The Annals of Mathematical Statistics
1971, Vol. 42, No. 4, 1285–1295

ON WEAK CONVERGENCE OF STOCHASTIC PROCESSES WITH MULTIDIMENSIONAL TIME PARAMETER

BY GEORG NEUHAUS
University of Münster

0. Summary and introduction. The well-known space D [0, 1] is generalized to k time dimensions and some properties of this space D_k are derived. Then, following the "classical" lines as presented in Billingsley [1], a Skorohod–metric, tightness criteria and some other results concerning weak convergence are given. The theory is applied to prove weak convergence of two generalizations of the one-dimensional empirical process and of the Kolmogorov–Smirnov test statistic of independence.

Stochastic processes with multidimensional time parameter and their weak convergence have been investigated by several authors. Dudley [4] established a theory of convergence of stochastic processes with sample functions in nonseparable metric spaces. Later on, Wichura [11] (see also Wichura [12]) modified the concepts of Dudley and developed them systematically. He applied his theory to a space which is with minor changes our space D_k. Weak convergence in the sense of Wichura [12] and ours differ usually, but both concepts coincide if the limit process has—with probability one—continuous sample functions only. From here it follows that the results of Dudley and Wichura concerning weak convergence of multivariate empirical processes are equivalent to ours.

At least two further authors proved the convergence of multivariate empirical processes, namely LeCam [8] and Bickel [1]. Our proof follows the classical approach of Parthasarathy [9] using an argument of Kuelbs [7] to carry over the proof from 1 to k dimensions. Kuelbs however deals properly with the "interpolated sum" process for two-dimensional time parameter.

The space D_k seems to be defined for the first time in connection with multivariate processes by Winkler [13], yet his investigations are not concerned with weak convergence. Another generalization of the space D[0, 1] and the Skorohod metric to functions on more general spaces than E_k is given in the paper [10] of Straf, in which there are applications to genuinely discontinuous limit processes.

1. The space D_k. In this section the space D_k is introduced and some of its properties are derived.

Before defining D_k we need some other definitions and notations. Let $E_k = [0, 1] \times \cdots \times [0, 1]$ be the k-dimensional unit cube in $\mathbb{R}_k$ with points $t = (t_1, \cdots, t_k)$, $t' = (t_1', \cdots, t_k')$, etc. and $|t| = \max\{|t_i| : i = 1, \cdots, k\}$ be the maximum norm in $\mathbb{R}_k$. The set consisting of the 2^k vertices of E_k is denoted by $\mathscr{P}$.

$$\mathscr{P} = \{\rho = (\rho_1, \cdots, \rho_k) : \rho_i = 0, 1 \,\forall\, i\}.$$

Received January 22, 1970; revised January 4, 1971.

For $t \in E_k$ and $\rho \in \mathscr{P}$ we define "quadrants" $Q(\rho, t)$ and $\tilde{Q}(\rho, t)$ in E_k with vertex t by

$$`\tilde{Q}'(\rho, t) \equiv `\tilde{I}'(\rho_1, t_1) \times \cdots \times `\tilde{I}'(\rho_k, t_k)$$

where the intervals $I(\tau, u)$ and $\tilde{I}(\tau, u)$ are given by

$$\begin{aligned} I(\tau, u) &= [0, u) \quad \text{for} \quad \tau = 0, \\ &= (u, 1] \quad \text{for} \quad \tau = 1, \end{aligned}$$

and

$$\begin{aligned} \tilde{I}(\tau, u) &= [0, u) \quad \text{for} \quad \tau = 0 \quad \text{and} \quad u < 1, \\ &= [0, 1] \quad \text{for} \quad \tau = 0 \quad \text{and} \quad u = 1, \\ &= \quad \varnothing \quad \text{for} \quad \tau = 1 \quad \text{and} \quad u = 1, \\ &= [u, 1] \quad \text{for} \quad \tau = 1 \quad \text{and} \quad u < 1. \end{aligned}$$

The following properties are immediate consequences of the above definitions:

$$\begin{aligned} &Q(\rho, t) \subset \tilde{Q}(\rho, t) \subset \bar{Q}(\rho, t),^1 \\ &Q(\rho, t) = \varnothing \Leftrightarrow \tilde{Q}(\rho, t) = \varnothing, \\ &\tilde{Q}(\rho, t) \cap \tilde{Q}(\rho', t) = \varnothing \quad \text{if} \quad \rho \neq \rho', \\ &\textstyle\sum_{\rho \in \mathscr{P}} \tilde{Q}(\rho, t) = E_k, \qquad t \in E_k. \end{aligned}$$

For every $t \in E_k$ there exists one and only one $\rho = \rho(t) \equiv \sigma \in \mathscr{P}^2$ with $\tilde{Q}(\sigma, t) \ni t$. The quadrants $\tilde{Q}(\sigma, t)$ and $Q(\sigma, t)$ are called continuity quadrants in t. This terminology will become clear in Definition 1.1. Clearly $\mathring{Q}(\sigma, t) \neq \varnothing$ and $\tilde{Q}(\sigma, t) = \bar{Q}(\sigma, t)$.

Now we generalize the notion of one-sided limits for functions defined on $[0, 1]$: Let f be a function defined on E_k. If for $t \in E_k$, $\rho \in \mathscr{P}$ with $Q(\rho, t) \neq \varnothing$ and every sequence $\{t_n\} \subset Q(\rho, t)$ with $t_n \to t$ the sequence $\{f(t_n)\}$ converges, then we denote (the necessarily unique) limit by $f(t+0_\rho)$ and call it ρ-limit of f in t or shortly "quadrant limit".

DEFINITION 1.1. The space D_k is the set of all functions $f: E_k \to \mathbb{R}$ for which the ρ-limit of f in t exists for every $\rho \in \mathscr{P}$, $t \in E_k$ with $Q(\rho, t) \neq \varnothing$ and which are "continuous from above" in the sense that $f(t) = f(t+0_\sigma)$, $t \in E_k$.

Apparently D_1 equals the function space $D[0, 1]$ of Skorohod, so that we have a natural generalization of functions with discontinuities of the first kind.

The requirement of continuity from above effects the existence of quadrant limits not only for sequences in $Q(\rho, t)$ but also in $\tilde{Q}(\rho, t)$. This is an easy consequence of the fact that $\tilde{Q}(\rho, t)$ may be characterized by

$$\tilde{Q}(\rho, t) = \{t' \in E_k : \overline{Q(\sigma(t'), t') \cap Q(\rho, t)} \ni t'\}.$$

---

[1] For $A \subset E_k$ $\bar{A}$ denotes the closure and $\mathring{A}$ the interior of A in the $\mathbb{R}_k$-topology.

The only discontinuities of functions in D_k are "jumps":

DEFINITION 1.2. A function $f \in D_k$ has a jump in $t \in E_k$ iff

$$H_f(t) \equiv \max \{|f(t+0_\rho)-f(t+0_{\rho'})| : \rho, \rho' \in \mathscr{P}\} > 0.$$

We call $H_f(t)$ the magnitude of jump.

A function $f \in D_k$ is continuous at $t \in E_k$ if and only if $H_f(t) = 0$. For subsets $J \subset \mathbb{N}_k \equiv \{1, \cdots, k\}$ we define a magnitude of jump with respect to J by $H_f(t, J) \equiv \max \{|f(t+0_\rho)-f(t+0_{\rho'})| : \rho, \rho' \in \mathscr{P}, \rho_j = \rho_j', j \in J\}$. Clearly $H_f(t) = H_f(t, \varnothing)$ and $H_f(t, \mathbb{N}_k) = 0$.

Let the numbers $u_j \in [0, 1], j \in J$, be fixed and put $L(\{u_j\}, J) \equiv \{t \in E_k : t_j = u_j, j \in J\}$. Elements of $D[0, 1]$ have the property that there can exist only finite many jumps of size greater than a given positive number. The following lemma gives a k-dimensional version of this statement. The proof is straightforward.

LEMMA 1.3. *For $\varepsilon > 0$ and $f \in D_k$ every system of points t on $L(\{u_j\}, J)$ with pairwise different components $t_j, j \in \mathbb{N}_k - J$, and $H_f(t, J) \geqq \varepsilon$ is finite.*

A function $f: E_k \to \mathbb{R}$ belongs to C_k, the space of continuous functions an E_k, if and only if $\lim_{\delta \to 0} \omega_f(\delta) = 0$, where ω_f denotes the "modulus of continuity" $\omega_f(\delta) \equiv \sup \{|f(t)-f(t')| : |t-t'| < \delta\}$.

We shall introduce another "modulus" ω_f', which characterizes the space D_k in the same manner.

Let points $t_1, \cdots, t_r \in E_k$ be given. The collection of all rectangles R of the following form (1.4) is called the "partition generated by the t_j" and denoted by $\mathscr{R} = \mathscr{R}(t_1, \cdots, t_r)$:

$$R = [u_1, u_1' \rangle \times \cdots \times [u_k, u_k' \rangle \tag{1.4}$$

with $u_j, u_j' \in K_j \equiv \{t_{1_j}, \cdots, t_{r_j}\} \cup \{0, 1\}$, $u_j < u_j'$ and $(u_j, u_j') \cap K_j = \varnothing$. Here "$\rangle$" means ")" or "]" if the right end-point of the interval is < 1 or $= 1$.

We define the permeability $m(\mathscr{R})$ of $\mathscr{R}$ to be the length of the shortest side among all rectangles $R \in \mathscr{R}$. Now the modulus ω_f' for D_k is defined by

$$\omega_f'(\delta) \equiv \inf_{\mathscr{R}} \max_R \omega_f(R), \qquad 0 < \delta < 1, \tag{1.5}$$

with $\omega_f(B) \equiv \sup \{|f(t)-f(s)| : t, s \in B\}$. The "inf" in (1.5) is to be taken w.r.t. all $\mathscr{R}$ with $m(\mathscr{R}) > \delta$ and the "max" w.r.t. all $R \in \mathscr{R}$. The modulus ω_f' characterizes D_k.

THEOREM 1.4. *A function $f: E_k \to \mathbb{R}$ belongs to D_k if and only if*

$$\lim_{\delta \downarrow 0} \omega_f'(\delta) = 0. \tag{1.6}$$

The crucial considerations for the "only if" part are summarized in

LEMMA 1.5. *Let $f \in D_k$ be given. For every $\varepsilon > 0$ there exist a $\delta > 0$ and a partition $\mathscr{R}$ such that for points $t, t' \in R$, $R \in \mathscr{R}$, with $|t-t'| < \delta$ the inequality $|f(t)-f(t')| < \varepsilon$ holds.*

---

[2] We write $a \equiv b$ or $b \equiv a$ if b is by definition equal to a.

PROOF. According to Lemma 1.3 with $J = \emptyset$, it follows that there exists a finite set S_1 in E_k such that every $t \in E_k$ with $H_f(t) \geqq \varepsilon$ has at least one component in common with one of the points in S_1. Now we define finite sets $S_2, \cdots, S_k$ by induction. For fixed j with $1 \leqq j < k$ let H_j consist of all points having at least j components in common with one of the points in S_j. A second application of Lemma 1.3 yields the existence of a finite set $S \subset H_j$ with the property that every point $t \in H_j$ coinciding in at least j components, say $t_{i_1}, \cdots, t_{i_j}$, with some point $s \in S$ and having a "jump" $H(t, \{i_1, \cdots, i_j\}) \geqq \varepsilon$ has a further component $t_{i_{j+1}}$ in common with s. Putting $S_{j+1} \equiv S$ the partition $\mathscr{R} = \mathscr{R}(\bigcup_{j=1}^k S_j)$ will satisfy the assertion of the lemma. This follows from the fact that, if there does not exist a $\delta > 0$ as stated above, we can find a rectangle $R \in \mathscr{R}$, sequences $\{t_n{}^l\}_n$, $l = 1, 2$, in R with $|f(t_n{}^1) - f(t_n{}^2)| \geqq \varepsilon \; \forall n$ and a point $t \in \bar{R}$ such that $\{t_n{}^l\}_n$ converges to t in the way that each component converges either monotonically in the strict sense, or is identically constant. Then the construction of $\mathscr{R}$ implies $\{t_n{}^1\}_n, \{t_n{}^2\}_n \in \tilde{Q}(\rho, t)$ for some $\rho \in \mathscr{P}$ which contradicts the existence of $f(t+0_\rho)$. □

REMARK. It is clear that by enlarging the number of generating points of $\mathscr{R}$ we can assume in the above lemma that $\sup \{|t-t'| : t, t' \in R\} < \delta$, $R \in \mathscr{R}$.

COROLLARY 1.6. *Every $f \in D_k$ is bounded.*

The "only if" part of Theorem 1.4 is a trivial consequence of Lemma 1.5 and the Remark. As regards the "if" part, let $t \in E_k$, $\rho \in \mathscr{P}$ and sequences $\{t_n\} \subset Q(\rho, t)$ be given with $t_n \to t$. The existence of $f(t+0_\rho)$ will follow if we show that $\{f(t_n)\}$ is a Cauchy sequence. For $\varepsilon > 0$ we conclude from (1.6) that there exists a partition $\mathscr{R}$ with $\max \{\omega_f(R) : R \in \mathscr{R}\} < \varepsilon$. Then we can find a rectangle $R_t \in \mathscr{R}$ and an integer n_0 with $t_n \in R_t$ for $n \geqq n_0$. But then $|f(t_n) - f(t_m)| < \varepsilon$ for $n, m \geqq n_0$. The continuity from above follows in a similar manner.

For a function $f: E_k \to \mathbb{R}$ let $||f||$ denote the supremum-norm. It can easily be shown that $||f_n - f|| \to 0$ with $f_n \in D_k$ implies $f \in D_k$. Clearly, primitive functions f of the form $f = \sum_{R \in \mathscr{R}} a_R \cdot I_R$ where $\mathscr{R}$ is a partition, I_R the indicator of R and $a_R \in \mathbb{R}$, $R \in \mathscr{R}$, belong to D_k as well as their uniform limits. On the other hand, according to Lemma 1.5, every $f \in D_k$ is the uniform limit of such primitive functions. Consequently, D_k can be characterized as the class of all uniform limits of primitive functions of the above form.

Applying the arguments of Billingsley [2], page 110, 111, to each of the k components of $t \in E_k$ one obtains for functions $f \in D_k$

$$\omega_f{}'(\delta) \leqq \omega(2\delta), \qquad 0 < \delta < \tfrac{1}{2}, \tag{1.7}$$

and for continuous functions

$$\omega_f(\delta) \leqq 2k\omega_f{}'(\delta), \qquad 0 < \delta < 1. \tag{1.8}$$

2. The Skorohod topology in D_k. In this section we generalize the well-known Skorohod metric on $D[0, 1]$ to a metric d on D_k and record some results on it. Since that generalization consists only in an extension from one to k dimensions and, since D_k generalizes $D[0, 1]$ in such a way that Theorem 1.4 holds in the same form as in $D[0, 1]$, we can carry over all the proofs of Billingsley [2], page 111–118, by componentwise argumentation. For completeness and easy reference we will only record the results.

Let Λ denote the class of all strictly increasing continuous mappings μ from $[0, 1]$ onto itself. For $\lambda = (\lambda_1, \cdots, \lambda_k) \in \Lambda_k = \Lambda \times \cdots \times \Lambda$, k times, and $t = (t_1, \cdots, t_k) \in E_k$ we write $\lambda t = (\lambda_1 t_1, \cdots, \lambda_k t_k)$.

The Skorohod distance $d(f, g)$ of functions $f, g \in D_k$ is defined to be the infimum of all $\varepsilon > 0$ such that there exists a $\lambda = \lambda(\varepsilon) \in \Lambda_k$ with

$$\sup\{|\lambda t - t| : t \in E_k\} \leqq \varepsilon \tag{2.1}$$

and

$$\sup\{|f(t) - g(\lambda t)| : t \in E_k\} \leqq \varepsilon. \tag{2.2}$$

The distance function d is a metric on D_k, called Skorohod-metric, which generates the Skorohod- or d-topology on D_k. A necessary and sufficient condition for a sequence $\{f_n\} \subset D_k$ to converge to some $f \in D_k$ in the d-topology, shortly $f_n \to_d f$, is that there exists a sequence $\{\lambda_n\} \subset \Lambda_k$ with

$$\lim f_n(\lambda_n t) = f(t) \qquad \text{uniformly in} \quad t, \tag{2.3}$$

and

$$\lim \lambda_n t = t \qquad \text{uniformly in} \quad t. \tag{2.4}$$

The metric space (D_k, d) is separable. A countable, dense subset $\vartheta \subset D_k$ is obtained as follows. For $n \in \mathbb{N}$ let $\mathscr{R}_n$ be the partition generated by the set of points $t = (t_1, \cdots, t_k) \in E_k$ with $t_i \in \{0, 1/n, 2/n, \cdots, 1\}$, $i \in \mathbb{N}_k$, and ϑ_n be the countable class of functions $\sum_{R \in \mathscr{R}_n} a_R I_R$ where a_R are rational numbers. Then $\vartheta = \bigcup_n \vartheta_n$ is countable, dense in D_k.

For $\mu \in \Lambda$ define

$$|||\mu||| = \sup\left\{\left|\log \frac{\mu u - \mu v}{u - v}\right| : u, v \in [0, 1], u \neq v\right\}. \tag{2.5}$$

Now, replacing the condition (2.1) by

$$|||\lambda_i||| \leqq \varepsilon, \qquad i = 1, \cdots, k, \tag{2.6}$$

we get another metric d_0 on D_k.

The metrics d and d_0 are equivalent and (D_k, d_0) is a complete metric space. Just as in Billingsley [2], page 113, we have the following relations between d and d_0:

$$d(f, g) \leqq 2 d_0(f, g) \qquad \text{if} \quad d_0(f, g) < \tfrac{1}{4} \tag{2.7}$$

and

$$d_0(f, g) \leqq 4\delta + \omega_f{}'(\delta) \qquad \text{if} \quad d(f, g) < \delta^2 \quad \text{and } 0 < \delta < \tfrac{1}{4}. \tag{2.8}$$

On C_k the d-, d_0- and $\|\cdot\|$-topology coincide.

Furthermore, relative compact sets in (D_k, d) have the same Arzela–Ascoli form of characterization as in the one-dimensional case.

THEOREM 2.1. *A set $A \subset D_k$ has compact closure in the d-topology of D_k if and only if*

$$\sup_{f \in A} \|f\| < \infty \tag{2.9}$$

and

$$\lim_{\delta \to 0} \sup_{f \in A} \omega_f'(\delta) = 0. \tag{2.10}$$

For points $t_j \in E_k$, $j = 1, \cdots, r$, and functions $f: E_k \to \mathbb{R}$ the projection $\pi_{t_1, \cdots, t_r}$ is defined by $\pi_{t_1, \cdots, t_r} f = (f(t_1), \cdots, f(t_r))$. As in the one-dimensional case a projection π_t is continuous at $f \in D_k$ if and only if f is continuous at t. The Borel σ-algebra $\mathscr{L}_d$ on (D_k, d) is the smallest σ-algebra making the projections π_t, $t \in T$, measurable. Here T denotes an arbitrary dense subset of E_k.

This shows that the finite dimensional sets $\pi_{t_1, \cdots, t_r}^{-1} H$, with H running through the Borel σ-algebra in $\mathbb{R}_r$ and $t_1, \cdots, t_r \in T$, $r \in \mathbb{N}$, form a determining class in the sense of Billingsley [2], page 15.

3. Weak convergence of measures on $(D_k, \mathscr{L}_d)$. For a probability measure P on $(D_k, \mathscr{L}_d)$ let T_P be the set of all $t \in E_k$ for which π_t is continuous except on a set of P-measure zero. Then $T_P = \{t \in E_k: P(J_t) = 0\}$ with $J_t = \{f \in D_k: \pi_t$ discontinuous at $f\} = \{f \in D_k: H_f(t) > 0\} = \{f \in D_k: f$ discontinuous at $t\}$ (see Section 2). J_t is an element of $\mathscr{L}_d$.

For fixed P there are countably many points $\{t_n\}$ in E_k such that from $P(J_t) > 0$ it follows that t has one component in common with some t_n. The proof of this fact is analogous to that in Billingsley [2], page 124, taking account of Lemma 1.3. Thus: All points t with $P(J_t) > 0$ are concentrated on countably many (proper) hyperplanes. T_P contains $\mathscr{P}$ and is dense in E_k. If all $t_1, \cdots, t_r$ lie in T_P, then $\pi_{t_1, \cdots, t_r}$ is P–a.e. continuous. From these remarks we see by standard calculations that a relatively sequentially compact sequence P_n, $n = 0, 1, 2, \cdots$, of probability measures on $(D_k, \mathscr{L}_d)$ converges weakly, $P_n \to_{\mathscr{D}} P_0$, if (3.1) holds:

$$P_n^{t_1, \cdots, t_r} \to_{\mathscr{D}} P_0^{t_1, \cdots, t_r}, \qquad t_1, \cdots, t_r \in T_{P_0}, r \in \mathbb{N}. \tag{3.1}$$

Here we have used notation $P^{t_1, \cdots, t_r} \equiv P\pi_{t_1, \cdots, t_r}^{-1}$ and the fact that for probability measure P and Q on $\mathscr{L}_d$ the set $T_P \cap T_Q$ is dense in E_k.

The Arzela–Ascoli-like characterization of compact sets in (D_k, d) (Theorem 2.1.) leads to the following criterion.

THEOREM 3.1. *A sequence $\{P_n\}$ of probability measures on $(D_k, \mathscr{L}_d)$ is relatively sequentially compact if and only if* (3.2) *and* (3.3) *hold*:

(3.2) *For every $\varepsilon > 0$ there exists a M_ε, $0 < M_\varepsilon < \infty$, with*

$$P_n(\{f \in D_k: \|f\| > M_\varepsilon\}) \leqq \varepsilon, \qquad n \geqq 1$$

and

$$\lim_{\delta \to 0} \limsup_{n \to \infty} P_n(\{f \in D_k: \omega_f'(\delta) \geqq \varepsilon\}) = 0, \qquad \varepsilon > 0. \tag{3.3}$$

In our later applications we can show even more, namely

(3.4) $\{P_n^{t_1,\cdots,t_r}\}_n$ converges in distribution, $t_1,\cdots,t_r \in E_k, r \in \mathbb{N}$,

and

(3.5) $\lim_{\delta\to 0}\limsup_{n\to\infty} P_n(\{f \in D_k : \omega_f(\delta) \geqq \varepsilon\}) = 0, \quad \varepsilon > 0.$

Using that projections π_t, $t \in E_k$, are $\mathscr{L}_d$-measurable, one can prove that the subsets of D_k occurring in (3.2), (3.3) and (3.5) are elements of $\mathscr{L}_d$.

Though (3.4) is often satisfied in applications, let us define the weaker condition

(3.6) $\{P_n^0 = \{P_n\pi_0^{-1}\}$ converges in distribution.

Then, using the same arguments as in the proof of Theorem 15.5 of Billingsley [2] we can show: Condition (3.5) implies (3.3) and from (3.5 and 3.6) condition (3.2) follows. Consequently, (3.5 and 3.6) entail (3.2 and 3.3). Furthermore, (3.5) then causes that every weak limit P of a subsequence $\{P_{n'}\}$ must fulfill $P(C_k) = 1$. Replacing (3.6) by (3.4) in (3.5 and 3.6), it follows that all subsequences $\{P_{n'}\}$ must have the same limit (for this it is used that the finite dimensional sets form a determining class). Altogether we have: (3.4) and (3.5) entail the existence of a probability P on $(D_k, \mathscr{L}_d)$ with $P(C_k) = 1$ such that $P_n \to_{\mathscr{D}} P$.

As was shown by Dudley and Wichura, see e.g. Wichura [12], Theorem 2, conditions (3.4) and (3.5) are necessary and sufficient for a sequence $\{P_n\}$ of probabilities on $(D_k, \mathscr{L}_d)$ to converge weakly (in their sense) to some probability P on $(D_k, \mathscr{L}_d)$ with $P(C_k) = 1$. In this special case, however, where the limiting measure gives probability one to C_k our concept of weak convergence coincides with that of Dudley and Wichura.

4. Random variables with values in D_k. Let $(\Omega, \mathscr{A}, \mathbf{P})$ be a probability space and X a measurable map from $(\Omega, \mathscr{A})$ into $(D_k, \mathscr{L}_d)$. Then X is called a D_k-valued random variable (D_k-rv). According to the fact that $\mathscr{L}_d$ is generated by the finite-dimensional sets, a map $X: \Omega \to D_k$ is a D_k-rv if and only if $X(t) \equiv X_t \equiv \pi_t \circ X$ is a real valued random variable for every $t \in E_k$. A D_k-rv is a stochastic process in the usual sense (with "time" space E_k) and vice versa every stochastic process in the usual sense with almost all realizations in D_k can be regarded as a D_k-rv. The same remarks hold for C_k instead of D_k. The statements in Section 3 may be rewritten in terms of D_k- or C_k-valued random variables in an obvious way, cf. Billingsley [1], page 22.

We shall consider some examples of D_k-valued random variables. For that purpose let $U_j = (U_{j1}, \cdots, U_{jk})$, $j \in \mathbb{N}$, be independent identically distributed random vectors where all the U_{ji} are uniformly distributed over (0, 1) with $U_{ji} \in (0, 1)$. Then the distribution function F of U_j satisfies a Lipschitz condition

(4.1) $$|F(t)-F(t')| \leqq k|t-t'|.$$

REMARK. In our considerations there is no loss of generality in assuming that the marginals of U_j are uniform. Otherwise—since F is continuous—this can be achieved by suitable componentwise transformations.

Now, for $n \notin \mathbb{N}$ we define D_k-valued random variables X_n^F and Y_n by

$$X_n^F(t) = \frac{1}{n^{\frac{1}{2}}}\left[\sum_{j=1}^{n} I_{[0,t]}(U_j) - nF(t)\right], \qquad t \in E_k, \tag{4.2}$$

where $[0, t]$ for $t = (t_1, \cdots, t_k) \in E_k$ denotes the Cartesian product $[0, t_1] \times \cdots \times [0, t_k]$, and

$$Y_n(t) = \frac{1}{n^{\frac{1}{2}}}\left[\sum_{j=1}^{n} \prod_{i=1}^{k} (I_{[0,t_i]}(U_{ji}) - t_i)\right], \qquad t \in E_k. \tag{4.3}$$

In the one-dimensional case ($k = 1$) X_n^F and Y_n are identical and they are equal to the well-known one-dimensional empirical process.

If the components in each U_j are independent, i.e. if $F(t) = t_1 \cdots t_k$, there is an explicit relationship between Y_n and $X_n = X_n^F$: Let L denote an ordered subset of $\mathbb{N}_k$ and $|L|$ the number of elements in L. Define for $t \in E_k$ the vector t_L replacing t_i in t by 1 if $i \notin L$. Now, an easy calculation leads to:

$$Y_n(t) = \textstyle\sum_L F(t_L) \cdot X_n(t_{\mathbb{N}_k - L}) \cdot (-1)^{|L|}, \tag{4.4}$$

where the sum extends over all ordered subsets of $\mathbb{N}_k$. From relation (4.4) it follows at once that if (3.5) is satisfied for X_n, then it is satisfied for Y_n, too.

For points $t_1, \cdots, t_r$ in E_k a standard calculation shows:

$$(X_n^F(t_1), \cdots, X_n^F(t_r)) \to_{\mathscr{D}} \mathscr{N}(0, \gamma_{X^F}(t_1, \cdots, t_r)), \tag{4.5}$$

where the right-hand side denotes an r-dimensional normal distribution with expectation vector 0 and covariance matrix

$$\gamma_{X^F}(t_1, \cdots, t_r) = (F(t_\nu \wedge t_\mu) - F(t_\nu)F(t_\mu))_{\nu,\mu=1,\cdots,r};$$

here $t_\nu \wedge t_\mu$ denotes the vector $(\min(t_{\nu 1}, t_{\mu 1}), \cdots, \min(t_{\nu k}, t_{\mu k}))$. Analogue considerations for Y_n are only possible if the components in each U_j are independent, i.e. $F(t) = t_1 \cdots t_k$. In this case we have

$$(Y_n(t_1), \cdots, Y_n(t_r)) \to_{\mathscr{D}} \mathscr{N}(0, \gamma_Y(t_1, \cdots, t_r)), \tag{4.6}$$

where

$$\gamma_Y(t_1, \cdots, t_r) = (\textstyle\prod_{i=1}^{k} (t_{\nu i} \wedge t_{\mu i} - t_{\nu i} \cdot t_{\mu i}))_{\nu,\mu=1,\cdots,r}.$$

In the next section we shall prove that (3.5) is fulfilled for $\{X_n^F\}$. This implies, according to (4.5) and the results at the end of Section 3, that $\{X_n^F\}$ converges weakly to a D_k-valued random variable X^F with $P(X^F \in C_k) = 1$ and with finite dimensional distributions given by the right side of (4.5). Because of (4.4) and (4.6) an analogue statement is true for $\{Y_n\}$ with some limiting process Y.

5. Convergence of the empirical processes X_n^F and Y_n. In order to verify condition (3.5) for X_n^F we shall extend the "classical" proof for the one-dimensional case (see Parthasarathy [9], page 262) using some special "k-dimensional" considerations where the classical proof does not work.

Since $X_n^F(t)+n^{\frac{1}{2}}F(t)$ is a monotonically increasing function in each component of t, it follows as in Pathasarathy [9], page 262, that the modulus of continuity $\omega_{X_n^F}(\delta)$ may be bounded essentially by a modulus of continuity where the "sup" is to be taken only over points from $L(r)$, where $L(r)$ is the set of all points $(l_1, \cdots, l_k)r^{-1}$ with $l_i \in \mathbb{N}_r \cup \{0\}$, $i \in \mathbb{N}_k$. More precisely, we have

$$\omega_{X_n^F}(\delta) \leqq 6 \sup |X_n^F(l_1) - X_n^F(l_2)| + 4kn^{\frac{1}{2}}2^{-m} \tag{5.1}$$

where the "sup" is to be taken over all $l_1, l_2 \in L(2^m)$ with $|l_1 - l_2| \leqq \delta + 2^{-m+1}$. For $m_n = \max\{m : n \geqq 2^m\}$ we have

$$n2^{-m_n} \geqq 1, \qquad n \geqq 1, \qquad \text{and} \quad n^{\frac{1}{2}}2^{-m_n} \to_n 0. \tag{5.2}$$

Consequently, (3.5) will follow if we show that

$$\lim_{\delta \to 0} \limsup_{n \to \infty} \mathbf{P}(\sup |X_n^F(l_1) - X_n^F(l_2)| > \varepsilon) = 0 \quad \text{for all} \quad \varepsilon > 0, \tag{5.3}$$

where the "sup" is to be taken over $l_1, l_2 \in L(2^{m_n})$ with $|l_1 - l_2| \leqq \delta$. To carry out the proof of (5.3) we shall need two auxiliary results.

LEMMA 5.1. *Let $\delta > 0$, $m(\delta) = \max\{r \in \mathbb{N} : \delta \cdot 2^r \leqq 1\}$, $m \in \mathbb{N}$ with $m > m(\delta)$ and $f : E_k \to \mathbb{R}$ be given. Then for points $t_1, t_2 \in L(2^m)$ with $|t_1 - t_2| \leqq 2^{-m(\delta)}$ the following inequality holds*

$$|f(t_1) - f(t_2)| \leqq 4k \cdot \sum_{r=m(\delta)}^{m} \sum_{\mu=1}^{k} \sup_j |f(j) - f(j + e_\mu 2^{-r})|, \tag{5.4}$$

where the "sup" *is to be taken over all $j \in L(2^r)$ with $j + e_\mu 2^{-r} \in E_k$ (e_μ denotes the μ-*th unit-vector of *E_k).*

LEMMA 5.2. *Let C be a positive constant and p, n, h positive numbers which fulfill simultaneously the conditions*

$$1/n \leqq h \leqq 1, \qquad p \leqq C \cdot h. \tag{5.5}$$

Then for the central moments $\mu_l(n, p)$ of a binomial distribution $B(n, p)$, there exists a constant K_l (independent of p, n and h) such that (5.6) *holds*

$$|\mu_l(n, p)| \leqq K_l \cdot n^{l/2} h^{l/2}. \tag{5.6}$$

Before giving comments to the proof of the lemmas, we first prove (5.3) using the lemmas. Applying (5.4) with $m = m_n$ and $f = X_n^F$, we get

$$\sup\{|X_n^F(l_1) - X_n^F(l_2)| : l_1, l_2 \in L(2^{m_n}), |l_1 - l_2| \leqq \delta\}$$
$$\leqq 4k \sum_{\mu=1}^{k} \sum_{r=m(\delta)}^{m_n} \sup |X_n^F(j) - X_n^F(j + e_\mu 2^{-r})|.$$

If the left-hand side exceeds ε, then there exists a μ such that

$$\sum_{r=m(\delta)}^{m_n} \sup_j |X_n^F(j) - X_n^F(j + e_\mu 2^{-r})| > \varepsilon(4k^2)^{-1}.$$

Let a be a real number with $0 < a < 1$ and $2 \cdot a^{2(k+1)} > 1$. Hence we can find at least one $r \in \{m(\delta), \cdots, m_n\}$ with

$$\sup_j |X_n^F(j) - X_n^F(j + e_\mu 2^{-r})| > (1-a)a^{r-m(\delta)} \cdot \varepsilon(4k^2)^{-1}.$$

It follows that

$$\mathbf{P}(\sup|X_n^F(l_1)-X_n^F(l_2)|>\varepsilon) \leqq \sum_{\mu=1}^{k}\sum_{r=m(\delta)}^{m_n}\mathbf{P}(\sup_j|X_n^F(j)-X_n^F(j+e_\mu 2^{-r})| > (1-a)a^{r-m(\delta)}\varepsilon(4k^2)^{-1})$$

$$\leqq \sum_{\mu=1}^{k}\sum_{r=m(\delta)}^{m_n}\sum_{j_1=1}^{2^r}\cdots\sum_{j_\mu=0}^{2^r-1}\cdots\sum_{j_k=1}^{2^r}\mathbf{P}[|X_n^F(j)-X_n^F(j+e_\mu 2^{-r})| > (1-a)a^{r-m(\delta)}(\varepsilon/4k^2)]$$

$$\leqq \sum_{\mu=1}^{k}\sum_{r=m(\delta)}^{m_n}\sum_{j_1=1}^{2^r}\cdots\sum_{j_\mu=0}^{2^r-1}\cdots\sum_{j_k=1}^{2^r}\left(\frac{4k^2}{\varepsilon(1-a)a^{r-m(\delta)}}\right)^{2(k+1)} \cdot E(X_n^F(j)-X_n^F(j+e_\mu 2^{-r}))^{2(k+1)}.$$

Since $n^{\frac{1}{2}}(X_n^F(j+e_\mu 2^{-r})-X_n^F(j))$ has a centred $B(n,p)$ distribution with $p = F(j+e_\mu 2^{-r})-F(j) \leqq k2^{-r}$, (see 4.1)), we can apply Lemma 5.2 with $C = k$ and $h = 2^{-r} \geqq 2^{-m_n} \geqq n^{-1}$. Then the above chain of inequalities can be extended to

$$\leqq \sum_{\mu=1}^{k}\sum_{r=m(\delta)}^{m_n}\sum_{j_1=1}^{2^r}\cdots\sum_{j_\mu=0}^{2^r-1}\cdots\sum_{j_k=1}^{2^r}\left(\frac{4k^2}{\varepsilon(1-a)a^{r-m(\delta)}}\right)^{2(k+1)} K_{2(k+1)}\cdot\frac{1}{2^{r(k+1)}}$$

$$= \sum_{r=m(\delta)}^{m_n}\left[k\cdot\left(\frac{4k^2}{\varepsilon(1-a)}\right)^{2(k+1)} K_{2(k+1)}\right]\cdot\frac{1}{2^r}\left(\frac{1}{a^{(r-m(\delta))}}\right)^{2(k+1)}.$$

If H denotes the expression in the square brackets we therefore get

$$\mathbf{P}(\sup|X_n^F(l_1)-X_n^F(l_2)|>\varepsilon) \leqq H\cdot\frac{1}{2^{m(\delta)}}\sum_{r=m(\delta)}^{m_n}\frac{1}{2^{(r-m(\delta))}}\cdot\left(\frac{1}{a^{r-m(\delta)}}\right)^{2(k+1)}$$

$$\leqq H\cdot\frac{1}{2^{m(\delta)}}\sum_{\rho=0}^{\infty}\left(\frac{1}{a^{2(k+1)}}\cdot\frac{1}{2}\right)^{\rho}.$$

Because of $2\cdot a^{2(k+1)} > 1$ the infinite series converges and since moreover, $m(\delta)\to\infty$ for $\delta\to 0$, the proof of (5.3) and therefore that of $X_n^F \to_{\mathscr{D}} X^F$ and of $Y_n \to_{\mathscr{D}} Y$ is complete.

As regards the proof of Lemma 5.1, the relevant considerations for the case $k = 2$ can be found in the proof of Lemma 1 of Kuelbs [7]. The extension of the method of Kuelbs to k dimensions, $k > 2$, then is a natural one but rather long to describe.

Lemma 5.2 can be proved by induction on l: For $l = 1$ (5.6) is obviously satisfied. Let the assertion be true for $l' \leqq l-1$, $l \geqq 2$. Then according to Kendall–Stuart [5], page 122,

$$\mu_l(n,p) = \sum_{j=0}^{l-2}[npq\mu_j(n,p)-p\mu_{j+1}(n,p)]\binom{l-1}{j}$$

and, consequently,

$$|\mu_l(n,p)| \leqq C\cdot n^{l/2}\cdot\sum_{j=0}^{l-2}[K_j h^{(j+2)/2}n^{(j+2-l)/2}+K_{j+1}h^{(j+3)/2}n^{(j-l+1)/2}]\binom{l-1}{j}.$$

Since $1/n \leqq h \leqq 1$ we get (5.6) with $K_l \equiv C\cdot\sum_{j=0}^{l-2}\binom{l-1}{j}(K_j+K_{j+1})$.

As another application we use the weak convergence of X_n and Y_n for $k = 2$ to prove that under the null hypothesis the well-known Kolmogorov–Smirnov test statistic of independence converges in distribution (in the case of continuous distribution function), cf. Blum, Kiefer, Rosenblatt [3]. After a standard reduction we may assume that independent random vectors $U_j = (U_{j1}, U_{j2})$, $j \geqq 1$, are given with uniform distribution on E_2 and $U_{ji} \in (0, 1)\ \forall i, \forall j$. Then the Kolmogorov–Smirnov statistic for $n \in \mathbb{N}$ is $||Z_n||$, where Z_n is defined by

$$(5.9)\quad Z_n(t) = n^{\frac{1}{2}}\left[\frac{1}{n}\sum_{j=1}^{n} I_{[0,t_1]}(U_{j1})\cdot I_{[0,t_2]}(U_{j2}) - \frac{1}{n^2}\sum_{j=1}^{n} I_{[0,t_1]}(U_{j1})\cdot \sum_{j=1}^{n} I_{[0,t_2]}(U_{j2})\right]$$

for $t = (t_1, t_2) \in E_2$. Z_n is a D_2-valued random variable and can be rewritten as

$$(5.10)\qquad Z_n(t) = Y_n(t) - \frac{1}{n^{\frac{1}{2}}} X_n(t_{L_1})X_n(t_{L_2}), \qquad t \in E_2,$$

with $L_1 \equiv \{1\}$, $L_2 \equiv \{2\}$(t_L was defined before (4.4)). Now, from $X_n \to_{\mathscr{D}} X$ it follows that the last term in (5.10) tends to zero in probability. Hence Z_n and Y_n converge weakly to the same limit, namely Y, so that $||Z_n|| \to_{\mathscr{D}} ||Y||$.

REFERENCES

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

[2] Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

[3] Blum, J. R. and Kiefer, J. and Rosenblatt, M. (1961). Distribution free tests of independence based on the sample distribution function. *Ann. Math. Statist.* **32** 485–498.

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

[5] Kendall, G. M. and Stuart, A. (1963). *The Advanced Theory of Statistics* **1**. Charles Griffin, London.

[6] Kiefer, J. and Wolfowitz, J. (1958). On the deviations of the empiric distribution function of vector chance variables. *Trans. Amer. Math. Soc.* **87** 173–186.

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

[8] LeCam, L. (1959). Remarques sur les variables aleatoires dans les espace vectoriels non separables. Le Calcul des Probabilities et ses Applications, C.N.R.S., Paris.

[9] Parthasarathy, K. R. (1967). *Probability Measures on Metric Spaces.* Academic Press, New York–London.

[10] Straf, M. L. (1969). A general Skorohod space and its application to the weak convergence of stochastic processes with several parameters. Ph.D. thesis, Univ. of Chicago.

[11] Wichura, M. J. (1968). On the weak convergence of non-Borel probabilities on a metric space. Ph.D. thesis, Columbia Univ.

[12] Wichura, M. J. (1969). Inequalities with applications to the weak convergence of random processes with multi-dimensional time parameters. *Ann. Math. Statist.* **40** 681–687.

[13] Winkler, W. (1962). Stetigkeitseigenschaften der Realisierungen Gaußscher zufälliger Felder. *Trans. Third Prague Conf.* 831–839.