictly increasing and continuous maps of the closed unit interval onto itself. Billingsley's metric $d(\cdot, \cdot)$ is defined, for x and y in $D[0, 1]$, to be the infimum of those positive ε such that there is a λ in Λ for which $\sup_t|\lambda(t) - t| \leqq \varepsilon$ and $\sup_t|x(t) - y(\lambda t)| \leqq \varepsilon$. Although $d(\cdot, \cdot)$ is not a complete metric, it is topologically complete. By imposing a constraint on λ stronger than $\sup_t|\lambda(t) - t| \leqq \varepsilon$, we may construct a complete metric which is equivalent to $d(\cdot, \cdot)$ on $D[0, 1]$. (We shall investigate this question in general.)

The Skorohod topology is strong enough so that convergence to a continuous function agrees with uniform convergence to that function. Thus, if we have a sequence of probability measures P_n converging weakly to a probability measure P concentrated on the subspace $C[0, 1]$ of continuous functions, as is the case for the probability measure of Brownian motion, then each measurable function h continuous with respect to the supremum metric is continuous almost everywhere P on $D[0, 1]$ with the result that the random variables induced by h are weakly convergent.

2. Stochastic processes with several parameters

Prior to the work of Lévy, the study of random functions of several variables had been undertaken by divers scholars with an empirical point of view. Lévy ([18] and [19]) filled the lacuna between theory and practice with his study of Brownian motion with several parameters.

The weak convergence, however, of stochastic processes with several parameters treated as random elements of spaces that are concrete generalizations of $C[0, 1]$ and $D[0, 1]$ has only recently been studied. Dudley [7] introduced a space of possibly discontinuous functions of several variables provided with a σ-field generated by the spheres of the uniform metric and developed a separate, but parallel, theory of weak convergence for probability measures on nonseparable metric spaces in the application to random functions of his space. His work has been complemented by that of Wichura [33]. Their conditions for weak convergence, however, apply only to processes converging to an essentially continuous limit. Such problems as those which they considered have also been explored by Le Cam [17], who investigated an omnibus central limit theorem for random elements of nonseparable linear spaces.

It is our purpose to place the analysis of convergence of stochastic processes with several parameters in our bailiwick of the classical theory of weak convergence of probability measures on a separable and complete metric space. Our method is to define a metric on Dudley's function space which generalizes Skorohod's for $D[0, 1]$.

To achieve this purpose, we study in general a space $D(T)$ of functions with possible jumps on an arbitrary space T, and, taking our cue from Skorohod, present an associated collection of metrics for $D(T)$. Exploring these metrics, we determine which ones make the function space separable and complete, and characterize the relative compactness of sequences of probability measures on the Borel sets of $D(T)$. When $T = [0, 1]$, our work reduces to results known for $D[0, 1]$; when $T \subset R^k$, this study provides our method for the analysis of stochastic processes with several parameters.

3. A general Skorohod space

We begin with an arbitrary space T over which our functions are defined and some group Λ of one to one functions from T onto itself. Generic elements of Λ are denoted by λ, μ, and γ; and the identity transformation, by e. On Λ we define a norm with respect to group composition, that is, some real valued function $\|\cdot\|$ on Λ for which (i) $\|\lambda\|$ is nonnegative and $\|\lambda\| = 0$ if and only if $\lambda = e$, (ii) $\|\lambda \circ \mu\| \leqq \|\lambda\| + \|\mu\|$, and (iii) $\|\lambda^{-1}\| = \|\lambda\|$.

Lemma 3.1. *The norm $\|\cdot\|$ induces a right invariant metric $d_\Lambda(\cdot, \cdot)$ on Λ defined by $d_\Lambda(\lambda, \mu) = \|\lambda \circ \mu^{-1}\|$.*

Proof. The norm assumes only nonnegative real values; thus, so must $d_\Lambda(\cdot, \cdot)$. If $d_\Lambda(\lambda, \mu)$ is zero, then $\lambda \circ \mu^{-1}$ is the identity, implying $\lambda = \mu$. Symmetry follows from $\|\lambda \circ \mu^{-1}\| = \|\mu \circ \lambda^{-1}\|$, while the triangle inequality follows from

$\|\lambda \circ \gamma^{-1}\| = \|\lambda \circ \mu^{-1} \circ \mu \circ \gamma^{-1}\| \leqq \|\lambda \circ \mu^{-1}\| + \|\mu \circ \gamma^{-1}\|$. Finally, the metric is right invariant since $d_\Lambda(\lambda \circ \gamma, \mu \circ \gamma) = \|\lambda \circ \gamma \circ \gamma^{-1} \circ \mu^{-1}\| = \|\lambda \circ \mu^{-1}\| = d_\Lambda(\lambda, \mu)$. *Q.E.D.*

It is not necessary for our analysis that Λ with the metric $d_\Lambda(\cdot, \cdot)$ be a topological group. But there is so much structure here that one can show that the metric space $\{\Lambda, d_\Lambda(\cdot, \cdot)\}$ is a topological group if and only if inversion is a continuous operation and if and only if composition on the left is continuous at the identity.

It is worthwhile to know that, whether or not $\{\Lambda, d_\Lambda(\cdot, \cdot)\}$ is a topological group, the norm is continuous: $\|\lambda\|$ is the distance from e to λ and the distance from a set in a metric space is a Lipschitz continuous function.

Let us look at some examples. Suppose that $\{T, \rho(\cdot, \cdot)\}$ is a compact metric space. Let Λ be some group of homeomorphisms from T onto itself. We define on Λ a *supremum norm* $\|\cdot\|_s$ by

$$\|\lambda\|_s = \sup_{p \in T} \rho(\lambda p, p). \tag{3.1}$$

It is easy to show that $\|\cdot\|_s$ is indeed a norm on Λ.

In this same context, we introduce a *slope norm* $\|\cdot\|_t$ on Λ defined by

$$\|\lambda\|_t = \sup_{\{p, q \in T:\, p \neq q\}} \left| \log \frac{\rho(\lambda p, \lambda q)}{\rho(p, q)} \right|. \tag{3.2}$$

Here, however, we restrict our attention to a subgroup Λ_t of Λ consisting of those homeomorphisms in Λ with finite slope norm:

$$\Lambda_t = \{\lambda \in \Lambda : \|\lambda\|_t < \infty\}. \tag{3.3}$$

At first glance, $\|\cdot\|_t$ appears to be a norm on Λ_t. Indeed, it assumes nonnegative real values and

$$\begin{aligned} \|\lambda^{-1}\|_t &= \sup_{p \neq q} \left| \log \frac{\rho(\lambda^{-1} p, \lambda^{-1} q)}{\rho(p, q)} \right| = \sup_{p \neq q} \left| \log \frac{\rho(p, q)}{\rho(\lambda p, \lambda q)} \right| \\ &= \sup_{p \neq q} \left| \log \frac{\rho(\lambda p, \lambda q)}{\rho(p, q)} \right| = \|\lambda\|_t. \end{aligned} \tag{3.4}$$

In addition, the triangle inequality holds:

$$\begin{aligned} \|\lambda \circ \mu\|_t &= \sup_{p \neq q} \left| \log \frac{\rho(\lambda \circ \mu p, \lambda \circ \mu q)}{\rho(\mu p, \mu q)} \cdot \frac{\rho(\mu p, \mu q)}{\rho(p, q)} \right| \\ &\leqq \|\lambda\|_t + \|\mu\|_t. \end{aligned} \tag{3.5}$$

What may prevent $\|\cdot\|_t$ from being a norm is that, from $\|\lambda\|_t = 0$, we may only conclude that λ is an isometry. If the group of homeomorphisms Λ were judiciously chosen to have no isometries other than the identity, then $\|\cdot\|_t$ would be a norm on Λ_t. Such a choice for Λ, other than the trivial one $\Lambda = \{e\}$, may easily be made when T is the unit interval with the usual metric: merely require that each homeomorphism have zero as a fixed point. In the general case, we may circumvent the problem of choosing such a Λ by introducing a *modified slope norm* $\|\cdot\|_m$ defined for each λ in Λ_t by

$$\|\lambda\|_m = \|\lambda\|_s + \|\lambda\|_t. \tag{3.6}$$

When T is the compact unit interval with the usual metric and Λ is some group of homeomorphisms of T onto itself which have zero as a fixed point, we define a *diffeomorphism norm* on Λ by

$$\|\lambda\|_d = \begin{cases} \sup_t |\log \lambda'(t)| & \text{if the derivative, } \lambda'(t), \text{ exists} \\ +\infty & \text{otherwise} \end{cases} \tag{3.7}$$

and a corresponding subgroup Λ_d by

$$\Lambda_d = \{\lambda \in \Lambda : \|\lambda\|_d < \infty\}. \tag{3.8}$$

LEMMA 3.2. *With Λ some group of homeomorphisms of the compact unit interval onto itself which have zero as a fixed point and with Λ_d defined by* (3.8), *the diffeomorphism norm is a norm on Λ_d.*

PROOF. For each λ in Λ_d, $\|\lambda\|_d$ is finite and nonnegative, and $\|\lambda\|_d = 0$ if and only if $\lambda'(t) \equiv 1$, which means that $\lambda = e$. It follows from the chain rule that

$$\begin{aligned} \|\lambda \circ \mu\|_d &= \sup_t |\log \lambda'(\mu t)\mu'(t)| \\ &\leqq \sup_t |\log \lambda'(\mu t)| + \sup_t |\log \mu'(t)| \\ &= \|\lambda\|_d + \|\mu\|_d; \end{aligned} \tag{3.9}$$

and from the inverse function theorem, that

$$\begin{aligned} \|\lambda^{-1}\|_d &= \sup_t |\log \lambda^{-1\prime}(t)| = \sup_t |\log \lambda^{-1\prime}(\lambda t)| \\ &= \sup_t |-\log \lambda'(t)| = \|\lambda\|_d. \end{aligned} \tag{3.10}$$

Q.E.D.

Each λ in Λ_d is a diffeomorphism, that is, a differentiable homeomorphism. To be precise, a diffeomorphism is required to have an inverse which is differentiable as well, but in our case this condition is guaranteed by the inverse function theorem.

An interesting property of the diffeomorphism norm is given by the following.

LEMMA 3.3. *For diffeomorphisms λ and μ of the compact unit interval,*

$$\|\lambda \circ \mu^{-1}\|_d = \sup_t |\log \lambda'(t) - \log \mu'(t)|. \tag{3.11}$$

In particular, convergence of the diffeomorphisms λ_n to λ with respect to the metric induced by the diffeomorphism norm is equivalent to the uniform convergence of their derivatives λ_n' to λ'.

PROOF. It follows from the chain rule and the inverse function theorem that

$$\|\lambda \circ \mu^{-1}\|_d = \sup_t |\log \lambda'(\mu^{-1}t)\mu^{-1\prime}(t)| \tag{3.12}$$

$$= \sup_t \left|\log \frac{\lambda'(\mu^{-1}t)}{\mu'(\mu^{-1}t)}\right| = \sup_u |\log \lambda'(u) - \log \mu'(u)|.$$

Suppose $\lim_n \|\lambda_n \circ \lambda^{-1}\|_d = 0$. Since $\sup_n \sup_t |\log \lambda_n'(t)| < \infty$ and the exponential function is uniformly continuous on bounded intervals, (3.11) implies that λ_n' converges uniformly to λ'. Conversely, suppose $\sup_t |\lambda_n'(t) - \lambda'(t)| = 0$. Since $\sup_t |\log \lambda'(t)| < \infty$, $\lambda'(t)$ is bounded away from zero and infinity; so then must be $\lambda_n'(t)$ for all sufficiently large n. Since the logarithm function is uniformly continuous on bounded intervals that are bounded away from zero,

$$\text{(3.13)} \qquad \lim_n \|\lambda_n \circ \lambda^{-1}\|_d = \lim_n \sup_t |\log \lambda_n'(t) - \log \lambda'(t)| = 0.$$

Q.E.D.

Returning to our general setting, let $\mathscr{D}$ be a collection of finite partitions of T, directed by refinement ($\Delta_1 \geqq \Delta_2$ means that Δ_1 is finer than Δ_2) and invariant under Λ; that is, $\lambda \in \Lambda$ and $\Delta = \{A_v\} \in \mathscr{D}$ imply $\lambda(\Delta) = \{\lambda(A_v)\} \in \mathscr{D}$. The class of all functions assuming a constant real value on each cell A_v of some partition $\Delta = \{A_v\}$ in $\mathscr{D}$ is denoted by $I_{\mathscr{D}}$ and called the class of *simple functions*. In addition, we denote by $B(T)$ the class of all real valued bounded functions on T.

With our general group Λ and its associated norm $\|\cdot\|$, we define a Skorohod distance $d(\cdot, \cdot)$ between two elements x and y of $B(T)$ by

$$\text{(3.14)} \qquad d(x, y) = \inf_\varepsilon \{\varepsilon > 0; \text{ there is a } \lambda \in \Lambda \text{ with } \|\lambda\| < \varepsilon \text{ for which } \sup_t |x(t) - y(\lambda t)| < \varepsilon\}.$$

Note that this distance does not exceed $\sup_t |x(t) - y(t)|$, the uniform distance between x and y and, when $\Lambda = \{e\}$, they are the same.

Finally, we define our function space $D(T)$ to be the class of those bounded functions on T which lie at distance zero from our class of simple functions $I_{\mathscr{D}}$ in terms of the Skorohod distance:

$$\text{(3.15)} \qquad D(T) = \{x \in B(T) : d(x, I_{\mathscr{D}}) = 0\}.$$

Even if we allowed $B(T)$ to contain unbounded functions or extended real valued functions, $D(T)$ would be unchanged: $d(x, I_{\mathscr{D}}) = 0$ implies that x is bounded.

LEMMA 3.4. *The space $D(T)$ with the Skorohod distance is a (pseudo) metric space.*

PROOF. Since $d(x, y) \leqq \sup_p |x(p)| + \sup_p |y(p)|$, $d(x, y)$ is finite; clearly, it is also nonnegative. Symmetry of $d(\cdot, \cdot)$ follows from

$$\text{(3.16)} \qquad \sup_p |x(p) - y(\lambda p)| = \sup_p |y(p) - x(\lambda^{-1} p)|, \qquad \|\lambda\| = \|\lambda^{-1}\|.$$

Finally, the triangle inequality follows from $\|\lambda \circ \mu\| \leqq \|\lambda\| + \|\mu\|$ and

$$\text{(3.17)} \qquad \sup_p |x(p) - z(\lambda \circ \mu p)| \leqq \sup_p |x(p) - y(\mu p)| + \sup_p |y(p) - z(\lambda p)|.$$

Q.E.D.

We adopt the convention of not distinguishing between two functions x and y for which $d(x, y) = 0$ and thus work with the resulting metric space of equivalence classes. In our applications of this theory, we shall have, in fact, $x = y$ whenever $d(x, y) = 0$.

The definition (3.15) of $D(T)$ shows us that $I_{\mathscr{D}}$ is dense in $D(T)$. What is more, since $\mathscr{D}$ is invariant under Λ, each x in $D(T)$ may be uniformly approximated by a sequence of simple functions in $I_{\mathscr{D}}$.

For each x in $B(T)$ and $\Delta = \{A_v\}$ in $\mathscr{D}$, let

$$\omega_x^{\cdot}(\Delta) = \max_v \sup \{|x(s) - x(t)| : s, t \in A_v\}. \tag{3.18}$$

With x in $B(T)$, we associate a modulus $f_x(\eta, \Delta)$ defined for each $\eta > 0$ and Δ in $\mathscr{D}$ by

$$f_x(\eta, \Delta) = \inf \{\omega_x^{\cdot}(\lambda\Delta) : \lambda \in \Lambda, \|\lambda\| < \eta\}. \tag{3.19}$$

It is not difficult to see that, for each x in $B(T)$, $\lim_\Delta f_x(\eta, \Delta) = 0$ for all $\eta > 0$ if and only if $\lim_\Delta \omega_x^{\cdot}(\Delta) = 0$, with limits taken along the direction of refinement. In particular, if $f_x(\eta, \Delta)$ converges to zero for some particular $\eta > 0$, then it does so for all $\eta > 0$.

Our modulus characterizes elements of $D(T)$ in the following sense.

THEOREM 3.5. *For each x in $B(T)$, $x \in D(T)$ if and only if* $\lim_\Delta f_x(\eta, \Delta) = 0$ *for all $\eta > 0$.*

PROOF. Let $\varepsilon > 0$ and $x \in D(T)$ be given. Choose from $I_{\mathscr{D}}$ a y such that $d(x, y) < \varepsilon/4$. Then there is a λ in Λ with $\|\lambda\| < \varepsilon/2$ for which

$$\sup_p |x(p) - y(\lambda p)| < \varepsilon/2. \tag{3.20}$$

Since y is constant on each cell A_v of some partition $\Delta = \{A_v\}$ in $\mathscr{D}$, $y \circ \lambda$ is constant on the cells $\lambda^{-1}(A_v)$. So for each v, the fluctuation of x over the cell $\lambda^{-1}(A_v)$ is less than ε, that is, $\omega_x^{\cdot}(\lambda^{-1}\Delta) < \varepsilon$. Therefore, $\Delta' \in \mathscr{D}$ and $\Delta' \geqq \lambda^{-1}\Delta$ imply $\omega_x^{\cdot}(\Delta') < \varepsilon$; so $\omega_x^{\cdot}(\Delta)$ converges to zero and thus $\lim_\Delta f_x(\eta, \Delta) = 0$ for all $\eta > 0$.

Conversely, if $\lim_\Delta f_x(\eta, \Delta) = 0$ for all $\eta > 0$, then by our previous remarks, $\lim_\Delta \omega_x^{\cdot}(\Delta) = 0$. So for each $\varepsilon > 0$, there is a partition $\Delta = \{A_v\} \in \mathscr{D}$ for which $\omega_x^{\cdot}(\Delta) < \varepsilon$. Choose a point p_v in each cell A_v and let y be the function in $I_{\mathscr{D}}$ which assumes the constant value $x(p_v)$ on the cell A_v. Then, $|x(p) - y(p)| < \varepsilon$ for all $p \in A_v$ and for all v, so $\sup_p |x(p) - y(p)| < \varepsilon$, implying $d(x, y) < \varepsilon$. Since $\varepsilon > 0$ was arbitrary, $d(x, I_{\mathscr{D}}) = 0$, which means that $x \in D(T)$. *Q.E.D.*

The proof of this theorem shows that a bounded function x belongs to $D(T)$ if and only if it is the uniform limit of a sequence of simple functions in $I_{\mathscr{D}}$. This fact also follows from our definition of $D(T)$ and from the fact that $x \in I_{\mathscr{D}}$ and $\lambda \in \Lambda$ imply $x \circ \lambda \in I_{\mathscr{D}}$.

We now prove a lemma describing a relation between our modulus and the Skorohod metric which will have an application later in determining the measurability of sets defined in terms of our modulus and in proving the characterization of compactness for our function space.

LEMMA 3.6. *With respect to the Skorohod metric on $D(T)$, for a fixed $\eta > 0$ and $\Delta \in \mathscr{D}$, the modulus $f_x(\eta, \Delta)$ is an upper semicontinuous function of x.*

PROOF. Let $\varepsilon > 0$ be given. We will show there is a $\beta > 0$ such that $d(x, y) < \beta$ implies $f_y(\eta, \Delta) \leqq f_x(\eta, \Delta) + \varepsilon$. To this end, first choose a λ in Λ with $\|\lambda\| < \eta$ for which $\omega_x^{\bullet}(\lambda\Delta) < f_x(\eta, \Delta) + \varepsilon/2$. Next, choose $\beta > 0$ such that $\|\lambda\| + \beta < \eta$ and $\beta < \varepsilon/4$. Then $d(x, y) < \beta$ implies there is a μ in Λ with $\|\mu\| < \beta$ for which

$$\sup_p |x(p) - y \circ \mu(p)| < \frac{\varepsilon}{4}. \tag{3.21}$$

Therefore, $\sup_p |x \circ \lambda(p) - y \circ \mu \circ \lambda(p)| < \varepsilon/4$, and thus,

$$\omega_y^{\bullet}(\mu \circ \lambda\Delta) \leqq \omega_x^{\bullet}(\lambda\Delta) + \frac{\varepsilon}{2} \leqq f_x(\eta, \Delta) + \varepsilon. \tag{3.22}$$

Moreover, $\|\mu \circ \lambda\| \leqq \|\mu\| + \|\lambda\| \leqq \beta + \|\lambda\| < \eta$, so $f_y(\eta, \Delta) \leqq \omega_y^{\bullet}(\mu \circ \lambda\Delta) \leqq f_x(\eta, \Delta) + \varepsilon$. *Q.E.D.*

Completeness of our general Skorohod space is assured by the following condition.

THEOREM 3.7. *If $\{\Lambda, d_\Lambda(\cdot, \cdot)\}$ is complete, then so is $\{D(T), d(\cdot, \cdot)\}$.*

PROOF. We show each Cauchy sequence in $D(T)$ contains a convergent subsequence. Let $\{x_n\}$ be a Cauchy sequence in $\{D(T), d(\cdot, \cdot)\}$. Without loss of generality, we may assume that $\Sigma_n d(x_n, x_{n+1}) < \infty$. So for each n there is a μ_n in Λ with $\Sigma_n \|\mu_n\| < \infty$ and

$$\sum_n \sup_p |x_n(p) - x_{n+1}(\mu_n p)| < \infty. \tag{3.23}$$

Let $\gamma_{n,m} = \mu_{n+m} \circ \cdots \circ \mu_n$. Since $d_\Lambda(\gamma_{n,m+1}, \gamma_{n,m}) = \|\mu_{n+m+1}\|$, for each n the sequence $\{\gamma_{n,m}\}$ is Cauchy in $\{\Lambda, d_\Lambda(\cdot, \cdot)\}$ and so must converge to a λ_n in Λ. By continuity of the norm,

$$\|\lambda_n\| = \lim_m \|\gamma_{n,m}\| \leqq \sum_{m \geqq n} \|\mu_m\|; \tag{3.24}$$

thus, $\|\lambda_n\|$ converges to zero. Since $\lambda_n = \lambda_{n+1} \circ \mu_n$,

$$\begin{aligned} \sup_p |x_n \circ \lambda_n^{-1}(p) - x_{n+1} \circ \lambda_{n+1}^{-1}(p)| &= \sup_p |x_n \circ \mu_n^{-1}(p) - x_{n+1}(p)| \\ &= \sup_p |x_n(p) - x_{n+1} \circ \mu_n(p)|; \end{aligned} \tag{3.25}$$

therefore, $\{x_n \circ \lambda_n^{-1}\}$ is a Cauchy sequence with respect to the metric of uniform convergence and so must converge uniformly to a limit x. Consequently, $x \in D(T)$ and $d(x, x_n)$ converges to zero. *Q.E.D.*

The converse to this theorem does not hold, as a suitable example with $\mathscr{D}$ consisting of the space T alone illustrates. An example in which we may apply this theorem is the following.

THEOREM 3.8. *With respect to the modified slope norm, the induced metric space $\{\Lambda_t, d'_{\Lambda_t}(\cdot, \cdot)\}$ of those homeomorphisms λ of a compact metric space onto itself for which $\|\lambda\|_m < \infty$ is complete.*

PROOF. We show that every Cauchy sequence contains a convergent subsequence. We suppose then that $\{\lambda_n\}$ is a Cauchy sequence with respect to $\|\cdot\|_m$, and without loss of generality, we may presume that

$$M = \sum_n d'_{\Lambda_t}(\lambda_n, \lambda_{n+1}) = \sum_n \|\lambda_n \circ \lambda_{n+1}^{-1}\|_m < \infty. \tag{3.26}$$

Thus, $\{\lambda_n\}$ is Cauchy with respect to the uniform norm and, therefore, converges uniformly to a limit function λ. We shall show that λ is a homeomorphism and that $\lambda \in \Lambda_t$.

First, λ is continuous, since the λ_n are each continuous and converge uniformly to λ.

Next, we show that $\|\lambda\|_t$ is finite. Now,

$$\|\lambda_n\|_t = \|\lambda_n^{-1}\|_t \leqq \|\lambda_1\|_t + \sum_{k<n} \|\lambda_k \circ \lambda_{k+1}^{-1}\|_t \leqq \|\lambda_1\|_t + M. \tag{3.27}$$

So for all distinct p and q in T,

$$\left| \log \frac{\rho(\lambda_n p, \lambda_n q)}{\rho(p, q)} \right| \leqq \|\lambda_n\|_t \leqq M + \|\lambda_1\|_t. \tag{3.28}$$

Letting n approach infinity, we obtain

$$\left| \log \frac{\rho(\lambda p, \lambda q)}{\rho(p, q)} \right| \leqq M + \|\lambda_1\|_t \tag{3.29}$$

for all distinct p and q in T. Thus, $\|\lambda\|_t \leqq M + \|\lambda_1\|_t < \infty$.

Now we show that λ is one to one. Otherwise, there are points p and q in T with $\rho(p, q) > 0$ for which $\lambda p = \lambda q$. This, however, would imply that $\|\lambda\|_t = \infty$ which contradicts the conclusion of the previous paragraph.

Next, we show that λ is onto. Otherwise, there exists a point p in T which does not belong to $\lambda(T)$. Since λ is continuous, $\lambda(T)$ is compact; so p is neither in nor on the boundary of $\lambda(T)$. Thus, for some $\varepsilon > 0$, $\rho(p, \lambda(T)) \geqq \varepsilon$. Choosing an n such that $\|\lambda \circ \lambda_n^{-1}\|_s < \varepsilon/2$, we have $\rho(\lambda \circ \lambda_n^{-1} p, p) < \varepsilon/2$, but $\lambda(\lambda_n^{-1} p) \in \lambda(T)$ contradicts $\rho(p, \lambda(T)) \geqq \varepsilon$.

Finally, continuity of the inverse will necessarily follow, since every one to one continuous map of a compact space onto a Hausdorff space is a homeomorphism. That $\|\lambda\|_t$ is finite then implies that $\lambda \in \Lambda_t$.

It remains to show that $d'_{\Lambda_t}(\lambda_n, \lambda)$ converges to zero. We already know that $\lim_n \|\lambda_n \circ \lambda^{-1}\|_s = 0$. For $k > n$,

$$\|\lambda_k \circ \lambda_n^{-1}\|_t = \|\lambda_k \circ \lambda_{k-1}^{-1} \circ \lambda_{k-1} \circ \cdots \circ \lambda_{n+1} \circ \lambda_n\|_t \leqq \sum_{n \leqq j < k} \|\lambda_{j+1} \circ \lambda_j^{-1}\|_t; \tag{3.30}$$

so, for distinct p and q in T,

$$\left| \log \frac{\rho(\lambda_k \circ \lambda_n^{-1} p, \lambda_k \circ \lambda_n^{-1} q)}{\rho(p, q)} \right| \leqq \sum_{n \leqq j < k} \|\lambda_{j+1} \circ \lambda_j^{-1}\|_t. \tag{3.31}$$

Letting k approach infinity, taking the supremum over all distinct p and q in T, and finally letting n become large, we obtain

$$\lim_n \|\lambda_n \circ \lambda^{-1}\|_t = \lim_n \|\lambda \circ \lambda_n^{-1}\|_t = 0. \tag{3.32}$$

Therefore, $\lim_n \|\lambda_n \circ \lambda^{-1}\|_m = 0$, which implies that $\lim_n d'_{\Lambda_t}(\lambda_n, \lambda) = 0$. *Q.E.D.*

Thus, we see that by imposing convergence of the slope norm as well as that of the supremum norm, we assure that our limit function λ is a homeomorphism. An analogous condition is used by Paul [21] in application to problems of differential geometry.

We discuss now the diffeomorphism norm on Λ_d of Lemma 3.2, where $T = [0, 1]$. Recall that $\|\lambda\|_d = \sup_t |\log \lambda'(t)|$.

THEOREM 3.9. *With respect to the diffeomorphism norm, the space* $\{\Lambda_d, d_{\Lambda_d}(\cdot, \cdot)\}$ *of those diffeomorphisms* λ *of the unit interval onto itself for which* $\lambda(0) = 0$ *and* $\|\lambda\|_d < \infty$ *is complete.*

PROOF. Suppose $\{\lambda_n\}$ is a Cauchy sequence; that is, $\lim_{n,m} \|\lambda_n \circ \lambda_m^{-1}\|_d = 0$. By Lemma 3.3, $\{\log \lambda_n'(t)\}$ is a Cauchy sequence with respect to the uniform (supremum) norm and thus converges uniformly to a limit $g(t)$, say. Since

$$\sup_t |g(t)| \leqq \sup_n \sup_t |\log \lambda_n'(t)| < \infty \tag{3.33}$$

and since the exponential function is uniformly continuous on bounded intervals, we must have $\lambda_n'(t)$ converging uniformly to $\exp\{g(t)\}$ on $[0, 1]$. Thus, $\{\lambda_n\}$ is a sequence of differentiable functions on $[0, 1]$ for which $\lambda_n(0) = 0$ and λ_n' converges uniformly on $[0, 1]$. We may then conclude (see [24], p. 140) that λ_n converges uniformly to a differentiable function λ and that $\lambda'(t) = \lim_n \lambda_n'(t) = \exp\{g(t)\}$. Thus,

$$\|\lambda\|_d = \sup_t |\log \lambda'(t)| = \sup_t |g(t)| < \infty. \tag{3.34}$$

This implies that the derivative of λ is bounded away from zero, so λ must be strictly increasing on $[0, 1]$ and thus one to one. Moreover, $\lambda(1) = \lim_n \lambda_n(1) = 1$, so λ is onto. All this implies that λ is a differentiable homeomorphism of $[0, 1]$ onto itself with finite diffeomorphism norm and therefore belongs to Λ_d. As we have seen, $\log \lambda_n'(t)$ converges uniformly to $g(t) = \log \lambda'(t)$. By Lemma 3.3, this implies $\lim_n \|\lambda_n \circ \lambda^{-1}\|_d = 0$. *Q.E.D.*

Another example with $T = [0, 1]$ is provided by the slope norm $\|\cdot\|_t$ defined by (3.2) in which the metric is $\rho(s, t) = |s - t|$.

THEOREM 3.10. *When* Λ_t *is the group of those homeomorphisms* λ *of the compact unit interval onto itself for which* $\lambda(0) = 0$ *and* $\|\lambda\|_t < \infty$, *then the induced metric space* $\{\Lambda_t, d_{\Lambda_t}(\cdot, \cdot)\}$ *is complete.*

PROOF. For $t \in [0, 1]$, $\log(\lambda t/t) \leqq \|\lambda\|_t$ implies $\lambda t - t \leqq \exp\{\|\lambda\|_t\} - 1$. Since $\log(t/\lambda t) = -\log(\lambda t/t)$, we similarly conclude that $t - \lambda t \leqq \exp\{\|\lambda\|_t\} - 1$; therefore,

$$\|\lambda\|_s \leqq \exp\{\|\lambda\|_t\} - 1. \tag{3.35}$$

Thus, convergence with respect to the slope norm implies convergence with respect to the supremum norm. If $\{\lambda_n\}$ is Cauchy with respect to $\|\cdot\|_t$, then (3.35) implies that the sequence is Cauchy with respect to $\|\cdot\|_s$ and so it is Cauchy with respect to the modified slope norm $\|\cdot\|_m = \|\cdot\|_s + \|\cdot\|_t$. By Theorem 3.8, there is a $\lambda \in \Lambda_t$ for which

$$\lim_n \|\lambda_n \circ \lambda^{-1}\|_t \leqq \lim_n \|\lambda_n \circ \lambda^{-1}\|_m = 0. \tag{3.36}$$

Thus, $\{\Lambda_t, d_{\Lambda_t}(\cdot, \cdot)\}$ is complete. *Q.E.D.*

Once completeness of $D(T)$ is established, compactness may be characterized in terms of our modulus by a theorem that smacks of the Arzelà-Ascoli characterization in the space of continuous functions.

THEOREM 3.11. *When $\{D(T), d(\cdot, \cdot)\}$ is complete, a set $A \subset D(T)$ has compact closure if and only if*

$$\sup_{x \in A} \sup_t |x(t)| < \infty \tag{3.37}$$

and

$$\lim_\Delta \sup_{x \in A} f_x(\eta, \Delta) = 0 \quad \text{for all } \eta > 0. \tag{3.38}$$

PROOF. We first show the sufficiency of (3.37) and (3.38). Since $\{D(T), d(\cdot, \cdot)\}$ is complete, it suffices to show that A is totally bounded. Let $\varepsilon > 0$ be given. By (3.38) with $\eta = \varepsilon$, there is a $\Delta = \{A_v\}$ such that

$$\sup_{x \in A} \inf \{\omega_x^\bullet(\lambda\Delta) : \|\lambda\| < \varepsilon\} < \varepsilon. \tag{3.39}$$

Let $B = \sup_{x \in A} \sup_{p \in T} |x(p)|$ and let H be an ε net in $[-B, B]$. Let E consist of those functions x in $D(T)$ which assume on each cell of Δ a constant value from H. Clearly E is finite. We show that every element of A is within a distance 3ε, in terms of the Skorohod metric, from E.

If $x \in A$, then there is a $\lambda \in \Lambda$ with $\|\lambda\| < \varepsilon$ for which $\omega_x^\bullet(\lambda\Delta) < 2\varepsilon$. Now

$$\sup_{p, q \in A_v} |x \circ \lambda(p) - x \circ \lambda(q)| = \sup_{p, q \in \lambda(A_v)} |x(p) - x(q)| \tag{3.40}$$

implies $\omega_x^\bullet(\lambda\Delta) = \omega_{x \circ \lambda}^\bullet(\Delta)$, so that $\omega_{x \circ \lambda}^\bullet(\Delta) < 2\varepsilon$. Choosing y in E such that $|y(p) - x \circ \lambda(p)| \leqq 3\varepsilon$ for all $p \in A_v$ and for all v, we have

$$\sup_p |y(p) - x \circ \lambda(p)| < 3\varepsilon, \tag{3.41}$$

which, in addition to $\|\lambda\| < \varepsilon$, implies $d(y, x) < 3\varepsilon$. So E is a 3ε net for A, and thus A is totally bounded.

Now for the necessity of the conditions. If $\bar{A}$, the closure of A, is compact, then A is bounded, so

$$\sup_{x \in A} \sup_t |x(t)| = \sup_{x \in A} d(x, 0) < \infty, \tag{3.42}$$

where 0 is the function identically zero. Theorem 3.5 shows that, for each $x \in A$, $\lim_\Delta f_x(\eta, \Delta) = 0$ for all $\eta > 0$; moreover the net $f_x(\eta, \Delta)$ for each x in A and

$\eta > 0$ converges monotonely to zero. This is all we need, since the upper semicontinuity of our modulus (Lemma 3.6) implies, by Dini's theorem, that the convergence is uniform on compacta. *Q.E.D.*

Separability of our function space is equivalent to the existence of a countable collection of partitions linearly ordered under refinement, along which we may take the limit in determining all the elements of $D(T)$ with our modulus.

THEOREM 3.12. *The space $\{D(T), d(\cdot, \cdot)\}$ is separable if and only if there exists in $\mathscr{D}$ a countable subcollection $\{\Delta_k\}$ such that for all bounded functions x, $x \in D(T)$ if and only if $\lim_k f_x(\eta, \Delta_k) = 0$ for all $\eta > 0$.*

PROOF. It is not difficult to see that without loss of generality, the subcollection $\{\Delta_k\}$ may be required to be linearly ordered under the direction of refinement ($\Delta_{k+1} \geqq \Delta_k$).

Assume then that there exists such a countable collection satisfying the hypothesis of our theorem and linearly ordered under refinement. Let K be the set of those $x \in D(T)$ which assume a constant rational value on the cells of some partition Δ_k. Clearly, K is countable. For $x \in D(T)$ and $\varepsilon > 0$, choose k such that $f_x(\varepsilon, \Delta_k) \leqq \varepsilon$. Then there exists a λ with $\|\lambda\| < \varepsilon$ such that $\omega_x^{\bullet}(\lambda\Delta_k) < 2\varepsilon$. Thus $\omega_{x\circ\lambda}^{\bullet}(\Delta_k) = \omega_x^{\bullet}(\lambda\Delta_k) < 2\varepsilon$; so there exists an x_ε in K such that

$$\sup_t |x \circ \lambda(t) - x_\varepsilon(t)| < 2\varepsilon, \tag{3.43}$$

which implies $d(x, x_\varepsilon) \leqq 2\varepsilon$. Thus, K is dense.

Conversely, let $D(T)$ be separable with a countable dense subset $\{x_n\}$. Choose Δ_1 such that $f_{x_1}(1, \Delta_1) < 1$, and at each stage k choose Δ_k finer than Δ_{k-1} and such that $f_{x_n}(1/k, \Delta_k) < 1/k$ for all $n = 1, \cdots, k$. Then, for each n and $\eta > 0$, $\lim_k f_{x_n}(\eta, \Delta_k) = 0$, since for a given $\varepsilon > 0$, we need only choose $k > \max \{n, 1/\varepsilon, 1/\eta\}$ to assure

$$f_{x_n}(\eta, \Delta_k) \leqq f_{x_n}\left(\frac{1}{k}, \Delta_k\right) < \frac{1}{k} < \varepsilon. \tag{3.44}$$

We now show that $\lim_k f_x(\eta, \Delta_k) = 0$ for all x in $D(T)$. Let $x \in D(T)$, $\eta > 0$, and $\varepsilon > 0$ be given. Since $\{x_n\}$ is dense, we may choose an x_n for which $d(x_n, x) < \min \{\varepsilon, \eta\}$. Thus, there is a γ in Λ with $\|\gamma\| < \eta$ such that

$$\sup_t |x(t) - x_n \circ \gamma(t)| < \varepsilon. \tag{3.45}$$

This implies that

$$f_x(\eta, \Delta_k) \leqq \inf_{\{\lambda: \|\lambda\| < 2\eta\}} \omega_{x_n}^{\bullet}(\lambda\Delta) + 2\varepsilon, \tag{3.46}$$

for if there is a λ with $\|\lambda\| < \eta$ for which $\omega_x^{\bullet}(\lambda\Delta_k) \leqq c$, then $\mu = \gamma \circ \lambda$ satisfies $\|\mu\| \leqq \|\gamma\| + \|\lambda\| < 2\eta$ and

$$\sup_t |x_n \circ \mu(t) - x \circ \lambda(t)| = \sup_t |x_n \circ \gamma(t) - x(t)| < \varepsilon, \tag{3.47}$$

implying $\omega_{x_n}^{\bullet}(\mu\Delta_k) \leqq c + 2\varepsilon$. So, by choosing k such that $f_{x_n}(2\eta, \Delta_k) < \varepsilon$, we have $f_x(\eta, \Delta_k) \leqq 3\varepsilon$. This shows, since $f_x(\eta, \Delta_k)$ is monotonely nonincreasing as

k increases, that $\lim_k f_x(\eta, \Delta_k) = 0$. To make the proof of the characterization in our theorem complete, we need only note that if $\lim_k f_x(\eta, \Delta_k) = 0$, then $\lim_\Delta f_x(\eta, \Delta) = 0$ which implies that $x \in D(T)$. *Q.E.D.*

We shall refer to the linearly ordered collection $\{\Delta_k\}$ of this theorem as a *countable determining collection.*

4. Weak convergence of probability measures on D(T)

Just as one may translate the Arzelà-Ascoli characterization of compactness for the space $C[0, 1]$ into a characterization of tightness for a family of probability measures on the Borel sets of that space, so we may do likewise with Theorem 3.11 for the space $D(T)$. Every probability measure on the Borel sets of a separable and complete metric space is tight ($\sup \{P(K): \text{compact } K\} = 1$). The hypothesis of completeness cannot be suppressed, as an example of Le Cam in [3] illustrates. A discussion of the problem of whether one may suppress separability, and other problems related to it, may be found in Appendix III of [3]. Under the sole assumption of completeness, we require that each probability measure in a sequence be tight in order to characterize the tightness of the sequence.

THEOREM 4.1. *Let $\{D(T), d(\cdot, \cdot)\}$ be a complete metric space and let $\{P_n\}$ be a sequence of tight probability measures on the Borel sets $\mathscr{B}$ of $D(T)$. The sequence $\{P_n\}$ is tight if and only if*

(i) *for each $b > 0$, there exists an a such that $P_n\{x: \sup_t |x(t)| > a\} \leqq b$ for all but a finite number of n and*

(ii) *for each $b > 0$, $\eta > 0$, and $\varepsilon > 0$, there exists a $\Delta \in \mathscr{D}$ such that $P_n\{x: f_x(\eta, \Delta) \geqq \varepsilon\} \leqq b$ for all but a finite number of n.*

PROOF. Since $\sup_t |x(t)| = d(x, 0)$, the set $\{x: \sup_t |x(t)| > a\}$ is open, and therefore in $\mathscr{B}$. Moreover, for a fixed $\eta > 0$ and Δ in $\mathscr{D}$, $f_x(\eta, \Delta)$ is an upper semicontinuous function of x (Lemma 3.6); thus, $\{x: f_x(\eta, \Delta) \geqq \varepsilon\}$ is closed and therefore in $\mathscr{B}$.

Suppose that $\{P_n\}$ is tight. Given $b > 0$, $\eta > 0$, and $\varepsilon > 0$, choose a compact set K such that $P_n(K) > 1 - b$ for all n. Then, $K \subset \{x: \sup_t |x(t)| \leqq a\}$ for some a and $K \subset \{x: f_x(\eta, \Delta) < \varepsilon\}$ for some Δ in $\mathscr{D}$.

Conversely, suppose $\{P_n\}$ satisfies (i) and (ii). Given $b > 0$, choose a such that $B = \{x: \sup_t |x(t)| \leqq a\}$ implies $P_n(B) > 1 - (b/2)$ for all but a finite number of n. For each integer k, choose a $\Delta_k \in \mathscr{D}$ such that $B_k = \{x: f_x(1/k, \Delta_k) < 1/k\}$ implies $P_n(B_k) \geqq 1 - (b/2^{k+1})$ for all but a finite number of n. Since each individual P_n is tight, the previous paragraph shows that we may choose a and Δ_k so that both $P_n(B) > 1 - (b/2)$ and $P_n(B_k) \geqq 1 - (b/2^{k+1})$ hold for all n. Let $K = B \cap (\bigcap_k B_k)$. We show that K has compact closure. Since $K \subset B$, $\sup_{x \in K} \sup_t |x(t)| \leqq a$. Now, let $\eta > 0$ and $\varepsilon > 0$ be given and choose $k > \max \{1/\eta, 1/\varepsilon\}$. Since $K \subset B_k$, there is a Δ_k such that

$$f_x(\eta, \Delta_k) \leqq f_x\left(\frac{1}{k}, \Delta_k\right) < \frac{1}{k} < \varepsilon \tag{4.1}$$

for all $x \in K$. Thus,

$$\lim_{\Delta} \sup_{x \in \bar{K}} f_x(\eta, \Delta) = 0 \tag{4.2}$$

for all $\eta > 0$. This implies by Theorem 3.5, that $\bar{K}$ is compact. Since $P_n(\bar{K}) \geqq P_n(K) > 1 - b$ for all n, the sequence $\{P_n\}$ is tight. *Q.E.D.*

When $D(T)$ is separable and complete, tightness is equivalent to the relative weak compactness of $\{P_n\}$. Moreover, the criteria for compactness of a set of functions and tightness of a sequence of probability measures are somewhat simplified in that the limits involved may be taken along a sequence rather than along a net. Such a modification of Theorem 3.11 yields the following form of Theorem 4.1.

THEOREM 4.2. *Let $\{D(T), d(\cdot, \cdot)\}$ be a separable and complete metric space with a countable determining collection of partitions $\{\Delta_k\}$* (see Theorem 3.12). *Let $\{P_n\}$ be a sequence of probability measures on the Borel sets $\mathscr{B}$ of $D(T)$. The sequence $\{P_n\}$ is tight if and only if*

$$\lim_{a \to \infty} \limsup_{n} P_n\{x : \sup_{t} |x(t)| > a\} = 0 \tag{4.3}$$

and

$$\lim_{k} \limsup_{n} P_n\{x : f_x(\eta, \Delta_k) \geqq \varepsilon\} = 0 \tag{4.4}$$

for all $\eta > 0$ and $\varepsilon > 0$.

5. Weak convergence of stochastic processes with multidimensional parameter

We now apply the results of the previous two sections to an investigation of stochastic processes with several parameters. These processes are envisioned as random functions of a parameter lying in some subset of Euclidean space.

Once we extricate a problem from its dependencies on many characteristics of the real line, generalizations to higher dimensions become transparent. For this reason, we investigate the case in which our parameter lies in a subset of two dimensional Euclidean space R^2. For brevity, many proofs and details are omitted. Particulars may be found in [29]. Once the function space and its metric are developed, the formulations for problems of measurability and of weak convergence follow from the classical theory and from the theory for a single dimension without much difficulty. To illustrate this relation we try to follow the manner of Billingsley [3] in our presentation.

For our underlying space T, we take the unit square, without the north and east boundaries,

$$T = [0, 1)^2 = \{\mathbf{t} = (t_1, t_2) : 0 \leqq t_i < 1, i = 1, 2\}, \tag{5.1}$$

equipped with the relativized topology of Euclidean space, generated by the norm $\|\mathbf{t}\|_E = \max_i \{t_i\}$. The compact unit square $\bar{T}$ is denoted by $[0, 1]^2$. We define four partial orders for points in R^2 by

(a) $\mathbf{s} \leqq_{NE} \mathbf{t}$ if $s_1 \leqq t_1$ and $s_2 \leqq t_2$, read $\mathbf{t}$ is northeast of $\mathbf{s}$;
(b) $\mathbf{s} \leqq_{SE} \mathbf{t}$ if $s_1 \leqq t_1$ and $s_2 > t_2$, read $\mathbf{t}$ is southeast of $\mathbf{s}$;
(c) $\mathbf{s} \leqq_{SW} \mathbf{t}$ if $s_1 > t_1$ and $s_2 > t_2$, read $\mathbf{t}$ is southwest of $\mathbf{s}$; and
(d) $\mathbf{s} \leqq_{NW} \mathbf{t}$ if $s_1 > t_1$ and $s_2 \leqq t_2$, read $\mathbf{t}$ is northwest of $\mathbf{s}$.

Working in the plane has the advantage of employing the language of geographical orientation for our partial orders, but this has a drawback: $\mathbf{t}$ is northeast of $\mathbf{s}$ does not imply $\mathbf{s}$ is southwest of $\mathbf{t}$ as the case $\mathbf{t} = \mathbf{s}$ illustrates. In n dimensions, we work with the analogous 2^n partial orders.

For a subset $E \subset R^2$, a *monotone path* approaching a point $\mathbf{t}$ in E is a sequence of points $\{\mathbf{t}_n\}$ in E converging monotonely to $\mathbf{t}$ with respect to one of the four partial orders (a), (b), (c), (d). A real valued function x on E has *limits along monotone paths* (lamp) in E, if, for each point $\mathbf{t}$ in E, $\lim_n x(\mathbf{t}_n)$ exists along each monotone path $\{\mathbf{t}_n\}$ in E approaching $\mathbf{t}$. We call such an x a *lamp function* on E. A function x is *continuous from above* if, for each $\mathbf{t}$ and monotone path $\{\mathbf{t}_n\}$ approaching $\mathbf{t}$ from the northeast ($\mathbf{t} \leqq_{NE} \mathbf{t}_{n+1} \leqq_{NE} \mathbf{t}_n$), $\lim_n x(\mathbf{t}_n) = x(\mathbf{t})$. For example, a distribution function of a bivariate random vector is, under usual conventions, a continuous from above lamp function on R^2. One property of a continuous from above lamp function is that, for each point $\mathbf{t}$, the limit of $x(\mathbf{s})$ exists as $\mathbf{s}$ approaches $\mathbf{t}$ while contained in one of the quadrants $Q_i = \{\mathbf{s}: \mathbf{t} \leqq_i \mathbf{s}\}$, $i =$ NE, SE, SW, NW.

Let Λ_s be the class of those homeomorphisms of the compact unit interval onto itself which have zero as a fixed point. Let Λ_t be the class of those homeomorphisms in Λ_s which have a finite slope norm

$$\Lambda_t = \{\lambda \in \Lambda_s : \|\lambda\|_t < \infty\}; \tag{5.2}$$

and let Λ_d be the class of those homeomorphisms in Λ_s which have finite diffeomorphism norm

$$\Lambda_d = \{\lambda \in \Lambda_s : \|\lambda\|_d < \infty\}. \tag{5.3}$$

It follows from the mean value theorem and the definition of the derivative that, for $\lambda \in \Lambda_d$,

$$\|\lambda\|_d = \sup_t |\log \lambda'(t)| = \sup_{s \neq t} \left| \log \frac{\lambda(s) - \lambda(t)}{s - t} \right| = \|\lambda\|_t. \tag{5.4}$$

(Since λ is a homeomorphism for which $\lambda(0) = 0$, its derivative and difference quotient are positive.) We then have $\Lambda_d \subset \Lambda_t \subset \Lambda_s$.

Define groups Λ_i^2 of homeomorphisms of $[0, 1]^2$ by $\Lambda_i^2 = \Lambda_i \times \Lambda_i$, $i = s, t, d$, where the image of a point $\mathbf{t} = (t_1, t_2)$ in $[0, 1]^2$ under a homeomorphism $\boldsymbol{\lambda} = (\lambda_1, \lambda_2) \in \Lambda_i^2$ is the point $\boldsymbol{\lambda}\mathbf{t} = (\lambda_1 t_1, \lambda_2 t_2)$. For each integer k. let Δ_k be the partition of T having cells $\{A_{i,j}(k)\}$ for $i, j = 1, \cdots, 2^k$, defined by

$$A_{i,j}(k) = \left\{\mathbf{t}: \frac{i-1}{2^k} \leqq t_1 < \frac{i}{2^k}, \frac{j-1}{2^k} \leqq t_2 < \frac{j}{2^k}\right\}. \tag{5.5}$$

We then define our class of partitions $\mathscr{D}$ to be all images of partitions Δ_k by homeomorphisms in Λ_s^2

$$\mathscr{D} = \{\lambda\Delta_k: \lambda \in \Lambda_s^2, k = 0, 1, \cdots\}. \tag{5.6}$$

In this manner, $\mathscr{D}$ is invariant under Λ_i^2 with $i = s, t, d$.

Each finite collection of points $0 = p_0 < p_1 < \cdots < p_a = 1, 0 = q_0 < q_1 < \cdots < q_b = 1$ gives rise to a partition $\Delta(p_0, \cdots, p_a; q_0, \cdots, q_b)$ of T having cells

$$A_{i,j}(p_0, \cdots, p_a; q_0, \cdots, q_b) = \{\mathbf{t}: p_{i-1} \leqq t_1 < p_i, q_{j-1} \leqq t_2 < q_j\} \quad i = 1, \cdots, a; j = 1, \cdots, b. \tag{5.7}$$

We call the partition $\Delta(p_0, \cdots, p_a; q_0, \cdots, q_b)$ a δ grid if $p_i - p_{i-1} > \delta$ and $q_j - q_{j-1} > \delta$ for all $i = 1, \cdots, a, j = 1, \cdots, b$. By choosing an integer $k > \log_2(\max\{a, b\})$, we may find homeomorphisms λ_1 and λ_2 of $[0, 1]$ for which $\lambda_1(p_i) = i/2^k$, $i = 0, 1, \cdots, a-1$, and $\lambda_2(q_j) = j/2^k$, $j = 0, 1, \cdots, b-1$. In this manner, $\lambda = (\lambda_1, \lambda_2)$ is in Λ_s^2 and takes $\Delta(p_0, \cdots, p_a; q_0, \cdots, q_b)$ onto a coarsening of Δ_k: $\Delta_k \geqq \lambda\Delta(p_0, \cdots, p_a; q_0 \cdots, q_b)$. Since $\lambda^{-1}(\Delta_k) \in \mathscr{D}$, we have shown that $\mathscr{D}$ is a cofinal subset of the family of all δ grids.

The class of simple functions $I_{\mathscr{D}}$, recall, consists of those functions assuming a constant value on each cell of some partition in $\mathscr{D}$. Our function space $D(T) = D[0, 1)^2$ consists of uniform limits of sequences of simple functions. What is more, the following can be shown.

THEOREM 5.1. *The space $D(T)$ consists of all continuous from above* lamp *functions on $[0, 1]^2$ restricted to T.*

In addition. each x in $D(T)$ is continuous except on at most countably many lines (hyperplanes in higher dimensions) $\{\mathbf{t}: t_i = \text{constant}\}$, $i = 1, 2$, since simple functions have this property. In particular, the set of continuity points of x is dense in T.

We have defined three groups of homeomorphisms on $[0, 1]$, Λ_s, Λ_t, and Λ_d, associated with three norms, the supremum norm $\|\cdot\|_s$, the slope norm $\|\cdot\|_t$, and the diffeomorphism norm $\|\cdot\|_d$. To extend the definition of these norms to the three groups of homeomorphisms of the compact unit square, we define

$$\|\lambda\|_j = \max_{i=1,2} \{\|\lambda_i\|_j\} \quad \text{for } \lambda \in \Lambda_j^2, j = s, t, d. \tag{5.8}$$

By Lemma 3.1, the three groups with their associated norms induce, respectively, three metric spaces $\{\Lambda_i^2, d_{\Lambda,i}\}$, $i = s, t, d$. These, by (3.14), induce, respectively, three Skorohod metrics on $D(T)$: $d_2(\cdot, \cdot)$, $d_1(\cdot, \cdot)$, and $d_0(\cdot, \cdot)$. The convention we adopted, not to distinguish between two functions x and y in $D(T)$ with $d_2(x, y) = 0$, is redundant here: $d_2(x, y) = 0$ implies $x(\mathbf{t}) = y(\mathbf{t})$ at all continuity points $\mathbf{t}$ of x which, in turn, implies $x = y$. The same conclusion holds for $d_1(\cdot, \cdot)$ and $d_0(\cdot, \cdot)$.

Associated with each of these metrics on $D(T)$ is a modulus $f_x^{(i)}(\eta, \Delta)$, $i = 2, 1, 0$, defined by (3.19) with the appropriate group of homeomorphisms and its associated norm. Each of the moduli serve to characterize functions in $D(T)$ in the sense of Theorem 3.5.

Theorems 3.10 and 3.9 imply that $\{\Lambda_t^2, d_{\Lambda,t}\}$ and $\{\Lambda_d^2, d_{\Lambda,d}\}$ are complete. So, by Theorem 3.7, $\{D(T), d_1(\cdot, \cdot)\}$ and $\{D(T), d_0(\cdot, \cdot)\}$ are complete metric spaces. The space $\{D(T), d_2(\cdot, \cdot)\}$, however, is not complete: the sequence $x_n = I\{[0, 1/n)^2\}$ is Cauchy with respect to $d_2(\cdot, \cdot)$ but not convergent. Completeness, recall, is the hypothesis of Theorem 3.11 which characterizes compactness in $D(T)$.

In order to show that these three metrics on $D(T)$ are equivalent, we isolate a useful technical lemma, the proof of which may be left to one's imagination.

LEMMA 5.2. *Let $0 < \eta < 1$ be given and let k be an integer. Each $(1/k)$ grid can be brought onto a coarsening of a Δ_k partition by a λ in Λ_d^2 with $\|\lambda\|_d < \eta$, provided $k > 2\pi/\eta$. We may require, moreover, that $\|\lambda\|_s < 1/2^k < \eta$.*

THEOREM 5.3. *The metrics $d_i(\cdot, \cdot)$, $i = 2, 1, 0$, are equivalent on $D(T)$.*

PROOF. From (5.4), it follows that $\|\lambda\|_t = \|\lambda\|_d$ for $\lambda \in \Lambda_d^2$. Thus d_0 convergence implies d_1 convergence.

If $\lim_n \|\lambda_n\|_t = 0$ for $\lambda_n \in \Lambda_t^2$, then (3.35) implies $\lim_n \|\lambda_n\|_s = 0$. Thus, d_1 convergence implies d_2 convergence.

Our final task is to prove that d_2 convergence implies d_0 convergence. For this purpose, let $\varepsilon > 0$ and $\eta > 0$ be given. If $\lim_n d_2(x_n, y) = 0$, then a sequence $\lambda_n \in \Lambda_s^2$ with $\lim_n \|\lambda_n\|_s = 0$ may be found for which

$$\lim_n \sup_{\mathbf{t}} |x_n \circ \lambda_n(\mathbf{t}) - y(\mathbf{t})| = 0. \tag{5.9}$$

Now $y \in D(T)$ implies there is a $\Delta \in \mathscr{D}$ for which $\omega_y^{\bullet}(\Delta) < \varepsilon/2$; therefore, $\omega_{x_n}^{\bullet}(\lambda_n\Delta) = \omega_{x_n\lambda_n}^{\bullet}(\Delta) < \varepsilon$ for all sufficiently large n. Since $\lim_n \|\lambda_n\|_s = 0$, there is an integer $k > 2\pi/\eta$ for which $\lambda_n\Delta$ is a $(1/k)$ grid for all n. (Since Δ is, for some $\delta > 0$, a δ grid, we may choose n_0 so large that $n \geqq n_0$ implies $\lambda_n\Delta$ is a $(\delta/2)$ grid. Then choose $k > \max\{2\pi/\eta, 2/\delta\}$ and so large that $\lambda_n\Delta$ is a $(1/k)$ grid for $n = 1, \cdots, n_0$.) Lemma 5.2 implies that, for each n, a $\mu_n \in \Lambda_d^2$ may be found with $\|\mu_n\|_d < \eta$ for which $\Delta_k \geqq \mu_n \circ \lambda_n\Delta$. We then have $\mu_n^{-1}\Delta_k \geqq \lambda_n\Delta$ which implies $\omega_{x_n}^{\bullet}(\mu_n^{-1}\Delta_k) \leqq \omega_{x_n}^{\bullet}(\lambda_n\Delta) < \varepsilon$ for all n. Thus, since $\|\mu_n^{-1}\|_d = \|\mu_n\|_d < \eta$,

$$\sup_n f_{x_n}^{(0)}(\eta, \Delta_k) < \varepsilon. \tag{5.10}$$

Clearly, $\sup_n \sup_{\mathbf{t}} |x_n(\mathbf{t})| < \infty$. Since $\{D(T), d_0(\cdot, \cdot)\}$ is complete, Theorem 3.11 implies that the sequence is d_0 precompact. Thus, every subsequence of $\{x_n\}$ has a further subsequence which d_0 converges to some element of $D(T)$. Whatever be the d_0 convergent subsequence, its limit must be y, since d_0 convergence implies d_2 convergence. It follows that the entire sequence $\{x_n\}$ d_0 converges to y. *Q.E.D.*

We refer to the common topology of the metrics $d_i(\cdot, \cdot)$ as the *Skorohod topology*. When discussing a topological concept, we may, without ambiguity, suppress the subscript and use $d(\cdot, \cdot)$ to represent a metric of the Skorohod topology. An interesting property one may show of Skorohod convergence is the following.

LEMMA 5.4. *If* $\lim_n d(x_n, y) = 0$, *then* $\lim_n x_n(\mathbf{t}) = y(\mathbf{t})$ *at all continuity points* $\mathbf{t}$ *of* y. *Moreover, if* y *is uniformly continuous on* T, *then* x_n *converges uniformly to* y.

Moreover, the sequence of partitions $\{\Delta_k\}$ each having cells defined by (5.5) is a countable determining collection; so, by Theorem 3.12, the space $\{D(T), d(\cdot, \cdot)\}$ is separable.

For points $\mathbf{t}_1, \cdots, \mathbf{t}_k$ in T, let $\pi_{\mathbf{t}_1, \cdots, \mathbf{t}_k}$ be the projection from $D(T)$ onto k dimensional Euclidean space

$$(5.11) \qquad \pi_{\mathbf{t}_1, \cdots, \mathbf{t}_k}(x) = [x(\mathbf{t}_1), \cdots, x(\mathbf{t}_k)].$$

These evaluation maps are not necessarily continuous: $x_n = I\{[1/2 + 1/n, 1)^2\}$ converges to $x = I\{[1/2, 1)^2\}$ but $x_n(1/2, 1/2)$ does not converge to $x(1/2, 1/2)$. In general, however, $\pi_{\mathbf{t}}$ is continuous at x if and only if $\mathbf{t}$ is a continuity point of x.

Let $(R^k, \mathscr{B}^k)$ be a k dimensional Euclidean space with its Borel σ-field. The projection maps (5.11) are measurable maps from $(D(T), \mathscr{B})$ into $(R^k, \mathscr{B}^k)$. For a subset $T_0 \subset T$, we define a field $\mathscr{F}_{T_0}$ of sets of $D(T)$ by

$$(5.12) \qquad \mathscr{F}_{T_0} = \{\pi^{-1}_{\mathbf{t}_1, \cdots, \mathbf{t}_k}(B) : \mathbf{t}_i \in T_0, i = 1, \cdots, k; B \in \mathscr{B}^k, k = 1, \cdots\}.$$

We call $\mathscr{F}_T$ the field of *finite dimensional sets* or *cylinder sets* of $D(T)$. Since the projection maps are measurable, $\mathscr{F}_{T_0} \subset \mathscr{B}$, for each subset T_0 of T. What is more, the field of cylinder sets corresponding to a dense subset of T will generate the Borel σ-field $\mathscr{B}$ of the Skorohod topology.

If P is a probability measure on $\mathscr{B}$, the collection

$$(5.13) \qquad \{P\pi^{-1}_{\mathbf{t}_1, \cdots, \mathbf{t}_k} : \mathbf{t}_i \in T_0, i = 1, \cdots, k; k = 1, \cdots\}$$

is called the collection of *finite dimensional distributions of* P *for points in* T_0. When T_0 is dense, $\mathscr{F}_{T_0}$ generates $\mathscr{B}$, and, in that case, P is uniquely determined by its finite dimensional distributions for points in T_0 in the sense that it is the only probability measure on $\mathscr{B}$ giving rise to (5.13).

By a *stochastic process* or *random element* in $D(T)$, we mean a measurable map from some probability space into $(D(T), \mathscr{B})$. Measurability of such a map is characterized by the fact that evaluating the process at each point in T induces a random variable. As is common in discussing random variables, we shall typically suppress the dependency of the random element X on the element of the probability space and write X and $X(\mathbf{t})$ for the process and its evaluation at $\mathbf{t}$.

We call $\mathbf{u} \in T$ a *stochastic continuity point* of the process X in $D(T)$, if the probability measure P corresponding to X assigns measure one to the set in $\mathscr{B}$ defined by

$$(5.14) \qquad \{x \in D(T) : x \text{ is continuous at } \mathbf{u}\}.$$

We denote by S_X or S_P the set of stochastic continuity points of X in T. Since $\pi_{\mathbf{u}}$ is continuous at x in $D(T)$ if and only if x is continuous at $\mathbf{u}$, the set S_X consists of those points $\mathbf{u}$ at which $\pi_{\mathbf{u}}$ is continuous almost everywhere P. Moreover, S_X contains $\mathbf{0} = (0, 0)$ and its complement lies on at most countably many lines (hyperplanes in higher dimensions) $\{\mathbf{t}: t_i = \text{constant}\}$. In particular, S_X is dense in T.

For a random element X with corresponding probability measure P, we use X to implicitly define some events and use the symbol $\mathscr{P}$ to mean the probability or P measure of the event so defined. For example, with $A \in \mathscr{B}$ we may write $\mathscr{P}\{X \in A\}$ to mean $P(A)$ and write $\mathscr{P}\{f_X(\eta, \Delta) \geqq \varepsilon\}$ to mean $P\{x \in D(T): f_x(\eta, \Delta) \geqq \varepsilon\}$.

Many of our theorems stated for probability measures may be reformulated in terms of random elements and vice versa. To apply the theorems of weak convergence, we need not have all our random elements defined on a common probability space. But even if they were, there would be no loss in generality in discussing the weak convergence of a sequence of random elements $\{X_n\}$ to a random element X: it is a theorem of Varadarajan [31] that, if $\{P_n\}$ is a sequence of probability measures defined on the Borel sets of a complete and separable metric space each of which is induced by a measurable map on a common probability space and if P_n converges weakly to a probability measure P, then the limiting measure P may be induced by some measurable map defined on the same probability space.

Proving weak convergence of a sequence of probability measures $\{P_n\}$, $n = 0, 1, \cdots$, on the Borel sets of a complete and separable metric space typically involves two steps. First, we prove that our sequence is tight, thus assuring that every subsequence of it has a limit point in the (metrizable) topology of weak convergence. Second, we identify the limit by showing that every weakly convergent subsequence must converge to P_0. Then we may conclude that the entire sequence $\{P_n\}$ converges weakly to P_0. This second step—identifying the limiting probability measure—is handled in $C[0, 1]$, for example, by proving that the finite dimensional distributions of P_n converge weakly to those of P_0. In the space $D(T)$, this technique must be modified: the evaluation maps (5.11) may not be continuous; and so it may be possible, for some $\mathbf{t} \in T$, that $P_n \Rightarrow P_0$, but $P_n\pi_{\mathbf{t}}^{-1}$ does not converge weakly to $P_0\pi_{\mathbf{t}}^{-1}$. Nevertheless, if the sequence $\{P_n\}$ is tight and if the finite dimensional distributions of P_0 for points in S_{P_0}, its set of stochastic continuity points, are the weak limits of those of the P_n, then $P_n \Rightarrow P_0$.

Reformulating Theorem 4.2 in terms of random elements yields the following criteria for weak convergence.

THEOREM 5.5. *A sequence $\{X_n\}$ of stochastic processes in $D(T)$ is weakly convergent if and only if* (i) *the finite dimensional distributions of X_n for points in some dense subset of T are weakly convergent—that is, there exists a dense T_0 for which $[X_n(\mathbf{t}_1), \cdots, X_n(\mathbf{t}_k)]$ converges weakly to a random vector, say $(X_{\mathbf{t}_1}, \cdots, X_{\mathbf{t}_k})$, whenever $\{\mathbf{t}_1, \cdots, \mathbf{t}_k\}$ is a finite subset of T_0 and* (ii) *for all $\eta > 0$ and $\varepsilon > 0$.*

$$\lim_k \limsup_n \mathscr{P}\{f_{X_n}^{(0)}(\eta, \Delta_k) \geqq \varepsilon\} = 0. \tag{5.15}$$

Necessarily, X_n converges weakly to that stochastic process X for which the finite dimensional distribution of $[X(\mathbf{s}_1), \cdots, X(\mathbf{s}_k)]$, for $\mathbf{s}_1, \cdots, \mathbf{s}_k$ in T is that of *the weak limit (which perforce exists) of the sequence* $(X_{\mathbf{t}_{1,m}}, \cdots, X_{\mathbf{t}_{k,m}})$ *defined by some arbitrary monotone paths* $\{\mathbf{t}_{i,m}\}$, $m = 1, \cdots$, *approaching* $\mathbf{s}_i$ *from the northeast with* $\mathbf{t}_{i,m} \leqq_{\mathrm{SW}} \mathbf{s}_i$ *and* $\mathbf{t}_{i,m} \in T_0$ *for all* m *with* $i = 1, \cdots, k$.

That condition (4.3) of Theorem 4.2 is satisfied in the present context is a consequence of (i) and (ii), and the fact that if Δ_k is a partition with cells $\{A_v\}$ and $\{\mathbf{t}_v\}$ are points in T_0 with $\mathbf{t}_v \in A_v$, then

$$(5.16) \qquad \{\sup_{\mathbf{t}} |x(\mathbf{t})| > a\} \subset \{\max_v |x(\mathbf{t}_v)| \geqq a - \max(2\varepsilon, \eta)\} \cup \{f_x^{(0)}(\eta, \Delta_k) \geqq \varepsilon\}.$$

The formulation of this theorem makes it unnecessary to specify the limiting distribution or even any of its stochastic continuity points. Note that if $X_n \Rightarrow X$, $X_n(\mathbf{t})$ is not necessarily weakly convergent, and even if it were so its weak limit need not be $X(\mathbf{t})$; hence the use of the notation $X_{\mathbf{t}}$ for its hypothesized weak limit and the need to approach $X(\mathbf{s})$ by the $X_{\mathbf{t}}$ with $\mathbf{s}$ southwest of $\mathbf{t}$. The idea that a putative limiting measure may be determined by iterative weak limits of finite dimensional distributions for points in an arbitrary dense subset of the unit interval (from which weak convergence to that measure is a consequence of tightness) is an observation of Topsøe [30] for processes in $D[0, 1]$.

When the limiting process belongs, with probability one, to the closed subspace $C_u(T)$ of those uniformly continuous functions on T, we need not measure the fluctuations of the random elements in terms of $f_x^{(0)}(\eta, \Delta)$, but rather we may use the stronger modulus of uniformly continuous functions defined for $\delta > 0$ and $x \in D(T)$ by

$$(5.17) \qquad \omega_x(\delta) = \sup\{|x(\mathbf{s}) - x(\mathbf{t})| : \mathbf{s}, \mathbf{t} \in T, \|\mathbf{s} - \mathbf{t}\|_{\mathrm{E}} \leqq \delta\}.$$

THEOREM 5.6. *A sequence* $\{X_n\}$ *of stochastic processes is weakly convergent to a process which belongs, with probability one, to* $C_u(T)$ *if and only if* (i) *the finite dimensional distributions of* X_n *for points in some dense subset of* T *are weakly convergent and* (ii) *for all* $\varepsilon > 0$,

$$(5.18) \qquad \lim_{\delta \to 0} \limsup_n \mathscr{P}\{\omega_{X_n}(\delta) \geqq \varepsilon\} = 0.$$

Necessarily X_n *converges weakly to that stochastic process* X *whose finite dimensional distributions for points in* T *are the weak limits of those of the* X_n.

The function space $D(T) = D[0, 1)^2$ is, with minor changes, the one used by Dudley [7] and Wichura [33]. The differences are the deletion of the north and east boundaries of the unit square and a choice of the convention of continuity from the northeast. Following Dudley's style, Wichura christens the functions in $D[0, 1]^2$ continuous from above with limits from below. Both of these authors endow the function space with the supremum metric $\rho(\cdot, \cdot)$, with the result that it is not separable. They then define a mode of weak convergence of probability measures on the σ-field $\mathscr{U}$ generated by the open balls of $\rho(\cdot, \cdot)$ to a probability measure concentrated on a separable subspace or, what is the same, to a tight probability measure. This mode of convergence is the weak* convergence of the probability measures as elements of the dual to the Banach space, with the uniform norm, of those real, continuous, and bounded functions which are $\mathscr{U}$ measurable on $\{D[0, 1]^2, \rho(\cdot, \cdot)\}$ (see Dudley [8]). Our mode of convergence—the classical one for probability measures on a separable and complete metric space—is equivalent to the weak* topology on the dual to the Banach space, with the uniform norm, of all real, continuous, and bounded functions on $\{D[0, 1)^2, d_0(\cdot, \cdot)\}$. The Borel σ-field of our space is the same as $\mathscr{U}$ relativized to $D[0, 1)^2$, since both are generated by the field of finite dimensional sets.

When our theory is applied to stochastic processes of a single parameter by letting $T = [0, 1)$, some differences to the theory for processes in $D[0, 1]$ arise. A function x in $D[0, 1)$ may be extended to a function $\hat{x}$ in $D[0, 1]$ by merely prescribing some arbitrary real value for $\hat{x}(1)$. It is easy to see that a sequence of functions $\{\hat{x}_n\}$, in $D[0, 1]$ converges to $\hat{x}_0$ if and only if the corresponding sequence of restrictions in $D[0, 1)$, $\{x_n\}$, converges to x_0 with respect to one of the Skorohod metrics of $D[0, 1)$ and $\hat{x}_n(1)$ converges to $\hat{x}_0(1)$. Nevertheless, a sequence of stochastic processes $\{\hat{X}_n\}$ in $D[0, 1]$ need not converge weakly to $\hat{X}_0$ even though the sequence of restricted processes $\{X_n\}$ converges weakly to X_0 in $D[0, 1)$ and the random variables $\hat{X}_n(1)$ converge weakly to $\hat{X}_0(1)$. The $\hat{X}_n$, while not necessarily convergent to $\hat{X}_0$, are however relatively compact. Thus, the $\hat{X}_n$ converge weakly to $\hat{X}_0$ in $D[0, 1]$ if and only if the restricted processes X_n converge to X_0 in $D[0, 1)$ and the finite dimensional distributions of $\hat{X}_n$ for points in some dense subset of $[0, 1]$ which contains the point 1 converge weakly to those of $\hat{X}_0$.

A modulus proposed by Billingsley ([3], p. 110) serves to characterize functions, compactness, and tightness in $D[0, 1]$. For $\delta > 0$, let

$$\omega'_x(\delta) = \inf_{\{\Delta\}} \omega_x^{\cdot}(\Delta), \tag{5.19}$$

where the infimum extends over all δ grids that are partitions of $T = [0, 1)$. Billingsley's necessary and sufficient conditions for tightness of a sequence $\{\hat{X}_n\}$ of random elements of $D[0, 1]$ are

$$\begin{aligned} &\lim_{a\to\infty} \limsup_n \mathscr{P}\{\sup_{t\in[0,1]} |\hat{X}_n(t)| > a\} = 0 \\ &\lim_{\delta\to 0} \limsup_n \mathscr{P}\{\omega'_{X_n}(\delta) \geqq \varepsilon\} = 0 \quad \text{for all } \varepsilon > 0. \end{aligned} \tag{5.20}$$

Statements (5.19) and (5.20) have interpretations in $D[0, 1)^2$ if t in $[0, 1]$ is replaced by t in $[0, 1)^2$. It is clear from the definitions of our metrics and from Billingsley's development in $D[0, 1]$ that (5.20) offers another characterization of tightness for random elements of $D[0, 1)^2$. Indeed, that these are sufficient conditions for tightness follows from Theorem 4.2 and the fact that Lemma 5.2 implies $f_x^{(0)}(\eta, \Delta_k) \leqq \omega'_x(1/k)$ for each x in $D[0, 1)^2$ whenever $1/k < \eta/2\pi$. While the modulus (5.19) is useful for this purpose, it is peculiar to functions on Euclidean spaces.

The essential trick in $D[0, 1)^2$ is just to perturb each point in the unit square one coordinate at a time and parallel to an axis. Concurrent with this research and independent of it, Neuhaus [20] applied this idea to the multidimensional version of Billingsley's metric, which is our $d_1(\cdot, \cdot)$. The minor novelty of the group of diffeomorphisms and its associated norm makes it easier to prove that our $d_0(\cdot, \cdot)$ is a metric—that fact becomes a consequence of the chain rule and the inverse function theorem. Moreover, completeness follows from the classical relation of uniform convergence to differentiability. An analogue of Lemma 5.2 would be easier to prove if the graph of the homeomorphism were allowed to have lines meeting at angles; but it is intuitive that we can smoothen the corners to construct a diffeomorphism which does the job as well. Just as intuitive, and what can be shown with the same chicanery, is that there is no difference at all between $d_0(\cdot, \cdot)$ and $d_1(\cdot, \cdot)$; not only are they topologically equivalent, they are equal.

6. Some examples

The practicality of the theory of weak convergence of probability measures on a function space is reflected in its application to the asymptotic analysis of nonparametric tests. Reciprocally, the relevance to the Kolmogorov-Smirnov statistic was an important stimulus for the development of the theory. We describe briefly a few examples of nonparametric tests which may be envisioned as functionals of stochastic processes with several parameters.

The statistic that first comes to mind is the multivariate Kolmogorov-Smirnov one. Let F_n be the empirical distribution function of a random sample $\{\mathbf{X}_1, \cdots, \mathbf{X}_n\}$ taken from a population with a continuous bivariate distribution function F concentrated on $T = [0, 1)^2$. The function

$$X_n(\mathbf{t}) = \sqrt{n}[F_n(\mathbf{t}) - F(\mathbf{t})] \tag{6.1}$$

is a random element of $D(T)$. If f is a measurable function on $[0, 1]^2$ whose square is integrable with respect to F, then the central limit theorem implies that

$$\int_{[0,1]^2} f(\mathbf{t})\, dX_n(\mathbf{t}) = \sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n} f(\mathbf{X}_i) - E[f(\mathbf{X}_1)]\right] \tag{6.2}$$

converges weakly to a normal random variable having mean zero and variance Var $[f(\mathbf{X}_1)]$. Since indicator functions are square integrable, and so are linear combinations of them, it follows from a theorem of Varadarajan [32] and also from the Cramér-Wold device (see [3], p. 49) that the finite dimensional distributions of X_n converge weakly to those of multivariate normal random vectors with the origin as mean vector. If we write, for points $\mathbf{s} = (s_1, s_2)$ and $\mathbf{t} = (t_1, t_2)$ in T, $\mathbf{s} \wedge \mathbf{t} = (\min\{s_1, t_1\}, \min\{s_2, t_2\})$, then the covariances of the limiting random vector are given by Cov $[X(\mathbf{s}), X(\mathbf{t})] = F(\mathbf{s} \wedge \mathbf{t}) - F(\mathbf{s})F(\mathbf{t})$.

Dudley ([6] and [7]) has shown that there is a stochastic process X in $D(T)$ which has the finite dimensional distributions of our limiting random vector and is uniformly continuous with probability one. For each functional h which is continuous with respect to the supremum norm on $D(T)$, the random variables $h(X_n)$ converge weakly to $h(X)$. Although Dudley has employed a different mode of weak convergence for the X_n, this conclusion is equivalent to their weak convergence to X as random elements of our Skorohod space. Using other methods, Bickel and Wichura ([1], [2], and [33]) and Neuhaus [20] have also proved the weak convergence of the X_n.

It follows from this weak convergence, that the bivariate Kolmogorov-Smirnov statistic $\sup_{\mathbf{t}}|X_n(\mathbf{t})|$ converges weakly to $\sup_{\mathbf{t}}|X(\mathbf{t})|$. As a consequence, F_n converges uniformly to F in probability. This fact in turn implies that the bivariate Cramér-von Mises statistic $\int|X_n(\mathbf{t})|^2\, dF_n(\mathbf{t})$ converges weakly to $\int|X(\mathbf{t})|^2\, dF(\mathbf{t})$.

The asymptotic distribution of the multivariate Kolmogorov-Smirnov statistic was first investigated by Kiefer and Wolfowitz [13]. They found a uniform bound for the tails of the distributions of the statistics $\sup_{\mathbf{t}}|X_n(\mathbf{t})|$. Moreover, they showed that the asymptotic distribution exists and that it can be approximated by taking the maximum of the k random variables formed by evaluating X_n at k points on a lattice and then letting n and k become large, respectively. Later, Kiefer [12] improved their bound for the tails of the distribution of $\sup_{\mathbf{t}}|X_n(\mathbf{t})|$. As for the multivariate Cramér-von Mises statistic, a limit theorem for a variant of it was considered by Rosenblatt [23]; and Dugué [9] investigated the limiting characteristic function of such a statistic.

The bivariate Kolmogorov-Smirnov statistic is unfortunately not distribution free (see Simpson [25]). Bickel [1], however, has proposed a two sample multivariate Smirnov test which is distribution free. For simplicity, we describe this test when samples of sizes m and n are drawn, respectively, from populations with continuous distribution functions F and G concentrated on the unit square. From the $m + n$ points in the pooled sample, there are $\binom{m+n}{m}$ ways to select m points and form an empirical distribution function F_m with them and an empirical distribution function G_n with the remaining n points. We construct a random function $X_{m,n}$ which assumes each of the possible $\binom{m+n}{m}$ functions $(mn/(m + n))^{1/2}[F_m - G_n]$ with equal probability. Bickel's test rejects the hypothesis that the two samples we have drawn come from the same population if the value of $\sup_{\mathbf{t}}|X_{m,n}(\mathbf{t})|$ obtained when F_m and G_n are the empirical distribution functions of our first and second samples, respectively, is significantly high when ranked among the possible $\binom{m+n}{m}$ values of $\sup_{\mathbf{t}}|X_{m,n}(\mathbf{t})|$. Using Dudley's terminology, Bickel shows, moreover, that if m and n become large in such a manner that $m/(m + n)$ converges to a number α in $[0, 1]$, then the stochastic process $X_{m,n}$ converges weakly to a Gaussian process X. If $H = \alpha F + (1 - \alpha)G$, then the finite dimensional distributions of X are multivariate normal ones with the origin as a mean vector and with covariances given by $\operatorname{Cov}[X(\mathbf{s}), X(\mathbf{t})] = H(\mathbf{s} \wedge \mathbf{t}) - H(\mathbf{s})H(\mathbf{t})$.

Another application may be made to tests of independence. Let $F_{1,n}$ and $F_{2,n}$ be the marginals of the empirical distribution function F_n of a random sample from a population with a continuous distribution function F which is concentrated on the unit square and is the product of its marginals: $F(t_1, t_2) = F_1(t_1)F_2(t_2)$. Blum, Kiefer, and Rosenblatt [4] have studied two statistics which, for a bivariate population, are functionals of the process

$$Z_n(\mathbf{t}) = \sqrt{n}[F_n(\mathbf{t}) - F_{1,n}(t_1)F_{2,n}(t_2)]. \tag{6.3}$$

Taking their cue from Kolmogorov and Smirnov, the authors look at $\sup_{\mathbf{t}} |Z_n(\mathbf{t})|$ and, in the spirit of Cramér and von Mises, they investigate $\int Z_n^2 \, dF_n$, which they show to be equivalent to a test proposed by Hoeffding [11].

The weak convergence of the Z_n is a consequence of that of the X_n. Consider for the moment the empirical distribution functions $F_{1,m}$ and $F_{2,n}$ of independent random samples of sizes m and n taken from populations with continuous distribution functions F_1 and F_2 concentrated on the unit interval. If m and n become large in such a manner that $m/(m+n)$ converges to an α in $[0, 1]$, then the process

$$H_{m,n}(\mathbf{t}) = (mn/(m+n))^{1/2}[F_{1,m}(t_1)F_{2,n}(t_2) - F_1(t_1)F_2(t_2)] \tag{6.4}$$

is weakly convergent. This fact is a consequence of the weak convergence of the univariate analogue of X_n and the fact that, with $p = m/(m+n) = 1 - q$,

$$\begin{aligned} H_{m,n}(\mathbf{t}) = {} & \sqrt{p}[qF_{1,m}(t_1) + pF_1(t_1)]\sqrt{n}[F_{2,n}(t_2) - F_2(t_2)] \\ & + \sqrt{q}[pF_{2,n}(t_2) + qF_2(t_2)]\sqrt{m}[F_{1,m}(t_1) - F_1(t_1)]. \end{aligned} \tag{6.5}$$

The weak convergence of Z_n is a consequence of the relation $Z_n(\mathbf{t}) = X_n(\mathbf{t}) - 2H_{n,n}(\mathbf{t})$, the weak convergence of its finite dimensional distributions, and the following lemma.

LEMMA 6.1. *If, of weakly convergent sequences of stochastic processes* $\{X_n\}$ *and* $\{Y_n\}$ *in* $D[0, 1)^2$, *the first converges weakly to a process that is uniformly continuous with probability one, then the sequence* $\{Z_n\}$ *defined by* $Z_n(\mathbf{t}) = X_n(\mathbf{t}) + Y_n(\mathbf{t})$ *is relatively compact.*

PROOF. Tightness of the sequence $\{Z_n\}$ is a consequence of the relation

$$\mathscr{P}\{\sup_{\mathbf{t}} |Z_n(\mathbf{t})| > a\} \leqq \mathscr{P}\left\{\sup_{\mathbf{t}} |X_n(\mathbf{t})| > \frac{a}{2}\right\} + \mathscr{P}\left\{\sup_{\mathbf{t}} |Y_n(\mathbf{t})| > \frac{a}{2}\right\} \tag{6.6}$$

and the fact that, for each $\delta > 1/2^{k-2}$, $\eta > 0$, and $\varepsilon > 0$,

$$\{f_{Z_n}^{(0)}(\eta, \Delta_k) \geqq \varepsilon\} \subset \left\{f_{Y_n}^{(0)}(\eta', \Delta_k) \geqq \frac{\varepsilon}{2}\right\} \cup \left\{\omega_{X_n}(\delta) \geqq \frac{\varepsilon}{2}\right\}, \tag{6.7}$$

provided $\eta' \in (0, \eta)$ is chosen so that $\|\lambda\|_d < \eta'$ implies $\|\lambda\|_s < \delta/4$. Such a choice is assured to be possible by (3.35) and (5.4). *Q.E.D.*

The process $H_{m,n}(\mathbf{t})$ is of some interest in itself. For example, the normalized Mann-Whitney form of the Wilcoxon statistic is the integral of the indicator function of the triangle $\{\mathbf{t}: 0 < t_2 < t_1 < 1\}$ with respect to $H_{m,n}(\mathbf{t})$.

Weak convergence criteria useful for processes constructed from multivariate empirical distribution functions have been investigated by Bickel and Wichura [2]. They have developed multivariate analogues of fluctuation inequalities proposed by Billingsley in [3].

Many problems of this theory remain, not the least of which is the computation of distribution functions for functionals of a Gaussian process of several parameters. Nevertheless, the theory of weak convergence promises to be an advantageous factotum for the asymptotic analysis of nonparametric tests.

Our second example which illustrates the flexibility of the modulus $f_x(\eta, \Delta)$ is an application to the weak convergence of stochastic point processes—a theory which we shall explore in detail elsewhere. For the moment however, we show that in the case of many special limiting point processes, weak convergence of the finite dimensional distributions is equivalent to weak convergence of the stochastic point process. Consider $X_n(\mathbf{t})$ defined by a stochastic point process to be the number of points in the closed rectangle with the origin as a southwest and $\mathbf{t}$ as a northeast corner. Let A_k be the event that a point process has no points occurring within a distance of $1/k$ from the boundary of the unit square and that when the square is partitioned by vertical and horizontal strips of width $1/k$, no two points occur within a common strip or in adjacent strips. If $X_n \in A_k$, then one may form a $1/(k + 1)$ grid Δ with the south and west boundaries of $T = [0, 1)^2$ and those vertical and horizontal lines in T which contain a point of the realization of the process defining X_n. Each point of the process will lie on the southwest corner of some cell of the resulting partition Δ; therefore, $\omega_{\dot{X}_n}(\Delta) = 0$ whenever $X_n \in A_k$. Lemma 5.2 then implies for a given $0 < \eta < 1$ that

$$\{X_n \in A_k\} \subset \{f_{X_n}^{(0)}(\eta, \Delta_{k+1}) = 0\} \tag{6.8}$$

for all sufficiently large k. Thus, condition (5.15) of Theorem 5.5 will be satisfied if

$$\lim_k \limsup_n \mathscr{P}\{X_n \in A_{2^k}\} = 1. \tag{6.9}$$

Now $\mathscr{P}\{X_n \in A_{2^k}\}$ is determined by finite dimensional distributions of X_n for points in the closed unit square and if these converge weakly to those of a process X_0 defined by a stochastic point process, the condition for weak convergence reduces to $\lim_k \mathscr{P}\{X_0 \in A_{2^k}\} = 1$. Since $A_{2^k} \subset A_{2^{k+1}}$, this last requirement may be written as $\mathscr{P}\{X_0 \in \cup_k A_{2^k}\} = 1$, which means that, with probability one, realizations of the limiting stochastic point process have neither multiplicities nor points on the boundary of the unit square. For example, the Poisson process satisfies this requirement. Essentially, this argument means that, relativized to a suitable subspace in which these processes lie, Skorohod convergence to a function in $\cup_k A_{2^k}$ follows from pointwise convergence at continuity points of the limiting function.

As our final example, we develop an anlogue to stochastic processes with stationary and independent increments and prove a theorem emulating the one of Donsker for the weak convergence of partial sum processes. Let $\{X_{i,j}(n)\}$ be a triangular array of random variables, the nth row of which is composed of n^2 independent random variables $X_{i,j}(n)$, $i, j = 1, \cdots, n$, with a common distribution that may vary from row to row. Denote the row sums by $S_n = \Sigma_{i,j} X_{i,j}(n)$.

For a random variable X, write its characteristic function as $\phi_X(u)$, and set $\phi_n(u) = \phi_{X_{i,j(n)}}(u)$ so that $\phi_{S_n}(u) = \phi_n^{n^2}(u)$. For the greatest integer less than or equal to a real number t, write $[t]$; and for a vector $\mathbf{t} = (t_1, t_2)$, interpret $[\mathbf{t}]$ as $([t_1], [t_2])$. Corresponding to each triangular array $\{X_{i,j}(n)\}$ is the sequence of partial sum processes Y_n in $D[0, 1)^2$ defined by

$$Y_n(\mathbf{t}) = \sum_{i,j} \{X_{i,j}(n): (i,j) \leqq_{\mathrm{NE}} [n\mathbf{t}]\} = \sum_{i=1}^{[nt_1]} \sum_{j=1}^{[nt_2]} X_{i,j}(n). \tag{6.10}$$

Let $T = [0, 1)^2$, let $\mu(\cdot)$ be Lebesgue measure on $\mathscr{B}_T$, the Borel sets of T, and let $\mathscr{A}$ be the class of finite unions of disjoint rectangles in T which have sides parallel to the axes and have boundaries on the north and east but are open to the south and west. For points $\mathbf{s} \leqq_{\mathrm{NE}} \mathbf{t}$ in T, we define $R(\mathbf{s}, \mathbf{t})$ to be that rectangle in $\mathscr{A}$ which has $\mathbf{s}$ and $\mathbf{t}$ as southwest and northeast corners, respectively. If $\mathbf{s}$ and $\mathbf{t}$ are collinear, $R(\mathbf{s}, \mathbf{t})$ is the empty set. For $x \in D(T)$, let $H_x(R(\mathbf{s}, \mathbf{t}))$ be the second difference of x about the rectangle $R(\mathbf{s}, \mathbf{t})$

$$H_x(R(\mathbf{s}, \mathbf{t})) = x(t_1, t_2) - x(t_1, s_2) - x(s_1, t_2) + x(s_1, s_2). \tag{6.11}$$

We extend the domain of the function H_x to $\mathscr{A}$ by requiring that, for disjoint rectangles $R(\mathbf{s}_i, \mathbf{t}_i)$, $\mathbf{s}_i \leqq_{\mathrm{NE}} \mathbf{t}_i$, $i = 1, \cdots, n$.

$$H_x\left(\bigcup_{i=1}^{n} R(\mathbf{s}_i, \mathbf{t}_i)\right) = \sum_{i=1}^{n} H_x(R(\mathbf{s}_i, \mathbf{t}_i)). \tag{6.12}$$

For $\mathbf{p} \in R^2$ and $A \in \mathscr{A}$, we define $A + \mathbf{p}$, the *translation of A by $\mathbf{p}$ within T*, by

$$A + \mathbf{p} = \begin{cases} \{\mathbf{t} + \mathbf{p} : \mathbf{t} \in A\} & \text{if this set is contained in } T \\ A & \text{otherwise.} \end{cases} \tag{6.13}$$

We shall say a stochastic process X in $D(T)$ has *stationary increments* if, for each $A \in \mathscr{A}$, the random variables $\{H_X(A + \mathbf{p})\}$, $\mathbf{p} \in R^2$, have a common distribution. A process with *independent increments* is a random element X for which, with $\{A_i\}$ a finite number of disjoint sets in $\mathscr{A}$, the random variables $\{H_X(A_i)\}$ are independent. The partial sum processes corresponding to a triangular array of random variables are stochastic processes in $D(T)$ with independent increments.

THEOREM 6.2. *If the row sums of our triangular array of random variables $\{X_{i,j}(n)\}$ have a limiting distribution with characteristic function $f(u)$, then the corresponding partial sum processes Y_n converge weakly to a stochastic process Y in $D(T)$ with stationary and independent increments and with*

$$\phi_{H_Y(A)}(u) = [f(u)]^{\mu(A)} \tag{6.14}$$

for each $A \in \mathscr{A}$.

Our proof of this theorem employs a common technique used in limit theorems for sums of random variables without moments: we truncate the $X_{i,j}(n)$. Our starting point is the assumption of the analogue of Donsker's theorem which states that Theorem 6.2 is true when the $X_{i,j}(n)$ satisfy Lindeberg's condition for the convergence of their row sums to a normal law and where the limiting process has Wiener measure; that is, W is the stochastic process in $D[0, 1)^2$ with stationary and independent increments, which is uniformly continuous with probability one, and for which $W(t_1, t_2)$ is a normal random variable with mean zero and variance $t_1 t_2$ (see the works of Kuelbs [15], Wichura ([33] and [34]), and Bickel and Wichura [2]).

The basic idea of our method, which has its roots in the work of Lévy, was used by Skorohod [26], who presented a rationale for the weak convergence of partial sum processes in $D[0, 1]$. (Later, and by other methods, Skorohod proved this result as a theorem dealing with the weak convergence of somewhat more general processes [28].) We split each partial sum process into a sum of two parts: a process formed with the truncated variables and another formed with their remainders. The first process differs little from one which, by the analogue of Donsker's theorem, converges to a Gaussian process. The second converges to a compound Poisson process. By employing a weak compactness lemma of Feller, we obviate the need to assume *a priori* the existence of our limit process in $D(T)$.

For $s > 0$, define a *selection function* β_s by

$$\beta_s(t) = \begin{cases} t & \text{if } |t| \leqq s \\ 0 & \text{otherwise} \end{cases} \tag{6.15}$$

and a *remainder function* ρ_s by $\rho_s(t) = t - \beta_s(t)$. Let

$$p_{s,n} = \mathscr{P}\{|X_{1,1}(n)| > s\} = \mathscr{P}\{\rho_s(X_{1,1}(n)) \neq 0\},$$

$$\sigma_n^2(s) = \operatorname{Var}\Big[\sum_{i,j} \beta_s(X_{i,j}(n))\Big] = n^2 \operatorname{Var}[\beta_s(X_{1,1}(n))], \tag{6.16}$$

$$\alpha_n^2(s) = n^2 E[\beta_s(X_{1,1}(n))]^2,$$

$$\gamma_{s,n} = E[\beta_s(X_{1,1}(n))].$$

The following lemma is, with minor changes, that of Feller ([10], pp. 299–300).

LEMMA 6.3. *Suppose that $X_{1,1}(n)$ converges in probability to zero. In order that there exist constants a_n such that $\{S_n - a_n\}$ be tight, it is necessary and sufficient that, to each $\varepsilon > 0$, there is a b and a real valued function M for which*

$$n^2 p_{b,n} < \varepsilon, \quad \text{for all } n, \tag{6.17}$$

$$\sup_n \sigma_n^2(s) < M(s). \tag{6.18}$$

Necessarily, $\{S_n - n^2\gamma_{s,n}\}$ is tight for each $s > 0$.

Unfortunately, Feller's proof of the sufficiency of (6.17) and (6.18) is incorrect. Nevertheless, the theorem is true, as we now shall verify. Our notation differs from Feller's; in particular, our row sums S_n are composed of n^2 summands.

PROOF. Only the sufficiency remains to be established. Choose an $s > 0$ and let $V_n = \Sigma_{i,j}[\beta_s(X_{i,j}(n)) - \gamma_{s,n}]$ and $Z_n = \Sigma_{i,j}\, \rho_s(X_{i,j}(n))$, with the result that $S_n - n^2\gamma_{s,n} = V_n + Z_n$. It suffices to show that both $\{V_n\}$ and $\{Z_n\}$ are tight.

That $\{V_n\}$ is tight follows from Chebyshev's inequality and (6.18). To verify that $\{Z_n\}$ is tight, let $\varepsilon > 0$ be specified and choose b to satisfy (6.17). By our hypothesis and the dominated convergence theorem, $\gamma_{b,n}$ must converge to zero; so, we may as well assume that $|\gamma_{b,n}| < s/2$. In addition to (6.17), we show that $n^2 p_{s,n}$ remains uniformly bounded for each s and, in particular for our choice of s. Only when s is less than b is there a problem and, in this case,

$$\{|X_{1,1}(n)| > s\} \subset \left\{|\beta_b(X_{1,1}(n)) - \gamma_{b,n}| > \frac{s}{2}\right\} \cup \{|X_{1,1}(n)| > b\}. \tag{6.19}$$

Thus,

$$\begin{aligned} n^2 p_{s,n} &\leqq n^2 \mathscr{P}\left\{|\beta_b(X_{1,1}(n)) - \gamma_{b,n}| > \frac{s}{2}\right\} + \varepsilon \\ &\leqq \left(\frac{4}{s^2}\right)\sigma_n^2(b) + \varepsilon \leqq \left(\frac{4}{s^2}\right) M(b) + \varepsilon. \end{aligned} \tag{6.20}$$

If $N_{n,s}$ is the number of $X_{i,j}(n), i, j = 1, \cdots, n$, which exceed s in absolute value, then $N_{n,s}$ has a binomial distribution with a uniformly bounded expectation. We may then choose an integer m for which $\mathscr{P}\{N_{n,s} > m\} < \varepsilon/2$ for all sufficiently large n. For that m, we have

$$\mathscr{P}\{|Z_n| > \alpha\} \leqq \mathscr{P}\left\{\max_{i,j} |X_{i,j}(n)| > \frac{\alpha}{m}\right\} + \mathscr{P}\{N_{n,s} > m\}, \tag{6.21}$$

since, if there are at most m random variables among the $X_{i,j}(n)$ which exceed s in absolute value, but all of the $|X_{i,j}(n)|$ are less than α/m,

$$|Z_n| \leqq \sum_{i,j} |\rho_s(X_{i,j}(n))| \leqq m \max_{i,j} |X_{i,j}(n)| \leqq \alpha. \tag{6.22}$$

Now,

$$\begin{aligned} \mathscr{P}\left\{\max_{i,j} |X_{i,j}(n)| \geqq \frac{\alpha}{m}\right\} &= 1 - \left[1 - \mathscr{P}\left\{|X_{1,1}(n)| \geqq \frac{\alpha}{m}\right\}\right]^{n^2} \\ &= 1 - (1 - p_{\alpha/m,n})^{n^2}, \end{aligned} \tag{6.23}$$

and this term by (6.17) may be made uniformly small for a sufficiently large choice of α. If α is chosen to assure $\mathscr{P}\{\max_{i,j} |X_{i,j}(n)| \geqq \alpha/m\} \leqq \varepsilon/2$ for all n, then (6.21) implies $\mathscr{P}\{|Z_n| > \alpha\} \leqq \varepsilon$. That is, $\{Z_n\}$ is tight. *Q.E.D.*

PROOF OF THEOREM 6.2. If the sequence of row sums $\{S_n\}$ is weakly convergent, then $X_{1,1}(n) \underset{P}{\to} 0$ (see [10], p. 300), and Lemma 6.3 implies that, for each $s > 0$, $\sup_n n^2\gamma_{s,n} < \infty$. Since the dominated convergence theorem implies $\lim_n \gamma_{s,n} = 0$,

$$\lim_n\, [\alpha_n^2(s) - \sigma_n^2(s)] = \lim_n n^2\gamma_{s,n}^2 = 0. \tag{6.24}$$

Now, $\{\alpha_n^2\}$ is a sequence of nondecreasing functions which, by (6.18), are uniformly bounded on (0, 1). It follows by Helley's theorem that there is a subsequence $\{n'\} \subset \{n\}$ along which $\alpha_{n'}^2(s)$ converges for each $s \in (0, 1)$ to a bounded nondecreasing function $\sigma^2(s)$. Necessarily, $\lim_{n'} \sigma_{n'}^2(s) = \sigma^2(s)$. In proving relative compactness, we may begin this analysis with an arbitrary subsequence and refer to a further subsequence if necessary, so we may as well assume that $n' = n$.

Suppose, first, that $\lim_{s \to 0} \sigma^2(s) = 0$. With $\varepsilon > 0$ specified, choose $s > 0$ for which $\sigma_n^2(s) < \varepsilon^3/4^4$ for all sufficiently large n and set

$$
\begin{aligned}
V_n(\mathbf{t}) &= \sum_{i,j} \{[\beta_s(X_{i,j}(n)) - \gamma_{s,n}] : (i, j) \leqq_{\mathrm{NE}} [n\mathbf{t}]\}, \\
X_n(\mathbf{t}) &= \sum_{i,j} \{\rho_s(X_{i,j}(n)) : (i, j) \leqq_{\mathrm{NE}} [n\mathbf{t}]\},
\end{aligned} \tag{6.25}
$$

and $G_n(\mathbf{t}) = [nt_1][nt_2]\, \gamma_{s,n}$, with the result that $Y_n(\mathbf{t}) = V_n(\mathbf{t}) + X_n(\mathbf{t}) + G_n(\mathbf{t})$. Since $\sup_n n^2 \gamma_{s,n} < \infty$, we may always refer to a subsequence along which G_n converges to a uniformly continuous function. As in the proof of Lemma 6.1, it suffices to show that $V_n + X_n$ satisfies (5.15) for all $\eta > 0$.

By an inequality of Wichura [34] for multidimensional partial sum processes, analogous to one of Doob ([5], p. 317) for the submartingales, we have

$$
\begin{aligned}
\mathscr{P}\left\{\sup_{\mathbf{t}} |V_n(\mathbf{t})| \geqq \frac{\varepsilon}{2}\right\} &\leqq \frac{4E[\sup_{\mathbf{t}} |V_n(\mathbf{t})|]^2}{\varepsilon^2} \\
&\leqq \frac{4^3 \sigma_n^2(s)}{\varepsilon^2}.
\end{aligned} \tag{6.26}
$$

By our choice of s, this last term is less than $\varepsilon/2$ for all sufficiently large n. Thus, the V_n process is, with large probability, uniformly small.

The behavior of the X_n process is a different matter. A sample path of this process may change its value only at points $\mathbf{t} = (i/n, j/n)$ and does so at such a $\mathbf{t}$ when and only when $\rho_s(X_{i,j}(n)) \neq 0$. For a rectangle $R(\mathbf{s}, \mathbf{t}) \in \mathscr{A}$ having coordinates $\mathbf{s} \leqq_{\mathrm{NE}} \mathbf{t}$ for its southwest and northeast corners, let $N_n(R(\mathbf{s}, \mathbf{t}))$ be the number of jumps of the X_n process in $R(\mathbf{s}, \mathbf{t})$; $N_n(R(\mathbf{s}, \mathbf{t}))$ is the number of nonzero members among the $\{\rho_s(X_{i,j}(n))\}$ with $(i, j) \leqq_{\mathrm{SW}} [n\mathbf{s}]$ and $(i, j) \leqq_{\mathrm{NE}} [n\mathbf{t}]$. Thus, $N_n(R(\mathbf{s}, \mathbf{t}))$ has a binomial distribution with sample size parameter

$$
([nt_1] - [ns_1])([nt_2] - [ns_2]) = n^2 \mu(R(\mathbf{s}, \mathbf{t})) + o(n^2) \tag{6.27}
$$

and probability of success parameter $p_{s,n}$.

Our proof of Lemma 6.3 shows that $n^2 p_{s,n}$ is uniformly bounded for each s and, in particular, for our choice of s. Since we may refer to a subsequence if necessary, we assume that $n^2 p_{s,n}$ converges to some number $\lambda \geqq 0$. If $\lambda > 0$, then $N_n(R(\mathbf{s}, \mathbf{t}))$ has, in the limit, a Poisson distribution with mean $\lambda \mu(R(\mathbf{s}, \mathbf{t}))$. Disjoint rectangles $R(\mathbf{s}_i, \mathbf{t}_i)$ give rise to independent random variables $N_n(R(\mathbf{s}_i, \mathbf{t}_i))$; thus, by our remarks on stochastic point processes, $N_n(R(\mathbf{0}, \cdot))$ converges weakly to a Poisson process. Moreover, in our remarks we have shown to each

$\eta > 0$ there is a k for which

$$\limsup_n \mathscr{P}\{f^{(0)}_{N_n(R(0,\cdot))}(\eta, \Delta_k) = 0\} = 1. \tag{6.28}$$

Therefore, with that same k,

$$\limsup_n \mathscr{P}\{f^{(0)}_{X_n}(\eta, \Delta_k) = 0\} = 1. \tag{6.29}$$

If $\lambda = 0$, then the X_n process is uniformly small with large probability. Indeed, the probability that it is identically zero is $(1 - p_{s,n})^{n^2}$ which converges to one. In this case, (6.29) holds as well.

Condition (5.15) of Theorem 5.5 now follows from the relation

$$\left\{\sup_{\mathbf{t}} |V_n(\mathbf{t})| < \frac{\varepsilon}{2}\right\} \cap \{f^{(0)}_{X_n}(\eta, \Delta_k) = 0\} \subset \{f^{(0)}_{V_n + X_n}(\eta, \Delta_k) < \varepsilon\}, \tag{6.30}$$

which implies

$$\mathscr{P}\{f^{(0)}_{V_n + X_n}(\eta, \Delta_k) \geqq \varepsilon\} \leqq \mathscr{P}\left\{\sup_{\mathbf{t}} |V_n(\mathbf{t})| \geqq \frac{\varepsilon}{2}\right\} + \mathscr{P}\{f^{(0)}_{X_n}(\eta, \Delta_k) > 0\}. \tag{6.31}$$

Of these last two terms, the first is less than $\varepsilon/2$ for all sufficiently large n by our choice of s, and (6.29) implies that so is the second.

Now we consider the case when $\lim_{s \to 0} \sigma^2(s) = \sigma^2 > 0$, which corresponds to the limit distribution having a normal component. Choose a sequence $\{s_n\}$ of points in (0, 1) converging to zero from above and along which $\lim_n \sigma_n^2(s_n) = \sigma^2$. Necessarily, $\lim_n \alpha_n^2(s_n) = \sigma^2$, since $\sigma_n^2(s_n) \leqq \alpha_n^2(s_n)$ and, for each $s > 0$,

$$\limsup_n \alpha_n^2(s_n) \leqq \lim_n \alpha_n^2(s) = \sigma^2(s). \tag{6.32}$$

Let

$$W_n(\mathbf{t}) = \sum_{i,j} \{[\beta_{s_n}(X_{i,j}(n)) - \gamma_{s_n,n}] \colon (i,j) \leqq_{\mathrm{NE}} [n\mathbf{t}]\} \tag{6.33}$$

and

$$U_n(\mathbf{t}) = \sum_{i,j} \{\rho_{s_n}(X_{i,j}(n)) + \gamma_{s_n,n} \colon (i,j) \leqq_{\mathrm{NE}} [n\mathbf{t}]\} \tag{6.34}$$

with the result that $Y_n(\mathbf{t}) = W_n(\mathbf{t}) + U_n(\mathbf{t})$.

The random variables $\beta_{s_n}(X_{i,j}(n)) - \gamma_{s_n,n}$ satisfy Lindeberg's condition for convergence of their row sums to a normal law: for each $\delta > 0$,

$$\frac{n^2}{\sigma_n^2(s_n)} \int I\{|\beta_{s_n}(X_{1,1}(n)) - \gamma_{s_n,n}| > \delta\sigma_n(s_n)\}[\beta_{s_n}(X_{1,1}(n)) - \gamma_{s_n,n}]^2 \, d\mathscr{P}, \tag{6.35}$$

tends to zero, since $\sigma_n(s_n)$ converges to $\sigma > 0$ and both s_n and $\gamma_{s_n,n}$ converge to zero. The analogue of Donsker's theorem implies that $[1/\sigma_n(s_n)]W_n$ converges weakly to Wiener measure. In particular, W_n converges weakly to a process that is uniformly continuous with probability one.

Since $S_n = Y_n(1, 1)$ and $W_n(1, 1)$ are both weakly convergent, $\{\rho_{s_n}[X_{i,j}(n)] + \gamma_{s_n,n}\}$ is a triangular array with row sums $U_n(1, 1)$ which are tight. Moreover, for each $s > 0$ and all sufficiently large n,

$$\begin{aligned}(6.36)\qquad & n^2E\{\beta_s[\rho_{s_n}(X_{1,1}(n)) + \gamma_{s_n,n}]\}^2 \\ & \leqq n^2E\{\beta_{2s}[\rho_{s_n}(X_{1,1}(n))] + \gamma_{s_n,n}\}^2 \\ & = n^2E\{\beta_{2s}(X_{1,1}(n)) - [\beta_{s_n}(X_{1,1}(n)) - \gamma_{s_n,n}]\}^2 \\ & = \alpha_n^2(2s) + \sigma_n^2(s_n) - 2\alpha_n^2(s_n) + 2n^2\gamma_{2s,n}\gamma_{s_n,n}.\end{aligned}$$

Taking limits first as n becomes large and then as s converges to zero, we see that this triangular array satisfies the case we have handled previously. Therefore, U_n satisfies (5.15) for each $\eta > 0$ and $\varepsilon > 0$. It follows, as in Lemma 6.1, that so does Y_n.

It remains to show that Y_n has weakly convergent finite dimensional distributions in order to apply Theorem 5.5. This is a consequence of the weak convergence of S_n, the fact that Y_n has independent increments, and the fact that, if $R(\mathbf{s}, \mathbf{t})$ is a rectangle in $\mathscr{A}$, then

$$(6.37)\qquad \lim_n \phi_{H_{Y_n}(R(\mathbf{s},\mathbf{t}))}(u) = \lim_n \big(\phi_n(u)\big)^{n^2\mu(R(\mathbf{s},\mathbf{t}))+o(n^2)} = \big(f(u)\big)^{\mu(R(\mathbf{s},\mathbf{t}))}.$$

To determine the finite dimensional distributions of the limiting process Y, let $\mathbf{s} \leqq_{\mathrm{NE}} \mathbf{t}$ be points in T and choose rectangles $R(\mathbf{s}_k, \mathbf{t}_k)$ whose corners approach those of $R(\mathbf{s}, \mathbf{t})$ from the northeast strictly, in the sense that each corner of $R(\mathbf{s}, \mathbf{t})$ lies southwest of the respective corner of $R(\mathbf{s}_k, \mathbf{t}_k)$. Then,

$$(6.38)\qquad \phi_{H_Y(R(\mathbf{s},\mathbf{t}))}(u) = \lim_k \lim_n \phi_{H_{Y_n}(R(\mathbf{s}_k,\mathbf{t}_k))}(u) = \big(f(u)\big)^{\mu(R(\mathbf{s},\mathbf{t}))}.$$

Thus, since each Y_n has independent increments, so must Y. It follows, that Y has stationary increments as well.

To complete the proof of this theorem, we note that, for A a union of disjoint rectangles $R(\mathbf{s}_i, \mathbf{t}_i)$ in $\mathscr{A}$, $i = 1, \cdots, k$,

$$(6.39)\qquad \phi_{H_Y(A)}(u) = \prod_{i=1}^k \phi_{H_Y(R(\mathbf{s}_i,\mathbf{t}_i))}(u) = \big(f(u)\big)^{\Sigma\mu(R(\mathbf{s}_i,\mathbf{t}_i))} = \big(f(u)\big)^{\mu(A)}.$$

Q.E.D.

A process Y in $D(T)$ is *continuous in probability* if each point in T is a stochastic continuity point of Y; that is, $S_Y = T$. The limit process of Theorem 6.2 enjoys this property. It suffices to show, since Y is with probability one a continuous from above lamp function, that, whenever $\{\mathbf{t}_k\}$ is a monotone path approaching a point $t \in T$ with respect to one of the four partial orders (a), (b), (c), (d) then $Y(\mathbf{t}_k)$ converges in probability to $Y(\mathbf{t})$. We handle the case when $\{\mathbf{t}_k\} = \{(t_1(k), t_2(k))\}$ is a monotone path approaching $\mathbf{t}$ from the southeast. Let $R_1(k) = R((0, t_2(k)), \mathbf{t})$ and $R_2(k) = R((t_1, 0), \mathbf{t}_k)$, so that $\lim_k \mu(R_i(k)) = 0$, $i = 1, 2$. Then

$$(6.40)\qquad \phi_{Y(\mathbf{t})-Y(\mathbf{t}_k)}(u) = \phi_{H_Y(R_1(k))}(u)\,\phi_{H_Y(R_2(k))}(-u),$$

which, by (6.14), converges to one and implies $Y(\mathbf{t}) - Y(\mathbf{t}_k)$ converges in probability to zero.

Since each infinitely divisible random variable is the weak limit of the row sums of a triangular array properly defined, Theorem 6.2 implies the existence of a stochastic process Y in $D(T)$ continuous in probability and with stationary and independent increments for which (6.14) holds for each $A \in \mathscr{A}$, when $f(u)$ is the characteristic function of the infinitely divisible law.

The purpose of $f(u)$ is to identify the limit. In $D[0, 1]$, there is a natural way to identify the limiting process: we specify that it vanishes at zero and that its value at one is a random variable whose distribution is the limiting law of the row sums of a triangular array. Theorem 6.2 with n random variables in each row of the triangular array carries over to $D[0, 1]$ by our previous remarks on the relation of $D[0, 1)$ and $D[0, 1]$: the sequence of partial sum processes $Y_n(t) = \Sigma_{i=1}^{[nt]} X_i(n)$ in $D[0, 1]$ is relatively compact and the weak convergence of their finite dimensional distributions for points in $[0, 1]$ determines their limit. Since the value at one of a stochastic process in $D[0, 1]$ with stationary and independent increments is evidently an infinitely divisible law, we have as a corollary the Khintchine theorem: if the row sums of a triangular array of random variables have a limiting distribution, then it is infinitely divisible.

◇ ◇ ◇ ◇ ◇

This work was initiated as a thesis presented to the Department of Statistics at the University of Chicago. The author is grateful for the counsel of Mr. Patrick Billingsley, his advisor, and Mr. Michael Wichura and for the financial assistance provided by the Warner-Lambert Research Institute and the Shell Companies Foundation.

Added in proof. While this manuscript was in press, the author became aware of an article by N. N. Chentsov: "Limit theorems for some classes of random functions," *Vsesoi͡uznoe soveschaine po teorii veroi͡atnosteĭ i matematicheskoĭ statistike* (Proc. All-Union Conf. Theory Prob. Math. Stat.) Erevan, 19–25 September 1958, Izdat. Akad. Nauk Armianskoĭ, SSR, Erevan, 1960, pp. 280–285. An English translation appears in *Selected Translations in Mathematical Statistics and Probability*, Vol. 9 (1971), pp. 37–42, in which Chentsov has extended his fluctuation inequalities to processes of several variables. Although the function space and the mode of convergence he briefly outlines differs from ours, his hypothesis, similar to one developed by Bickel and Wichura [2], offers sufficient conditions for tightness of a sequence of random elements of our space.

REFERENCES

[1] P. J. Bickel, "A distribution free version of the Smirnov two sample test in the p-variate case," *Ann. Math. Statist.*, Vol. 40 (1969), pp. 1–23.

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

[3] P. Billingsley, *Convergence of Probability Measures*, New York, Wiley, 1968.

[4] J. R. BLUM, J. KIEFER, and M. ROSENBLATT, "Distribution free tests of independence based on the sample distribution function," *Ann. Math. Statist.*, Vol. 32 (1961), pp. 485–498.

[5] J. L. DOOB, *Stochastic Processes*, New York, Wiley, 1953.

[6] R. M. DUDLEY, "Gaussian processes on several parameters," *Ann. Math. Statist.*, Vol. 36 (1965), pp. 771–788.

[7] ———, "Weak convergence of probabilities on nonseparable metric spaces and empirical measures on Euclidean spaces," *Illinois J. Math.*, Vol. 10 (1966), pp. 109–126.

[8] ———, "Measures on non-separable metric spaces," *Illinois J. Math.*, Vol. 11 (1967), pp. 449–453.

[9] D. DUGUÉ, "Characteristic functions of random variables connected with brownian motion of the von Mises multidimensional ω_n^2," *Multivariate Analysis—II, Proceedings of the Second International Symposium on Multivariate Analysis held at Wright State University*, Dayton, Ohio, *June 17–22, 1968*, (edited by Paruchuri R. Krishnaiah), New York, Academic Press, 1969, pp. 289–301.

[10] W. FELLER, *An Introduction to Probability Theory and Its Applications*, Vol. 2, New York, Wiley, 1966.

[11] W. HOEFFDING, "A non-parametric test of independence," *Ann. Math. Statist.*, Vol. 19 (1948), pp. 546–557.

[12] J. KIEFER, "On large deviations of the empiric distribution function of vector chance variables and a law of the iterated logarithm," *Pacific J. Math.*, Vol. 11 (1961), pp. 649–660.

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

[14] A. N. KOLMOGOROV, "On Skorohod convergence," *Theor. Probability Appl.*, Vol. 1 (1956), pp. 213–222.

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

[16] L. LE CAM, "Convergence in distribution of stochastic processes," *Univ. California Publ. Statist.*, Vol. 2 (1957), 207–236.

[17] ———, "Remarques sur les variables aléatoires dans les espaces vectoriels non séparables," *Le Calcul des Probabilités et ses Applications*, Paris. Centre National de la Recherche Scientifique, 1959, pp. 39–54.

[18] P. LÉVY, "Le mouvement brownien dépendent de plusieurs paramètres," *C. R. Acad. Sci., Paris*, Vol. 220 (1945), pp. 420–422.

[19] ———, *Processus Stochastiques et Mouvement Brownien*, Paris, Gauthier-Villars, 1965 (revised ed.).

[20] G. NEUHAUS, "Zur Theorie der Konvergenz stochastischer Prozesse mit mehrdimensionalem Zeitparameter," thesis, Westfälischen Wilhelms-Universität zu Münster, 1969.

[21] J. L. PAUL, "Sequences of homeomorphisms which converge to homeomorphisms," *Pacific J. Math.*, Vol. 24 (1968), pp. 143–152.

[22] Y. V. PROHOROV, "Convergence of random processes and limit theorems in probability theory," *Theor. Probability Appl.*, Vol. 1 (1956), pp. 157–214.

[23] M. ROSENBLATT, "Limit theorems associated with variants of the von Mises statistic," *Ann. Math. Statist.*, Vol. 23 (1952), pp. 617–623.

[24] W. RUDIN, *Principles of Mathematical Analysis*, New York, McGraw-Hill, 1964 (2nd ed.).

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

[26] A. V. SKOROHOD, "On the limiting transition from a sequence of sums of independent random quantities to a homogeneous random process with independent increments," *Dokl. Akad. Nauk SSSR*, Vol. 104 (1955), pp. 364–367. (In Russian.)

[27] ———, "Limit theorems for stochastic processes," *Theor. Probability Appl.*, Vol. 1 (1956), pp. 261–290.

[28] ———, "Limit theorems for stochastic processes with independent increments," *Theor. Probability Appl.*, Vol. 2 (1957), pp. 138–171.

[29] M. L. Straf, "A general Skorohod space and its application to the weak convergence of stochastic processes with several parameters," thesis, University of Chicago, 1969.

[30] F. Topsøe, "A criterion for weak convergence of measures with an application to convergence of measures on $D[0, 1]$," *Math. Scand.*, Vol. 25 (1969), pp. 97–104.

[31] V. S. Varadarajan, "On a problem in measure-spaces," *Ann. Math. Statist.*, Vol. 29 (1958), pp. 1275–1278.

[32] ———, "A useful convergence theorem," *Sankhyā*, Vol. 20 (1958), pp. 221–222.

[33] M. Wichura, "On the weak convergence of non-Borel probabilities on a metric space," thesis, Columbia University, 1968.

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