SOME RESULTS FOR THE WEIGHTED EMPIRICAL PROCESS CONCERNING THE LAW OF THE ITERATED LOGARITHM AND WEAK CONVERGENCE

By

Alfred Joseph Vanderzanden

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Statistics and Probability

1980

ABSTRACT

SOME RESULTS FOR THE WEIGHTED EMPIRICAL PROCESS CONCERNING THE LAW OF THE ITERATED LOGARITHM AND WEAK CONVERGENCE

By

Alfred Joseph Vanderzanden

In this paper we establish two main results for the weighted empirical process. The first result is a functional law of the iterated logarithm when the underlying random variables are i.i.d. Uniform [0,1]. The second result is the weak convergence of the weighted empirical process to a Gaussian process with almost sure continuous sample paths when the underlying random variables represent an array of row independent random vectors taking values in the k-dimensional unit cube $[0,1]^k$.

To

Joyce

ACKNOWLEDGEMENTS

I wish to express my sincere thanks to Professors Hira Koul and Joel Zinn for their guidance and especially their patience during the preparation of this dissertation.

I would also like to thank Professors Roy Erickson and Joel Shapiro for their review of my work.

Finally, special thanks go to Mrs. Noralee Burkhardt for her excellent typing of the manuscript.

TABLE OF CONTENTS

1. SUMMARY

The weighted empirical process deserves special recognition among stochastic processes. It serves as a fundamental tool in the study of statistics based on ranks, as they occur in nonparametric statistics, because of the ability to express these rank statistics in terms of the weighted empirical process (see Koul (1970a) and Koul and Staudte (1972)). Furthermore, some statistical procedures have recently been proposed by Sinha and Sen (1979) and Koul (1980) which involve the weighted empirical process directly.

In this paper we establish two main results for the weighted empirical process. The first result is a functional law of the iterated logarithm when the underlying random variables are i.i.d. Uniform $[0,1]$. This appears in Section 2 as Theorem 2.2.1 and extends the work of Finkelstein (1971) using some ideas found in James (1975) and Kuelbs (1976). The second result appears in Section 3 as Theorem 3.2.2 and establishes the weak convergence of the weighted empirical process to a Gaussian process with almost sure continuous sample paths when the underlying random variables represent an array of row independent random vectors taking values in the k-dimensional unit cube $[0,1]^k$. Theorem 3.2.2 extends the work of Koul (1970), Withers (1975), and Shorack (1980) using the fluctuation inequalities of Bickel and Wichura (1971) which we present in Section 4.

2. A FUNCTIONAL LAW OF THE ITERATED LOGARITHM FOR THE WEIGHTED EMPIRICAL PROCESS

2.1. Introduction

Let $Y_1, Y_2, \ldots$ be a sequence of random variables defined on a probability space (Ω, F, P) such that $P(Y_i \in [0,1]) = 1$ for each $i = 1,2,\ldots$. Furthermore, let $c_1, c_2, \ldots$ be a sequence of real numbers and define

$$(2.1.1) \qquad \sigma_N^2 = \sum_{i=1}^{N} c_i^2 , \quad N = 1,2,\ldots$$

In this section the weighted empirical process is defined by

$$(2.1.2) \qquad V_N(t) = \sum_{i=1}^{N} c_i[I(Y_i \leq t) - P(Y_i \leq t)], \quad t \in [0,1].$$

We also define the "normalized" weighted empirical process by

$$(2.1.3) \qquad X_N(t) = V_N(t) / \sqrt{2\, \sigma_N^2 \log \log \sigma_N^2} , \quad t \in [0,1].$$

Functional laws of the iterated logarithm have been established for the X_N process by Finkelstein (1971) when $Y_1, Y_2, \ldots$ are i.i.d. Uniform [0,1] and $c_i = 1$ for all i; and also by Philipp (1977) when $Y_1, Y_2, \ldots$ are strictly stationary strongly mixing Uniform [0,1] and $c_i = 1$ for all i. James (1975) and Wellner (1977) have extended Finkelstein's (1971) result to empirical processes of the form wX_N where $w(t)$, $t \in [0,1]$ is a suitable weight function, $Y_1, Y_2, \ldots$ are i.i.d. Uniform [0,1], and $c_i = 1$ for all i. In this paper we extend Finkelstein's (1971) result in a way different from James and Wellner. We still require $Y_1, Y_2, \ldots$ to be i.i.d. Uniform $[0,1]$ but we allow the weights $c_1, c_2, \ldots$ to be arbitrary real numbers satisfying two regularity conditions.

Before stating the main result we introduce some notation adopted from Kuelbs (1976).

If (M,d) is a metric space and $A \subseteq M$ we define the distance from $x \in M$ to A by $d(x,A) = \inf\{d(x,y) : y \in A\}$. If $\{x_N\}$ is a sequence of points in M, then $C(\{x_N\})$ denotes the cluster set of $\{x_N\}$. That is, $C(\{x_N\})$ is the set of all possible limit points of the sequence $\{x_N\}$. We write $\{x_N\} \twoheadrightarrow A$ if both $\lim_{N\to\infty} d(x_N, A) = 0$ and $C(\{x_N\}) = A$.

If $\Gamma(s,t)$, $s,t \in T \subseteq (-\infty,\infty)$ is a nonnegative definite real-valued function, we define $H(\Gamma)$ to be the reproducing kernel Hilbert space generated by the kernel Γ and $\|\cdot\|_{H(\Gamma)}$ denotes the associated norm on $H(\Gamma)$. For an extensive discussion of reproducing kernel Hilbert spaces see Aronszajn (1950).

2.2. The law of the iterated logarithm for weighted empirical processes

Let the space of real-valued functions on $[0,1]$ which are right continuous on $[0,1]$ and have left limits on $(0,1]$ be denoted by $D[0,1]$. Endow the space $D[0,1]$ with the metric generated by the supremum norm

(2.2.1) $\quad \|x\|_\infty = \sup\{|x(t)| : t \in [0,1]\}, \quad x \in D[0,1],$

and let $\mathcal{D}$ denote the σ-field generated by the $\|\cdot\|_\infty$-open balls of $D[0,1]$.

Theorem 2.2.1. If $Y_1, Y_2, \ldots$ are i.i.d. Uniform $[0,1]$ random variables and $c_1, c_2, \ldots$ are any real numbers satisfying

$$(2.2.2) \qquad \lim_{N\to\infty} \sigma_N^2 = \infty \quad \text{and} \quad \lim_{N\to\infty} \left(\max_{1\le i\le N} c_i^2 \right) \frac{\log\log \sigma_N^2}{\sigma_N^2} = 0,$$

then with respect to $(D[0,1], \mathcal{D}, \|\cdot\|_\infty)$ we have

$$(2.2.3) \qquad P(X_N \twoheadrightarrow B) = 1$$

where X_N is the normalized weighted empirical process (2.1.3) and $B = \{x \in H(\Gamma) : \|x\|_{H(\Gamma)} \le 1\}$ with $\Gamma(s,t) = (s \wedge t) - st$, $s,t \in [0,1]$.

Theorem 2.2.1 will follow from Lemma 1.1 in Kuelbs (1976) once we establish Lemma 2.2.1 and Lemma 2.2.2 which we now state.

Lemma 2.2.1. Suppose the assumptions of Theorem 2.2.1 are satisfied and let T denote any finite subset of $[0,1]$, then with respect to $(R^T, \|\cdot\|_{R^T})$ we have

$$(2.2.4) \qquad P(X_N^T \twoheadrightarrow B^T) = 1$$

where R^T is the space of all real-valued functions defined on T, $\|x\|_{R^T} = \max_{t\in T} |x(t)|$ for all $x \in R^T$, X_N^T is the process (2.1.3) restricted to T, $B^T = \{x \in H(\Gamma^T) : \|x\|_{H(\Gamma^T)} \le 1\}$, and Γ^T is the restriction of $\Gamma(s,t) = (s \wedge t) - st$ to $T \times T$.

Before stating Lemma 2.2.2 we introduce some additional notation adopted from Kuelbs (1976).

If $T = \{t_0, t_1, \ldots, t_m\}$ where $0 = t_0 < t_1 < \ldots < t_m = 1$ and if $x \in D[0,1]$, then we define $\Lambda_T(x)$ to be the continuous polygonal function such that

$$(2.2.5) \qquad \Lambda_T(x)(t) = \left(\frac{t_i - t}{t_i - t_{i-1}}\right)x(t_{i-1}) + \left(\frac{t - t_{i-1}}{t_i - t_{i-1}}\right)x(t_i)$$

for $t \in [t_{i-1}, t_i]$ and $i = 1,2,\ldots,m$.

Lemma 2.2.2. Suppose the assumptions of Theorem 2.2.1 are satisfied, $T = \{t_0,t_1,\ldots,t_m\}$ where $0 = t_0 < t_1 <\ldots< t_m = 1$, and $\max_{1\le i\le m} |t_i - t_{i-1}| \le \frac{1}{2}$, then

$$(2.2.6) \qquad P(\|X_N - \Lambda_T(X_N)\|_\infty \le \varphi_T \text{ for all sufficiently large } N) = 1$$

where X_N is the process (2.1.3), $\varphi_T = 2 \max_{1\le i\le m}\sqrt{\frac{8\rho(t_i - t_{i-1})}{1 - (t_i - t_{i-1})}}$, and $\rho > 1$.

Lemmas 2.2.1 and 2.2.2 will be proved shortly, but we first show they imply Theorem 2.2.1. With this in mind let T_M, $M = 1,2,\ldots$ denote any sequence of increasing finite subsets of $[0,1]$ such that the points in T_M satisfy

$$(2.2.7) \qquad \begin{aligned} &0 = t_{M0} < t_{M1} <\ldots< t_{M,m_M} \quad \text{for } M = 1,2,\ldots \\ &\text{and} \quad \lim_{M\to\infty} \max_{1\le i\le m_M} |t_{Mi} - t_{M,i-1}| = 0. \end{aligned}$$

If the assumptions of Theorem 2.2.1 are satisfied, then Lemma 2.2.2 gives $P(\|X_N - \Lambda_{T_M}(X_N)\|_\infty \le \varphi_{T_M}$ for all sufficiently large $N) = 1$ for each $M = 1,2,\ldots$ Hence, it follows that $P(\overline{\lim_{N\to\infty}} \|X_N - \Lambda_{T_M}(X_N)\|_\infty \le \varphi_{T_M}$ for each $M) = 1$. Using (2.2.7) we have $\lim_{M\to\infty} \varphi_{T_M} = 0$ so that

$$(2.2.8) \qquad P(\overline{\lim_{M\to\infty}}\ \overline{\lim_{N\to\infty}} \|X_N - \Lambda_{T_M}(X_N)\|_\infty = 0) = 1 .$$

Relation (2.2.8) shows that condition (ii.c) of Lemma 1.1 in Kuelbs (1976) is satisfied.

Since $\Gamma(s,t) = (s \wedge t) - st$ is continuous on $[0,1] \times [0,1]$, Lemma 3 in Oodaira (1972) implies that $B = \{x \in H(\Gamma) : \|x\|_{H(\Gamma)} \leq 1\}$ is compact in $(C[0,1], \|\cdot\|_\infty)$ where $C[0,1]$ is the space of all real-valued continuous functions on $[0,1]$ and $\|x\|_\infty = \sup\{|x(t)| : t \in [0,1]\}$ for all $x \in C[0,1]$. Furthermore, from the Theorem on page 351 in Aronszajn (1950) it is easy to show that

$$(2.2.9) \quad \{x \in H(\Gamma^T) : \|x\|_{H(\Gamma^T)} \leq 1\} = \{x = y^T : y \in H(\Gamma) \text{ and } \|y\|_{H(\Gamma)} \leq 1\},$$

where Γ can be any nonnegative definite function on $[0,1] \times [0,1]$ and T any finite subset of $[0,1]$. Hence, Lemma 2.2.1 applied to each T_M, $M = 1,2,\ldots$ shows that condition (i.) of Lemma 1.1 in Kuelbs (1976) is satisfied. Thus, Theorem 2.2.1 now follows from Lemma 1.1 in Kuelbs (1976).

<u>Proof of Lemma 2.2.1.</u>

Let $T = \{t_1, t_2, \ldots, t_m\}$ denote any finite subset of $[0,1]$. Let $Y_1, Y_2, \ldots$ and $c_1, c_2, \ldots$ be as in Theorem 2.2.1 and define

$$Z_i = (Z_i(t_1), Z_i(t_2), \ldots, Z_i(t_m)), \quad i = 1,2,\ldots$$

where $Z_i(t) = c_i[I(Y_i \leq t) - P(Y_i \leq t)]$, $t \in [0,1]$, $i = 1,2,\ldots$ Then $X_N^T = (X_N(t_1), X_N(t_2), \ldots, X_N(t_m)) = \sum_{i=1}^N Z_i / \sqrt{2\sigma_N^2 \log\log \sigma_N^2}$ and $\frac{1}{\sigma_N^2} \sum_{i=1}^N \mathrm{Cov}(Z_i) = (\Gamma(t_i,t_j))_{i,j=1}^m = \Gamma^T$ where $\Gamma(s,t) = (s \wedge t) - st$. Lemma 2.2.1 now follows immediately from the multivariate law of the iterated logarithm, Thereom 1 in Berning (1979), applied to Z_i, $i = 1,2,\ldots$ $\square$

The proof of Lemma 2.2.2 depends on several results which we present in the following Lemmas.

Lemma 2.2.3. If $T = \{t_0, t_1, \ldots, t_m\}$ where $0 = t_0 < t_1 < \ldots < t_m = 1$ and $X(t)$, $t \in [0,1]$ is any stochastic process, then

$$\text{(2.2.10)} \quad \|X - \Lambda_T(X)\|_\infty \le \max_{1 \le i \le m} \Big[\sup_{t \in [t_{i-1}, t_i]} |X(t) - X(t_{i-1})| + \sup_{t \in [t_{i-1}, t_i]} |X(t_i) - X(t)| \Big]$$

where $\Lambda_T(X)$ is the continuous polygonal function defined in (2.2.5).

Proof. For any $t \in [0,1]$ there exists $i \in \{1,2,\ldots,m\}$ such that $t_{i-1} \le t \le t_i$. Hence,

$$|X(t) - \Lambda_T(X)(t)| = \Big|\frac{t_i - t}{t_i - t_{i-1}} [X(t) - X(t_{i-1})] + \frac{t - t_{i-1}}{t_i - t_{i-1}} [X(t) - X(t_i)]\Big|$$

$$\le |X(t) - X(t_{i-1})| + |X(t) - X(t_i)|$$

and (2.2.10) now follows immediately. □

Lemma 2.2.4. Let $N \in \{1,2,\ldots\}$ be fixed. Suppose $Y_1, Y_2, \ldots, Y_N$ are independent real-valued random variables defined on a probability space (Ω, F, P) and $c_1, c_2, \ldots, c_N$ are any real numbers. Define

$$\text{(2.2.11)} \quad M(t) = \sum_{i=1}^{N} [c_i / P(Y_i \notin (a,t])][I(Y_i \in (a,t]) - P(Y_i \in (a,t])]$$

and

$$\text{(2.2.12)} \quad R(t) = \sum_{i=1}^{N} [c_i / P(Y_i \notin (t,b])][I(Y_i \in (t,b]) - P(Y_i \in (t,b])]$$

where a and b are any two real numbers.

Then $\{M(t), F_1(t), t \in T_1\}$ is a martingale and $\{R(t), F_2(t), t \in T_2\}$ is a reversed martingale where $T_1 \subseteq [a,\infty)$

and $T_2 \subseteq (-\infty,b]$ are such that $M(t)$ is well-defined for all $t \in T_1$ and $R(t)$ is well-defined for all $t \in T_2$. $F_1(t)$ and $F_2(t)$ are the σ-fields defined in (2.2.13) and (2.2.16), respectively.

Proof. We first show that $\{M(t),\ t \in T_1\}$ is a martingale with respect to the nondecreasing family of σ-fields

$$\{F_1(t),\ t \in T_1\} \tag{2.2.13}$$

where $F_1(t)$ is the σ-field generated by the class of sets of the form $\{Y_i \in (a,s]\}$ for some $i \in \{1,2,\ldots,N\}$ and some $s \in [a,t]$.

Let $s,t \in T_1$ with $s \leq t$. Then it is easy to show that for each $i \in \{1,2,\ldots,N\}$

$$P(Y_i \in (a,t] \mid F_1(s)) = I(Y_i \in (a,s]) + \frac{P(Y_i \in (s,t])}{P(Y_i \notin (a,s])} I(Y_i \notin (a,s]). \tag{2.2.14}$$

Using (2.2.14) we obtain

$$\begin{aligned} &P(Y_i \notin (a,s])[P(Y_i \in (a,t] \mid F_1(s)) - P(Y_i \in (a,t])] \\ &\quad = P(Y_i \notin (a,t])[I(Y_i \in (a,s]) - P(Y_i \in (a,s])]. \end{aligned} \tag{2.2.15}$$

From (2.2.15) and (2.2.11) it now follows that

$$\begin{aligned} E(M(t) \mid F_1(s)) &= \sum_{i=1}^{N} [c_i/P(Y_i \notin (a,t])][P(Y_i \in (a,t] \mid F_1(s)) - \\ &\qquad P(Y_i \in (a,t])] \\ &= \sum_{i=1}^{N} [c_i/P(Y_i \notin (a,s])][I(Y_i \in (a,s]) - P(Y_i \in (a,s])] \\ &= M(s). \end{aligned}$$

This shows $\{M(t), F_1(t), t \in T_1\}$ is a martingale.

Next we show $\{R(t), t \in T_2\}$ is a reversed martingale with respect to the nonincreasing family of σ-fields

$$(2.2.16) \qquad \{F_2(t), t \in T_2\}$$

where $F_2(t)$ is the σ-field generated by the class of sets of the form $\{Y_i \in (s,b]\}$ for some $i \in \{1,2,\dots,N\}$ and some $s \in [t,b]$.

Let $s,t \in T_2$ with $s \le t$. Then it is easy to show that for each $i \in \{1,2,\dots,N\}$

$$(2.2.17) \quad P(Y_i \in (s,b] | F_2(t)) = I(Y_i \in (t,b]) + \frac{P(Y_i \in (s,t])}{P(Y_i \notin (t,b])} I(Y_i \notin (t,b]).$$

Using (2.2.17) we obtain

$$(2.2.18) \quad P(Y_i \notin (t,b])[P(Y_i \in (s,b] | F_2(t)) - P(Y_i \in (s,b])]$$
$$= P(Y_i \notin (s,b])[I(Y_i \in (t,b]) - P(Y_i \in (t,b])].$$

Finally, from (2.2.18) and (2.2.12) it follows that

$$\begin{aligned} E(R(s)|F_2(t)) &= \sum_{i=1}^{N} [c_i/P(Y_i \notin (s,b])][P(Y_i \in (s,b] | F_2(t)) - \\ &\qquad\qquad - P(Y_i \in (s,b])] \\ &= \sum_{i=1}^{N} [c_i/P(Y_i \notin (t,b])][I(Y_i \in (t,b]) - P(Y_i \in (t,b])] \\ &= R(t). \end{aligned}$$

This shows $\{R(t), F_2(t), t \in T_2\}$ is a reversed martingale. □

<u>Lemma 2.2.5</u>. Let $N \in \{1,2,\dots\}$ be fixed. Suppose $Y_1, Y_2, \dots, Y_N$ are independent real-valued random variables defined on a probability space (Ω, F, P) and $c_1, c_2, \dots, c_N$ are any real numbers. Define

the random variable X by

$$(2.2.19)\quad X = \sum_{i=1}^{N} [c_i/P(Y_i \notin (a,b])][I(Y_i \in (a,b]) - P(Y_i \in (a,b])]$$

where a and b are any real numbers such that $a \leq b$ and X is well-defined. Then for any $\alpha > 0$ we have

$$(2.2.20)\qquad E\exp(\alpha X) \leq \exp\{\alpha^2 f(\alpha\beta)\mathrm{Var}(X)\}$$

where

$$(2.2.21)\quad \beta = \max_{1\leq i\leq N} [|c_i| \max\{1, P(Y_i \in (a,b])/P(Y_i \notin (a,b])\}]$$

and $f(x)$, $x \in (-\infty,\infty)$ is the positive, strictly increasing, continuous function defined by

$$(2.2.22)\qquad f(x) = \begin{cases} (e^x - 1 - x)/x^2 & \text{if } x \neq 0 \\ 1/2 & \text{if } x = 0 . \end{cases}$$

<u>Proof</u>. For each $i = 1,2,\ldots,N$ let

$$X_i = [c_i/P(Y_i \notin (a,b])][I(Y_i \in (a,b]) - P(Y_i \in (a,b])].$$

Then it is clear that $X_1,X_2,\ldots,X_N$ are independent random variables with $EX_i = 0$ and $|X_i| \leq \beta$, $i = 1,2,\ldots,N$ where β is defined in (2.2.21). With f as defined in (2.2.22) we have $e^x = 1 + x + x^2 f(x)$ so that for all $\alpha > 0$ and $i = 1,2,\ldots,N$ we have

$$\begin{aligned} E\exp(\alpha X_i) &= E[1 + \alpha X_i + \alpha^2 X_i^2 f(\alpha X_i)] \\ &= 1 + \alpha^2 EX_i^2 f(\alpha X_i) \\ &\leq 1 + \alpha^2 f(\alpha\beta)\, EX_i^2 \\ &\leq \exp\{\alpha^2 f(\alpha\beta)\,\mathrm{Var}(X_i)\} \quad \text{since } 1 + x \leq e^x. \end{aligned}$$

Therefore,

$$\begin{aligned} E\exp(\alpha X) &= E\exp(\alpha\sum_{i=1}^{N} X_i) \\ &= \prod_{i=1}^{N} E\exp(\alpha X_i) \\ &\le \prod_{i=1}^{N} \exp\{\alpha^2 f(\alpha\beta)\,\mathrm{Var}(X_i)\} \\ &= \exp\{\alpha^2 f(\alpha\beta)\,\mathrm{Var}(X)\}. \qquad \square \end{aligned}$$

<u>Lemma 2.2.6</u>. Suppose $\lambda(N)$, $N = 1,2,\ldots$ is a nondecreasing sequence of positive numbers and $\{U_N(t),\ t \in T\}$, $N = 1,2,\ldots$ is a sequence of independent stochastic processes defined on a probability space (Ω,F,P) and taking values in the space of real-valued functions defined on $T \subseteq (-\infty,\infty)$. Define $\{W_N(t),\ t \in T\}$ by $W_N(t) = \sum_{i=1}^{N} U_i(t)$ and assume there is a countable subset $\{t_j,\ j = 1,2,\ldots\}$ of T such that $\sup\{|W_N(t)|,\ t \in T\} = \sup\{|W_N(t_j)|,\ j = 1,2,\ldots\}$. Then for any positive integers $N_1 \le N_2$ and $\varepsilon \ge \varphi(N_1,N_2)$ we have

$$(2.2.23)\quad P\Big(\max_{N_1\le N\le N_2}\ \sup_{t\in T}\frac{|W_N(t)|}{\lambda(N)} > \varepsilon\Big) \le 2P\Big(\sup_{t\in T}\frac{|W_{N_2}(t)|}{\lambda(N_2)} > \frac{\varepsilon}{2}\,\frac{\lambda(N_1)}{\lambda(N_2)}\Big)$$

where

$$(2.2.24)\quad \varphi^2(N_1,N_2) = \sup_{t\in T}\frac{8}{\lambda^2(N_1)}\,[\mathrm{Var}(W_{N_2}(t)) - \mathrm{Var}(W_{N_1}(t))].$$

<u>Proof</u>. Let $\varepsilon > 0$. Since $\lambda(N)$, $N = 1,2,\ldots$ is a nondecreasing sequence of positive numbers, the event

$$\{\max_{N_1\le N\le N_2}\ \sup_{t\in T}\frac{|W_N(t)|}{\lambda(N)} > \varepsilon\}$$

is contained in the event A where

$$A = \{\max_{N_1 \le N \le N_2} \sup_{t \in T} |W_N(t)| > \varepsilon\, \lambda(N_1)\}$$

(2.2.25)

$$= \{\max_{N_1 \le N \le N_2} \sup_{j \ge 1} |W_N(t_j)| > \varepsilon\, \lambda(N_1)\} \quad .$$

For $N \in \{N_1,\ldots,N_2\}$ and $j \in \{1,2,\ldots\}$ define

$$(2.2.26) \quad B_{Nj} = \{\max_{N_1 \le n \le N} \max_{1 \le i < j} |W_n(t_i)| \le \varepsilon\, \lambda(N_1) \quad \text{and} \quad |W_N(t_j)| > \varepsilon\, \lambda(N_1)\}$$

and

$$(2.2.27) \qquad C_{Nj} = \{|W_{N_2}(t_j) - W_N(t_j)| \le \tfrac{1}{2}\, \varepsilon\, \lambda(N_1)\}.$$

It is clear that for each $N \in \{N_1,\ldots,N_2\}$ the family $\{B_{Nj},\ j = 1,2,\ldots\}$ consists of pairwise disjoint events. Furthermore, since B_{Nj} depends only on $\{U_i(t),\ t \in T\}$, $i = 1,2,\ldots,N$ and C_{Nj} depends only on $\{U_i(t),\ t \in T\}$, $i = N+1,\ldots,N_2$, we have for each $N \in \{N_1,\ldots,N_2\}$ that the families $\{B_{Nj},\ j = 1,2,\ldots\}$ and $\{C_{Nj},\ j = 1,2,\ldots\}$ are statistically independent since the processes $U_1, U_2, \ldots$ are independent. Extending Loéve's Lemma for Events on page 246 in Loéve (1963) to countable collections of events we obtain

$$(2.2.28) \quad [\inf\{P(C_{Nj}),\ N_1 \le N \le N_2,\ j \ge 1\}]\, P\Big(\bigcup_{N=N_1}^{N_2} \bigcup_{j=1}^{\infty} B_{Nj}\Big) \le P\Big(\bigcup_{N=N_1}^{N_2} \bigcup_{j=1}^{\infty} B_{Nj} C_{Nj}\Big).$$

It is easy to show $A = \bigcup_{N=N_1}^{N_2} \bigcup_{j=1}^{\infty} B_{Nj}$ where A is defined in (2.2.25). On the other hand, in the event $\bigcup_{N=N_1}^{N_2} \bigcup_{j=1}^{\infty} B_{Nj} C_{Nj}$

we have for some $N \in \{N_1,\ldots,N_2\}$ and some $j \in \{1,2,\ldots\}$ that $|W_N(t_j)| > \varepsilon\,\lambda(N_1)$ and $|W_{N_2}(t_j) - W_N(t_j)| \le \frac{1}{2}\,\varepsilon\,\lambda(N_1)$. Therefore, $|W_{N_2}(t_j)| > \frac{1}{2}\,\varepsilon\,\lambda(N_1)$ and it follows that

$$(2.2.29)\qquad \bigcup_{N=N_1}^{N_2}\ \bigcup_{j=1}^{\infty} B_{Nj}C_{Nj} \subseteq \{\sup_{t\in T} \frac{|W_{N_2}(t)|}{\lambda(N_2)} > \frac{\varepsilon}{2}\,\frac{\lambda(N_1)}{\lambda(N_2)}\}\ .$$

Furthermore, for each $N \in \{N_1,\ldots,N_2\}$ and $j \in \{1,2,\ldots\}$ Chebysev's inequality gives

$$1 - P(C_{Nj}) = P(|W_{N_2}(t_j) - W_N(t_j)| > \frac{\varepsilon}{2}\,\lambda(N_1))$$

$$\le (\frac{2}{\varepsilon\,\lambda(N_1)})^2 \mathrm{Var}(W_{N_2}(t_j) - W_N(t_j)).$$

Using the definition of W_N and the independence of U_N, $N = 1,2,\ldots$ we get $\mathrm{Var}(W_{N_2}(t_j) - W_N(t_j)) \le \mathrm{Var}(W_{N_2}(t_j)) - \mathrm{Var}(W_{N_1}(t_j))$. Therefore,

$$(2.2.30)\qquad \sup_{\substack{j\ge 1\\ N_1\le N\le N_2}} [1 - P(C_{Nj})] \le \sup_{t\in T} (\frac{2}{\varepsilon\,\lambda(N_1)})^2 [\mathrm{Var}(W_{N_2}(t)) - \mathrm{Var}(W_{N_1}(t))].$$

The expression on the right-hand side in (2.2.30) will be less than or equal to 1/2 if

$$(2.2.31)\qquad \varepsilon^2 \ge \sup_{t\in T} \frac{8}{\lambda^2(N_1)}\,[\mathrm{Var}(W_{N_2}(t)) - \mathrm{Var}(W_{N_1}(t))].$$

Hence, if (2.2.31) is satisfied, then (2.2.25), (2.2.28), (2.2.29), and (2.2.30) give

$$P(A) \le 2\,P(\sup_{t\in T} \frac{|W_{N_2}(t)|}{\lambda(N_2)} > \frac{\varepsilon}{2}\,\frac{\lambda(N_1)}{\lambda(N_2)})$$

and the lemma is proved. $\square$

Proof of Lemma 2.2.2.

Let $0 \le a < b \le 1$ such that $b-a \le \frac{1}{2}$. Recall that $c_1, c_2, \ldots$ is a sequence of real numbers such that $\sigma_N^2 = \sum_{i=1}^N c_i^2 \to \infty$ and $(\max_{1 \le i \le N} c_i^2) \dfrac{\log\log \sigma_N^2}{\sigma_N^2} \to 0$ as $N \to \infty$. Hence, $c_N^2/\sigma_{N-1}^2 \to 0$ as $N \to \infty$. Let $\rho > 1$, set $\lambda(N) = \sqrt{2\, \sigma_N^2 \log\log \sigma_N^2}$, and choose a positive integer N_0 such that

$$\lambda(N) > 0 \quad \text{and} \quad c_N^2/\sigma_{N-1}^2 \le \rho - 1 \quad \text{for all} \quad N \ge N_0.$$

Next choose a number n such that $\sigma_{N_0}^2 \le \rho^n$. Finally, for each $k = 1,2,\ldots$ define

$$\text{(2.2.32)} \qquad N(k) = \min\{N \ge N_0 : \sigma_N^2 > \rho^{n+k}\}.$$

It is easy to show that the sequence $\{N(k),\ k = 1,2,\ldots\}$ has the following properties:

$$\text{(2.2.33)} \quad N(k) < N(k+1) \quad \text{and} \quad \rho^{n+k} < \sigma_{N(k)}^2 \le \rho^{n+k+1} \quad \text{for} \quad k = 1,2,\ldots$$

$$\text{(2.2.34)} \quad \lim_{k\to\infty} \frac{\sigma_{N(k+1)}^2}{\sigma_{N(k)}^2} = \rho = \lim_{k\to\infty} \frac{\lambda^2(N(k+1))}{\lambda^2(N(k))}.$$

Let $V_N(t) = \sum_{i=1}^N c_i[I(Y_i \le t) - P(Y_i \le t)]$ be the weighted empirical process (2.1.2) where $Y_1, Y_2, \ldots$ are i.i.d. Uniform$[0,1]$ random variables. Furthermore, let $\{W_N(t),\ t \in [a,b]\}$ denote either $\{V_N(t) - V_N(a),\ t \in [a,b]\}$ or $\{V_N(b) - V_N(t),\ t \in [a,b]\}$. We now apply Lemma 2.2.6 to $\{W_N(t),\ t \in [a,b]\}$, $\lambda(N) = \sqrt{2\, \sigma_N^2 \log\log \sigma_N^2}$, $N_1 = N(k)$ and $N_2 = N(k+1)$ for $k = 1,2,\ldots$ to obtain

$$(2.2.35)\quad P\Big(\max_{N(k)\le N\le N(k+1)}\ \sup_{t\in[a,b]} \frac{|W_N(t)|}{\lambda(N)} > \varepsilon\Big)$$

$$\le 2\,P\Big(\sup_{t\in[a,b]} |W_{N(k+1)}(t)| > \frac{\varepsilon}{2}\,\lambda(N(k))\Big)$$

provided

$$(2.2.36)\quad \varepsilon^2 \ge \sup_{t\in[a,b]} \frac{8}{\lambda^2(N(k))}\,[\mathrm{Var}(W_{N(k+1)}(t)) - \mathrm{Var}(W_{N(k)}(t))].$$

Using the fact that $Y_1, Y_2, \ldots$ are i.i.d. Uniform [0,1], it is easy to see that the right-hand side of (2.2.36) is less than or equal to $8[\sigma^2_{N(k+1)} - \sigma^2_{N(k)}](b-a)/\lambda^2(N(k))$. Hence, (2.2.35) will hold if

$$(2.2.37)\qquad \varepsilon^2 \ge 8(b-a)[\sigma^2_{N(k+1)} - \sigma^2_{N(k)}]/\lambda^2(N(k)).$$

Using Lemma 2.2.4 we obtain that

$$M_N(t) = \frac{V_N(t) - V_N(a)}{1-(t-a)}\ ,\ t \in [a,b] \text{ is a martingale and}$$

$$R_N(t) = \frac{V_N(b) - V_N(t)}{1-(b-t)}\ ,\ t \in [a,b] \text{ is a reversed martingale.}$$

In the case $W_N(t) = V_N(t) - V_N(a)$ we have for all $\alpha > 0$ and $\delta > 0$

$$P\Big(\sup_{t\in[a,b]} |W_N(t)| > \delta\Big) \le P\Big(\sup_{t\in[a,b]} |M_N(t)| > \delta\Big)$$

$$\le P\Big(\sup_{t\in[a,b]} M_N(t) > \delta\Big) + P\Big(\sup_{t\in[a,b]} (-M_N(t)) > \delta\Big)$$

$$(2.2.38) = P\Big(\sup_{t\in[a,b]} \exp(\alpha M_N(t)) > \exp(\alpha\delta)\Big)$$

$$+\ P\Big(\sup_{t\in[a,b]} \exp(-\alpha M_N(t)) > \exp(\alpha\delta)\Big)$$

$$\le \exp(-\alpha\delta)[E\exp(\alpha M_N(b)) + E\exp(-\alpha M_N(b))].$$

Inequality (2.2.38) follows from Theorem 3.2, page 353 in Doob (1953) since $\exp(\alpha M_N(t))$ is a submartingale for all $\alpha \in (-\infty,\infty)$.

Applying Lemma 2.2.5 with $X = M_N(b)$ and $X = -M_N(b)$ (2.2.38) can be continued to yield

$$(2.2.39)\quad P(\sup_{t\in[a,b]} |W_N(t)| > \delta) \le 2 \exp\{-\alpha\delta + \alpha^2 f(\alpha\beta_N)\mathrm{Var}(M_N(b))\}$$

where

$$(2.2.40)\qquad \beta_N = \max_{1\le i\le N} |c_i|$$

and

$$\mathrm{Var}(M_N(b)) = \sigma_N^2(b-a)/[1 - (b-a)].$$

Hence, (2.2.35), (2.2.37) and (2.2.39) give

$$(2.2.41)\quad \begin{aligned} &P(\max_{N(k)\le N\le N(k+1)} \sup_{t\in[a,b]} \frac{|W_N(t)|}{\lambda(N)} > \varepsilon) \\ &\le 4 \exp\{-\frac{\alpha\varepsilon\lambda(N(k))}{2} + \alpha^2 f(\alpha\beta_{N(k+1)})\sigma_{N(k+1)}^2 \frac{b-a}{1-(b-a)}\} \end{aligned}$$

for $k = 1,2,\ldots$, $\alpha > 0$, and $\varepsilon^2 \ge 8(b-a)[\sigma_{N(k+1)}^2 - \sigma_{N(k)}^2]/\lambda^2(N(k))$ where $W_N(t) = V_N(t) - V_N(a)$. In the same way, (2.2.41) can be shown to hold if $W_N(t) = V_N(b) - V_N(a)$.

Now set

$$(2.2.42)\qquad \alpha = \frac{\lambda(N(k+1))}{\sigma_{N(k+1)}^2} \frac{\varepsilon}{2\sqrt{\rho}} \frac{1-(b-a)}{b-a} \quad \text{in (2.2.41)}$$

and define

$$(2.2.43)\qquad \gamma_k = \frac{\sqrt{\rho}\,\lambda(N(k))}{\lambda(N(k+1))} - f\left(\frac{\beta_{N(k+1)}\lambda(N(k+1))}{\sigma_{N(k+1)}^2} \frac{\varepsilon}{2\sqrt{\rho}} \frac{1-(b-a)}{b-a}\right).$$

Then the right-hand expression in (2.2.41) can be written as

$$(2.2.44)\quad 4\exp\{-\frac{\lambda^2(N(k+1))}{\sigma^2_{N(k+1)}}\ \frac{\varepsilon^2}{4\rho}\,\frac{1-(b-a)}{b-a}\,\gamma_k\}\ .$$

The assumptions imposed on the sequence $c_1, c_2, \ldots$ and (2.2.34) imply $\lim_{k\to\infty}\gamma_k = \frac{1}{2}$ and $\lim_{k\to\infty} 8[\sigma^2_{N(k+1)} - \sigma^2_{N(k)}]/\lambda^2(N(k)) = 0$. Hence, if $\varepsilon^2 > \frac{8\rho(b-a)}{1-(b-a)}$, then for all sufficiently large k we have $\varepsilon^2 > 8(b-a)[\sigma^2_{N(k+1)} - \sigma^2_{N(k)}]/\lambda^2(N(k))$ and (2.2.44) is less than or equal to

$$(2.2.45)\quad 4\exp\{-\lambda^2(N(k+1))/\sigma^2_{N(k+1)}\} = 4[\log \sigma^2_{N(k+1)}]^{-2}$$

$$\leq 4[\log \rho^{n+k+1}]^{-2} = 4[(n+k+1)\log \rho]^{-2}\ .$$

Since the series $\sum_{k=1}^{\infty} 4[(n+k+1)\log\rho]^{-2} < \infty$, the Borel-Cantelli lemma, (2.2.41), (2.2.44) and (2.2.45) give

$$(2.2.46)\quad 0 = P(\max_{N(k)\leq N\leq N(k+1)}\ \sup_{t\in[a,b]} \frac{|W_N(t)|}{\lambda(N)} > \varepsilon \text{ for infinitely many } k)$$

$$= P(\sup_{t\in[a,b]} \frac{|W_N(t)|}{\lambda(N)} > \varepsilon \text{ for infinitely many } N).$$

Therefore, for each $\rho > 1$, $0 \leq a < b \leq 1$, $b-a \leq \frac{1}{2}$ and $\varepsilon^2 > \frac{8\rho(b-a)}{1-(b-a)}$ we have

$$(2.2.47)\quad P(\sup_{t\in[a,b]} \frac{|W_N(t)|}{\lambda(N)} \leq \varepsilon \quad \text{for all sufficiently large } N) = 1$$

where $W_N(t)$ is either $V_N(t) - V_N(a)$ or $V_N(b) - V_N(t)$. Lemma 2.2.2 now follows from (2.2.47) and Lemma 2.2.3. □

3. WEAK CONVERGENCE OF THE WEIGHTED EMPIRICAL PROCESS WITH MULTIDIMENSIONAL PARAMETER

3.1. Introduction

For each $N = 1,2,\ldots$ let c_{Ni}, $i = 1,2,\ldots,N$ be any real numbers and let $Y_{Ni} = (Y_{Ni1}, Y_{Ni2}, \ldots, Y_{Nik})$, $i = 1,2,\ldots,N$ be k-variate $(k \geq 1)$ random vectors taking values in the k-dimensional unit cube $[0,1]^k$. In this section we define the weighted empirical process by

$$(3.1.1) \quad V_N(t) = \sum_{i=1}^{N} c_{Ni}[I(Y_{Ni} \leq t) - P(Y_{Ni} \leq t)], \quad t \in [0,1]^k$$

where, as usual, if $x = (x_1, x_2, \ldots, x_k)$ and $y = (y_1, y_2, \ldots, y_k)$, then we write $x \leq y$ if and only if $x_i \leq y_i$ for all $i = 1,2,\ldots,k$. We also define the "normalized" weighted empirical process by

$$(3.1.2) \qquad Z_N(t) = V_N(t)/\sigma_N, \quad t \in [0,1]^k$$

where

$$(3.1.3) \qquad \sigma_N^2 = \sum_{i=1}^{N} c_{Ni}^2 .$$

Our goal is to establish sufficient conditions for the Z_N process (3.1.2) to converge weakly in the generalized Skorohod metric space (D_k, d) as defined in Bickel and Wichura (1971). To be sure, weak convergence of the Z_N process has been studied by many authors under a variety of conditions. Therefore, so that our result can be put in perspective with other established results, we shall briefly indicate what has already been done.

To begin with, when $k = 1$, $(\max_{1 \leq i \leq N} c_{Ni}^2)/\sigma_N^2 \to 0$, and for each $N = 1,2,\ldots$, Y_{Ni}, $i = 1,2,\ldots,N$ are statistically independent, Koul (1969), Koul (1970b), Withers (1975), and Shorack (1980) each prove that Z_N converges weakly in (D_k,d). Conditions imposed by these authors on the distribution funtions of the Y_{Ni} vary, but the least restrictive condition is stated in Withers (1975) and Shorack (1980), namely

$$(3.1.4) \quad \lim_{\delta \to 0} \overline{\lim_{N\to\infty}} \sup_{t\in[0,1-\delta]} \frac{1}{\sigma_N^2} \sum_{i=1}^{N} c_{Ni}^2 P(Y_{Ni} \in (t,t+\delta]) = 0.$$

Shorack (1973) also proves Z_N converges weakly, but is limited to the case $c_{Ni} = 1$ and an assumption much stronger than (3.1.4) is imposed.

Several authors have studied the weak convergence of Z_N when Y_{Ni}, $i = 1,2,\ldots,N$ are not independent but satisfy specific "mixing" conditions. For example, when $k = 1$ see Billingsley (1968), Sen (1971), Deo (1973), Yokoyama (1973), Yoshihara (1974), Withers (1975), Mehra and Rao (1975), and Koul (1977). The first five authors only consider the case $c_{Ni} = 1$ and assume Y_{Ni}, $i = 1,2,\ldots$ are identically distributed with a continuous distribution function or with a Uniform $[0,1]$ distribution. Withers (1975) and Koul (1977) both assume

$$(3.1.5) \qquad \sup_{N \geq 1} N(\max_{1 \leq i \leq N} c_{Ni}^2)/\sigma_N^2 < \infty.$$

Furthermore, Withers (1975) assumes (3.1.4) holds along with some other regularity conditions. Koul (1977) assumes the average of the distributions $P(Y_{Ni} \leq t)$, $i = 1,2,\ldots,N$ is Uniform [0,1]. Mehra and Rao (1975) assume each Y_{Ni} is Uniform [0,1] and either $(\max_{1\leq i\leq N} c_{Ni}^2)/\sigma_N^2 \to 0$ or $\sup_{N\geq 1} N^{\delta}(\max_{1\leq i\leq N} c_{Ni}^2)/\sigma_N^2 < \infty$ for some $\delta > 0$ depending on the kind of "mixing" condition assumed.

Weak convergence of Z_N with respect to metrics stronger than the usual Skorohod metric d has been studied in the case $k = 1$ by Pyke and Shorack (1968), O'Reilly (1974), Mehra and Rao (1975), Withers (1976), and Shorack (1980). These authors require $c_{Ni} = 1$ or (3.1.5) except when Y_{Ni}, $i = 1,2,\ldots,N$ are either i.i.d. Uniform [0,1] or identically distributed as Uniform [0,1] and satisfy a certain kind of "mixing" condition in which case $(\max_{1\leq i\leq N} c_{Ni}^2)/\sigma_N^2 \to 0$ suffices.

Among those authors who have studied the weak convergence of Z_N when $k \geq 2$ we have Bickel and Wichura (1971), Neuhaus (1971), Sen (1974), Rüschendorf (1974), Neuhaus (1975), Yoshihara (1975/76), and Rüschendorf (1976). The first six authors each limited their study to the case $c_{Ni} = 1$. Bickel and Wichura (1971) also assumed Y_{Ni}, $i = 1,2,\ldots$ were i.i.d. with a continuous distribution function while Neuhaus (1971) assumed Y_{Ni}, $i = 1,2,\ldots$ were i.i.d. with a distribution function satisfying a Lipschitz condition. Sen (1974) and Rüschendorf (1974) both assumed Y_{Ni}, $i = 1,2,\ldots$ satisfied a certain "mixing" condition. Sen (1974) also assumed Y_{Ni}, $i = 1,2,\ldots$ were identically distributed and had Uniform [0,1] marginal distributions while Rüschendorf (1974) assumed

$$\sup_{N \geq 1} \sup_{1 \leq i \leq N} P(Y_{Ni} \in A) \leq \mu(A)$$

for some measure μ on $[0,1]^k$ with continuous marginals.

Neuhaus (1975), in addition to assuming $c_{Ni} = 1$, assumed Y_{Ni}, $i = 1,2,\ldots$ were independent and the average of the distribution functions of Y_{Ni}, $i = 1,\ldots,N$ had Uniform $[0,1]$ marginals. Yoshihara (1975/76) assumed Y_{Ni}, $i = 1,2,\ldots$ satisfied a certain "mixing" condition and were stationary in addition to assuming $c_{Ni} = 1$.

Rüschendorf (1976) is the only author that has studied the weak convergence of the multiparameter weighted empirical process with general weights c_{Ni}. Most of the results obtained by Rüschendorf (1976) depend on his Lemma 2.1 appearing on page 913 in the same article. From Rüschendorf's description of the proof of Lemma 2.1 it appears to this writer that the proof is incorrect. Hence, at this time no further comment will be made concerning the results in Rüschendorf (1976).

In this paper we extend the results of Koul (1970b), Withers (1975), and Shorack (1980) to the multidimensional parameter weighted empirical process in the case where Y_{Ni}, $i = 1,2,\ldots,N$ are statistically independent (see Theorems 3.2.1 and 3.2.2).

3.2. Weak Convergence of the Weighted Empirical Process

For $k = 1,2,\ldots$ let (D_k,d) denote the (separable) metric space of real-valued functions defined on $[0,1]^k$ which are "continuous from above, with limits from below" as defined in Bickel and Wichura (1971). Furthermore, let C_k denote the set of all continuous real-valued functions defined on $[0,1]^k$. If $t = (t_1,t_2,\dots,t_k)$ is a point in R^k we define the norm of t by $\|t\| = \max\{|t_j| : j = 1,2,\dots,k\}$. Also, if $\delta > 0$ and $x \in D_k$, then we define $w_\delta(x)$ to be the usual modulus of continuity, namely

(3.2.1) $$w_\delta(x) = \sup\{|x(s) - x(t)| : s,t \in [0,1]^k \text{ and } \|s - t\| \le \delta\}.$$

The main results of this section concern the "normalized" weighted empirical process (3.1.2) and are stated as Theorem 3.2.1 and Theorem 3.2.2 .

Theorem 3.2.1. Let $Z_N(t)$, $t \in [0,1]^k$ denote the process in (3.1.2) and assume

(3.2.2) $Y_{N1}, Y_{N2}, \dots, Y_{NN}$ are statistically independent for each $N = 1,2,\dots,$

(3.2.3) $$\lim_{N\to\infty} \left(\max_{1\le i\le N} c_{Ni}^2 \right)/\sigma_N^2 = 0 ,$$

and for each $j = 1,2,\dots,k$

(3.2.4) $$\lim_{\delta\to 0} \overline{\lim_{N\to\infty}} \sup_{x\in[0,1]} \frac{1}{\sigma_N^2} \sum_{i=1}^{N} c_{Ni}^2 \, P(x < Y_{Nij} \le x + \delta) = 0 .$$

Then for all $\varepsilon > 0$

(3.2.5) $$\lim_{\delta\to 0} \overline{\lim_{N\to\infty}} P(w_\delta(Z_N) \ge \varepsilon) = 0 .$$

Theorem 3.2.2. If in addition to the assumptions of Theorem 3.2.1 we also have

(3.2.6) $$\lim_{N\to\infty} \operatorname{Cov}(Z_N(s), Z_N(t)) = \Gamma(s,t)$$

for all $s,t \in [0,1]^k$, then

Z_N converges weakly in (D_k,d) to a zero mean Gaussian process Z having covariance Γ and $P(Z \in C_k) = 1$.

The proof of Theorem 3.2.1 is quite similar to the proof of Theorem 2.2 in Koul (1970b). The main tools used in proving Theorem 3.2.1 are the fluctuation inequalities of Bickel and Wichura (1971) [see Lemma 4.1 and Theorem 4.1 in Section 4 of this paper] while Koul (1970b) uses the fluctuation inequalities in Billingsley (1968).

Theorem 3.2.1 will be proved after first establishing four lemmas. The first lemma, Lemma 3.2.1, provides a necessary and sufficient condition for (3.2.5) to hold. This condition (3.2.9) is more convenient to work with than (3.2.5) when applying Bickel and Wichura's (1971) fluctuation inequalities. The second lemma, Lemma 3.2.2, provides sufficient moment inequalities to justify the use of the fluctuation inequalities in Bickel and Wichura (1971). Finally, the third and fourth lemmas, Lemma 3.2.3 and Lemma 3.2.4, provide inequalities from which Theorem 3.2.1 will follow easily. Theorem 3.2.2 will then follow from Theorem 3.2.1 and an easy application of the multivariate version of the Lindeberg-Feller Central Limit Theorem.

Keeping the preceding remarks in mind let us for each Borel set A in R^k define

$$(3.2.7) \quad Z_N(A) = \frac{1}{\sigma_N} \sum_{i=1}^{N} c_{Ni}[I(Y_{Ni} \in A) - P(Y_{Ni} \in A)].$$

Furthermore, for each $\delta > 0$ and $j = 1,2,\ldots,k$ let $A(j,\delta)$ denote the class of all subsets $A \subseteq [0,1]^k$ having the following form

$$(3.2.8)\quad A = [0,t_1] \times \ldots \times [0,t_{j-1}] \times (x,y] \times [0,t_{j+1}] \times \ldots \times [0,t_k]$$

where $0 \leq y-x \leq \delta$.

Lemma 3.2.1. Let $Z_N(t)$, $t \in [0,1]^k$ denote the process in (3.1.2). Then

$$(3.2.5)\qquad \lim_{\delta\to 0} \overline{\lim_{N\to\infty}} P(w_\delta(Z_N) \geq \varepsilon) = 0 \quad \text{for all} \quad \varepsilon > 0$$

if and only if

$$(3.2.9)\qquad \lim_{\delta\to 0} \overline{\lim_{N\to\infty}} P(w_\delta^{(j)}(Z_N) \geq \varepsilon) = 0 \quad \text{for all} \quad \varepsilon > 0 \quad \text{and} \quad j = 1,2,\ldots,k$$

where

$$(3.2.10)\qquad w_\delta^{(j)}(Z_N) = \sup\{|Z_N(A)| : A \in A(j,\delta)\},$$

$Z_N(A)$ is defined in (3.2.7) and $A(j,\delta)$ is the class of sets of the form (3.2.8).

Proof. First observe that if $s = (s_1,s_2,\ldots,s_k)$ and $t = (t_1,t_2,\ldots,t_k)$ are any two points in $[0,1]^k$ with $\|s - t\| \leq \delta$ and $u = (s_1 \vee t_1,\ldots,s_k \vee t_k)$, then $s \leq u$, $t \leq u$, $\|s - u\| \leq \delta$, and $\|t - u\| \leq \delta$. Hence, by the triangle inequality we have

$$|Z_N(s) - Z_N(t)| \leq |Z_N(s) - Z_N(u)| + |Z_N(t) - Z_N(u)|$$

and it follows that

$$(3.2.11)\qquad w'_\delta(Z_N) \leq w_\delta(Z_N) \leq 2\, w'_\delta(Z_N)$$

where

$$(3.2.12)\quad w'_\delta(Z_N) = \sup\{|Z_N(s) - Z_N(t)| : s,t \in [0,1]^k,\ s \leq t,\ \|s - t\| \leq \delta\}.$$

It is now clear that (3.2.5) will hold if and only if

$$\text{(3.2.13)}\quad \lim_{\delta\to 0}\ \overline{\lim_{N\to\infty}}\ P(w'_\delta(Z_N) \geq \varepsilon) = 0 \quad \text{for all} \quad \varepsilon > 0.$$

We now show (3.2.13) is equivalent to (3.2.9).

Let $s = (s_1,\ldots,s_k)$ and $t = (t_1,\ldots,t_k)$ be any points in $[0,1]^k$ with $s \leq t$ and $\|s - t\| \leq \delta$. Then (3.1.2) and (3.2.7) give

$$\text{(3.2.14)}\qquad Z_N(t) - Z_N(s) = Z_N([0,t] \setminus [0,s]).$$

It is also clear from (3.2.7) that if A and B are disjoint Borel sets in R^k, then $Z_N(A \cup B) = Z_N(A) + Z_N(B)$. Hence, if we define A_j, $j = 1,2,\ldots,k$ by

$$\text{(3.2.15)}\quad A_j = [0,s_1] \times\ldots\times [0,s_{j-1}] \times (s_j,t_j] \times [0,t_{j+1}] \times\ldots\times [0,t_k],$$

then $A_1,A_2,\ldots,A_k$ is a partition of the set $[0,t] \setminus [0,s]$ and it follows from (3.2.14) that

$$\text{(3.2.16)}\qquad Z_N(t) - Z_N(s) = \sum_{j=1}^{k} Z_N(A_j)\ .$$

Since $A_j \in A(j,\delta)$, $j = 1,2,\ldots,k$ (3.2.16) gives

$$\text{(3.2.17)}\qquad w'_\delta(Z_N) \leq \sum_{j=1}^{k} w_\delta^{(j)}(Z_N).$$

On the other hand if $A \in A(j,\delta)$, then A has the form

$$A = [0,u_1] \times\ldots\times [0,u_{j-1}] \times (x,y] \times [0,u_{j+1}] \times\ldots\times [0,u_k]$$

where $0 \leq y-x \leq \delta$.

If $s = (s_1, s_2, \ldots, s_k)$ and $t = (t_1, t_2, \ldots, t_k)$ are defined by $s_j = x$, $t_j = y$ and $s_i = u_i = t_i$ for $i \neq j$, then $s, t \in [0,1]^k$, $s \leq t$, $\|s - t\| \leq \delta$, and

$$A = [0,t] \setminus [0,s].$$

Hence, $Z_N(A) = Z_N([0,t] \setminus [0,s]) = Z_N(t) - Z_N(s)$ and it follows that

$$w_\delta^{(j)}(Z_N) \leq w_\delta'(Z_N) \quad \text{for all} \quad j = 1,2,\ldots,k \ . \tag{3.2.18}$$

Lemma 3.2.1 now follows from (3.2.13), (3.2.17), and (3.2.18). □

<u>Lemma 3.2.2</u>. Let $Y_1, Y_2, \ldots, Y_N$ be statistically independent k-variate random vectors taking values in R^k, let $c_1, c_2, \ldots, c_N$ be any real numbers, and for each Borel set A in R^k define

$$Z_N(A) = \sum_{i=1}^{N} c_i [I(Y_i \in A) - P(Y_i \in A)]$$

and

$$\mu_N(A) = \sum_{i=1}^{N} c_i^2 \, P(Y_i \in A) \ .$$

Then

$$E|Z_N(A)|^2 |Z_N(B)|^2 \leq 3 \, \mu_N^2(A \cup B) \quad \text{if } A \text{ and } B \text{ are disjoint;} \tag{3.2.19}$$

and

$$E|Z_N(A)|^4 \leq 3 \, \mu_N^2(A) + (\max_{1 \leq i \leq N} c_i^2) \mu_N(A). \tag{3.2.20}$$

<u>Proof</u>. For each Borel set A in R^k and $i = 1,2,\ldots,N$ define

$$X(A_i) = I(Y_i \in A) - P(Y_i \in A)$$

and

$$P(A_i) = P(Y_i \in A) \ .$$

Now let A and B denote any two Borel sets in R^k. Since $Y_1, Y_2, \ldots, Y_N$ are independent and $E\,X(A_i) = 0$ for all $i = 1,2,\ldots,N$ and all Borel sets A, we have

$$E|Z_N(A)|^2|Z_N(B)|^2 = \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{\ell=1}^{N} c_i c_j c_k c_\ell\, E[X(A_i)X(A_j)X(B_k)X(B_\ell)]$$

$$(3.2.21) = \sum_{i=1}^{N} c_i^4\, EX^2(A_i)X^2(B_i) + \sum_{\substack{i=1\\ i\neq j}}^{N}\sum_{j=1}^{N} c_i^2c_j^2[EX^2(A_i)][EX^2(B_j)]$$

$$+ 2\sum_{\substack{i=1\\ i\neq j}}^{N}\sum_{j=1}^{N} c_i^2c_j^2[E\,X(A_i)X(B_i)][E\,X(A_j)X(B_j)].$$

Since $|X(A_i)| \leq 1$ and $E\,X^2(A_i) \leq P(A_i)$, (3.2.21) gives the following result when $A = B$.

$$E|Z_N(A)|^4 = \sum_{i=1}^{N} c_i^4E\,X^4(A_i) + 3\sum_{\substack{i=1\\ i\neq j}}^{N}\sum_{j=1}^{N} c_i^2c_j^2[E\,X^2(A_i)][E\,X^2(A_j)]$$

$$\leq (\max_{1\leq i\leq N} c_i^2)\sum_{i=1}^{N} c_i^2P(A_i) + 3[\sum_{i=1}^{N} c_i^2P(A_i)]^2 .$$

This proves (3.2.20).

If A and B are disjoint, then $E\,X(A_i)X(B_i) = -P(A_i)P(B_i)$ and $E\,X^2(A_i)X^2(B_i) \leq P(A_i)P(B_i)$ so that

$$E|Z_N(A)|^2|Z_N(B)|^2 \leq \sum_{i=1}^{N} c_i^4P(A_i)P(B_i) + 3\sum_{\substack{i=1\\ i\neq j}}^{N}\sum_{j=1}^{N} c_i^2c_j^2P(A_i)P(B_j)$$

$$\leq 3[\sum_{i=1}^{N} c_i^2P(A_i)][\sum_{i=1}^{N} c_i^2P(B_i)].$$

Hence, (3.2.19) follows. □

Lemma 3.2.3. Let $Y_1, Y_2, \ldots, Y_N$ be statistically independent k-variate random vectors taking values in $[0,1]^k$; let $c_1, c_2, \ldots, c_N$

be any real numbers; for each Borel set A in R^k define $Z_N(A) = \sum_{i=1}^{N} c_i[I(Y_i \in A) - P(Y_i \in A)]$; for each $j = 1,2,\ldots,k$ and each Borel set B in R define $Z_{Nj}(B) = \sum_{i=1}^{N} c_i[I(Y_{ij} \in B) - P(Y_{ij} \in B)]$ and $\mu_{Nj}(B) = \sum_{i=1}^{N} c_i^2 P(Y_{ij} \in B)$ where $Y_i = (Y_{i1}, Y_{i2}, \ldots, Y_{ik})$ for $i = 1,2,\ldots,N$; and for each $j = 1,2,\ldots,k$, $\delta > 0$, and $x \in [0,1]$ let $A(j,\delta,x)$ denote the class of all subsets A of the form $A = [0,t_1] \times\ldots\times [0,t_{j-1}] \times (x,t_j] \times [0,t_{j+1}] \times\ldots\times [0,t_k]$ where $t_j \in [x, x+\delta]$ and $t_i \in [0,1]$ for $i \neq j$. Then for each $\varepsilon > 0$, $\delta > 0$, $j = 1,2,\ldots,k$, and $x \in [0,1]$

$$(3.2.22)\quad P(\sup\{|Z_N(A)| : A \in A(j,\delta,x)\} \geq \varepsilon) \leq P(|Z_{Nj}((x,x+\delta])| \geq \tfrac{\varepsilon}{2}) + \frac{3kC_k(2,4)}{(\varepsilon/2k)^4}\, \mu_{Nj}^2((x,x+\delta])$$

where $C_k(2,4)$ is a constant depending only on k.

<u>Proof</u>. Let $\varepsilon > 0$, $\delta > 0$, $j \in \{1,2,\ldots,k\}$, and $x \in [0,1]$. For each $t = (t_1,\ldots,t_k)$ in R^k define

$$Z_N(t) = \sum_{i=1}^{N} c_i[I(Y_i \leq t) - P(Y_i \leq t)],$$

$$t_x = (t_1,\ldots,t_{j-1},x,t_{j+1},\ldots,t_k),$$

and

$$(3.2.23)\qquad X_N(t) = Z_N(t) - Z_N(t_x)\ .$$

Let us first observe that the fluctuation inequalities in Bickel and Wichura (1971) (see Section 4, Lemma 4.1 and Theorem 4.1 in this paper) can be applied to the stochastic process $X_N(t)$, $t \in T$ where $T = T_1 \times T_2 \times\ldots\times T_k$,

$$T_j = [x, x+\delta] \ , \quad \text{and} \quad T_i = [-\delta, 1+\delta] \quad \text{for} \quad i \neq j.$$

If $t = (t_1,\ldots,t_k)$ is a point in the lower boundary of T, $\ell(T)$ (see Section 4 (4.2)), then either $t_j = x$ or $t_i = -\delta$ for some $i \neq j$. If $t_j = x$, then $t = t_x$ and $X_N(t) = 0$. If $t_i = -\delta$, then $X_N(t) = 0$ since $Y_1,\ldots,Y_N$ take values in $[0,1]^k$.

Therefore, $X_N(t) = 0$ for all $t \in \ell(T)$ so that Lemma 4.1 in Section 4 can be applied to give

$$(3.2.24) \qquad \sup_{t \in T} |X_N(t)| \leq |X_N(b)| + k \max_{1 \leq j \leq k} M_j''(X_N)$$

where $b = (b_1,\ldots,b_k)$, $b_j = x + \delta$, and $b_i = 1 + \delta$ for $i \neq j$. Since $Y_1, Y_2,\ldots,Y_N$ take values in $[0,1]^k$, it is easy to see from (3.2.23) that

$$(3.2.25) \qquad \sup_{t \in T} |X_N(t)| = \sup\{|X_N(t)| : t \in T \cap [0,1]^k\}$$

and

$$(3.2.26) \qquad X_N(b) = \sum_{i=1}^{N} c_i[I(Y_{ij} \in (x, x+\delta]) - P(Y_{ij} \in (x, x+\delta])]$$

$$= Z_{Nj} \ ((x, x+\delta]) \ .$$

Furthermore, if $t = (t_1,\ldots,t_k)$ is a point in $T \cap [0,1]^k$, then (3.2.23) gives

$$(3.2.27) \qquad X_N(t) = Z_N(A) = \sum_{i=1}^{N} c_i[I(Y_i \in A) - P(Y_i \in A)]$$

where $A = [0,t_1] \times\ldots\times [0,t_{j-1}] \times (x,t_j] \times [0,t_{j+1}] \times\ldots\times [0,t_k]$. Combining (3.2.25) and (3.2.27) yields

$$(3.2.28) \qquad \sup_{t \in T} |X_N(t)| = \sup\{|Z_N(A)| : A \in A(j,\delta,x)\}$$

If $B = (s,t] \cap T$ is a block in T (see (4.3) in Section 4) then by Lemma 4.2, Section 4 the increment (see (4.7) in Section 4) of $Z_N(t)$, $t \in T$ around B is $Z_N(B) = \sum_{i=1}^{N} c_i[I(Y_i \in B) - P(Y_i \in B)]$. Furthermore, the increment of $Z_N(t_x)$, $t \in T$ around B is zero. Hence, the increment of $X_N(t)$, $t \in T$ around B is

$$(3.2.29) \quad X_N(B) = Z_N(B) = \sum_{i=1}^{N} c_i[I(Y_i \in B) - P(Y_i \in B)]$$

and Lemma 3.2.2 gives

$$(3.2.30) \quad P(\min\{|X_N(A)|,|X_N(B)|\} \geq \lambda) = P(\min\{|Z_N(A)|,|Z_N(B)|\} \geq \lambda)$$

$$\leq \frac{1}{\lambda^4} E|Z_N(A)|^2|Z_N(B)|^2 \leq \frac{3}{\lambda^4} \mu_N^2(A \cup B)$$

for all $\lambda > 0$ and every pair A, B of disjoint neighboring blocks in T where $\mu_N(A) = \sum_{i=1}^{N} c_i^2 P(Y_i \in A)$. Theorem 4.1 in Section 4 can now be applied to yield

$$(3.2.31) \quad P(\max_{1 \leq j \leq k} M_j''(X_N) \geq \lambda) \leq \frac{3kC_k(2,4)}{\lambda^4} \mu_N^2(T \setminus \ell(T))$$

for all $\lambda > 0$ where $C_k(2,4)$ is a constant defined in (4.13) in Section 4.

Since $Y_1, Y_2, \ldots, Y_N$ take values in $[0,1]^k$, it is easy to see that

$$(3.2.32) \quad \mu_N(T \setminus \ell(T)) = \sum_{i=1}^{N} c_i^2 P(Y_{ij} \in (x, x + \delta]) = \mu_{Nj}((x, x + \delta]).$$

Finally, Lemma 3.2.3 follows from (3.2.24), (3.2.26), (3.2.28), (3.2.31), and (3.2.32). $\square$

<u>Lemma 3.2.4</u>. Suppose the assumptions of Lemma 3.2.3 hold. Then for each $\varepsilon > 0$, $\delta \in (0,1)$, and $j = 1,2,\ldots,k$

(3.2.33) $P(\sup\{|Z_N(A)| : A \in A(j,\delta)\} \geq \varepsilon) \leq$

$$\frac{1}{\varepsilon^4}[d_k \sup_{x\in[0,1]} \mu_{Nj}((x, x+\delta]) + 6^4 \max_{1\leq i\leq N} c_i^2] \sum_{i=1}^{N} c_i^2$$

where d_k is a constant depending only on k.

Proof. Let $\varepsilon > 0$, $\delta \in (0,1)$, and $j \in \{1,2,\ldots,k\}$. Furthermore, define $m(\delta) = \min\{m \in \{1,2,\ldots\} : 1 \leq m\delta\}$. For each $t = (t_1,\ldots,t_k)$ in R^k define

$$Z_N(t) = \sum_{i=1}^{N} c_i[I(Y_i \leq t) - P(Y_i \leq t)]$$

and

$$t_x = (t_1,\ldots,t_{j-1},x,t_{j+1},\ldots,t_k) \quad \text{for} \quad x \in R.$$

If $A(j,\delta)$ is the class of sets defined in (3.2.8) and $A \in A(j,\delta)$, then $A \subset [0,1]^k$ and has the form $A = [0,t_1] \times\ldots\times [0,t_{j-1}] \times (x,y] \times [0,t_{j+1}] \times\ldots\times [0,t_k]$ where $0 \leq y-x \leq \delta$. Clearly, $x \in [m_1\delta, (m_1+1)\delta]$ and $y \in [m_2\delta, (m_2+1)\delta]$ for some integers $m_1,m_2 \in \{0,1,\ldots,m(\delta)-1\}$ satisfying either $m_1 = m_2$ or $m_2 = m_1 + 1$. Hence,

$$\begin{aligned}(3.2.34)\quad |Z_N(A)| &= |Z_N(t_y) - Z_N(t_x)| \\ &\leq |Z_N(t_y) - Z_N(t_{m_2\delta})| + |Z_N(t_{m_2\delta}) - Z_N(t_{m_1\delta})| \\ &\quad + |Z_N(t_{m_1\delta}) - Z_N(t_x)| \\ &\leq 3 \max_{0\leq i<m(\delta)} \sup\{|Z_N(B)| : B \in A(j,\delta,i\delta)\}\end{aligned}$$

where $A(j,\delta,x)$ is the class of sets defined in Lemma 3.2.3.

Therefore, (3.2.34) and Lemma 3.2.3 give

$P(\sup\{|Z_N(A)| : A \in A(j,\delta)\} \geq \varepsilon) \leq$

$$P(\max_{0\leq i<m(\delta)} \sup\{|Z_N(B)| : B \in A(j,\delta,i\delta)\} \geq \frac{1}{3}\varepsilon) \leq$$

$$\sum_{i=0}^{m(\delta)-1} P(\sup\{|Z_N(B)| : B \in A(j,\delta,i\delta)\} \geq \frac{1}{3}\varepsilon) \leq$$

(3.2.35)

$$\sum_{i=0}^{m(\delta)-1} [P(|Z_{Nj}((i\delta,(i+1)\delta])| \geq \frac{1}{6}\varepsilon) + \frac{3kC_k(2,4)}{(\varepsilon/6k)^4}\mu_{Nj}^2((i\delta,(i+1)\delta])].$$

Applying Chebysev's inequality and Lemma 3.2.2 to Z_{Nj} gives the following upper bound for the first term inside the brackets in (3.2.35)

$$(3.2.36)[1/(\varepsilon/6)^4][3\ \mu_{Nj}^2((i\delta,(i+1)\delta]) + (\max_{1\leq i\leq N} c_i^2)\mu_{Nj}((i\delta,(i+1)\delta])].$$

Hence, the expression (3.2.35) is bounded above by

$$(3.2.37)\quad \frac{d_k}{\varepsilon^4}\mu_{Nj}((0,1])\max_{0\leq i<m(\delta)}\mu_{Nj}((i\delta,(i+1)\delta]) + \frac{6^4}{\varepsilon^4}(\max_{1\leq i\leq N} c_i^2)\mu_{Nj}((0,1])$$

where $d_k = 3kC_k(2,4)(6k)^4 + 3\cdot 6^4$ and $\mu_{Nj}((0,1]) \leq \sum_{i=1}^{N} c_i^2$. Lemma 3.2.4 now follows from (3.2.35) and (3.2.37). $\square$

Proof of Theorem 3.2.1.

For each $N = 1,2,\ldots$ apply Lemma 3.2.4 with Y_{Ni} in place of Y_i and c_{Ni}/σ_N in place of c_i to get

$$P(w_\delta^{(j)}(Z_N) \geq \varepsilon) \leq \frac{1}{\varepsilon^4}[d_k \sup_{x\in[0,1]}\mu_{Nj}((x,x+\delta]) + \frac{6^4(\max_{1\leq i\leq N} c_{Ni}^2)}{\sigma_N^2}]$$

for each $N = 1,2,\ldots$, $j = 1,2,\ldots,k$, $\delta \in (0,1)$, $\varepsilon > 0$ where $w_\delta^{(j)}(Z_N)$ is defined by (3.2.10) in Lemma 3.2.1, d_k is a constant depending only on k, $\sigma_N^2 = \sum_{i=1}^{N} c_{Ni}^2$, and $\mu_{Nj}((x,x+\delta]) = \frac{1}{\sigma_N^2}\sum_{i=1}^{N} c_{Ni}^2 P(x < Y_{Nij} \leq x+\delta)$.

Theorem 3.2.1 now follows from (3.2.3), (3.2.4), and Lemma 3.2.1. □

Proof of Theorem 3.2.2.

By Theorem 2, page 683 in Wichura (1969), the result in Theorem 3.2.2 will follow if each of the following two conditions hold:

(i) for each finite subset T of $[0,1]^k$, the distribution of $Z_N(t)$, $t \in T$ converges weakly to a multivariate Normal distribution,

and

(ii) for each $\varepsilon > 0$, $\lim_{\delta \to 0} \overline{\lim_{N \to \infty}} P(w_\delta(Z_N) \geq \varepsilon) = 0$.

Condition (ii) holds as a result of Theorem 3.2.1 while condition (i) follows from (3.2.6), (3.2.3) and an easy application of the multivariate version of the Lindeberg-Feller Central Limit Theorem. □

4. APPENDIX

The paper by Bickel and Wichura (1971) provides the key tool for proving the weak convergence of the weighted empirical process. In this appendix we present some notation, terminology, and results that can be found in Bickel and Wichura (1971); although we occasionally make statements in a slightly more general form than Bickel and Wichura.

To begin with, let k denote a positive integer and $T = T_1 \times T_2 \times \ldots \times T_k$ where for each $j = 1,2,\ldots,k$, T_j is either a finite subset of $(-\infty,\infty)$ or a closed bounded interval in $(-\infty,\infty)$. Furthermore, let

(4.1) $\quad a_j = \inf T_j \quad$ and $\quad b_j = \sup T_j \quad$ for $\quad j = 1,2,\ldots,k.$

The <u>lower boundary</u> of T is defined to be the set

(4.2) $\quad \ell(T) = \{t = (t_1,\ldots,t_k) \in T : t_j = a_j \text{ for some } j = 1,2,\ldots,k\}.$

A <u>block</u> in T is a set B of the form

(4.3) $\quad B = (s,t] \cap T \quad$ where $\quad s,t \in T,\ \ s \leq t,$

and $(s,t] = (s_1,t_1] \times \ldots \times (s_k,t_k]$. We say two disjoint blocks $A = (s,t] \cap T$ and $B = (u,v] \cap T$ are <u>neighbors</u> if s agrees with u and t agrees with v except in the j^{th} coordinate (for some $j = 1,2,\ldots,k$) where either $s_j \leq t_j = u_j \leq v_j$ or $u_j \leq v_j = s_j \leq t_j$.

We next introduce a stochastic process $X(t)$, $t \in T$ whose state space F is a normed linear space having norm $|\cdot|$. For $j = 1,2,\dots,k$ and $t \in T_j$ we define the stochastic process $X_t^{(j)}$ having parameter set $T^{(j)} = T_1 \times\dots\times T_{j-1} \times T_{j+1} \times\dots\times T_k$ by

$$(4.4) \qquad X_t^{(j)}(s) = X(s_1,\dots,s_{j-1},t,s_{j+1},\dots,s_k)$$

where $s = (s_1,\dots,s_{j-1},s_{j+1},\dots,s_k) \in T^{(j)}$.

For $j = 1,2,\dots,k$ and $s,t,u \in T_j$ with $s \le t \le u$ we define

$$(4.5) \quad m_j(s,t,u)(X) = \min\{\|X_s^{(j)} - X_t^{(j)}\|_\infty, \|X_t^{(j)} - X_u^{(j)}\|_\infty\}$$

and

$$(4.6) \qquad M_j''(X) = \sup\{m_j(s,t,u)(X) : s,t,u \in T_j \text{ and } s \le t \le u\}.$$

Finally, we define the <u>increment</u> of X around the block $B = (s,t] \cap T$ by

$$(4.7) \qquad X(B) = \sum_{\delta\in\{0,1\}^k} (-1)^{k-(\delta_1+\dots+\delta_k)} X(s + \delta(t-s))$$

where $\delta = (\delta_1,\delta_2,\dots,\delta_k)$ and

$$s + \delta(t-s) = (s_1 + \delta_1(t_1 - s_1),\dots,s_k + \delta_k(t_k - s_k)).$$

We now state two results from Bickel and Wichura (1971) which will be used to prove Theorem (3.2.1).

<u>Lemma 4.1</u>. If $X(t) = 0$ for $t \in \ell(T)$, then

$$(4.8) \qquad \sup_{t\in T} |X(t)| \le |X(b)| + k \max_{1\le j\le k} M_j''(X)$$

where $b = (b_1,\dots,b_k)$ and $b_j = \sup T_j$, $j = 1,\dots,k$.

<u>Proof</u>. See (1) on page 1657 in Bickel and Wichura (1971). □

Theorem 4.1 (see Theorem 1, page 1658 in Bickel and Wichura (1971)). Assume $X(t) = 0$ for $t \in \ell(T)$ and

$$(4.9) \quad P(\min\{|X(A)|, |X(B)|\} \geq \lambda) \leq \mu^{\beta}(A \cup B)/\lambda^{\gamma}$$

for some numbers $\gamma > 0$, $\beta > 1$, and some nonnegative finite measure μ on T, and all $\lambda > 0$ and every pair A,B of disjoint neighboring blocks in T. Then for all $\lambda > 0$ we have

$$(4.10) \quad P(M_j''(X) \geq \lambda) \leq \frac{C_k(\beta,\lambda)}{\lambda^{\gamma}} \mu^{\beta}(T \setminus \ell(T)), \quad j = 1,2,\ldots,k$$

$$(4.11) \quad P(\max_{1 \leq j \leq k} M_j''(X) \geq \lambda) \leq \frac{kC_k(\beta,\gamma)}{\lambda^{\gamma}} \mu^{\beta}(T \setminus \ell(T))$$

where

$$(4.12) \quad C_1(\beta,\gamma) = 2^{\beta+\gamma}[1 - (\tfrac{1}{2})^{\frac{\beta-1}{1+\gamma}}]^{-(1+\gamma)}$$

and

$$(4.13) \quad C_k(\beta,\gamma) = C_1(\beta,\gamma)[1 + (k-1)C_{k-1}^{1/1+\gamma}(\beta,\gamma)]^{1+\gamma}, \quad k = 2,3,\ldots$$

Proof. With one minor change (see remark (2) below) Theorem 4.1 follows from the proof of Theorem 1 in Bickel and Wichura (1971). □

Remarks concerning Theorem 4.1:

(1) The inequalities in Theorem 1 (Bickel and Wichura (1971)) which are analogous to (4.10) and (4.11) in Theorem 4.1 have $\mu(T)$ appearing instead of $\mu(T \setminus \ell(T))$. Furthermore, Bickel and Wichura assume $\mu(\ell(T)) = 0$. However, with one minor change in the definition of F in Step 3 of Bickel and Wichura's proof

of Theorem 1, it is seen that the assumption $\mu(\ell(T)) = 0$ is superfluous and also that the inequalities hold with $\mu(T \setminus \ell(T))$ in place of $\mu(T)$. This gives a slightly sharper inequality with one less assumption imposed on μ.

(2) The change referred to in the above remark is to define F in Step 3, page 1660 in Bickel and Wichura as follows:

Let F be linear over $[t_{j-1}, t_j]$, $j = 1,2,\ldots,m$ with $F(t_0) = 0$, $F(t_1) = \mu(\{t_1\})$, and $F(t_j) - F(t_{j-1}) = \mu(\{t_{j-1}\}) + \mu(\{t_j\})$ for $j = 2,\ldots,m$.

(3) The constants $C_k(\beta,\gamma)$ appearing in Theorem 4.1 are the same as $K_q(\beta,\gamma)$ in Bickel and Wichura with $q = k$.

Finally, in Lemma 4.2 below we determine a more convenient form for the increment (see (4.7)) of the weighted empirical process around a block B.

<u>Lemma 4.2</u>. Let $Y_1, Y_2, \ldots, Y_N$ be k-variate random vectors taking values in R^k, let $c_1, c_2, \ldots, c_N$ be any real numbers, and for each $t \in R^k$ define

$$(4.14) \qquad Z_N(t) = \sum_{i=1}^{N} c_i[I(Y_i \leq t) - P(Y_i \leq t)].$$

Then for any block $B = (s,t]$ in R^k, the increment of Z_N around the block B is

$$(4.15) \qquad Z_N(B) = \sum_{i=1}^{N} c_i[I(Y_i \in B) - P(Y_i \in B)].$$

<u>Proof</u>. For $i = 1,2,\ldots,N$ define the stochastic processes $X_i(t)$, $t \in R^k$ by $X_i(t) = I(Y_i \leq t)$.

If $B = (s,t]$ is any block in R^k, then by (4.7) the increment of X_i around the block B is

$$X_i(B) = \sum_{\delta\in\{0,1\}^k} (-1)^{k-(\delta_1+\ldots+\delta_k)} X_i(s + \delta(t-s)). \tag{4.16}$$

We shall prove by induction on k that

$$X_i(B) = I(Y_i \in B) . \tag{4.17}$$

When $k = 1$, the right-hand side of (4.16) is

$$X_i(t) - X_i(s) = I(Y_i \leq t) - I(Y_i \leq s) = I(Y_i \in (s,t])$$

so that (4.17) is established for $k = 1$. Suppose that (4.17) holds for some $k \in \{1,2,\ldots\}$. We now show (4.17) also holds for $k + 1$.

If $\delta = (\delta_1,\delta_2,\ldots,\delta_{k+1})$, $s = (s_1,s_2,\ldots,s_{k+1})$, $t = (t_1,t_2,\ldots,t_{k+1})$, and $Y_i = (Y_{i1},Y_{i2},\ldots,Y_{i,k+1})$, then define $\delta_0 = (\delta_1,\delta_2,\ldots,\delta_k)$, $s_0 = (s_1,s_2,\ldots,s_k)$, $t_0 = (t_1,t_2,\ldots,t_k)$, and $Y_{i0} = (Y_{i1},Y_{i2},\ldots,Y_{ik})$. Hence, (4.16) gives

$$\begin{aligned}
X_i(B) &= \sum_{\delta\in\{0,1\}^{k+1}} (-1)^{k+1-(\delta_1+\ldots+\delta_{k+1})} I(Y_i \leq s + \delta(t-s)) \\
&= \sum_{\delta_0\in\{0,1\}^k} (-1)^{k-(\delta_1+\ldots+\delta_k)} I(Y_{i0} \leq s_0 \\
&\qquad + \delta_0(t_0 - s_0))I(Y_{i,k+1} \leq t_{k+1}) \\
&\quad - \sum_{\delta_0\in\{0,1\}^k} (-1)^{k-(\delta_1+\ldots+\delta_k)} I(Y_{i0} \leq s_0 \\
&\qquad + \delta_0(t_0 - s_0))I(Y_{i,k+1} \leq s_{k+1}) \\
&= I(Y_{i,k+1} \in (s_{k+1},t_{k+1}])I(Y_{i0} \in (s_0,t_0]) \\
&= I(Y_i \in (s,t]).
\end{aligned}$$

Hence, (4.17) holds for all $k = 1,2,\ldots$

Finally, if we define the process $X(t)$, $t \in R^k$ by

$$X(t) = \sum_{i=1}^{N} c_i\, X_i(t) = \sum_{i=1}^{N} c_i I(Y_i \leq t),$$

then

$$Z_N(t) = X(t) - E(X(t)), \quad t \in R^k. \tag{4.18}$$

If $B = (s,t]$ is a block in R^k and $W(B)$ denotes the increment of $E(X(t))$ around B, then

$$Z_N(B) = X(B) - W(B)\ . \tag{4.19}$$

Since $E\, Z_N(t) = 0$ for all $t \in R^k$ we have

$$E\, Z_N(B) = 0$$

so that (4.19) gives $E\, X(B) = W(B)$ and

$$Z_N(B) = X(B) - E\, X(B). \tag{4.20}$$

The lemma now follows from (4.20) and (4.17) after observing $X(B) = \sum_{i=1}^{N} c_i\, X_i(B)$. $\square$

BIBLIOGRAPHY

BIBLIOGRAPHY

Aronszajn, N. (1950). The Theory of Reproducing Kernels. Trans. Amer. Math. Soc. 68 337-404.

Berning, J.A. (1979). On the Multivariate Law of the Iterated Logarithm. Ann. Probability 7 980-988.

Bickel, P.J. and Wichura, M.J. (1971). Convergence Criteria for Multiparameter Stochastic Processes and Some Applications. Ann. Math. Statist. 42 1656-1670.

Billingsley, P. (1968). Convergence of Probability Measures. Wiley, New York.

Chow, Y.S. and Teicher, H. (1978). Probability Theory, Independence, Interchangeability, Martingales. Springer-Verlag, New York.

Deo, C.M. (1973). A Note On Empirical Processes of Strong-Mixing Sequences. Ann. Probability 1 870-875.

Doob, J.L. (1953). Stochastic Processes. Wiley, New York.

Finkelstein, Helen (1971) The Law of the Iterated Logarithm for Empirical Distributions. Ann. Math. Statist. 42 607-615.

James, B.R. (1975). A Functional Law of the Iterated Logarithm for Weighted Empirical Distributions. Ann. Probability 3 762-772.

Koul, H.L. (1969). Asymptotic Behavior of Wilcoxon Type Confidence Regions in Multiple Linear Regression. Ann. Math. Statist. 40 1950-1979.

Koul, H.L. (1970a). A Class of ADF Tests for Subhypothesis in the Multiple Linear Regression. Ann. Math. Statist. 41 1273-1281.

Koul, H.L. (1970b). Some Convergence Theorems for Ranks and Weighted Empirical Cumulatives. Ann. Math. Statist. 41 1768-1773.

Koul, H.L. (1977). Behavior of Robust Estimators in the Regression Model with Dependent Errors. Ann. Statist. 5 681-699.

Koul, H.L. (1980). Some Weighted Empirical Inferential Procedures for a Simple Regression Model. To appear.

Koul, H.L. and Staudte, R.G., Jr. (1972). Weak Convergence of Weighted Empirical Cumulatives Based on Ranks. Ann. Math. Statist. 43 832-841.

Kuelbs, J. (1976). A Strong Convergence Theorem for Banach Space Valued Random Variables. Ann. Probability 4 744-771.

Kuelbs, J. (1977). Kolmogorov's Law of the Iterated Logarithm for Banach Space Valued Random Variables. Illinois J. of Mathematics 21 748-800.

Loéve, M. (1963). Probability Theory, 3rd Ed. Van Nostrand-Reinhold, New Jersey.

Mehra, K.L. and Rao, M.S. (1975). Weak Convergence of Generalized Empirical Processes Relative to d_q under Strong Mixing. Ann. Probability 3 979-991.

Neuhaus, G. (1971). On Weak Convergence of Stochastic Processes with Multidimensional Time Parameter. Ann. Math. Statist. 42 1285-1295.

Neuhaus, G. (1975). Convergence of the Reduced Empirical Process for Non i.i.d. Random Vectors. Ann. Statist. 3 528-531.

Oodaira, H. (1972). On Strassen's Version of the Law of the Iterated Logarithm for Gaussian Processes. Z. Wahr. 21 289-299.

O'Reilly, N.E. (1974). On the Weak Convergence of Empirical Processes in Sup-norm Metrics. Ann. Probability 2 642-651.

Philipp, Walter (1977). A Functional Law of the Iterated Logarithm for Empirical Distribution Functions of Weakly Dependent Random Variables. Ann. Probability 5 319-350.

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.

Rüschendorf, L. (1974). On the Empirical Process of Multivariate, Dependent Random Variables. J. Multivariate Anal. 4 469-478.

Rüschendorf, L. (1976). Asymptotic Distributions of Multivariate Rank Order Statistics. Ann. Statist. 4 912-923.

Sen, P.K. (1971). A Note on Weak Convergence of Empirical Processes for Sequences of φ-mixing Random Variables. Ann. Math. Statist. 42 2131-2133.

Sen, P.K. (1974). Weak Convergence of Multidimensional Empirical Processes for Stationary φ-mixing Processes. Ann. Probability 2 147-154.

Shorack, Galen R. (1973). Convergence of Reduced Empirical and Quantile Processes with Application to Functions of Order Statistics in the Non-i.i.d. Case. Ann. Statist. 1 146-152.

Shorack, G.R. (1980). The Weighted Empirical Process of Row Independent r.v.'s with Arbitrary d.f.'s. To appear.

Singh, R.S. (1975). On the Glivenko-Cantelli Theorem for Weighted Empiricals Based on Independent Random Variables. Ann. Probability 3 371-374.

Sinha, A.N. and Sen, P.K. (1979). Progressively Censored Tests for Clinical Experiments and Life Testing Problems Based on Weighted Empirical Distributions. Comm. Statist., Part A, Theor. Meth. 8 871-897.

VanZuijlen, M.C.A. (1976). Empirical Distributions and Rank Statistics. Mathematisch Centrum, Amsterdam.

Wellner, J.A. (1977). A Martingale Inequality for the Empirical Process. Ann. Probability 5 303-308.

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.

Withers, C.S. (1975). Convergence of Empirical Processes of Mixing Random Variables on [0,1]. Ann. Statist. 3 1101-1108.

Withers, C.S. (1976). On the Convergence of Empirical Processes of Mixing Variables. Australian J. of Statistics 18 76-83.

Yokoyama, R. (1973). Weak Convergence of Empirical Processes for Strong Mixing Sequences of Random Variables. Sci. Rep. Tokyo Kyoiku Daigaku Sec. A 12 36-39.

Yoshihara, K. (1974). Extensions of Billingsley's Theorems on Weak Convergence of Empirical Processes. Z. Wahr. 29 87-92.

Yoshihara, K. (1975/76). Weak Convergence of Multidimensional Empirical Processes for Strong Mixing Sequences of Stochastic Vectors. Z. Wahr. 33 133-137.