AD 665557

WEAK CONVERGENCE OF PROBABILITY MEASURES ON PRODUCT SPACES WITH APPLICATIONS TO SUMS OF RANDOM VECTORS

BY

DONALD L. IGLEHART

TECHNICAL REPORT NO. 109

MARCH 4, 1968

SUPPORTED BY THE ARMY, NAVY, AIR FORCE AND NASA UNDER
CONTRACT Nonr-225(53)(NR-042-002)
WITH THE OFFICE OF NAVAL RESEARCH

DEPARTMENT OF OPERATIONS RESEARCH

AND

DEPARTMENT OF STATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA

WEAK CONVERGENCE OF PROBABILITY MEASURES ON PRODUCT SPACES WITH APPLICATIONS TO SUMS OF RANDOM VECTORS

By

Donald L. Iglehart

TECHNICAL REPORT No. 109

March 4, 1968

Supported by the Army, Navy, Air Force and NASA under

Contract Nonr-225(53) (NR-042-002)

with the Office of Naval Research*

*This research was partiall, supported by the Office of Naval Research under Contract Nonr-401(55) (NR-047-055) at Cornell University.

Gerald J. Lieberman, Project Director

DEPARTMENT OF OPERATIONS RESEARCH

and

DEPARTMENT OF STATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA

NON-TECHNICAL SUMMARY

The central limit theorem deals with the asymptotic distribution of the sum of a large number of scalar-valued random variables. If the number of summands in the sum is large, the distribution of the sum is approximately normal (Gaussian). A functional central limit theorem considers the asymptotic distribution of certain functionals of, say, the first n partial sums of random variables. For example, the maximum or minimum of the first n sums could be considered. Again, if n is large a good approximation for the distribution of the functional would be available. Both the central limit theorem and the functional central limit theorem are available for sequences of independent, identically distributed random variables as well as certain dependent sequences. Finally, results of this sort are also available for sums of a random number of random variables.

This paper extends the functional central limit theorem to the case of the sums of vector-valued random variables. For example, an approximation is given for the distribution of the length of the vector arising from the sum of the first n random vectors. A rich class of other functionals can also be handled along with the consideration of sums of a random number of random vectors.

WEAK CONVERGENCE OF PROBABILITY MEASURES ON PRODUCT SPACES WITH APPLICATIONS TO SUMS OF RANDOM VECTORS[1/]

By

Donald L. Iglehart
Stanford University

1. Introduction

Billingsley [1] has given an excellent treatment of the subject of weak convergence of probability measures. The principal application he considers is functional central limit theorems for sums of random variables. Our objective in this paper is to point out a simple observation which allows one to obtain functional central limit theorems for sums of random vectors.

The classical central limit theorem is concerned with the asymptotic behavior of the distribution of $S_n = X_1 + \cdots + X_n$ $(S_0=0)$, where the X_i's are random variables satisfying certain conditions. A functional central limit theorem, on the other hand, treats the asymptotic behavior of the distribution of $f(S_i : 0 \leq i \leq n)$ for a certain class of functionals f. The phrase functional central limit theorem, which we shall use, has been proposed in [1] to replace the less suggestive terminology of an invariance principle.

Historically the development of functional central limit theorems began with Erdös and Kac [5], [6] and was extended and generalized by

---

1/ This work was supported by Office of Naval Research Contracts Nonr-401(55) at Cornell University and Nonr-225(53) at Stanford University.

Donsker [4], Billingsley [2], [3], Prohorov [9], and Skorohod [10], [11]. This development has taken place in the context of weak convergence of probability measures on metric spaces. This part of the theory was developed primarily by Prohorov [9] and Skorohod [10].

The metric spaces of concern in this paper will be product spaces. If $C[0,1]$ is the space of continuous functions on $[0,1]$ with the uniform metric, then we let $C^k = C[0,1] \times \cdots \times C[0,1]$ be the product of k copies of $C[0,1]$ with the product topology. If $D[0,1]$ is the space of right-continuous functions on $[0,1]$ having left limits with the Skorohod topology (to be defined later), then $D^k = D[0,1] \times \cdots \times D[0,1]$ is the product of k copies of $D[0,1]$ with the product topology. The spaces C^k and D^k shall be of special interest since they are the natural spaces in which to consider probability measures induced by random vectors.

The principal result of this paper is to obtain necessary and sufficient conditions for a sequence of probability measures on C^k or D^k to converge to a probability measure. These conditions are then applied to obtain functional central limit theorems for sums of random vectors in a variety of situations. In particular, we consider sums of independent, identically distributed random vectors as well as sums of stationary, φ-mixing random vectors. These results are then extended to sums of a random number of random vectors. Application is also mentioned to the k-dimensional random walk induced by the multi-urn Ehrenfest model.

This paper is organized into the following sections: Section 2 deals with various preliminaries on weak convergence; Section 3 contains a theorem giving necessary and sufficient conditions for a sequence of probability measures on C^k to convergence weakly; Section 4 gives the corresponding conditions for measures on D^k; Section 5 applies these results to sums of random vectors; and Section 6 applies the results to the multi-urn Ehrenfest model.

2. Preliminaries on Weak Convergence

Let S be a metric space and $\mathcal{S}$, the class of Borel sets, be the σ-field generated by the open sets of S. If P_n and P are probability measures on $\mathcal{S}$ which satisfy

$$\lim_{n \to \infty} \int_S f dP_n = \int_S f dP$$

for every bounded, continuous, real-valued function f on S, we shall say that P_n converges weakly to P as $n \to \infty$ and write $P_n \Longrightarrow P$. In the case where $S = R^k$, k-dimensional Euclidean space, weak convergence is equivalent to ordinary weak convergence of the distribution functions associated with P_n to that associated with P. However, for the function spaces we shall consider weak convergence is a deeper concept.

The notion of tightness introduced by Prohorov [9] plays a key role in the weak convergence of probability measures. A family Π of probability measures on the metric space S is said to be tight if for every $\epsilon > 0$ there exists a compact set K_ϵ such that $P(k_\epsilon) > 1 - \epsilon$ for all P in Π. The main result which makes this a useful concept is a theorem in [9]. This theorem requires the notion of Π being relatively compact. A family Π is said to be relatively compact if every sequence of elements of Π contains a convergent subsequence (the limit need not belong to Π). The theorem is as follows.

Theorem 1 (Prohorov [9]). (i) *If* Π *is tight, then it is relatively compact.* (ii) *If* S *is a complete separable metric space and* Π *is relatively compact, then it is tight.*

It is convenient to list here two other results and a definition which we shall need. The first is an analog for sequences of measures of a property of sequences of numbers; cf. [1], Theorem 2.3.

Theorem 2. *We have* $P_n \Rightarrow P$ *if and only if each subsequence* $\{P_{n'}\}$ *contains a further subsequence* $\{P_{n''}\}$ *such that* $P_{n''} \Rightarrow P$.

Let h be a measurable mapping of S into another metric space S' (with σ-field $\mathcal{S}'$ of Borel sets). Each probability measure P on $(S,\mathcal{S})$ induces on $(S',\mathcal{S}')$ a unique probability measure $Ph^{-1}(A) = P(h^{-1}A)$ for $A \in \mathcal{S}'$. Let D_h be the set of discontinuities of h. The next theorem is an analog of the Mann-Wald theorem for the Euclidean case; cf. [1] Theorem 5.1.

Theorem 3. *If* $P_n \Rightarrow P$ *and* $P(D_h) = 0$, *then* $P_n h^{-1} \Rightarrow Ph^{-1}$.

For a metric space S (with Borel sets $\mathcal{S}$) a class of sets $\mathcal{U} \subset \mathcal{S}$ is said to be a *determining class* if for any two probability measures P and Q on $\mathcal{S}$, the fact that $P(A) = Q(A)$ for all $A \in \mathcal{U}$ implies that $P \equiv Q$. Of course, if $\mathcal{U} \subset \mathcal{S}$ is a field and $\mathcal{B}(\mathcal{U})$, the σ-field generated by $\mathcal{U}$, equals $\mathcal{S}$, then $\mathcal{U}$ is a determining class by virtue of the Carathéodory extension theorem.

We shall be concerned with separable product spaces $S = S_1 \times \cdots \times S_k$

(endowed with the product topology) and product σ-field $\mathcal{S}_1 \times \cdots \times \mathcal{S}_k$, where each S_i is a separable metric space with an associated class of Borel sets $\mathcal{S}_i$. Let $\mathcal{S}$ be the class of Borel sets generated by the open sets of S. Since S is separable, it is known that $\mathcal{S} = \mathcal{S}_1 \times \cdots \mathcal{S}_k$; cf. [1], p. 468. For a given probability measure P on $\mathcal{S}$ we define the marginal measures P^i $(i = 1,\ldots,k)$ by $P^i(A) = P(S_1 \times \cdots \times S_{i-1} \times A \times S_{i+1} \times \cdots \times S_k)$ for $A \epsilon \mathcal{S}_i$. For a family Π of probability of probability measures on S the notion of tightness can be stated in terms of the tightness of the families Π^i $(i = 1,\ldots,k)$ of marginal measures. This result which is elementary was stated for $k = 2$ as problem 5, p. 79 of [1]. For the sake of completeness we indicate a proof.

Lemma 1. Let Π be a family of probability measures on $(S,\mathcal{S})$ and Π^i $(i = 1,\ldots,k)$ be the corresponding families of marginal measures on $(S_i,\mathcal{S}_i)$ $(i = 1,\ldots,k)$. Then Π is tight on $(S,\mathcal{S})$ if and only if each Π^i is tight on $(S_i,\mathcal{S}_i)$.

Proof. (Sufficiency). Let each Π^i be tight on $(S_i,\mathcal{S}_i)$. Then for each $\epsilon > 0$ there exists a compact set K_ϵ^i such that $P^i(K_\epsilon^i) > 1 - \epsilon/k$ for every $P^i \in \Pi^i$. Let $K_\epsilon = K_\epsilon^1 \times \cdots \times K_\epsilon^k$. Since K_ϵ is the product of compact sets, it is compact in the product topology by Tychonoff's theorem. Furthermore[2] $K_\epsilon^c \subset \bigcup_{i=1}^{k} (S_1 \times \cdots \times S_{i-1} \times (K_\epsilon^i)^c \times S_{i+1} \times \cdots \times S_k)$ and hence $P(K_\epsilon^c) \leq \sum_{i=1}^{k} P^i(K_\epsilon^i) \leq \epsilon$. Hence $P(K_\epsilon) > 1 - \epsilon$ for every $P \epsilon \Pi$

[2] The complement of a set A is denoted A^c.

and thus Π is tight.

(Necessity). Suppose Π is tight. Then for each $\epsilon > 0$ there exists a compact set K_ϵ such that $P(K_\epsilon) > 1 - \epsilon$ for every $P \epsilon \Pi$. For any point $\underset{\sim}{x} = (x^1, \ldots, x^k) \epsilon S$ let $\pi^i(\underset{\sim}{x}) = x^i$ be the projection function. Since each π^i is continuous in the product topology, $\pi^i(K_\epsilon)$ which we shall denote K_ϵ^i, is compact. Hence $P^i(K_\epsilon^i) =$ $P(S_1 \times \cdots \times S_{i-1} \times K_\epsilon^i \times S_{i+1} \times \cdots \times S_k) \geq P(K_\epsilon) > 1 - \epsilon$ for each $i = 1, \ldots, k$ and thus each Π^i is tight.

The next lemma is one that is useful for establishing determining classes for product spaces. As a result in measure theory it must be well-known, however, we could not find a reference in which it is stated in this manner. Before stating the lemma we introduce some notation. Let $(S_1, \mathcal{S}_1)$ and $(S_2, \mathcal{S}_2)$ be pairs of metric spaces and associated classes of Borel sets. For families $\mathcal{U}_1 \subset \mathcal{S}_1$ and $\mathcal{U}_2 \subset \mathcal{S}_2$, the family of sets $\mathcal{U}_1 \otimes \mathcal{U}_2 \equiv \{A_1 \times A_2 : A_i \epsilon \mathcal{S}_i\}$ consists of rectangles in $S_1 \times S_2 \equiv S$. By $(\mathcal{U}_1 \otimes \mathcal{U}_2)^*$ we shall mean the family of sets in S formed by taking finite sums of sets in $\mathcal{U}_1 \otimes \mathcal{U}_2$. With this notation the product σ-field $\mathcal{S}_1 \times \mathcal{S}_2$ (equal to $\mathcal{S}$, the Borel sets of S, when S is separable) can be expressed as $\mathcal{B}\{(\mathcal{S}_1 \otimes \mathcal{S}_2)^*\}$, while $(\mathcal{S}_1 \otimes \mathcal{S}_2)^*$ is a field.

Lemma 2. Let $\mathcal{U}_i \subset \mathcal{S}_i$ $(i = 1, \ldots, k)$ be fields which generate $\mathcal{S}_i$ $(\mathcal{B}(\mathcal{U}_i) = \mathcal{S}_i)$. Then if $S = S_1 \times \cdots \times S_k$ is separable, the family of sets $(\mathcal{U}_1 \otimes \cdots \otimes \mathcal{U}_k)^* \equiv \mathcal{U}$ is a field which generates $\mathcal{S} = \mathcal{S}_1 \times \cdots \times \mathcal{S}_k$.

Proof. The family $\mathcal{U}$ is simply the product field and hence a field; cf. Loève [8], p. 61. Since $\mathcal{U} \subset (\mathcal{S}_1 \otimes \cdots \otimes \mathcal{S}_k)^*$ $\mathcal{B}(\mathcal{U}) \subset \mathcal{S}$. To show that $\mathcal{S} \subset \mathcal{B}(\mathcal{U})$ it will suffice to show that $\mathcal{S}_1 \otimes \cdots \otimes \mathcal{S}_k \subset \mathcal{B}(\mathcal{U}_1 \otimes \cdots \otimes \mathcal{U}_k)$, since $\mathcal{B}(\mathcal{S}_1 \otimes \cdots \otimes \mathcal{S}_k) = \mathcal{S}$ and $\mathcal{B}(\mathcal{U}_1 \otimes \cdots \otimes \mathcal{U}_k) = \mathcal{B}(\mathcal{U})$. Let $A_i \epsilon \mathcal{S}_i$, then the rectangles $A_i' \equiv S_1 \times \cdots \times S_{i-1} \times A_i \times S_{i+1} \times \cdots \times S_k$ all belong to $\mathcal{S}_1 \otimes \cdots \otimes \mathcal{S}_k$. Since $\mathcal{S}_i = \mathcal{B}(\mathcal{U})$, $A_i' \epsilon$ $\mathcal{B}(S_1 \otimes \cdots \otimes S_{i-1} \otimes \mathcal{U}_i \otimes S_{i+1} \otimes \cdots \otimes S_k) \subset \mathcal{B}(\mathcal{U}_1 \otimes \cdots \otimes \mathcal{U}_k)$. But $A_1 \times \cdots \times A_k = \bigcap_{i=1}^{k} A_i' \in \mathcal{B}(\mathcal{U}_1 \otimes \cdots \otimes \mathcal{U}_k)$, which shows that $\mathcal{S}_1 \otimes \cdots \otimes \mathcal{S}_k \subset \mathcal{B}(\mathcal{U}_1 \otimes \cdots \otimes \mathcal{U}_k)$ and completes the proof.

In the next two sections we shall apply these results to the product spaces C^k and D^k.

3. Weak Convergence of Probability Measures on C^k

Let C be the space of all continuous real-valued functions on the closed unit interval $[0,1]$ with the metric of uniform convergence, $\rho(x,y) = \sup_{0 \leq t \leq 1} |x(t)-y(t)|$, and $\mathcal{C}$ denote the class of Borel sets. Now let C^k be the product of k copies of C, and endow C^k the product topology. We shall assume that the metric on C^k is defined for $\underset{\sim}{x}, \underset{\sim}{y} \epsilon C^k$, as $\rho^k(\underset{\sim}{x},\underset{\sim}{y}) = \max_{1 \leq i \leq k} \{\rho(x^i,y^i)\}$, where $\underset{\sim}{x} = (x^1,\ldots,x^k)$ and $\underset{\sim}{y} = (y^1,\ldots,y^k)$. Since C is a complete separable metric space so is C^k. Hence the class of Borel sets $\mathcal{C}^k$ of C^k equals the product σ-field $\mathcal{C} \times \cdots \times \mathcal{C}$. Furthermore, by Theorem 1 relative compactness of a family of probability measures on $(C^k,\mathcal{C}^k)$ is equivalent to tightness of the family.

For points $0 \leq t_1 < \cdots < t_\ell \leq 1$ let $\pi_{t_1 \cdots t_\ell}$ be the mapping that carries $\underset{\sim}{x} \epsilon C^k$ into $(\underset{\sim}{x}(t_1),\ldots,\underset{\sim}{x}(t_\ell)) \epsilon R^{k \times \ell}$, $k \times \ell$-dimensional Eudlidean space. Define $\mathcal{F}^k$, the class of finite-dimensional sets, to be all sets of the form $\pi^{-1}_{t_1 \ldots t_\ell} A$ for $A \in \mathcal{R}^{k \times \ell}$ (the Borel sets of $R^{k \times \ell}$ under the Euclidean metric), where $\ell \geq 1$ and $0 \leq t_1 < \cdots < t_\ell \leq 1$. Since $\pi_{t_1 \ldots t_\ell}$ is continuous, $\mathfrak{F}^k \subset \mathcal{C}^k$. For any probability measure P on $(C^k, \mathcal{C}^k)$, the measures $P\pi^{-1}_{t_1 \ldots t_\ell}$ on $(R^{k \times \ell}, \mathcal{R}^{k \times \ell})$ are called the finite-dimensional measures of P. The distributions corresponding to these measures are called the finite-dimensional distributions of P. For $k = 1$, it is well-known that $\mathfrak{F}^1$ is a determining class; cf. [1], p. 36. In the next lemma we show that $\mathfrak{F}^k$ is a determining class by applying Lemma 2.

Lemma 3. The class $\mathfrak{F}^k$ is a determining class.

Proof. The proof that $\mathfrak{F}^1$ is a determining class is accomplished by showing that $\mathfrak{F}^1$ is a field and that $\mathcal{B}(\mathfrak{F}^1) = \mathcal{C}$. From Lemma 2 we have that $(\mathfrak{F}^1 \otimes \cdots \otimes \mathfrak{F}^1)^* \equiv \mathfrak{F}$ is a field and that $\mathcal{B}(\mathfrak{F}) = \mathcal{C}^k$. It is easy to show that $\mathfrak{F}^k$ is a field and that $\mathfrak{F} \subset \mathfrak{F}^k$. Therefore $\mathcal{B}(\mathfrak{F}^k) = \mathcal{C}^k$ and thus $\mathfrak{F}^k$ is a determining class by the extension theorem.

The next theorem follows directly from Theorems 1-3 and Lemmas 1 and 3 by the same proof used in the case $k = 1$; cf. [1], Theorem 8.1.

Theorem 4. Let $\{P_n\}$ and P be probability measures on $(C^k, \mathcal{C}^k)$. Then (i) and (ii) are necessary and sufficient conditions for $P_n \Rightarrow P$:

(i) the finite-dimensional distributions of P_n converge weakly to those of P;

(ii) the families of marginal measures $\{P^i_n\}$ on $(C, \mathcal{C})$ are tight for $i = 1, \ldots, k$.

4. Weak Convergence of Probability Measures on D^k

We turn our attention now to the space D, the space of all real-valued functions $x(t)$ on $[0,1]$ that are right-continuous and have left limits:

(i) for $0 \le t < 1$, $x(t+) = \lim_{s \downarrow t} x(s)$ exists and $x(t) = x(t+)$;

(ii) for $0 < t \le 1$, $x(t-) = \lim_{s \uparrow t} x(s)$ exists.

This space is natural for studying stochastic processes having jump discontinuities and is somewhat harder to deal with than the space C. Skorohod [10] has introduced the following topology on D. Let Λ denote the class of strictly increasing, continuous mappings of $[0,1]$ onto itself. For $\lambda \epsilon \Lambda$, $\lambda(0) = 0$ and $\lambda(1) = 1$. The metric $d(x,y)$, for x and y in D, is defined to be the infimum of those positive ϵ for which there exists a $\lambda \epsilon \Lambda$ such that

$$\sup_t |\lambda t - t| \le \epsilon$$

and[3]

$$\sup_t |x(t) - y(\lambda t)| \le \epsilon \quad .$$

A sequence of elements $\{x_n\}$ belonging to D converges to x in the Skorohod topology if and only if there exists functions λ_n in Λ such that $\lim_{n \to \infty} x_n(\lambda_n t) = x(t)$ and $\lim_{n \to \infty} \lambda_n t = \lambda t$, both limits being uniform in $t \epsilon [0,1]$. With this metric D is separable, but not complete.

---

[3] For functions $\lambda \epsilon \Lambda$ we shall write λt for $\lambda(t)$.

Fortunately, there is another metric d_0 which is equivalent to d, in the sense that it generates the same Skorohod topology, but under which D is a complete separable metric space. For $\lambda \epsilon \Lambda$ let

$$\|\lambda\| = \sup_{s \neq t} \left|\log \frac{\lambda t - \lambda s}{t-s}\right|$$

and define $d_0(x,y)$ to be the infimum of those positive ϵ for which there exists a λ in Λ such that

$$\|\lambda\| \leq \epsilon$$

and

$$\sup_t |x(t)-y(\lambda t)| \leq \epsilon \quad .$$

From here on we shall assume that the metric on D is d_0.

Now let D^k the product of k copies of D, with the product topology. We shall assume that the metric D^k is defined as $d_0^k(\underset{\sim}{x},\underset{\sim}{y}) = \max_{1 \leq i \leq k} \{d_0(x^i,y^i)\}$. Again we have D^k a complete separable metric space and $\mathcal{B}^k$, the class of Borel sets of D^k, equal to the product σ-field $\mathcal{B} \times \cdots \mathcal{B}$. The mapping $\pi_{t_1 \ldots t_\ell}$ for $0 \leq t_1 < \cdots < t_\ell \leq 1$ carries $\underset{\sim}{x} \epsilon D^k$ into $(\underset{\sim}{x}(t_1),\ldots,\underset{\sim}{x}(t_\ell))\ R^{k \times \ell}$. These mappings are not everywhere continuous on D^k which complicates the analysis of measures on this space. However, $\pi_{t_1 \ldots t_\ell}$ is measurable and we can define the class $\mathfrak{F}^k$ of finite-dimensional sets in $\mathcal{B}^k$ as was done for $(C^k,\mathcal{C}^k)$; cf. [1], p. 236.

Let T_0 be a subset of $[0,1]$ and define $\mathfrak{F}^k_{T_0}$ to be the family of sets

$$\mathfrak{Z}^k_{T_0} = \{\pi^{-1}_{t_1 \ldots t_\ell} H : H \in \mathcal{R}^{k \times \ell};\ t_1, \ldots, t_\ell \in T_0;\ \ell \geq 1\} \quad .$$

For $k = 1$, $\mathfrak{Z}^1_{T_0}$ is a determining class provided T_0 contains 1 and is dense in $[0,1]$; cf. [1], p. 237. Using Lemma 2, we next show that $\mathfrak{Z}^k_{T_0}$ is a determining class. The proof is the same as that for Lemma 3 and therefore omitted.

Lemma 4. *If T_0 contains 1 and is dense in $[0,1]$, then $\mathfrak{Z}^k_{T_0}$ is a determining class.*

In order to prove a theorem for D^k comparable to Theorem 4 for C^k we define a subset of $[0,1]$, T_P, for every probability measure P on $(D^k, \mathcal{D}^k)$. A point $t \in T_P$ if and only if $P(J_t) = 0$, where

$$J_t = \{\underset{\sim}{x} : \underset{\sim}{x}(t) \neq \underset{\sim}{x}(t-)\} \quad .$$

The set J_t is the set of $\underset{\sim}{x}$'s for which π_t is discontinuous. Using Billingsley's ([1], p. 243) argument one can show that T_P contains 0 and 1 and its complement in $[0,1]$ is at most countable. With this preparation it is easy to prove the analog of Theorem 4; cf. [1], Theorem 15.1.

Theorem 5. *Let $\{P_n\}$ and P be probability measures on $(D^k, \mathcal{D}^k)$. Then* (i) *and* (ii) *are necessary and sufficient conditions for* $P_n \Rightarrow P$:

(i) $P_n \pi^{-1}_{t_1 \ldots t_\ell} \Rightarrow P \pi^{-1}_{t_1 \ldots t_\ell}$ *whenever* $t_1, \ldots, t_\ell \in T_P$;

(ii) <u>the families of marginal measures</u> $\{P_n^i\}$ on $(D,\mathcal{D})$, <u>for</u> $i = 1,\ldots,k$, <u>are tight</u>.

5. Applications to Sums of Random Vectors

Theorems 4 and 5 enable us to extend many of the results in [1] on functional central limit theorems for sums of random variables to the case of random vectors. These extensions follow almost immediately as we shall show in the following examples.

For the applications to be presented here it is convenient to introduce the following terminology used in [1]. Let X be a measurable mapping from a probability space $(\Omega,\mathcal{B},\mathcal{P})$ into a metric space S; measurability of X means $X^{-1}\mathcal{S} \subset \mathcal{B}$. We shall call X a random element of S. If $S = R^1$, we call X a random variable; if $S = R^k$, we call X a random vector; and if $S = C^k$ or D^k, we call X a random function. The <u>distribution</u> of X is the probability measure $P = \mathcal{P}X^{-1}$ on $(S,\mathcal{S})$. We shall say a sequence $\{X_n\}$ of random elements of S <u>converges in distribution</u> to the random element X, and write

$$X_n \Longrightarrow X \quad ,$$

if the distribution P_n of X_n converge weakly to the distribution P of X : $P_n \Longrightarrow P$. While this definition requires that the range S and topology be the same for the random elements $X, X_1, X_2,\ldots,$ the domains $(\Omega,\mathcal{B},\mathcal{P})$ may be different. This terminology does not give us anything new, but rather it simplifies the statement of many results. For convenience later we restate Theorem 3.

Theorem 3'. If X_n and h is a measurable mapping of S into S' satisfying $P(X \in D_h) = 0$, then $h(X_n) \Rightarrow h(X)$.

Let $\xi_1, \xi_2, \ldots$ be a sequence of independent identically distributed random vectors (dimension k) on some probability space $(\Omega, \mathcal{B}, P)$ with mean 0 and covariance matrix Σ, where Σ is positive definite. Define the partial sums $S_0 = 0$, $S_n = \xi_1 + \cdots + \xi_n$ for $n \geq 1$. With this set-up and the help of Theorems 4 and 5 it is easy to prove the vector equivalents of Donsker's [4] theorem in either $(C^k, \mathcal{C}^k)$ or $(D^k, \mathcal{D}^k)$. To this end let W be a random element with values in C^k and with k-dimensional Wiener measure as its distribution. Form the random elements X_n of C^k as follows:

$$(1) \qquad X_n(t,\omega) = \frac{\Sigma^{-\frac{1}{2}}}{\sqrt{n}} S_{[nt]}(\omega) + (nt-[nt]) \frac{\Sigma^{-\frac{1}{2}}}{\sqrt{n}} \xi_{[nt]+1}$$

for $[nt]n^{-1} \leq t < ([nt]+1)n^{-1}$, where $\Sigma^{-\frac{1}{2}}$ is the square root of Σ^{-1}; i.e., $\Sigma^{-1} = (\Sigma^{-\frac{1}{2}})' \Sigma^{-\frac{1}{2}}$. Then the functional central limit theorem becomes

Theorem 6. Let the random vectors $\xi_1, \xi_2, \ldots$ be independent and identically distributed with mean 0 and finite, positive definite covariance matrix Σ. Then the random elements defined by (1) satisfy

$$(2) \qquad X_n \Rightarrow W .$$

Proof. We shall apply the sufficient conditions of Theorem 4. First we show that the finite-dimensional distributions of $\underset{\sim}{X}_n$ converge to those of $\underset{\sim}{W}$. The argument used for this part is essentially that of [1], p. 130. For a single time point t we must show that

$$\underset{\sim}{X}_n(t) \Rightarrow \underset{\sim}{W}(t) \quad .$$

First observe that $|\underset{\sim}{X}_n(t) - \frac{\underset{\sim}{\Sigma}^{-\frac{1}{2}}}{\sqrt{n}} \underset{\sim}{S}_{[nt]}| \leq \frac{1}{\sqrt{n}} |\underset{\sim}{\Sigma}^{-\frac{1}{2}} \xi_{[nt]+1}|$, where for $\underset{\sim}{x} \in R^k$, $|\underset{\sim}{x}| = \{(x^1)^2 + \cdots + (x^k)^2\}^{\frac{1}{2}}$. Hence $|\underset{\sim}{X}_n(s) - \frac{\underset{\sim}{\Sigma}^{-\frac{1}{2}}}{\sqrt{n}} \underset{\sim}{S}_{[nt]}|$ goes to 0 in probability and thus by Theorem 4 of [1] it is sufficient to show

$$\frac{\underset{\sim}{\Sigma}^{-\frac{1}{2}}}{\sqrt{n}} \underset{\sim}{S}_{[nt]} \Rightarrow \underset{\sim}{W}(t) \quad .$$

But this follows from the Lindeberg-Levy central limit theorem and the Cramér-Wold device; [1], p. 93. For two time points s and t with $s < t$ we must show that

$$(\underset{\sim}{X}_n(s), \underset{\sim}{X}_n(t)) \Rightarrow (\underset{\sim}{W}(s), \underset{\sim}{W}(t))$$

which follows by Theorem 3' if we can show that

$$(\underset{\sim}{X}_n(s), \underset{\sim}{X}_n(t) - \underset{\sim}{X}_n(s)) \Rightarrow (\underset{\sim}{W}(s), \underset{\sim}{W}(t) - \underset{\sim}{W}(s)) \quad .$$

Again for similar reasons it suffices to show that

(3) $$\left(\frac{\Sigma^{-\frac{1}{2}}}{\sqrt{n}} S_{[ns]}, \frac{\Sigma^{-\frac{1}{2}}}{\sqrt{n}} S_{[nt]} - \frac{\Sigma^{-\frac{1}{2}}}{\sqrt{n}} S_{[ns]}\right) \Rightarrow (W(s), W(t) - W(s)) \quad .$$

Since the ξ_i's are independent, the components on the left are independent. Hence the above result for one time point and Theorem 3.2 of [1] establishes (3). For three or more time points the same method can be used. This completes the proof of condition (i) of Theorem 4.

To demonstrate condition (ii) we note that of the random element $X_n^i(t,\omega)$ of C has distribution P_n^i, the marginal measure of P_n (the distribution of X_n). Hence P_n^i converging weakly is equivalent to $X_n^i(t,\omega)$ converging in distribution. But $X_n^i \Rightarrow W$, the Wiener measure on $(C,\mathcal{C})$, by Donsker's theorem. Thus (2) follows from Theorem 4.

Consider now the random elements Y_n of D^k defined as

(4) $$Y_n(t,\omega) = \frac{\Sigma^{-\frac{1}{2}}}{\sqrt{n}} S_{[nt]} \quad .$$

Let W^k denote Wiener measure on $(C^k,\mathcal{C}^k)$. To extend W^k to $(D^k,\mathcal{D}^k)$ observe that $C^k \in \mathcal{D}^k$ and that the relative Skorohod topology in C^k coincides with the uniform topology, so hence $A \in \mathcal{D}^k$ implies that $A \cap C^k \in \mathcal{C}^k$. Therefore we can extend W^k to $(D^k,\mathcal{D}^k)$ by letting $W^k(A) = W^k(A \cap C^k)$ for $A \in \mathcal{D}^k$. From now on let W be a random element of D^k with the extended W^k for distribution. Then the same method used in Theorem 6, along with Theorem 5, yields

Theorem 7. Let the random vectors $\xi_1, \xi_2, \ldots$ be independent and identically distributed with mean $\underset{\sim}{0}$ and finite, positive definite covariance matrix $\underset{\sim}{\Sigma}$. Then the random elements defined by (4) satisfy

$$(5) \qquad \underset{\sim}{Y}_n \Longrightarrow \underset{\sim}{W} \ .$$

We now turn our attention to stationary φ-mixing sequences of random vectors. For this application we follow [1], Section 20, in which the functional central limit theorem is developed for the case $k = 1$ (sums of random variables). Let

$$(6) \qquad \ldots, \underset{\sim}{\xi}_{-1}, \underset{\sim}{\xi}_0, \underset{\sim}{\xi}_1, \ldots$$

be a strictly stationary sequence of k-dimensional random vectors defined on a probability space $(\Omega, \mathcal{B}, P)$. For $a \leq b$, define $\mathfrak{M}_a^b$ as the σ-field generated by the random vectors $\underset{\sim}{\xi}_a, \ldots, \underset{\sim}{\xi}_b$; define $\mathfrak{M}_{-\infty}^a$ as the σ-field generated by $\ldots, \underset{\sim}{\xi}_{a-1}, \underset{\sim}{\xi}_a$; and define $\mathfrak{M}_b^\infty$ as the σ-field generated by $\underset{\sim}{\xi}_b, \underset{\sim}{\xi}_{b+1}, \ldots$. For a non-negative function φ defined on the positive integers we shall say that the sequence $\{\underset{\sim}{\xi}_i\}$ is φ-mixing if for each k $(-\infty < k < \infty)$ and each n $(n \geq 1)$, $E_1 \in \mathfrak{M}_{-\infty}^k$ and $E_2 \in \mathfrak{M}_{k+n}^\infty$ implies that

$$(7) \qquad |P(E_1 \cap E_2) - P(E_1)P(E_2)| \leq \varphi(n) P(E_1) \ .$$

We shall be interested in functions φ for which $\varphi(n) \to 0$ as $n \to \infty$ at a particular rate. Thus for large n (7) implies that the future and the past are essentially independent. We mention two examples of φ-mixing sequences which are vector generalizations of those given by Billingsley.

Example 1. The sequence (6) is said to be m-dependent if the vectors $(\xi_1, \ldots, \xi_k)$ and $(\xi_{k+n}, \ldots, \xi_j)$ are independent for $n > m$. An m-dependent sequence is φ-mixing with $\varphi(n) = 0$ for $n > m$. Such an example can be obtained by forming the sequence

$$\xi_n = a_0 \zeta_n + a_1 \zeta_{n-1} + \cdots + a_m \zeta_{n-m} ,$$

where the a_i are constants and the ζ_i are independent, identically distributed random vectors.

Example 2. Let (ζ_n) be a stationary, irreducible, aperiodic, Markov process with finite state space, S. Let f be a mapping from S into R^k and define $\xi_n = f(\zeta_n)$. The sequence $\{\xi_n\}$ is φ-mixing with $\varphi(n) = a\rho^n$ $(a > 0,\ 0 < \rho < 1)$ by the same argument used by Billingsley. If S is infinite and $\{\zeta_n\}$ is a Markov process satisfying Doeblin's condition with one ergodic class and is aperiodic, $\{\xi_n\}$ is also φ-mixing.

As before we define $S_0 = 0$, $S_i = \xi_1 + \cdots + \xi_i$ and let Z_n be the random element of D^k defined as $Z_n(t:\omega) = S_{[nt]}(\omega)/n^{\frac{1}{2}}$. Then the functional central limit theorem for φ-mixing sequence can be stated as follows.

Theorem 8. Let $\{\xi_n\}$ be a strictly stationary φ-mixing sequence with $E[\xi_0] = 0$, $E[\xi_0\, \xi_0']$ finite, and $\sum_{n=1}^{\infty} \varphi_n^{\frac{1}{2}} < \infty$. Then the series

$$\Sigma = E[\xi_0\, \xi_0'] + 2 \sum_{\ell=1}^{\infty} E[\xi_0\, \xi_\ell'] \tag{8}$$

converges absolutely. If Σ is positive definite, then

$$\Sigma^{-\frac{1}{2}} Z_n \Longrightarrow W \quad . \tag{9}$$

Proof. The fact that (8) converges absolutely follows immediately from [1], Lemma 1, p. 319. To prove (9) we apply Theorem 5. First we must show that the finite-dimensional distributions of $\Sigma^{-\frac{1}{2}} Z_n$ converge to those of W. For a single time point we must show that

$$\Sigma^{-\frac{1}{2}} Z_n(t) \Longrightarrow W(t) \quad . \tag{10}$$

For any $s \in R^k$ the sequence of random variables $\{s \cdot \xi_i : i = 0, \pm 1, \ldots\}$ is strictly stationary and φ-mixing. Furthermore, $E[s \cdot \xi_0] = 0$ and the variance of $s \cdot \xi_0$ is finite. If Σ is positive definite, then $E[(s \cdot \xi_0)^2] + 2 \sum_{k=1}^{\infty} E[(s \cdot \xi_0)(s \cdot \xi_k)]$ is positive and finite. Hence an application of the Cramér-Wold device and Theorem 20.1 of [1] completes the proof of (10). For two or more time points we use the method of [1], p. 337-8. We shall illustrate the method by showing that

$$(U_n, V_n) \equiv (\Sigma^{-\frac{1}{2}} Z_n(t),\ \Sigma^{-\frac{1}{2}}(Z_n(1) - Z_n(t)))$$

$$\Longrightarrow (W(t), W(1) - W(t)) \quad .$$

Let $\{p_n\}$ be a sequence of positive integers going to infinity slowly enough that $n^{-1} p_n \to 0$, and let

$$\underset{\sim}{V}_n' = \underset{\sim}{\Sigma}^{-\frac{1}{2}}(\underset{\sim}{Z}_n(1)-\underset{\sim}{Z}_n(t+n^{-1}p_n)) \quad .$$

By stationarity $\underset{\sim}{V}_n - \underset{\sim}{V}_n'$ has the same distribution as $\underset{\sim}{\Sigma}^{-\frac{1}{2}}\underset{\sim}{S}_{p_n}/n^{\frac{1}{2}}$, which in turn converges in probability to $\underset{\sim}{0}$ by an application of the Cramér-Wold device, Chebyshev's inequality, and Lemma 3, p. 323 of [1]. Therefore by Theorem 4.1 of [1] we only need to show that

$$(11) \qquad (\underset{\sim}{U}_n,\underset{\sim}{V}_n') \Longrightarrow (\underset{\sim}{W}(t),\underset{\sim}{W}(1)-\underset{\sim}{W}(t)) \quad .$$

Since the $\underset{\sim}{\xi}_i$'s are φ-mixing, for $H_1, H_2 \in R^k$

$$|P\{\underset{\sim}{U}_n \in H_1, \underset{\sim}{V}_n \in H_2\} - P\{\underset{\sim}{U}_n \in H_1\} P\{\underset{\sim}{V}_n \in H_2| \leq \varphi(p_n) \to 0 \quad .$$

The random vectors $\underset{\sim}{W}(t)$ and $\underset{\sim}{W}(1) - \underset{\sim}{W}(t)$ are independent and hence a simple application of Theorem 3.1 (see also equation (4.15)) of [1] completes the proof of (11). The higher dimensional distributions can be handled in the same manner.

To show condition (ii) of Theorem 5 we must demonstrate that $(\underset{\sim}{\Sigma}^{-\frac{1}{2}}\underset{\sim}{Z}_n)^1$ converges weakly. If we let $\eta_k = \underset{\sim}{\Sigma}^{-\frac{1}{2}}\underset{\sim}{\xi}_k$, then it is easy to check that $E[\underset{\sim}{\eta}_0\underset{\sim}{\eta}_0'] + 2\sum_{\ell=1}^{\infty} E[\underset{\sim}{\eta}_0\underset{\sim}{\eta}_\ell] = I$, the $k \times k$ identity matrix. Furthermore, the sequence $\{\underset{\sim}{\eta}_k\}$ is strictly stationary, φ-mixing, and $E[\underset{\sim}{\eta}_0] = \underset{\sim}{0}$. Therefore, by the functional central limit theorem for φ-mixing sequences of random variables (cf., [1], Theorem 20.2) $(\eta_1^1 + \cdots + \eta_{[nt]}^1)/n^{\frac{1}{2}}$ converges to the one-dimensional Wiener process. But this sum is simply $(\underset{\sim}{\Sigma}^{-\frac{1}{2}}\underset{\sim}{Z}_n)^1$, and hence the proof of the theorem is complete.

As a final application of Theorem 5 to sums of random vectors, we shall mention functional central limit theorems for sums of a random number of random vectors. Again we let $\underset{\sim}{S}_n = \underset{\sim}{\xi}_1 + \cdots + \underset{\sim}{\xi}_n$ and define the random element $\underset{\sim}{X}_n$ in D^k by

$$\underset{\sim}{X}_n(t:\omega) = \frac{\underset{\sim}{\Sigma}^{-\frac{1}{2}}}{\sqrt{n}} \underset{\sim}{S}_{[nt]}(\omega) \quad ,$$

where $\underset{\sim}{\Sigma}$ is an appropriate positive definite matrix. Let $\nu_n(\omega)$ be a positive integer-valued random variable defined on the same probability space as the $\underset{\sim}{\zeta}_n$'s. Define

$$\underset{\sim}{Y}_n(t,\omega) = \frac{\underset{\sim}{\Sigma}^{-\frac{1}{2}}}{\sqrt{\nu_n(\omega)}} \underset{\sim}{S}_{[\nu_n(\omega)t]}(\omega) \quad .$$

Then Theorem 18.1 of [1] can be easily generalized to obtain

Theorem 9. *If* $\nu_n/a_n \xrightarrow{P} \theta$, *where* θ *is a positive constant and the* a_n *are constants going to infinity, then*

$$\underset{\sim}{X}_n \Rightarrow \underset{\sim}{W}$$

implies

$$\underset{\sim}{Y}_n \Rightarrow \underset{\sim}{W} \quad .$$

This result yields functional central limit theorems for random sums of independent, identically distributed random vectors (Theorem 7) and for random sums of φ-mixing sequences of random vectors (Theorem 8). Finally, Theorem 3' results in limit theorems for appropriate functionals of random sums.

While we have not attempted to carry out the details, it seems very likely that one could obtain vector versions of the functional central limit theorem in the case where the summands are in the domain of attraction of a stable law. For $k = 1$ these results have been obtained by Skorohod [11].

6. Application to the Multi-urn Ehrenfest Model

In the multi-urn Ehrenfest model N balls are distributed among $k + 1$ $(k \geq 2)$ urns. If we label the urns $0, 1, \ldots, k,$ then the system is said to be in state $\underset{\sim}{i} = (i_1, \ldots, i_k)$ when there are i_j balls in urn j $(j = 1, \ldots, k)$ and $N - \underset{\sim}{1} \cdot \underset{\sim}{i}$ balls in urn 0. At discrete epochs a ball is chosen at random from one of the $k + 1$ urns; each of the N balls has probability $1/N$ of being selected. The ball chosen is removed from its urn and placed in urn i $(i = 0, 1, \ldots, k)$ with probability p^i, where the p^i's are elements of a given vector $(p^0, \underset{\sim}{p})$, satisfying $p^i > 0$ and $\sum_{i=0}^{k} p^i = 1$. We shall let $\underset{\sim}{X}_N(\ell)$ denote the state of the system after the ℓ^{th} such rearrangement of balls. Define

$$\underset{\sim}{Y}_N([nt]) = (\underset{\sim}{X}_N([nt]) - N\underset{\sim}{p})/N^{\frac{1}{2}}$$

and let $X_N^i(0) = [N^{\frac{1}{2}} y_0^i + Np^i]$ with probability one, where $\underset{\sim}{y}_0 = (y_0^1, \ldots, y_0^k)$ is an arbitrary element of R^k. Our purpose here is to apply Theorem 5 to show that $\underset{\sim}{Y}_N \Longrightarrow \underset{\sim}{Y}$, where $\underset{\sim}{Y}$ is a k-dimensional analog of the Ornstein-Uhlenbeck process.

We showed in [7] that a continuous version of $\underset{\sim}{Y}_N$ converges weakly. Theorem 2 of that paper shows that the finite-dimensional distributions of $\underset{\sim}{Y}_N$ converge to those of $\underset{\sim}{Y}$. To complete the proof we must show that the marginals of $\underset{\sim}{Y}_N$ converge weakly to a probability measure. It is easy to see that the process $Y_N^i(\ell)$ is a one-dimensional random walk. Using Stone [12] it is a simple matter to show that $Y_N^i([Nt])$ converges weakly to an Ornstein-Uhlenbeck process. This completes the proof that $\underset{\sim}{Y}_N \Longrightarrow \underset{\sim}{Y}$ and considerably shortens the original proof given in [7].

Unfortunately, for more general random walks in k dimensions one is not likely to have the marginal processes be random walks, so that their weak convergence will present a more difficult problem.

REFERENCES

[1] BILLINGSLEY, P. (1967). Weak Convergence of Probability Measures. To appear.

[2] BILLINGSLEY, P. (1956). The invariance principle for dependent random variables. Trans. Amer. Math. Soc. **83** 250-268.

[3] BILLINGSLEY, P. (1962). Limit theorems for randomly selected partial sums. Ann. Math. Statist. **33** 85-92.

[4] DONSKER, M. (1951). An invariance principle for certain probability limit theorems. Mem. Amer. Math. Soc. **6**.

[5] ERDÖS, P. and KAC, M. (1946). On certain limit theorems in the theory of probability. Bull. Amer. Math. Soc. **52** 292-302.

[6] ERDÖS, P. and KAC, M. (1947). On the number of positive sums of independent random variables. Bull. Amer. Math. Soc. **53** 1011-1020.

[7] IGLEHART, D. (1968). Limit theorems for the multi-urn Ehrenfest model. To appear in the Ann. Math. Statist.

[8] LOÈVE, M. (1963). Probability Theory. (3rd edition). Van Nostrand, Princeton.

[9] PROHOROV, Yu. (1956). Convergence of random processes and limit theorems in probability theory. Theor. Probability Appl. **1** 157-214. (English translation.)

[10] SKOROHOD, A. (1956). Limit theorems for stochastic processes. Theor. Probability Appl. **1** 262-290. (English translation.)

[11] SKOROHOD, A. (1957). Limit theorems for stochastic processes with independent increments. Theor. Probability Appl. **2** 138-171.

[12] STONE, C. (1961). Limit theorems for birth and death processes and diffusion processes. Ph.D. thesis, Stanford University.