The Annals of Mathematical Statistics
1969, Vol. 40, No. 2, 681–687

INEQUALITIES WITH APPLICATIONS TO THE WEAK CONVERGENCE OF RANDOM PROCESSES WITH MULTI-DIMENSIONAL TIME PARAMETERS[1]

BY MICHAEL J. WICHURA

University of Chicago

0. Summary. Bounds on the distribution of the maximum of the absolute values of the "partial sums" constructed from a multi-dimensional array of independent random variables are derived and used to establish the weak convergence of certain random functions defined in terms of such "partial sums".

1. Statement of the results. Let q and $n_1, n_2, \cdots, n_q$ be positive integers, and let $D_{j_1,\cdots,j_q}(1 \leqq j_p \leqq n_p, 1 \leqq p \leqq q)$ be independent random variables with zero means and finite variances. Set

$$S_{k_1,\cdots,k_q} = \sum_{1 \leqq p \leqq q} \sum_{1 \leqq j_p \leqq k_p} D_{j_1,\cdots,j_q}$$

for $1 \leqq k_p \leqq n_p$, $1 \leqq p \leqq q$, and set

$$M = \max \{|S_{k_1,\cdots,k_q}| : 1 \leqq k_p \leqq n_p, 1 \leqq p \leqq q\},$$

$$\sigma^2 = ES^2_{n_1,\cdots,n_q} = \sum_{1 \leqq p \leqq q} \sum_{1 \leqq j_p \leqq n_p} ED^2_{j_1,\cdots,j_q}.$$

THEOREM 1. *In the above framework,*

$$EM^2 \leqq 4^q \sigma^2. \tag{1}$$

Moreover, if $\sigma^2 < a^2$, *then*

$$P\{M > 2^q a\} \leqq (1 - (\sigma/a)^2)^{-q} P\{|S_{n_1,\cdots,n_q}| > a\}, \tag{2a}$$

$$P\{M > 2^q a\} \leqq (1 - (\sigma/a)^2)^{-q} (\sigma/a)^2. \tag{2b}$$

Also, if $\sigma^2 < (2a)^2/(q-1)$, *then*

$$P\{M > 2^q a\} \leqq (1 - (\sigma/a)^2)^{-1}(1 - (q-1)(\sigma/2a)^2)^{-1} P\{|S_{n_1,\cdots,n_q}| > a\}, \tag{3a}$$

$$P\{M > 2^q a\} \leqq (1 - (q-1)(\sigma/2a)^2)^{-1}(\sigma/2a)^2. \tag{3b}$$

When σ^2 is small compared to a^2, the inequalities (3a) and (3b) are sharper than (2a) and (2b) respectively. When $q = 1$, (1) is a special case of an inequality of Doob ([2], p. 317), (2a) and (3a) reduce to a well known but nameless inequality (see [4], p. 219), and (3b) is just Kolmogorov's inequality. For the case $q = 2$, Galen Shorack has observed that a limited Hajek-Renyi type inequality extending (3b) holds:

Received 16 October 1968.
[1] This research was supported in part by Research Grant No. NSF GP 8026 from the Division of Mathematical, Physical and Engineering Sciences of the National Science Foundation, and, in part from the Statistics Branch, Office of Naval Research. Reproduction in whole or in part is permitted for any purpose of the United States Government.

PROPOSITION. *If* $c_1 \geqq c_2 \geqq \cdots \geqq c_{n_1} \geqq 0$ *and if the assumptions of Theorem* 1 *hold, then, provided that* $\sum_{k_1} \sigma_{k_1}^2 c_{k_1}^2 < 4a^2$, *where* $\sigma_{k_1}^2 = \sum_{k_2} ED_{k_1,k_2}^2$, *one has*

$$\text{(3b}') \quad P\{\max_{k_1,k_2} c_{k_1}|S_{k_1,k_2}| > 4a\} \leqq (\sum \sigma_{k_1}^2 c_{k_1}^2/4a^2)/(1 - (\sum \sigma_{k_1}^2 c_{k_1}^2/4a^2)).$$

It is not known whether a full Hajek-Renyi type inequality holds for general q.

Pyke and Shorack point out an application of (3b$'$) in their work [5] on random sample size Chernoff-Savage theorems. In this paper, we present an application of (2a), an inequality which, like (3a), is rather strong whenever the number a is sufficiently large and the distribution of $S_{n_1,\cdots,n_q}$ is sufficiently close (in the topology of weak convergence) to a distribution, such as the standard normal, whose tail probabilities decrease very rapidly. We first introduce some preliminary material concerning a mode of weak convergence of probability distributions over a certain function space.

Let $[0, 1]^q$ denote the unit cube in R^q. Let us agree to call a function $x: [0, 1]^q \rightarrow R$ (the real line) a step function if x is a linear combination of functions of the form

$$t \rightarrow I_{E_1 \times E_2 \times \cdots \times E_q}(t),$$

where each E_p is either a left closed, right open subinterval of the unit interval $[0, 1]$ or the singleton $\{1\}$, and where I_E denotes the indicator function of the set E. Let D_q be the uniform closure, in the space of all bounded functions from $[0, 1]^q$ to R, of the vector subspace of step functions. The functions belonging to D_q are continuous from above with limits from below, in a sense made explicit in [6]. Let $\mathcal{A}$ be the σ-algebra of subsets of D_q generated by the projection mappings $\pi_t: x \rightarrow x(t)$ $(t \varepsilon [0, 1]^q)$. The mode of weak convergence with which we work is given in the following

DEFINITION 1. Let P_n $(n \geqq 1)$ and P be probabilities on $(D_q, \mathcal{A})$. The sequence $(P_n)_{n \geqq 1}$ is said to converge weakly to P in the U-topology if for every function $f: D_q \rightarrow R$ which is both measurable between $\mathcal{A}$ and the Borel σ-algebra of R and continuous in the topology of uniform convergence on D_q, the sequence $(P_n f^{-1})_{n \geqq 1}$ of induced distributions converges weakly (in the usual sense; see [1] or [4]).

For each finite subset T of $[0, 1]^q$, let $\pi_T: D_q \rightarrow R^T$ be defined by $\pi_T(x) = (x(t))_{t \varepsilon T}$, and for each $\delta > 0$, let $w_\delta: D_q \rightarrow R$ be defined by

$$w_\delta(x) = \sup\{|x(t) - x(s)| : s, t \varepsilon [0, 1]^q, |t - s| < \delta\},$$

where $|u| = \max_p |u_p|$ for $u = (u_1, \cdots, u_q)$. Let

$$C_q = \{x \varepsilon D_q : \limsup_{\delta \rightarrow 0} w_\delta(x) = 0\}$$

be the subset of D_q consisting of all continuous functions mapping $[0, 1]^q$ into R. The next theorem is a direct consequence of Corollary 2.3 of [6] (weak convergence in the U-topology on D_q is the same kind of weak convergence discussed in [6], provided the limit is tight with respect to the compact sets of the topology of uniform convergence in D_q—in which case the limit can be extended to the Borel σ-algebra for the same topology; every probability on $(D_q, \mathfrak{A})$ concentrated in C_q has this tightness property).

THEOREM 2. *Let $(P_n)_{n \geqq 1}$ be a sequence of probabilities on $(D_q, \mathfrak{A})$. Then there exists a probability P on $(D_q, \mathfrak{A})$, giving probability one to C_q, to which the P_n converge weakly in the U-topology if and only if the following two conditions hold:*

(i) *for each finite subset T of $[0, 1]^q$, the "finite-dimensional" distributions $P_n \pi_T^{-1}$ converge weakly, and*

(ii) *for each $\epsilon > 0$,* $\lim_{\delta \to 0} \limsup_n P_n\{w_\delta > \epsilon\} = 0$.

In this case, the limiting distribution P is determined by the fact that its finite-dimensional distributions are the weak limits of the corresponding finite-dimensional distributions of the P_n.

Now, for each $n \geqq 1$, let $\beta_{j_1, \cdots, j_q}(n)$ $(0 \leqq j_p \leqq n, 1 \leqq p \leqq q)$ be any random variables. For each $t = (t_1, \cdots, t_q) \in [0, 1]^q$, put

$$X_n(t) = \sum_{1 \leqq p \leqq q} \sum_{0 \leqq j_p \leqq [nt_p]} \beta_{j_1, \cdots, j_q}(n), \tag{4}$$

where $[c]$ denotes the greatest integer in the number c. The random function $X_n = (X_n(t))_{t \in [0,1]^q}$ takes values in D_q and is measurable with respect to the σ-algebra $\mathfrak{A}$; it therefore induces a probability distribution on $\mathfrak{A}$ which we shall denote by $L(X_n)$. Here now is the promised application of Theorem 1:

THEOREM 3. *For each $n \geqq 1$, let the random variables $\beta_{j_1, \cdots, j_q}(n)$ be independent with zero means and finite variances, and let the random functions X_n be defined by means of formula (4). If*

(i) *for each finite subset T of $[0, 1]^q$, the distribution of $X_n \pi_T^{-1}$ converges weakly to a normal limit G_T, and*

(ii) *there exists a positive number C such that for each s and t in $[0, 1]^q$,*

$$\limsup_n E|X_n(t) - X_n(s)|^2 \leqq C|t - s|,$$

then the $L(X_n)$ converge weakly in the U-topology to a distribution G on $(D_q, \mathfrak{A})$ such that $G(C_q) = 1$ and such that $G\pi_T^{-1} = G_T$ for every finite subset T of $[0, 1]^q$.

As a direct consequence of Theorem 3, we obtain the following extension of the invariance principle for a sequence of independent identically distributed random variables:

COROLLARY 1. *Let $\alpha_{j_1, \cdots, j_q}$ $(1 \leqq j_p < \infty, 1 \leqq p \leqq q)$ be independent identically distributed random variables with mean zero and variance 1, and let the random functions X_n $(n \geqq 1)$ be defined by the formula*

$$X_n(t) = n^{-q/2} \sum_{1 \leqq p \leqq q} \sum_{1 \leqq j_p \leqq [nt_p]} \alpha_{j_1, \cdots, j_q}$$

$(t = (t_1, \cdots, t_q) \in [0, 1]^q)$. Then the $L(X_n)$ converge weakly in the U-topology to a distribution G on $(D_q, \mathfrak{A})$ such that

(a) $G(C_q) = 1$,

(b) *$G\pi_T^{-1}$ is normal for each finite subset T of $[0, 1]^q$,*

(c) *for each $s, t \in [0, 1]^q$,*

$$\int \pi_t \, dG = 0 \quad \int \pi_s \pi_t \, dG = \prod_{1 \leqq p \leqq q} \min (s_p, t_p).$$

With minor modifications, this last result was obtained in [3] for the case $q = 2$ under additional assumptions about the common distribution of the $\alpha_{j_1,\cdots,j_q}$. The distribution G defined by (a), (b), and (c) is simply a Brownian motion with time parameter t in $[0, 1]^q$.

2. Proof of Theorem 1. We prove (1) by induction on q. The proof is based on a submartingale inequality of Doob ([2], p. 317): namely, if $Y_1, \cdots, Y_n$ is a positive submartingale, then

$$E(\max_{1 \leqq m \leqq n} Y_m)^2 \leqq 4E(Y_n^2). \tag{5}$$

This inequality immediately implies (1) for $q = 1$ (put $Y_m = |S_m|$). Suppose now that (1) holds for $q - 1$ ($q \geqq 2$); we will show that it also holds for q.

To this end, set for $1 \leqq m \leqq n_q$,

$$S_{k_1,\cdots,k_{q-1}}(m) = \sum_{1 \leqq p < q} \sum_{1 \leqq j_p \leqq k_p} D_{j_1,\cdots,j_{q-1},m}$$
$$U_m = (S_{k_1,\cdots,k_{q-1}}(m))_{1 \leqq k_p \leqq n_p, 1 \leqq p < q} \, \varepsilon \, R^c$$

and put

$$V_m = (S_{k_1,\cdots,k_{q-1},m})_{1 \leqq k_p \leqq n_p,\, 1 \leqq p < q}$$
$$= U_1 + U_2 + \cdots + U_m \, \varepsilon \, R^c,$$

where
$$c = n_1 + \cdots + n_{q-1}.$$

Since the maximum norm of V_m is $\|V_m\| = \max_{1 \leqq k_p \leqq n_p, 1 \leqq p < q} |S_{k_1,\cdots,k_{q-1},m}|$, one obtains the relation

$$M = \max_{1 \leqq m \leqq n_q} \|V_m\|.$$

Now since the U_m's are independent random vectors with zero expectation, $(V_m)_{1 \leqq m \leqq n_q}$ is a martingale in R^c, and since $\|\cdot\|$ is a convex function, $(\|V_m\|)_{1 \leqq m \leqq n_q}$ is a positive submartingale. Doob's inequality (5) implies that

$$E(M^2) = E(\max_{1 \leqq m \leqq n_q} \|V_m\|)^2 \leqq 4E(\|V_{n_q}\|^2)$$

with

$$\|V_{n_q}\| = \max \{|S'_{k_1,\cdots,k_{q-1}}| : 1 \leqq k_p \leqq n_p,\ 1 \leqq p \leqq q - 1\}, \tag{6}$$

where

$$S'_{k_1,\cdots,k_{q-1}} = \sum_{1 \leqq p < q} \sum_{1 \leqq j_p \leqq k_p} (\sum_{1 \leqq j \leqq n_q} D_{j_1,\cdots,j_{q-1},j}).$$

The induction hypothesis now yields

$$E(M^2) \leqq 4 \cdot 4^{q-1} E(S'_{n_1,\cdots,n_{q-1}})^2$$
$$= 4^q E(S_{n_1,\cdots,n_{q-1},n_q})^2 = 4^q \sigma^2,$$

thereby completing the proof of (1).

Proceeding to the proof of (2a), define the stopping time τ over the event

$\{M > 2^q a\}$ by

$$\tau = \min \{m: \|V_m\| > 2^q a\}.$$

Then

$$\delta P(M > 2^q a) \leqq \sum_m P(\tau = m) P(\|V_{n_q} - V_m\| \leqq 2^{q-1} a)$$
$$= \sum_m P(\tau = m, \|V_{n_q} - V_m\| \leqq 2^{q-1} a) \leqq P(\|V_{n_q}\| > 2^{q-1} a),$$

where

$$\delta = \min_{1 \leqq m \leqq n_q} P(\|V_{n_q} - V_m\| \leqq 2^{q-1} a)$$
$$= 1 - \max_{1 \leqq m \leqq n_q} P(\|V_{n_q} - V_m\| > 2^{q-1} a).$$

Then since

$$\|V_{n_q} - V_m\| = \max_{1 \leqq k_p \leqq n_p, 1 \leqq p < q} |\sum_{1 \leqq p < q} \sum_{1 \leqq j_p \leqq k_p} (\sum_{m < j \leqq n_q} D_{j_1, \cdots, j_{q-1}, j})|,$$

the inequality (1) implies that

$$P(\|V_{n_q} - V_m\| > 2^{q-1} a) \leqq E\|V_{n_q} - V_m\|^2 / (4^{q-1} a^2) \leqq (\sigma/a)^2$$

and thus

$$P(M > 2^q a) \leqq (1 - (\sigma/a)^2)^{-1} P(\|V_{n_q}\| > 2^{q-1} a).$$

In view of (6), the inequality (2a) now follows by induction on q. Inequality (2b) is obtained from (2a) by applying Chebychev's inequality to the right hand side.

Inequalities (3a) and (3b) are also established by induction on q. Using the Kolmogorov type inequality (3b) for $q - 1$ instead of inequality (1) in the argument of the paragraph above, one obtains

$$P(M > 2^q a) \leqq [(1 - (q - 2)\sigma^2/(2a)^2)/$$
$$(1 - (q - 1)\sigma^2/(2a)^2)] P(\|V_{n_q}\| > 2^{q-1} a),$$

which, together with (3a) and (3b) for $q - 1$, implies (3a) and (3b) for q.

3. Proof of Theorem 3. Condition (i) of Theorem 3 implies condition (i) of Theorem 2. To see that (ii) of Theorem 2 is also satisfied, and hence that Theorem 3 holds, let $[t]_n = [nt]/n$ for t in $[0, 1]$ and let $[u]_n = ([u_1]_n, \cdots, [u_q]_n)$ for $u = (u_1, \cdots, u_q)$ in $[0, 1]^q$. For each $m, n \geqq 1$ and $i = (i_1, \cdots, i_q)$ such that $0 \leqq i_p \leqq m - 1$ for $1 \leqq p \leqq q$, put

$$A_i(m; n) = \prod_{1 \leqq p \leqq q} [[i_p/m]_n, [(i_p + 1)/m]_n].$$

Note that

$$[(i_p + 1)/m]_n - [i_p/m]_n > 1/m - 2/n.$$

Thus if $s = (s_1, \cdots, s_q)$ and $t = (t_1, \cdots, t_q)$ are any two points of $[0, 1]^q$ such that $|t - s| < 1/m$, then for all sufficiently large n, s and t lie either in the same

$A_i(m; n)$ or in a pair of $A_i(m; n)$'s with a common boundary point, since for each p, s_p and t_p lie either in one and the same interval, or a pair of contiguous intervals, of the form $[[i_p/m]_n, [(i_p + 1)/m]_n]$. It follows that, for each m,

$$\lim_{\delta\to 0} \limsup_n P(w_\delta(X_n) > 4\epsilon) \leqq \limsup_n \sum_{A(m;n)} P(\omega_{A(m;n)}(X_n) > \epsilon),$$

where in the summation $A(m; n)$ ranges over the m^q sets $A_i(m; n)$, and where

$$\omega_{A_i(m;n)}(X_n) = \sup_{t \varepsilon A_i(m;n)} |X_n(t) - X_n([i/m]_n)|.$$

Thus it suffices to show that, for each $\epsilon > 0$,

$$\lim_{m\to\infty} \limsup_n \sum_{A(m;n)} P(\omega_{A(m;n)}(X_n) > \epsilon) = 0. \tag{7}$$

To this end, momentarily fix i, m, and n. Set

$$K_p = n[(i_p + 1)/m]_n - n[i_p/m]_n, \qquad (1 \leqq p \leqq q),$$

$$t_p(k) = [i_p/m]_n + k/n, \quad \mu_p(k) = nt_p(k), \qquad (0 \leqq k \leqq K_p, 1 \leqq p \leqq q),$$

and for $0 \leqq k_p \leqq K_p$, $1 \leqq p \leqq q$, define

$$\begin{aligned} D_{k_1,\cdots,k_q} &= 0, \qquad \text{if } k_1 = k_2 = \cdots = k_q = 0; \\ &= \sum_{1\leqq \rho \leqq q, \rho\neq p} \sum_{0\leqq j_\rho \leqq \mu_\rho(0)} \beta_{j_1,\cdots,j_{p-1},\mu_p(k_p), j_{p+1},\cdots,j_q}(n), \\ &\qquad \text{if } k_1 = \cdots = k_{p-1} = k_{p+1} = \cdots = k_q = 0, k_p > 0; \\ &= \beta_{\mu_1(k_1),\cdots,\mu_q(k_q)}(n), \qquad \text{otherwise.} \end{aligned}$$

Then $\omega_{A_i(m;n)}(X_n)$ equals

$$\begin{aligned} &\max_{0\leqq k_p \leqq K_p, 1\leqq p\leqq q} |X_n((t_1(k_1), \cdots, t_q(k_q))) - X_n((t_1(0), \cdots, t_q(0)))| \\ &\qquad = \max_{0\leqq k_p\leqq K_p, 1\leqq p\leqq q} \left|\sum_{1\leqq p\leqq q} \sum_{0\leqq j_p\leqq k_p} D_{j_1,\cdots,j_q}\right|, \end{aligned}$$

and since the $D_{j_1,\cdots,j_q}$ are independent with zero means and finite variances, inequality (2a) implies that (provided m and n are sufficiently large)

$$P(\omega_{A_i(m;n)}(X_n) > 2^q\epsilon) \leqq (1 - ES_i^2(m; n)/\epsilon^2)^{-q} P(|S_i(m; n)| \geqq \epsilon), \tag{8}$$

where the random variable $S_i(m; n)$ equals

$$\begin{aligned} X_n([(i + 1)/m]_n) - X_n([i/m]_n) &= X_n((i + 1)/m) - X_n((i/m)), \\ &\qquad (i + 1 = (i_1 + 1, \cdots, i_q + 1)). \end{aligned}$$

From (i) and (ii) of Theorem 3, it follows that as $n \to \infty$, the distribution of $S_i(m; n)$ converges weakly to a normal distribution with mean zero and variance not exceeding

$$\liminf_n ES_i^2(m; n) \leqq \limsup_n ES_i^2(m; n) \leqq C/m \tag{9}$$

(see [1], p. 32). Letting $N(y)$ $(y > 0)$ denote the probability that a standard normal random variable exceeds y in absolute value, we find from (8) and (9) that for each $\epsilon > 0$ and all sufficiently large m,

$$\text{(10)} \quad \limsup_n \sum_{A(m;n)} P(\omega_{A(m;n)}(X_n) > 2^q\epsilon) \leqq m^q(1 - C\epsilon^{-2}m^{-1})^{-q}N(\epsilon m^{\frac{1}{2}}C^{-\frac{1}{2}}).$$

Since the right hand side of (10) tends to zero as $m \rightarrow \infty$, it follows that (7) does indeed hold, and the proof is complete.

REFERENCES

[1] BILLINGSLEY, PATRICK (1968). *Convergence of Probability Measures*. Wiley, New York.

[2] DOOB, J. L. (1953). *Stochastic Processes*. Wiley, New York.

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

[4] PARTASARATHY, K. R. (1967). *Probability Measures on Metric Spaces*. Academic Press, New York.

[5] PYKE, R. AND SHORACK, G. R. (1968). Weak convergence of a two-sample empirical process and a new approach to Chernoff-Savage theorems. *Ann. Math. Statist.* **39** 755–771.

[6] WICHURA, MICHAEL J. (1968). On the weak convergence of non-Borel probabilities on a metric space. Thesis, Columbia University.