The Annals of Mathematical Statistics
1971, Vol. 42, No. 5, 1656–1670

CONVERGENCE CRITERIA FOR MULTIPARAMETER STOCHASTIC PROCESSES AND SOME APPLICATIONS[1]

BY P. J. BICKEL AND M. J. WICHURA

University of California, Berkeley and University of Chicago

Chentsov–Billingsley type fluctuation inequalities for stochastic processes whose time parameter ranges over the q-dimensional unit cube are derived and used to establish weak convergence results for such processes.

1. Introduction. In his excellent recent book (1968), Billingsley has given several fluctuation inequalities for sums of random variables (Theorems 12.1, 12.2, 12.5, 12.6) leading to convergence criteria for sequences of stochastic processes $(X_n(t))_{t \in [0,1]}$ whose sample paths are right-continuous and have left-limits everywhere. These criteria, which may be viewed as generalizations of results of Kolmogorov and Chentsov (1956), have been applied by Billingsley to provide simple proofs of various classical results in the theory of weak convergence of one-parameter stochastic processes.

There has recently been considerable interest in questions of weak convergence of similar stochastic processes $(X_n(t))$, where t ranges over the unit cube in q-dimensional space. Situations in which such convergence arises include:

(i) Convergence of the normalized empirical cumulative distribution function for samples from a continuous distribution concentrating on the unit cube in R^q (Dudley (1966), Le Cam (1957)).

(ii) Convergence of the analogue of the partial sum process for two and higher dimensional "time" (Kuelbs (1968), Wichura (1969)).

(iii) Convergence of the normalized, randomly-stopped empirical cumulative for samples from a q-dimensional continuous distribution on the unit q-cube (Pyke (1968), Wichura (1968)).

(iv) Convergence of the normalized empirical cumulative for samples (drawn without replacement) from a finite population (Bickel (1969), Rosén (1967)).

In this paper we prove multidimensional analogues of Theorems 12.5 and 15.6 of Billingsley (1968) and apply them in the situations cited above. The fluctuation inequalities may be found in Section 2 in a format similar to that given in Billingsley (1968) pages 87–102, the convergence criteria in Section 3, and the applications in Section 4.

Other methods work, frequently more elegantly, in all of the above examples. However, as in the one-dimensional case, in situations where moments are "cheap" and the dependence structure formidable we feel that this approach will prove important. In particular we hope to show in a subsequent paper how these criteria

Received November 4, 1970.

[1] This paper was prepared with the partial support of ONR Contract N00014-67-A-011-0004, and also by NSF Grant NSF-6P16071 and ONR Contract N00014-67-A-0285-0009.

may be successfully applied to the problem of convergence of the normalized, randomly-stopped, empirical cumulative distribution of the normalized sample spacings from a uniform distribution on [0, 1]. The question of whether this sequence of processes converges weakly was posed by Pyke (1965).

2. Fluctuation inequalities. Let q be a positive integer, and let $T_1, \cdots, T_q$ be subsets of [0, 1], each of which contains 0 and 1, and is either a finite set or [0, 1] itself. Put $T = T_1 \times \cdots \times T_q$. Let $X = (X(t))_{t \in T}$ be a stochastic process whose state space is some linear space E (typically R^1) endowed with a norm, say $|\cdot|$; we assume that the sample paths of X are smooth enough to permit each of the supremal quantities defined below to be computed by running the time indices involved through countable dense subsets. For simplicity, we assume that X vanishes along the lower boundary, $\bigcup_{1 \leqq p \leqq q} T_1 \times \cdots \times T_{p-1} \times \{0\} \times T_{p+1} \times \cdots \times T_q$, of T. For each p and each $t \in T_p$ define $X_t^{(p)}: T_1 \times \cdots \times T_{p-1} \times T_{p+1} \times \cdots \times T_q \to E$ by

$$X_t^{(p)}(t_1, \cdots, t_{p-1}, t_{p+1}, \cdots, t_q) = X(t_1, \cdots, t_{p-1}, t, t_{p+1}, \cdots, t_q),$$

and for each $s \leqq t \leqq u$ in T_p, set

$$m_p(s, t, u) \equiv m_p(s, t, u)(X) = \min(\|X_t^{(p)} - X_s^{(p)}\|, \|X_u^{(p)} - X_t^{(p)}\|),$$

where $\|\cdot\|$ is the usual supremum norm. The quantities of primary concern to us here are the random variables

$$M_p'' \equiv M_p''(X) = \sup\{m_p(s, t, u): s \leqq t \leqq u \in T_p\}$$

$(1 \leqq p \leqq q)$ and

$$M'' \equiv M''(X) = \max_p M_p''.$$

For $p = 1$ and T finite, the modulus M'' is that of Billingsley (1968) (cf. (12.62)), which is very useful in studying the weak convergence of $D([0, 1])$-valued processes. Our goal in this section is to establish bounds on the tail probabilities of the M_p'''s, and thus also on those of M''.

In passing, we note that bounds on M'' give rise to bounds on the random variable

$$M \equiv \sup\{|X(t)|: t \in T\}$$

via the inequality (compare Billingsley (12.4))

$$M \leqq \textstyle\sum_{1 \leqq p \leqq q} M_p'' + |X(u)| \leqq qM'' + |X(u)|, \tag{1}$$

where $u = (1, \cdots, 1)$. To establish this inequality take any $t = (t_1, \cdots, t_q) \in T$, and set $u_p = (1, \cdots, 1, t_{p+1}, \cdots, t_q)(0 \leqq p \leqq q)$, so that $u_0 = t$ and $u_q = u$. The assumption that X vanishes along the lower boundary of T then yields

$$\begin{aligned}|X(u_{p-1})| &\leqq \min\{|X(u_{p-1})|, |X(u_p) - X(u_{p-1})|\} + |X(u_p)| \\ &\leqq M_p'' + |X(u_p)|\end{aligned}$$

for $1 \leqq p \leqq q$; together, these inequalities imply that $|X(t)|$ is majorized by the middle term of (1).

To describe the hypotheses under which we will derive the desired bounds, we will make use of the following notation and terminology. *A* **block** B in T is a subset of T of the form $(s, t] = \prod_p (s_p, t_p]$ with s and t in T; the pth-**face** of $B = (s, t]$ is $\prod_{\rho \neq p}(s_\rho, t_\rho]$. Disjoint blocks B and C are p-**neighbors** if they abut and have the same pth face; they are **neighbors** if they are p-neighbors for some p (for example, when $q = 3$, the blocks $(s, t] \times (a, b] \times (c, d]$ and $(t, u] \times (a, b] \times (c, d]$ are 1-neighbors ($s \leqq t \leqq u$ in T_1)). For each block $B = (s, t]$, let

$$X(B) = \sum_{\varepsilon_1 = 0,1} \cdots \sum_{\varepsilon_q = 0,1} (-1)^{q - \Sigma_p \varepsilon_p} X(s_1 + \varepsilon_1(t_1 - s_1), \cdots, s_q + \varepsilon_q(t_q - s_q))$$

be the **increment** of X around B; $X(\cdot)$ is a (random) finitely additive function on blocks. For each pair of neighboring blocks B, C, put

$$m(B, C) = \min\{|X(B)|, |X(C)|\},$$

$m(B, C)$ is small iff at least one of the increments $X(B)$ and $X(C)$ is small.

Now let $\beta > 1$ and $\gamma > 0$, and let μ be a finite nonnegative measure on T. Again for simplicity, we assume that μ assigns measure zero to the lower boundary of T. Say that (X, μ) **satisfies condition** (β, γ), and write $(X, \mu) \in \mathscr{C}(\beta, \gamma)$, if

$$P\{m(B, C) \geqq \lambda\} \leqq \lambda^{-\gamma}(\mu(B \cup C))^\beta \tag{2}$$

for all $\lambda > 0$ and every pair of neighboring blocks B and C in T. From Chebychev's inequality, one sees that (2) is implied by its moment version, namely $E(m(B, C))^\gamma \leqq (\mu(B \cup C))^\beta$, as well as by the frequently employed moment condition

$$E(|X(B)|^{\gamma_1}|X(C)|^{\gamma_2}) \leqq (\mu(B))^{\beta_1}(\mu(C))^{\beta_2}, \tag{3}$$

where γ_1, γ_2, β_1, and β_2 satisfy $\gamma_1 + \gamma_2 = \gamma$ and $\beta_1 + \beta_2 = \beta$. When $q = 1$ and T is finite, condition (2) is essentially (12.11) of Billingsley (with β here equal to 2α there, and γ here equal to 2γ there).

Define constants $K_q(\beta, \gamma)$ and $L_q(\beta, \gamma)$ inductively as follows: Put $\delta = 1/(1+\gamma)$, $\rho = 2^{-(\beta-1)\delta}$, $K(\beta, \gamma) = 2^\gamma(1-\rho)^{-1/\delta}$, $K_1(\beta, \gamma) = L_1(\beta, \gamma) = 2^\beta K(\beta, \gamma)$, and for $r \geqq 2$, $K_r(\beta, \gamma) = K_1(\beta, \gamma)([L_{r-1}(\beta, \gamma)(r-1)^\gamma]^\delta + 1)^{1/\delta}$, $L_r(\beta, \gamma) = rK_r(\beta, \gamma)$. Here is the main result, which for $q = 1$ is a variant both of Billingsley's Theorem 12.5 and Chentsov's (1956) Theorem 1.

THEOREM 1. *If* $(X, \mu) \in \mathscr{C}(\beta, \gamma)$, *then*

$$P\{M_p''(X) \geqq \lambda\} \leqq K_q(\beta, \gamma)\lambda^{-\gamma}(\mu(T))^\beta \qquad (1 \leqq p \leqq q) \tag{4}$$

$$P\{M''(X) \geqq \lambda\} \leqq L_q(\beta, \gamma)\lambda^{-\gamma}(\mu(T))^\beta \tag{5}$$

for all positive λ.

A few remarks should be made at this point. When $T = [0, 1]^q$ and μ is continuous, the factor 2^β may be dropped from the definition of $K_1(\beta, \gamma)$, thus giving smaller universal constants. For T finite and $q = 1$, Theorem 1 reduces to

Theorem 12.5 of Billingsley, except that Billingsley gives a different value, namely $2^{(2+\gamma-\beta)}K_1(\beta, \gamma)$, for the universal constant. The K's are quite large; for example, when $q = 1$ and, as in many applications, $\beta = 2$ and $\gamma = 4$, $K_1(\beta, \gamma)$ is approximately 1,750,000. Finally, the assumption that the process X and the measure μ vanish along the lower boundary of T can be removed, provided condition (β, γ) is strengthened so as to restrain the behavior of X over the lower boundary; what is needed is simply that (X', μ') satisfy condition (β, γ), where (slightly abusing our convention concerning time domains) $T' = T_1' \times \cdots \times T_q'$, $T_p' = \{-1\} \cup T_p$, and X' (resp. μ') equals X (resp. μ) over T and zero over $T' \sim T$ (note that $M_T''(X) \leqq M_T''(X')$).

PROOF OF THEOREM 1. The proof will be carried out in several steps, as follows: (i) $q = 1$, $T = [0, 1]$, $\mu =$ Lebesgue measure, (ii) $q = 1$, $T = [0, 1]$, μ atomless, (iii) $q = 1$, T finite, (iv) $q = 1$, $T = [0, 1]$, μ general, and (v) $q \geqq 2$.

Step 1. Here condition (β, γ) reads

$$P[\min\{|X(t)-X(s)|, |X(u)-X(t)|\} \geqq \lambda] \leqq \lambda^{-\gamma}(u-s)^{\beta} \tag{6}$$

for all $\lambda > 0$ and all $0 \leqq s \leqq t \leqq u \leqq 1$; we shall show that (6) implies

$$P\{M'' \geqq \lambda\} \leqq K(\beta, \gamma)\lambda^{-\gamma}$$

for all $\lambda > 0$.

Take any positive numbers θ_i, $i \geqq 0$, set

$$s_{i,n} = (n-1)2^{-i}, u_{i,n} = n2^{-i}, t_{i,n} = (s_{i,n}+u_{i,n})/2,$$

and define events

$$F_{i,n} = \{\min(|X(t_{i,n})-X(s_{i,n})|, |X(u_{i,n})-X(t_{i,n})|) < \lambda\theta_i\}$$
$$F_i = \bigcap\nolimits_{1 \leqq n \leqq 2^i} F_{i,n}$$
$$F = \bigcup\nolimits_{0 \leqq i < \infty} F_i.$$

If $F_{i,n}$ occurs, then one has a "favorable" comparison of the two increments involved, in the sense that at least one of them is "small."

On the one hand, the probability that all comparisons are favorable is high, i.e.

$$P(F^c) \leqq \textstyle\sum_i \sum_n P(F_{i,n}^c) \leqq \sum_i 2^i(\lambda\theta_i)^{-\gamma}2^{-i\beta} = \lambda^{-\gamma}\sum_{0 \leqq i < \infty} 2^{-\alpha i}\theta_i^{-\gamma}, \tag{7}$$

where $\alpha = \beta - 1 > 0$. On the other hand, whenever all comparisons are favorable, M'' is small, i.e.

$$F \subset \{M'' \leqq 2(\textstyle\sum_{0 \leqq i < \infty} \theta_i)\lambda\}. \tag{8}$$

To see this, let $S_i = \{n2^{-i}; 0 \leqq n \leqq 2^i\}$, let $\omega \in F$, and, referring to the definition of the F_i's, construct ω-dependent order-preserving maps $\psi_i : S_{i+1} \to S_i$ such that

$$|X(\psi_i(s))(\omega) - X(s)(\omega)| < \lambda\theta_i$$

for all $s \in S_{i+1}$ and all i. Piece the ψ_i's together to produce an (ω-dependent) order-preserving map ψ from $S = \bigcup_i S_i$ to $\{0, 1\}$ such that

$$|X(\psi(s))(\omega) - X(s)(\omega)| < \lambda \textstyle\sum_{0 \leqq i < \infty} \theta_i$$

for all $s \in S$. By the monotonicity of ψ, one must have $\psi(s) = \psi(t)$ or $\psi(t) = \psi(u)$ for any three points $s \leqq t \leqq u$ in S. Our assumption about the smoothness of the sample paths of X now implies that (8) holds.

From (7) and (8), one sees that M'' is likely to be small, i.e.

$$P\{M'' \geqq \lambda\} \leqq \lambda^{-\gamma} \inf_\xi f(\xi),$$

where $\xi = (\xi_i)_{i \geqq 0}$ ranges over all probability measures on $\{0, 1, 2, \cdots\}$ and where

$$f(\xi) = 2^\gamma \sum_{i \geqq 0} 2^{-\alpha i} \xi_i^{-\gamma}.$$

Elementary calculations show that f achieves its minimum at that ξ for which $\xi_i = \rho^i(1-\rho)$ (all i) and has there the value $K(\beta, \gamma)$.

Step 2. *Proof for* $q = 1$, $T = [0, 1]$, μ *having continuous distribution function* F. For F both continuous and strictly increasing, a transformation of the time scale making use of the well-defined inverse function of F reduces the present case to that treated in Step 1, and yields

$$P\{M'' \geqq \lambda\} \leqq K(\beta, \gamma)\lambda^{-\gamma}(F(1))^\beta \tag{9}$$

($F(0) = 0$ by assumption). For F merely continuous, first note that (9) holds with F replaced by $F + \varepsilon I$ (I = identity function), and then pass to the limit as $\varepsilon \downarrow 0$.

Step 3. *Proof for* $q = 1$, T *finite*. Let $0 = t_0 < t_1 < \cdots < t_m = 1$ be the points of T. Let $Y = (Y(u))_{0 \leqq u \leqq 1}$ be the process, defined on the same probability space as X, having right continuous sample paths constant over the intervals separating the t_i's and satisfying $Y(t_i) = X(t_i)$ for $0 \leqq i \leqq m$, i.e.

$$Y(u) = \sum_{0 \leqq i < m} X(t_i) I_{[t_i, t_{i+1})}(u) + X(t_m) I_{\{t_m\}}(u).$$

One has $M_T''(X) = M''_{[0,1]}(Y)$. Now look at $m(s, t, u)(Y)$. This quantity is zero unless

$$0 \leqq t_{i-1} \leqq s < t_i \leqq t < t_k \leqq u < t_{k+1}$$

for some $0 < i < k \leqq m$, in which case the hypotheses on X yield

$$\begin{aligned} P\{m(s,t,u)(Y) \geqq \lambda\} &\leqq \lambda^{-\gamma}(\textstyle\sum_{i \leqq j \leqq k} \mu(\{t_j\}))^\beta \\ &\leqq \lambda^{-\gamma}[\textstyle\sum_{i < j \leqq k} (\mu(\{t_j\}) + \mu(\{t_{j-1}\}))]^\beta \\ &\leqq \lambda^{-\gamma}[F(u) - F(s)]^\beta, \end{aligned}$$

where F is that continuous distribution function, satisfying $F(0) = 0$, which is linear over $[t_{j-1}, t_j]$ with

$$F(t_j) - F(t_{j-1}) = \mu(\{t_j\}) + \mu(\{t_{j-1}\})$$

for $1 \leqq j \leqq m$. Since $F(1) \leqq 2\mu(T)$, it follows from Step 2 that

$$P\{M''(X) \geqq \lambda\} \leqq \lambda^{-\gamma} K_1(\beta, \gamma)(\mu(T))^\beta.$$

Step 4. *Proof for* $q = 1$, $T = [0, 1]$ (μ *arbitrary*). Let $0 = t_0 < t_1 < \cdots < t_m = 1$ be points in T, put $U = \{t_0, \cdots, t_m\}$, $Y = (X(u))_{u \in U}$, and let ν be the measure on U such that

$$\nu(\{t_j\}) = \mu((t_{j-1}, t_j]), \quad \text{if } j \geqq 1$$
$$= 0, \quad \text{if } j = 0.$$

Since (X, μ) satisfies condition (β, γ), so does (Y, ν). Apply Step 3 to (Y, ν) and make a suitable passage to the limit to get the desired result.

Step 5. *Proof for* $q \geqq 2$. We now know that Theorem 1 is true when $q = 1$, i.e. when we are dealing with univariate time. The rest of the proof proceeds by induction on q for (4) and (5) simultaneously. Consider (4), with $p = 1$ for convenience. The key observation to be made is that the univariate-time version of Theorem 1 may be applied to the (function space valued) process $(X_t^{(1)})_{t \in T_1}$, once bounds on the increments of this process are found; these bounds will come to us from (1) and the induction hypothesis. More specifically, let $s \leqq t \leqq u$ in T_1, and define processes $Y = X_t^{(1)} - X_s^{(1)}$ and $Z = X_u^{(1)} - X_t^{(1)}$ having $T_2 \times \cdots \times T_q$ as index set. From (1), we have

$$M(Y) \leqq (q-1)M''(Y) + |Y(\mathbf{1})|, M(Z) \leqq (q-1)M''(Z) + |Z(\mathbf{1})|$$

(where $\mathbf{1} = (1, 1, \cdots, 1)$), so that

$$m_1(s,t,u)(X) = \min\{M(Y), M(Z)\} \leqq (q-1)[\max\{M''(Y), M''(Z)\}] + \min\{|Y(\mathbf{1})|, |Z(\mathbf{1})|\}.$$

Using the fact that the increment of Y around a block B in $T_2 \times \cdots \times T_q$ is the increment of X around the block $(s, t] \times B$ in T, one gets from the induction hypothesis that

$$P\{M''(Y) \geqq \lambda\} \leqq \lambda^{-\gamma} L_{q-1}(\beta,\gamma)(F(t) - F(s))^\beta,$$

where F is the distribution function of the marginal of μ on T_1. Similarly,

$$P\{M''(Z) \geqq \lambda\} \leqq \lambda^{-\gamma} L_{q-1}(\beta,\gamma)(F(u) - F(t))^\beta$$

while from the original hypothesis on X,

$$P[\min\{|Y(\mathbf{1})|, |Z(\mathbf{1})|\} \geqq \lambda] \leqq \lambda^{-\gamma}(F(u) - F(s))^\beta.$$

It follows easily from this, the estimate

$$P\{U + V \geqq \lambda\} \leqq P\{U \geqq \lambda\xi_1\} + P\{V \geqq \lambda\xi_2\}$$

(valid for any random variables U, V and positive numbers ξ_1, ξ_2 such that $\xi_1 + \xi_2 = 1$), and the relation

$$\inf\{C_1/\xi_1^\gamma + C_2/\xi_2^\gamma : \xi_1 + \xi_2 = 1\} = (C_1^\delta + C_2^\delta)^{1/\delta}$$

($\delta = 1/(1+\gamma)$) that

$$P\{m_1(s,t,u)(X) \geqq \lambda\} \leqq \lambda^{-\gamma}([(q-1)^\gamma L_{q-1}(\beta,\gamma)]^\delta + 1)^{1/\delta}(F(u) - F(s))^\beta.$$

In other words, the process $X^{(1)}$ meets the hypotheses of the theorem for the case of univariate time, so that (4) holds; of course, (4) implies (5). □

3. Convergence criteria. Let T denote the unit cube $[0, 1]^q$. Call a function $x : T \to R^1$ a step function if x is a linear combination of functions of the form

$$t \to I_{E_1 \times E_2 \times \cdots \times E_q}(t), \qquad \text{where}$$

each E_p is either a left-closed, right-open subinterval of [0, 1], or the singleton {1} and where I_E denotes the indicator of the set E. Let D_q be the uniform closure, in the space of all bounded functions from T to R^1, of the vector subspace of simple functions. The functions in D_q may be characterized by their continuity properties, as follows. If $t \in T$ and if, for $1 \leqq p \leqq q$, R_p is one of the relations $<$ and $\geqq$, let $Q_{R_1, \cdots, R_q}(t)$ denote the quadrant

$$\{(s_1, \cdots, s_q) \in T : s_p R_p t_p, 1 \leqq p \leqq q\}.$$

Then (see Neuhaus (1969), or Straf (1970), page 29) $x \in D_q$ iff for each $t \in T$, (a) $x_Q \equiv \lim_{s \to t, s \in Q} x(s)$ exists for each of the 2_q quadrants $Q = Q_{R_1, \cdots, R_q}(t)$, and (b) $x(t) = x_{Q \geqq, \cdots, \geqq}$. In this sense, the functions of $\mathscr{D}_q$ are "continuous from above, with limits from below."

One can introduce a metric topology on D_q which for $q = 1$ coincides with Skorohod's well-known and useful J_1-topology (see Billingsley (1968), for example). For this, let Λ be the group of all transformations $\lambda : T \to T$ of the form $\lambda(t_1, \cdots, t_q) = (\lambda_1(t_1), \cdots, \lambda_q(t_q))$, where each $\lambda_p : [0, 1] \to [0, 1]$ is continuous, strictly increasing, and fixes zero and one. Define the "Skorohod" distance between x and y in D_q to be

$$d(x, y) = \inf\{\min(\|x - y\lambda\|, \|\lambda\|) : \lambda \in \Lambda\},$$

where $\|x - y\lambda\| = \sup\{|x(t) - y(\lambda(t))| : t \in T\}$ and $\|\lambda\| = \sup\{|\lambda(t) - t| : t \in T\}$. With respect to the corresponding metric topology (S-topology), D_q is separable and topologically complete, and the Borel σ-algebra $\mathscr{D}_q$ coincides with the σ-algebra generated by the coordinate mappings (Billingsley (1968), Neuhaus (1969), Straf (1969)). Consequently, a stochastic process $(X(t))_{t \in T}$ taking values in D_q is $\mathscr{D}_q$-measurable.

We turn now to a discussion of weak convergence for D_q-valued processes. For simplicity we shall speak only of sequences of processes, but everything we say is true for generalized sequences, i.e. nets. A sequence $(X_n)_{n \geqq 1}$ of D_q-valued processes is said to converge weakly in the S-topology to a D_q-valued process X, written $X_n \to X$, if $Ef(X_n) \to Ef(X)$ for all S-continuous bounded functions $f: D_q \to R$. According to the general theory of weak convergence, $X_n \to X$ is equivalent to $f(X_n) \to f(X)$ (in the sense of weak convergence for real-valued random variables) for all $\mathscr{D}_q$-measurable functions $f: D_q \to R$ which are X-continuous in the S-topology (i.e., continuous almost surely with respect to the distribution of X). If X takes all its values in C_q, the subset of D_q consisting of continuous functions, then one has $f(X_n) \to f(X)$ even for $\mathscr{D}_q$-measurable functions f which are X-continuous with respect to the stronger topology of uniform convergence (see Billingsley (1968), Neuhaus (1969) and Straf (1969)).

A criterion for the weak convergence of D_q-valued processes can be given in terms of the weak convergence of the corresponding finite-dimensional distributions together with a tightness condition. To make this explicit, define $\pi_S: D_q \to R^S$ by $\pi_S(x) = (x(s))_{s\in S}$, for each finite set $S \subset T$. Let $\mathscr{T}$ be the collection of subsets of T of the form $U_1 \times \cdots \times U_q$, where each U_p contains zero and one and has countable complement. For each D_q-valued process X, put $T_X = \{t \in T, \pi_{\{t\}}$ is continuous with probability one with respect to the law of X on $(D_q, \mathscr{D}_q)\}$; one can show $T_X \in \mathscr{T}$ (Billingsley (1968), Neuhaus (1969), Straf (1969)). Finally, call a partition of T formed by finitely many hyperplanes parallel to the coordinate axes a δ-grid if each element of the partition is a "left-closed, right-open" rectangle of diameter at least δ, and define $w_\delta': D_q \to R$ by

$$w_\delta'(x) = \inf_\Delta \max_{G\in\Delta} \sup_{s,t\in G} |x(t) - x(s)|,$$

where the infimum extends over all δ-grids Δ in T. Following the development of Billingsley (1969), it is easy to prove the following fundamental result (confer Straf (1970) page 36):

THEOREM 2. *Let X_n, $n \geqq 1$, be D_q-valued processes. In order that the sequence (X_n) converge weakly, it is necessary and sufficient that*

(i) $(\pi_S(X_n))$ *converges weakly, for all finite subsets S of some member τ of $\mathscr{T}$, and*

(ii) $\text{plim}_\delta \lim_n w_\delta'(X_n) = 0$;

and then $X_n \to X$, where the distribution of the D_q-valued process X is determined by $\pi_S(X_n) \to \pi_S(X)$ for all finite $S \in \tau \cap T_X$. (Condition (ii) means $\lim_{\delta\downarrow 0} \limsup_n P\{w_\delta'(X_n) \geqq \varepsilon\} = 0$ for all $\varepsilon > 0$).

One can deduce (cf Theorem 14.4 and 15.4 of Billingsley (1968)) from this basic result the corollary below, which is sufficient for our purposes. First define $w_\delta'': D_q \to R$ by

$$w_\delta''(x) = \max_p w_\delta''^{(p)}(x)$$

where

$$w_\delta''^{(p)}(x) = \sup\{\min(\|x_t^{(p)} - x_s^{(p)}\|, \|x_u^{(p)} - x_t^{(p)}\|) : s \leqq t \leqq u, u - s \leqq \delta\}$$

$(1 \leqq p \leqq q)$. To motivate this definition, we note that the set-theoretic identity

$$D_q \equiv D(I^q, R) = D_1(I, D_{q-1})$$

is valid via any one of the correspondences $x(\cdot) \leftrightarrow x_\cdot^{(p)}(\cdot)$, provided on the right-hand side D_{q-1} is equipped with the supremum norm. This is easily proved (confer Straf (1970), page 32) by first considering step functions and then their uniform limits. The modulus w_δ'' can thus be viewed as a more or less natural generalization of Billingsley's modulus, of the same name, for $p = 1$. Another consequence of the above identity is that for any D_q-valued process X, $\lim_{t\uparrow 1} X_t^{(p)}$ exists uniformly over $[0, 1]^{q-1}$. This limit will be $X_1^{(p)}$ provided the finite-dimensional distributions of the $X_t^{(p)}$'s converge to those of $X_1^{(p)}$, as will be the case if, say, $X(s_1, \cdots, s_{p-1}, t, s_{p+1}, \cdots, s_q)$ converges to $X(s_1, \cdots, s_{p-1}, 1, s_{p+1}, \cdots, s_q)$ in probability for all choices of the s_j's. We shall say that X is continuous at the upper boundary of T if $\lim_{t \uparrow 1} X_t^{(p)} = X_1^{(p)}$ for each p, with probability one.

COROLLARY. *Let X_n, $n \geqq 1$, and X be D_q-valued processes, and suppose that X is continuous at the upper boundary of T. Then in order that $X_n \to X$, it is necessary and sufficient that*

$$\pi_S(X_n) \to \pi_S(X) \qquad \text{for all finite subsets } S \text{ of some member } \tau \text{ of } \mathscr{T},$$

$$\operatorname{plim}_\delta \lim_n w_\delta''(X_n) = 0. \tag{10}$$

PROOF. Here is the proof of the sufficiency. The proof uses induction on q. For $q = 1$, the corollary is just Theorem 15.4 of Billingsley (1968). Suppose now that the sufficiency part of the corollary is known to hold for $q-1$; we shall show that it holds for q. We have only to verify that condition (ii) of Theorem 2 holds. For each p, define $w_\delta'^{(p)}$ on D_q by

$$w_\delta{}^{(p)}(x) = \inf_{\Delta_p} \max_{G \in \Delta_p} \sup_{s,t \in G} \|x_t^{(p)} - x_s^{(p)}\|,$$

where the infimum here extends over all δ-grids Δ_p in [0, 1]. Clearly,

$$w_\delta' \leqq \sum_{1 \leqq p \leqq q} w_\delta'^{(p)}.$$

Moreover, a simple but tedious argument (cf Billingsley (1968), Theorems 14.4 and 15.4) shows that

$$w_{\delta/2}'^{(p)}(x) \leqq 2[w_\delta''^{(p)}(x) + L_\delta^{(p)}(x) + R_\delta^{(p)}(x)],$$

where

$$L_\delta^{(p)}(x) = \sup_{0 \leqq t < \delta} \|x_t^{(p)} - x_0^{(p)}\| \leqq 2[\|x_\delta^{(p)} - x_0^{(p)}\| + w_\delta''^{(p)}(x)]$$

$$R_\delta^{(p)}(x) = \sup_{\zeta < t \leqq 1} \|x_1^{(p)} - x_t^{(p)}\| \leqq 2[\|x_1^{(p)} - x_\zeta^{(p)}\| + w_\delta''^{(p)}(x)]$$

($\zeta = 1-\delta$).

Thus it suffices to show that the plim $\inf_\delta \lim_n$'s of

$$\|(X_n)_\delta^{(p)} - (X_n)_0^{(p)}\| \quad \text{and} \quad \|(X_n)_1^{(p)} - (X_n)_\zeta^{(p)}\|$$

are zero. As the arguments in both cases are similar, we shall discuss only the first case. Fix p, and set $Z_{n,\delta} = (X_n)_\delta^{(p)} - (X_n)_0^{(p)}$, $Z_\delta = X_\delta^{(p)} - X_0^{(p)}$; Z_δ and the $Z_{n,\delta}$'s are D_{q-1}-valued processes. We will show below that $Z_{n,\delta} \to Z_\delta$ for all but countably many δ's. Since $\|\cdot\| = d(\cdot, 0)$ is an S-continuous function on D_{q-1}, we will then have $\|Z_{n,\delta}\| \to \|Z_\delta\|$ for all but countably many δ's. But the identity $D_q = D_1(I, D_{q-1})$ implies that $\|Z_\delta\| \to 0$ as $\delta \to 0$. All this gives

$$\lim\inf_\delta \lim\sup_n P\{\|Z_{n,\delta}\| \geqq \varepsilon\} = 0$$

for all $\varepsilon > 0$, as desired.

It remains to show that $Z_{n,t} \to Z_t$ for all but countably many t. One finds easily that

(a) for any t, the process Z_t is continuous at the upper boundary of $[0, 1]^{q-1}$,

(b) one has $w_\delta''(Z_{n,t}) \leqq 2w_\delta''(X_n)$, where on the left-hand side w_δ'' is the modulus appropriate for D_{q-1}, and, with $\tau = U_1 \times \cdots \times U_q$,

(c) for any $t \in U_1$, one has $\pi_S(Z_{n,t}) \to \pi_S(Z_t)$ for all finite subsets S of $U_2 \times \cdots \times U_q$.

The $q-1$ dimensional version of the corollary now implies that for $t \in U_1$, $Z_{n,t} \to Z_t$. □

For these results to be useful in practice, one needs easily verifiable conditions which imply the somewhat awkward tightness condition (10). This is where the fluctuation inequality of the previous section comes into play. The following theorem extends Theorem 15.6 of Billingsley.

THEOREM 3. *Suppose that each X_n vanishes along the lower boundary of T, and that there exist constants $\beta > 1$, $\gamma > 0$ and a finite nonnegative measure μ on T with continuous marginals such that $(X_n, \mu) \in \mathscr{C}(\beta, \gamma)$ for each n. Then the tightness condition* (10) *is in force.*

PROOF. It is enough to show $\operatorname{plim}_\delta \lim_n w_\delta''^{(p)}(X_n) = 0$ for each p. For this, put

$$w(\sigma, \tau; n) = \sup\{\min(\|(X_n)_t^{(p)} - (X_n)_s^{(p)}\|, \|(X_n)_u^{(p)} - (X_n)_t^{(p)}\|) : \sigma \leqq s \leqq t \leqq u \leqq \tau\}.$$

Since

$$w_{\frac{1}{2}k}''^{(p)}(X_n) \leqq \max_{1 \leqq j \leqq k} w((2j-2)/2k, 2j/2k; n) + \max_{1 \leqq j < k} w((2j-1)/2k, (2j+1)/2k; n)$$

it suffices (cf Billingsley (1968) page 130) to show that

$$P\{w(\sigma, \tau; n) \geqq \varepsilon\} \leqq \varepsilon^{-\gamma} K_q(\beta, \gamma)(\mu_p((\sigma, \tau]))^\beta \tag{11}$$

where μ_p denotes the (continuous) marginal of μ on the pth edge of T. But (11) is an easy consequence of Theorem 1 and the fact that in the definition of $w(\sigma, \tau; n)$, X_n can be replaced by Y_n, where Y_n, defined on $T^* = [0, 1]^{p-1} \times [\sigma, \tau] \times [0, 1]^{q-p}$ so that $(Y_n)_t^{(p)} = (X_n)_t^{(p)} - (X_n)_\sigma^{(p)}$ for $\sigma \leqq t \leqq \tau$, vanishes along the lower boundary of T^* and has the same increments around blocks in T^* as does X_n. □

Actually, Theorem 3 is not flexible enough to apply to some of the simplest processes. The following extension will be useful. For each n, suppose that there exists a subset $T^n = T_1^n \times \cdots \times T_q^n$ of T such that

(a) T_p^n contains 0 and 1 for each n $(1 \leqq p \leqq q)$,

(b) $w_\delta''(X_n)$ may be computed using T^n as the time set (instead of T),

(c) T^n becomes dense in T as n grows large, and

(d) Condition (β, γ) holds for blocks whose corner points lie in T^n.

Then the conclusion to Theorem 3 holds; the proof is essentially the same (with the role of the equally spaced points $j/2k$, $0 \leqq j \leqq 2k$, in the estimate of $w_{\frac{1}{2}k}''^{(p)}(X_n)$ being taken over by almost equally spaced points from T_p^n). The theorem may be extended further by allowing μ to depend on n and to have discontinuous marginals, while requiring that the new μ_n's converge weakly to a limit μ having continuous marginals (under this condition, (11) holds with a lim $\sup_n$ prefixed to the left-hand side; inspection of the argument on page 130 of Billingsley (1968) shows that this is good enough). Finally we note that an analogue of Theorem 15.7 of Billingsley (1968) can be proved by essentially the same method, and thus that there is no loss of generality in considering only D_q-valued processes from the outset. Specifically one has

THEOREM 4. *Let $\mathscr{S}$ denote the class of finite subsets of T. Let $(\nu_S)_{S\in\mathscr{S}}$ be a consistent family of probabilities on the finite-dimensional spaces $(R^S, \mathscr{R}^S)$, $S \in \mathscr{S}$. Define ν on the algebra $\bigcup_{S\in\mathscr{S}}\pi_S^{-1}(\mathscr{R}^S)$ of subsets of R^T so that $\nu\pi_S^{-1} = \nu_S$ for all S in $\mathscr{S}$. Suppose that*

(i) $\nu\{x \in R^T: x(t) = 0\} = 1$, *if any coordinate of $t \in T$ is* 0,

(ii) $\nu\{x \in R^T: |x(t+h)-x(t)| \geqq \varepsilon\} \to 0$ *for all $\varepsilon > 0$, as h tends to 0 "from above,"*

(iii) $\nu\{x \in R^T: |x(s_1, \cdots, s_{p-1}, t, s_{p+1}, \cdots, s_q)-x(s_1, \cdots, s_{p-1}, 1, s_{p+1}, \cdots, s_q)| \geqq \varepsilon\} \to 0$ *as $t \to 1$, for all choices of p and of the s_j's, $j \neq p$, and for all $\varepsilon > 0$,*

(iv) *for some $\beta > 1$, $\gamma > 0$, and some measure μ on T having continuous marginals,*

$$\nu\{x \in R^T: \min(|x(B)|, |x(C)|) \geqq \lambda\} \leqq \lambda^{-\gamma}(\mu(B \cup C))^\beta$$

for all $\lambda > 0$ and all pairs of neighboring blocks B and C in T.
Then there exists a D_q-valued process whose finite dimensional distributions are the ν_S's.

4. Applications. Our purpose here is to illustrate the use of Theorem 3 in establishing weak convergence results. Accordingly, no fuss will be made about convergence of finite-dimensional distributions, which in most of the examples below is obvious. Some of the results have been deduced before, by a variety of different methods.

(I) *Partial sum processes.* For convenience, we work with 2-dimensional time. The following theorem extends the classic result of Prohorov (for $q = 1$) (cf Prohorov (1956) and Wichura (1969)). For each n, let $X_{i,j}^{(n)}(1 \leqq i \leqq I_n, 1 \leqq j \leqq J_n)$ be independent random variables with zero means and finite variances

$$\operatorname{Var}(X_{i,j}^{(n)}) = a_i^{(n)} b_j^{(n)}$$

such that

$$\textstyle\sum_i a_i^{(n)} = 1 = \sum_j b_j^{(n)}.$$

Put

$$A_i^{(n)} = \textstyle\sum_{g\leqq i} a_g^{(n)} \qquad B_j^{(n)} = \sum_{h\leqq j} b_h^{(n)},$$

and define D_2-valued processes S_n by

$$S_n(t) = \textstyle\sum_{i\leqq A^{(n)}(t)} \sum_{j\leqq B^{(n)}(t)} X_{i,j}^{(n)},$$

where $A^{(n)}(t)$ (resp. $B^{(n)}(t)$) is the largest $A_i^{(n)}$ (resp. $B_j^{(n)}$) not exceeding t_1 (resp. t_2) $(t = (t_1, t_2))$.

THEOREM 5. *If the* $X_{i,j}^{(n)}$ *satisfy Lindeberg's condition, namely*

$$\lim_n \left[\sum_i \sum_j \int_{[|X_{i,j}^{(n)}| \geqq \varepsilon]} (X_{i,j}^{(n)})^2 \, dP\right] = 0 \qquad \text{for all } \varepsilon > 0,$$

and if

$$\max_i a_i^{(n)} \to 0, \max_j b_j^{(n)} \to 0,$$

then $S_n \to S$, *where* S *is a Gaussian process with zero means and covariances*

$$\mathrm{Cov}(S(t_1, u_1), S(t_2, u_2)) = \min(t_1, t_2) \min(u_1, u_2)$$

(*i.e.* S *is a Brownian motion process on* $[0, 1]^2$).

PROOF. For each n, put $T^n = \{A_i^{(n)};\ 0 \leqq i \leqq I_n\} \times \{B_j^{(n)}:\ 0 \leqq j \leqq J_n\}$. If B and C are a pair of neighboring blocks with corner points in T^n, then by independence one has

$$E[S_n^2(B) S_n^2(C)] = \mathrm{Var}[S_n^2(B)] \mathrm{Var}[S_n^2(C)] = \lambda(B)\lambda(C)$$

where λ denotes Lebesgue measure on $[0, 1]^2$. Consequently inequality (3) holds for S_n with $\gamma_1 = 2 = \gamma_2$ and $\beta_1 = 1 = \beta_2$ (so that $\gamma = 4$, $\beta = 2 > 1$), and the theorem follows from the remarks after Theorem 3 (which is not itself directly applicable). □

L. LeCam informed us that in an unpublished work carried out several years ago, he used the methods of LeCam (1958) to prove a theorem, involving partial sum processes, which is analogous to the normal convergence criteria for sums of u.a.n. variables. This of course includes Theorem 4.

(II) *Sampling from finite populations.* Let $p_{1,N}, \cdots, p_{N,N}$ be N given points in $T = [0, 1]^q$. Suppose that m points are drawn at random without replacement from this population. Distribution free tests in the q-variate two sample problem involve comparing the distribution of the drawn points to that of the remaining ones (see Bickel (1969)). This is conveniently done in terms of the following process. Let H_N be the (non-random) distribution function of the uniform probability over $p_{1,N}, \cdots, p_{N,N}$, and let F_m (resp. G_n) be the (random) distribution function of the uniform measure over the m drawn (resp. $n = N - m$ undrawn) points. Define a D_q-valued process $X_{m,n}$ by

$$X_{m,n} = (mn/N)^{\frac{1}{2}}(F_m - G_n) = (mN/n)^{\frac{1}{2}}(F_m - H_N).$$

The convergence of the $X_{m,n}$ was studied by a different method in Bickel (1969); in particular, it was shown that if $H_N \to H$ as $N \to \infty$, then $X_{m,n} \to X$ as m and n tend to ∞, where X is a Gaussian process with zero means and covariances $\mathrm{Cov}(H(t), H(u)) = H(\min(t, u)) - H(t)H(u)$ (the minimum being computed coordinatewise). Here we show how Theorem 3 may be applied to establish the tightness condition (assuming H is continuous).

For any two neighboring blocks B and C in T, one has

$$E(X_{m,n}(B))^2(X_{m,n}(C))^2 = (N/mn)^2 E(N_B - mp_B)^2(N_C - mp_C)^2,$$

where N_B, N_C, and $N_D = N - N_B - N_C$ have a multiple hypergeometric distribution:

$$P\{N_B = i, N_C = j, N_D = k\} = \binom{Np_B}{i}\binom{Np_C}{j}\binom{Np_D}{k} \Big/ \binom{N}{m}$$

$(i+j+k = m)$ with $p_B = H_N(B), p_C = H_N(C), p_D = 1-p_B-p_C$. By the extended version of Theorem 3, it suffices to show that for $N \geqq 4$

$$E(N_B - mp_B)^2(N_C - mp_C)^2 \leqq 33(mn/N)^2 H_N(B)H_N(C). \tag{12}$$

For this note that given N_B, N_C is hypergeometric with parameters Nq_B (total population size), Np_C (sub-population size), and $m-N_B$ (sample size) $(q_B = 1-p_B)$. It follows that the left-hand side of (12) does not exceed

$$(p_C/q_B)^2 E(N_B-mp_B)^4 + (p_C p_D/(q_B{}^2(Nq_B-1)))[mnq_B{}^2 E(N_B-mp_B)^2 + q_B(m-n)E(N_B-mp_B)^3].$$

From David, Kendall and Barton (1966) page 216, one finds

$$E(N_B-mp_B)^4 = \frac{[Np_Bq_B(p_B{}^3+q_B{}^3)(N(N+1)-6mn)mn + 3N^2 p_B{}^2 q_B{}^2 mn(m-1)(n-1)]}{N(N-1)(N-2)(N-3)}$$

$$E(N_B-mp_B)^3 = Np_Bq_B(p_B-q_B)mn(n-m)/[N(N-1)(N-2)]$$

$$E(N_B-mp_B)^2 = mnp_Bq_B/(N-1).$$

Simple manipulations now yield (12).

(III) *Empirical distribution functions.* We lead into the next application of Theorem 3 with a central limit theorem for D_q-valued processes. Let $Z, Z_1, Z_2, \cdots$ be independent identically distributed D_q-valued processes. Suppose that Z vanishes along the lower boundary of $T = [0, 1]^q$, that $EZ(t) = 0$ for all t in T, and that there exists a continuous finite measure μ on T such that

$$EZ^2(B) \leqq \mu(B)$$

$$EZ^2(B)Z^2(C) \leqq \mu(B)\mu(C)$$

for all pairs of neighboring blocks B and C in T.

Define D_{q+1}-valued processes $X_n (n \geqq 1)$ by

$$X_n(s, t) = (n^{-\frac{1}{2}}) \sum_{j \leqq [ns]} Z_j(t)$$

$(s \in I = [0, 1],\ t \in T)$. Suppose that there exists a D_q-valued Gaussian process $X = (X(s,t))_{s \in I,\, t \in T}$ with zero means and covariances

$$\operatorname{Cov}(X(s_1, t_1), X(s_2, t_2)) = \min(s_1, s_2)\Gamma(t_1, t_2),$$

where $\Gamma(t_1, t_2) = \operatorname{Cov}(Z(t_1), Z(t_2))$. Assume that X is almost surely continuous along the upper boundary of $I \times T$. Such an X exists by Theorem 4 if, for example, Γ is continuous.

THEOREM 6. *In the present context,* $X_n \to X$.

PROOF. This follows easily from the remark following Theorem 3, inequality (3), and the inequalities below:

(i) $n^{-2}E[\sum_{i<\alpha\leqq j} Z_\alpha(B)]^2[\sum_{j<\alpha\leqq k} Z_\alpha(B)]^2 = [(j-i)/n]\mu(B)\cdot[(k-j)/n]\mu(B)$

$$\begin{aligned}(\text{ii})\quad & n^{-2}E[\textstyle\sum_{i<\alpha\leqq j} Z_\alpha(B)]^2[\sum_{i<\alpha\leqq j} Z_\alpha(C)]^2 \\ & \leqq n^{-2}[(j-i)EZ^2(B)Z^2(C)+(j-i)(j-i-1)EZ^2(B)EZ^2(C) \\ & \quad +2(j-i)(j-i-1)(E(Z(B)Z(C)))^2] \\ & \leqq 3[(j-i)/n]\mu(B)\cdot[(j-i)/n]\mu(C),\end{aligned}$$

holding for neighboring blocks B and C in T. □

In passing, we note that it is known that for function space valued processes, even so simple a central limit theorem as the assertion $X_n(1,\cdot) \to X(1,\cdot)$ is not valid without assumptions beyond those of independence and equi-distribution, zero means, and finite variances (see, e.g. Dudley and Strassen (1969)). Now let $(U_k)_{k>1}$ be a sequence of i.i.d. T-valued random variables having a continuous distribution, say Q. Define D_q-valued processes Z_k by

$$Z_k(t) = I_{C(t)}(U_k) - Q(C(t)),$$

where $C(t) = \prod_p[0, t_p](t = (t_1, \cdots, t_q))$, and define G_k by

$$G_k(t) = (1/k^{\frac{1}{2}})\textstyle\sum_{1\leqq j\leqq k} Z_k(t).$$

G_k is of course nothing but the normalized empirical distribution function based on $U_1, \cdots, U_k$. Define a D_{q+1}-valued process X_n by

$$X_n(s,t) = ([ns]/n)^{\frac{1}{2}}G_{[ns]}(t) = (n^{-\frac{1}{2}})\textstyle\sum_{j\leqq[ns]} Z_j(t)$$

($s \in [0, 1]$, $t \in T$). Since

$$\begin{aligned} EZ^2(B) &= \operatorname{Var}(Z(B)) = Q(B)(1-Q(B)) \leqq Q(B) \\ E(Z(B)Z(C))^2 &= Q^2(B^c)Q^2(C)Q(B)+Q^2(B)Q^2(C^c)P(C) \\ &\quad +Q^2(B)Q^2(C)(1-Q(B)-Q(C)) \\ &\leqq 3Q(B)Q(C),\end{aligned}$$

Theorem 5 implies that the X_n converge weakly (to a Gaussian process having continuous sample paths). In particular, the $G_n = X_n(1,\cdot)$ converge weakly. But of course much more than this is true. For example, using the methods of Billingsley (1968), Section 17, one can easily deduce (cf. Wichura (1968)) that G_{N_n} converges, to the limit of the G_n's, whenever $(N_n)_{n\geqq 1}$ is a sequence of positive, integer-valued random variables such that, for some sequence of constants $c_n \to \infty$, N_n/c_n converges in probability to a positive random variable (see Fernandez (1970) for a different approach).

REFERENCES

BICKEL, P. J. (1969). A distribution free version of the Smirnov test in the p-variate case. *Ann. Math. Statist.* **40** 1–23.

BILLINGSLEY, P. (1968). Convergence of probability measures. Wiley, New York.

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

DAVID, F. N., KENDALL, M. G., and BARTON, D. E. (1966). *Symmetric Function and Allied Tables.* Cambridge Univ. Press.

DUDLEY, R. (1966). Weak convergence of probabilities on nonseparable metric spaces and empirical measures on Euclidean spaces. *Illinois J. Math.* **10** 109–126.

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

FERNANDEZ, P. (1970). On the weak convergence of random sums of independent random elements. Ph.D. dissertation, Univ. of California, Berkeley.

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

LECAM, L. (1957). Convergence in distribution of stochastic processes. *Univ. California Publ. Statist.* **2** 207–236.

LECAM, L. (1958). Remarques sur les variables aléatoires dans les espaces vectoriels non separables. *Publ. Inst. Stat. Univ. Paris* **7** 39–53.

NEUHAUS, G. (1969). Zur theorie der Konvergenz stochasticher Prözessen mit mehrdimensionalem Zeitparameter. Ph.D. dissertation, Westphälischen Wilhelms-Universität zu Münster.

NEVEU, J. (1965). *Mathematical Foundations of the Calculus of Probability.* Holden–Day, San Francisco.

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

PYKE, R. (1965). Spacings. *J. Roy. Statist. Soc. Ser. B* **27** 395–449.

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

ROSÉN, B. (1964). Limit theorems for sampling from a finite population. *Ark. Math.* **5** 383–424.

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

STRAF, M. L. (1969). A general Skorohod space and its applications to the weak convergence of stochastic processes with several parameters. Ph.D. dissertation, Univ. of Chicago.

STRAF, M. L. (1970). Weak convergence of stochastic processes with several parameters. *Proc. Sixth Berkley Symp. Math. Statist. Prob.* (to appear).

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

WICHURA, M. J. (1969). Inequalities with applications to the weak convergence of random processes with multidimensional time parameters. *Ann. Math. Statist.* **40** 681–687.