ANNALES DE L'I. H. P., SECTION B

HEROLD DEHLING
MANFRED DENKER
WALTER PHILIPP

The almost sure invariance principle for the empirical process of U-statistic structure

Annales de l'I. H. P., section B, tome 23, nº 2 (1987), p. 121-134
<http://www.numdam.org/item?id=AIHPB_1987__23_2_121_0>

Ann. Inst. Henri Poincaré,
Vol. 23, n° 2, 1987, p. 121-134. *Probabilités et Statistiques*

The almost sure invariance principle for the empirical process of U-statistic structure

by

Herold DEHLING, Manfred DENKER and Walter PHILIPP

ABSTRACT. — We prove the generalization of Kiefer's almost sure invariance principle for the empirical distribution function to empirical processes of a certain class of dependent random variables having a U-statistic structure with non-degenerate kernel.

Key-words: U-statistics, empirical process, almost sure invariance principle.

RÉSUMÉ. — Nous montrons la généralisation d'un théorème de Kiefer concernant le principe d'invariance p. p. pour la répartition empirique dans le cas de certains processus empiriques qui ont une structure comme les U-statistiques.

1. INTRODUCTION

Empirical processes of U-statistic structure were first considered by Serfling [*7*] in order to study generalized L-statistics. (For further motivations to investigate these processes see also Janssen, Serfling, Veraverbeke [*5*]).

Let $X_1, X_2, \ldots$ be i. i. d. random variables and let $g: \mathbb{R}^m \to \mathbb{R}$ be a measurable kernel. As a generalization of the classical empirical distribution function (for the purpose of its probabilistic investigation or its statistical application) one considers the empirical distribution μ_n defined by the random variables

$$\{ g(X_{i_1}, \ldots, X_{i_m}) : 1 \leq i_1 < \ldots < i_m \leq n \}$$

i. e. for $t \in \mathbb{R}$ put

$$\mu_n(t) = \binom{n}{m}^{-1} \operatorname{card} \{ 1 \leq i_1, \ldots, i_m \leq n \mid g(X_{i_1}, \ldots, X_{i_m}) \leq t \}.$$

If $\mu(t) = E(\mu_n(t))$, then $\{ \sqrt{n}(\mu_n(t) - \mu(t)), n \geq 1 \}$ is said to be an empirical process of U-statistic structure.

Serfling [7] proved central limit theorems for functionals of the empirical distribution function (e. d. f.) of U-statistic structure. Silverman [9] extended these results by proving the weak convergence of $n^{1/2}(\mu_n - \mu)$ to some continuous Gaussian process in weighted sup-norm metrics in $D(-\infty, +\infty)$. Finally, we should mention that Glivenko-Cantelli results appear in Helmers, Janssen, Serfling [4].

Though the contents of this note is a little more general, our main result is the almost sure approximation of $n(\mu_n - E\mu_n)$ by a suitable Gaussian process with a remainder term of order $O(n^{\frac{1}{2}-\lambda})$, $\lambda > 0$. This result extends an earlier one by Csörgö, Horvath, Serfling [2] and Kiefer's classical result for the e. d. f. of i. i. d. random variables.

We would like to thank S. Csörgö, P. Janssen and R. Serfling for making their papers [2] [4] and [5] available to us.

2. NOTATION AND STATEMENT OF RESULTS

A Kiefer-process is a two-parameter mean-zero Gaussian process $K(s, t)$, $0 \leq s \leq 1$, $t \geq 0$ with the following covariance structure:

$$EK(s, t)K(s', t') = s(1 - s') \cdot (t \wedge t') \qquad \text{if} \quad s \leq s'.$$

Given a Kiefer-process $K(s, t)$, we define the Gaussian process $\gamma_n(x_1, \ldots, x_m)$ by

$$\gamma_n(x_1, \ldots, x_m) = \sum_{i=1}^{m} K(x_i, n) \prod_{j \neq i} x_j. \tag{2.1}$$

Our main result is:

THEOREM 1. — After possible redefinition on a larger probability space there exists a Kiefer process $K(s, t)$ such that with γ_n defined in (2.1) we have:

$$P\left(\sup_{-\infty \leqq t \leqq +\infty} \left| n(\mu_n(t) - \mu(t)) - \int_0^1 \dots \int_0^1 1\{ g(x_1, \dots, x_m) \leqq t \} \gamma_n(dx_1, \dots, dx_m) \right| \geqq Hn^{\frac{1}{2} - \lambda} \right) \leqq n^{-1/28}.$$

Moreover,

$$(2.2) \quad \sup_{-\infty \leq t \leq +\infty} \left| n(\mu_n(t) - \mu(t)) - \int_0^1 \dots \int_0^1 1\{ g(x_1, \dots, x_m) \leq t \} \gamma_n(dx_1, \dots, dx_m) \right| = O(n^{\frac{1}{2} - \lambda}) \quad \text{a. s.}$$

where $\lambda = \dfrac{1}{2\,400}$.

Remark. — In the classical case of a U-statistic

$$\binom{n}{m}^{-1} \sum_{1 \leqq i_1 < \dots < i_m \leqq n} g(X_{i_1}, \dots, X_{i_m})$$

our result gives the well known weak invariance principle for square integrable, non-degenerate kernels g. This can be shown by standard methods. First of all (2.2) shows the existence of a Gaussian process $\{ G(s, t) : 0 \leqq s \leqq 1, t \in \mathbb{R} \}$ such that $\sqrt{n}(\mu_{[ns]}(t) - \mu(t))$ converges weakly to $G(s, t)$. Then

$$\sqrt{n}\left(\binom{[ns]}{m}^{-1} \sum_{1 \leqq i_1 < \dots < i_m \leqq [ns]} (g(X_{i_1}, \dots, X_{i_m}) - Eg(X_1, \dots, X_m)) \right) =$$

$$= \int t\sqrt{n}(\mu_{[ns]} - \mu)(dt) \to \int tG(s, dt)$$

weakly. This last statement is obviously true replacing the integrand t by a bounded function and a well known L^2-approximation.

In the same way Theorem 1 can be applied to other functionals of $\mu_n - \mu$, which correspond to classical statistical procedures (e. g. quantiles). This has been discussed extensively in Serfling [7] and need not be repeated here, but we shall give one application for the law of the iterated logarithm at the end of this section. First of all we obtain a Functional Law of the Iterated Logarithm as a corollary of Theorem 1.

COROLLARY 2. — $\dfrac{\sqrt{n}}{\sqrt{2\log\log n}}(\mu_n(t) - \mu(t))$ is almost surely relatively compact in D[0, 1] and has the following set of limit functions:

$$\left\{ \sum_1^m \int \cdots \int 1\{ g(x_1, \ldots, x_m) \leqq t \} f(x_i) \prod_{j=1}^m dx_j : \int_0^1 f(x)dx = 0, \int_0^1 f^2(x)dx \leqq 1 \right\}.$$

Theorem 1 will be a corollary of an almost sure bound for D(ℝ)-valued degenerate U-statistics, which we shall formulate below. By D = D(ℝ) we denote, as usual, the space of all functions $f : \mathbb{R} \to \mathbb{R}$ which are right-continuous and have left limits. A measurable function $h : [0, 1]^m \to D$ is called a D-valued kernel of dimension m. h is said to be degenerate, if

$$\int_0^1 h(x_1, \ldots, x_m)dx_i = 0 \qquad \text{for all } x_j, \quad 1 \leq j \leq m, \quad j \neq i \tag{2.3}$$

and all $i = 1, \ldots, m$.

If $J \subset \{ 1, \ldots, m \}$ is an index set of cardinality l, we can define a kernel h_J of dimension l by

$$h_J(x_{i_1}, \ldots, x_{i_l}) = \int_0^1 \cdots \int_0^1 h(x_1, \ldots, x_m) \prod_{i \notin J} dx_i, \quad J = \{ i_1 < \ldots < i_l \}. \tag{2.4}$$

The following facts are well-known for ℝ-valued kernels, but carry over immediately to the D-valued case. Define h^* by

$$h^* = \sum_{k=1}^m (-1)^{m-k} \sum_{\substack{J \subset \{1, \ldots, m\} \\ |J| = k}} h_J. \tag{2.5}$$

Then h^* is a degenerate kernel and the (non-degenerate) kernel h can be written as a sum of degenerate kernels in the following way:

$$h - \int_0^1 \cdots \int_0^1 h(x_1, \ldots, x_m) \prod_{i=1}^m dx_i = \sum_{\substack{J \subset \{1, \ldots, m\} \\ |J| \geq 1}} h_J^* \circ P_J \tag{2.6}$$

where $P_J : \mathbb{R}^m \to \mathbb{R}^{|J|}$ denotes the obvious projection.

DEFINITION 3. — The U-statistic with kernel h is defined by

$$U_n(h) = \sum_{1 \le i_1 < \ldots < i_m \le n} h(X_{i_1}, \ldots, X_{i_m}), \qquad m \ge n.$$

THEOREM 4. — Let $h : [0, 1]^m \to D(\mathbb{R})$ be a kernel satisfying

(2.7) $h(x_1, \ldots, x_m)$ is a non-decreasing element of D for all $(x_1, \ldots, x_m) \in [0, 1]^m$

and

$$(2.8) \qquad 0 \le h_t(x_1, \ldots, x_m) \le 1.$$

Then, for the corresponding degenerate kernel h^* in (2.5), we have

$$(2.9) \qquad \limsup_{n \to \infty} \| n^{-m+1} (\log n)^{-1} U_n(h^*) \| \le K < \infty \quad \text{a. s.}$$

Remark. — $\|\cdot\|$ denotes the sup-norm in D and by $h_t(x_1, \ldots, x_m)$ we mean the value of $h(x_1, \ldots, x_m)$ at $t \in \mathbb{R}$.

For non-degenerate statistics we have the following corollary:

COROLLARY 5. — Let h: $[0, 1]^m \to$ D satisfy (2.7) and (2.8). Then there exists (after a possible enlargening of the basic probability space) a sequence $\{ Y_j(t), t \in \mathbb{R} \}_{j=1,2,\ldots}$ of independent, identically distributed Gaussian processes on $\mathbb{R}$ such that $\left(\lambda = \frac{1}{2\,400}\right)$

$$(2.10) \qquad \sup_{-\infty \leqq t \leqq +\infty} \Bigg| \sum_{1 \le i_1 < \ldots < i_m \le n} h_t(X_{i_1}, \ldots, X_{i_m}) - \binom{n}{m} \int_0^1 \ldots \int_0^1 h_t(x_1, \ldots, x_m) \prod_{i=1}^{m} dx_i - n^{-1} \binom{n}{m} \sum_{j=1}^{n} Y_j(t) \Bigg| = O(n^{m - \frac{1}{2} - \lambda}) \quad \text{a. s.}$$

As an application of our results, we obtain now the Law of the Iterated Logarithm for generalized L-statistics. Let

$$W_{n,1} \leqq W_{n,2} \leqq \ldots \leqq W_{n,n_{(m)}}$$

be the ordered values of $g(X_{i_1}, \ldots, X_{i_m})$ taken over all m-tuples $(i_1, \ldots, i_m)$ of distinct elements of $\{ 1, \ldots, n \}$. Let $c_{n,i}$, $1 \leqq i \leqq n_{(m)}$, be arbitrary constants. Statistics given by $\sum_{i \leqq n(m)} c_{n,i} W_{n,i}$ were termed GL-statistics by Serfling [7]. They generalize linear functions of order statistics (L-statistics) as well as U-statistics.

As noted in Serfling [7], a sufficiently large subclass of GL-statistics is given by

$$(2.11) \qquad T(\mu_n) = \int_0^1 J(t)\mu_n^{-1}(t)dt + \sum_{j=1}^{d} a_j \mu_n^{-1}(p_j)$$

where $\mu_n^{-1}(t) = \inf\{ x : \mu_n(x) \geqq t \}$ is the quantile function. Under certain restrictions, Serfling proves asymptotic normality of $n^{1/1}(T(\mu_n) - T(\mu))$. This is achieved through a linearization of $T(\mu_n) - T(\mu)$, given by

$$(2.12) \quad d_1(\mu, \mu_n - \mu) = -\int_{-\infty}^{\infty} (\mu_n(y) - \mu(y))J(\mu(y))dy + \sum_{j=1}^{d} a_j \frac{p_i - \mu_n(\mu^{-1}(p_j))}{m(\mu^{-1}(p_j))}$$

where $m = \mu'$. $d_1(\mu, \mu_n - \mu)$ is a U-statistic with kernel

$$(2.13) \quad A(x_1, \ldots, x_m) = -\int_{-\infty}^{\infty} (1\{g(x_1, \ldots, x_m) \leqq y\} - \mu(y))J(\mu(y))dy .$$

The asymptotic variance of $n^{1/2}(T(\mu_n) - T(\mu))$ is then given by

$$\sigma^2(T, \mu) = \operatorname{Var}\left(\sum_{i=1}^{m} E(A(X_1, \ldots, X_m) \mid X_i)\right).$$

THEOREM 6. — Let μ be twice differentiable and have positive derivatives at its p_j-quantiles. Let $J(t)$ vanish for t outside $[\alpha, \beta]$, where $0 < \alpha < \beta < 1$, and suppose that on $[\alpha, \beta]$ J is bounded and continuous Lebesgue- and μ^{-1}-a. e. Assume that $EA^2(X_1, \ldots, X_m) < \infty$ and $0 < \sigma^2(T, \mu) < \infty$. Then

$$\limsup_{n \to \infty} \frac{\sqrt{n}}{\sqrt{2 \log \log n}} (T(\mu_n) - T(\mu)) = \sigma(T, \mu) .$$

3. A MOMENT INEQUALITY FOR DEGENERATE U-STATISTICS

For a kernel $h : [0, 1]^m \to \mathbb{R}$ let $\| h \|_p$ denote the usual L^p-norm with respect to Lebesgue-measure λ^m.

LEMMA 7. — Let $h : [0,1]^m \to \mathbb{R}$ be degenerate with $\|h\|_{2p} < \infty$ for some integer p. Then

(3.1) $$\mathrm{E}(\mathrm{U}_n(h))^{2p} \leq$$
$$\leq n^{pm}\left\{ 2p^{pm}\left(\frac{2m}{e}\right)^{pm} \|h\|_{2m-1}^{2p} + n^{-1}p^{pm+1}\left(\frac{40m}{e}\right)^{pm+1} \|h\|_{2p}^{2p} \right\}.$$

Remark. — This lemma gives a precise estimate for the « o » term in Lemma B of Serfling ([*6*], p. 186) indicating the dependence on p and m also. Estimates of this type are known to be useful when no exponential inequality is available (but can not replace the latter completely, e. g. for moderate deviation results). However, for our purpose in the proof of Theorem 4 this lemma will be sufficient.

Proof.

$$\mathrm{E}(\mathrm{U}_n(h))^{2p}$$
$$= \sum_{\substack{1 \leq i_1,\dots,i_{2pm} \leq n \\ i_{km+1} < \dots < i_{(k+1)m} \\ k=0,\dots,2p-1}} \mathrm{E}\{ h(\mathrm{X}_{i_1}, \dots, \mathrm{X}_{i_m}) \cdot \dots \cdot h(\mathrm{X}_{i_{(2p-1)m+1}}, \dots, \mathrm{X}_{i_{2pm}}) \}$$
$$= \sum_{\substack{p_1,\dots,p_n \geq 0 \\ \cdot \sum_{i \leq n} p_i = 2pm}} \sum_{\substack{1 \leq i_1,\dots,i_{2pm} \leq n \\ i_{km+1} < \dots < i_{(k+1)m},\, k=0,\dots,2p-1 \\ \mathrm{card}\{j : i_j = i\} = p_i}} \mathrm{E}\{ h(\mathrm{X}_{i_1}, \dots, \mathrm{X}_{i_m}) \cdot \dots \cdot h(\mathrm{X}_{i_{(2p-1)m+1}}, \dots, \mathrm{X}_{i_{2pm}}) \}.$$

For each choice of $(p_1, \dots, p_n)$ the inner sum contains $\dfrac{(2pm)!}{p_1! \dots p_n!}$ elements. We divide the outer sum into three parts Σ^1, Σ^2 and Σ^3 where in Σ^1 we sum over those m-tuples $(p_1, \dots, p_n)$ with one $p_i = 1$ in Σ^2 over those with all $p_i \in \{0, 2\}$ and in Σ^3 over the remaining terms. Because of degeneracy, each integral in Σ^1 vanishes. In Σ^2, the outer sum has $\binom{n}{pm}$ summands, the inner sum $\dfrac{(2pm)!}{p_1! \dots p_n!}$. To estimate the expectations in Σ^2, we note that in the product

$$h(\mathrm{X}_{i_1}, \dots, \mathrm{X}_{i_m}) \cdot \dots \cdot h(\mathrm{X}_{i_{(2p-1)m+1}}, \dots, \mathrm{X}_{i_{2pm}})$$

each index occurs exactly twice. Hence we can subdivide this product into $2m - 1$ subproducts, each of them containing only independent random variables. Applying Hölder's inequality first and then independence, we obtain

$$E\{h(X_{i_1},\ldots,X_{i_m})\cdot\ldots\cdot h(X_{i_{(2p-1)m+1}},\ldots,X_{i_{2pm}})\}\le\|h\|_{2m-1}^{2p}.$$

Hence Stirling's formula yields:

$$\Sigma^2\le\binom{n}{pm}\frac{(2pm)!}{2^{pm}}\|h\|_{2m-1}^{2p}\le n^{pm}\frac{(2pm)!}{(pm)!\,2^{pm}}\|h\|_{2m-1}^{2p}$$
$$\le 2n^{pm}\left(\frac{2pm}{e}\right)^{pm}\|h\|_{2m-1}^{2p}.$$

In order to estimate Σ^3, note that only $pm-1$ of the p_i's in $(p_1,\ldots,p_n)$ can be non-zero. Counting the number of n-tuples with this restriction we find that there are $\binom{n}{pm-1}$ ways to choose the non-zero p_i's and at most 10^{pm-1} ways of solving the equation $\sum_{i=1}^{pm-1}q_i=2pm$ in the non-negative integers. Hence applying Hölder's inequality we find:

$$\Sigma^3\le\binom{n}{pm-1}10^{pm-1}(2pm)!\,\|h\|_{2p}^{2p}$$
$$\le n^{pm-1}\frac{(2pm)!\,pm}{(pm)!}10^{pm-1}\|h\|_{2p}^{2p}$$
$$\le n^{pm-1}p^{pm+1}\left(\frac{40m}{e}\right)^{pm+1}\|h\|_{2p}^{2p}.$$

4. PROOF OF THEOREM 4

To simplify the notation, we restrict ourselves to the case $m=3$. (This is the simplest general case for m. $m=1, 2$ are easier.) We define partitions $\{-\infty=t_0(n)<t_1(n)<\ldots<t_n(n)=+\infty\}$ such that

$$(4.1)\qquad E\,|\,h_{t_{i+1}}(X_1,X_2,X_3)-h_{t_i}(X_1,X_2,X_3)\,|\le\frac{1}{n}.$$

At this point we have to assume that the map $t\to Eh_t(X_1,X_2,X_3)$ is continuous. If this is not the case, a minor modification of the proof is necessary. Put:

$$v_n(t) = U_n(h_t^*), \qquad v_n^J(t) = U_n((h_J)_t^*), \qquad J \subset \{1, 2, 3\}$$
$$\mu_n(t) = U_n(h_t), \qquad \mu_n^J(t) = U_n((h_J)_t)$$
$$\mu(t) = E(h_t(X_1, X_2, X_3)).$$

Then we have:

$$\sup_{-\infty \leqq t \leqq +\infty} |v_n(t)| \leqq \sup_{i=1,\dots,n} |v_n(t_i(n))| + \sup_{-\infty \leqq t \leqq +\infty} |v_n(t) - v_n(t_{i(t,n)}(n))| \tag{4.2}$$

where

$$i(t, n) := \max\{ i : t_i(n) \leqq t \}. \tag{4.3}$$

We first estimate the first term on the r. h. s. of (4.2). Using Lemma 7 with $p = 3$ we find:

$$\begin{aligned} \text{(4.4)} \quad & P\{ \max_{i=1,\dots,n} |v_n(t_i(n))| > n^2 \log n \} \\ & \leqq n \cdot \sup_{-\infty \leqq t \leqq +\infty} P\{ |n^{-\frac{3}{2}} v_n(t)| > n^{\frac{1}{2}} \log n \} \\ & \ll n \cdot n^{-p} (\log n)^{-2p} \left\{ p^{3p} \left(\frac{6}{e}\right)^{3p} + n^{-1} p^{3p+1} \left(\frac{120}{e}\right)^{3p+1} \right\} \\ & \ll n^{-2}. \end{aligned}$$

While estimating the second term on the r. h. s. of (4.2) we put $t_i = t_{i(t,n)}(n)$ if no confusion arises from this. Using (2.5), (2.6) and the monotonicity of h_t we find:

$$\begin{aligned} \text{(4.5)} \quad v_n(t) - v_n(t_i) & \leqq \sum_{J \subset \{1,2,3\}} n^{3-|J|} (\mu_n^J(t_{i+1}) - \mu_n^J(t_i)) \\ & \leqq \sum_{J \subset \{1,2,3\}} \sum_{\substack{K \subset J \\ |K| \geqq 1}} n^{3-|K|} (v_n^K(t_{i+1}) - v_n^K(t_i)) + 8n^3(\mu(t_{i+1}) - \mu(t_i)). \end{aligned}$$

By (4.1), the last term is bounded by $8n^2$. For $K = \{1, 2, 3\}$ we can apply (4.4), since

$$\max_{i=1,\dots,n} |v_n^K(t_i) - v_n^K(t_{i-1})| \leqq 2 \max_{i=1,\dots,n} |v_n(t_i(n))|. \tag{4.6}$$

For $K = \{1, 2\}$ we find using Lemma 7 with $p = \log n$:

$$(4.7) \quad P\{ \max_{i=1,\dots,n} n \cdot | v_n^{(1,2)}(t_i) - v_n^{(1,2)}(t_{i-1}) | \geqq 2cn^2 \log n \}$$
$$\leqq n \cdot \sup_t P\{ n^{-1} | v_n^{(1,2)}(t) | \geqq c \log n \}$$
$$\ll n \cdot (\log n)^{-2p} c^{-2p} \left\{ p^{2p} \left(\frac{4}{e}\right)^{2p} + n^{-1} p^{2p+1} \left(\frac{80}{e}\right)^{2p+1} \right\}$$
$$\ll n n^{-2 \log c} n^{2 \log (4/e)} + n^{-2 \log c} n^{2 \log (80/e)}$$
$$\ll n^{-2} \qquad \text{if } c \text{ is large enough}.$$

For $K = \{1\}$ we find using Bernstein's inequality (see Bennett [1] and (4.1))

$$(4.8) \quad P\{ \max_{i=1,\dots,n} n^2 | v_n^{(1)}(t_i) - v_n^{(1)}(t_{i-1}) | \geqq cn^2 \log n \}$$
$$\leqq n \max_{i=1,\dots,n} P\{ | v_n^{(1)}(t_i) - v_n^{(1)}(t_{i-1}) | \geqq c \log n \}$$
$$\leqq n \exp\left\{ - \frac{c^2 \log^2 n}{1 + 2/3 \log n} \right\}$$
$$\ll n^{-2} \qquad \text{if } c \text{ is large enough}.$$

(4.2), (4.4)-(4.8) together with an application of the Borel-Cantelli Lemma give the desired result.

5. PROOF OF THEOREM 1

We shall prove Theorem 1 using Theorem 4 and the following almost sure invariance principle for independent identically formed $D(\mathbb{R})$-valued random elements:

PROPOSITION 8. — Let $\{ X_i, i \geq 1 \}$ be a sequence of independent identically distributed real-valued random variables and let $h : \mathbb{R} \to D(\mathbb{R})$ be a map satisfying

(5.1) *i)* $h(x)$ is a non-decreasing element of D for all $x \in \mathbb{R}$
ii) $0 \leq h_t(x) \leq 1$.

Then after a possible redefinition on a larger probability space there exists a sequence of independent, identically distributed Gaussian processes $\{ X(s, n) : s \in \mathbb{R} \}$, $n \geq 1$, such that

$$P\left(\max_{k \leq n} \sup_{-\infty \leqq s \leqq +\infty} \left| \sum_{i \leqq k} (h_s(X_i) - Eh_s(X_i) - X(s, i)) \right| \geqq Hn^{\frac{1}{2}-\lambda} \right) \leqq n^{-1/28}$$

for some constant $H > 0$ and $\lambda = 1/2\,400$.

In particular, with probability one

$$\sup_{-\infty \leqq s \leqq +\infty} \left| \sum_{i \leqq n} (h_s(X_i) - Eh_s(X_i) - X(s, i)) \right| = O(n^{\frac{1}{2}-\lambda}). \tag{5.2}$$

Proof. — $(D(\mathbb{R}), \|\cdot\|)$ is a (non-separable) Banach space and we can apply Theorem 6.1 of Dudley and Philipp [3]. Thus, it suffices to show that for each $m = 1, 2, \ldots$ one can find a partition

$$-\infty = t_0 < t_1 < \ldots < t_{m^4} = \infty$$

such that

$$P\left\{ \sup_{k=1,\ldots,m^4} \sup_{t \in [t_{k-1}, t_k]} n^{-\frac{1}{2}} \left| \sum_{i=1}^{n} (h_t(X_i) - Eh_t(X_i) - h_{t_k}(X_i) + Eh_{t_k}(X_i) \right| \geqq \frac{1}{m} \right\} \leqq 1/m \tag{5.3}$$

for $n \geqq m^4$.

The proof of this carries over almost verbatim from Dudley and Philipp, Theorem 7.4 and we shall omit it here.

We now introduce the map $h : [0, 1]^m \to D(\mathbb{R})$ by

$$h_t(x_1, \ldots, x_m) = 1\{ g(x_1, \ldots, x_m) \leqq t \}.$$

Here h_t is defined in Remark after Theorem 4. We shall apply Proposition 8 to the map $f_t(x) = \sum_{i=1}^{m} (h_{\{i\}})_t(x)$ as defined in (2.4). By (5.2) we can approximate $\sum_{j \leqq n} \{ f_t(X_j) - Ef_t(X_j) \}$ by $\sum_{j \leqq n} Y_j(t)$, where $Y_j(\cdot)$, $j = 1, 2, \ldots$ are i. i. d. Gaussian processes. We want to identify this limit process as a certain stochastic integral. Let $\{ B_k(t), 0 \leqq t \leqq 1 \}_{k \geqq 1}$ be a sequence of i. i. d. Brownian bridges on $[0, 1]$ and define

$$\mu_k^*(x_1, \ldots, x_m) = \sum_{i=1}^{m} B_k(x_i) \prod_{j \neq i} x_j \qquad 0 \leqq x_i \leqq 1. \tag{5.4}$$

It is easy to verify that

$$(5.5)\qquad \int_0^1 \cdots \int_0^1 h_t(x_1, \ldots, x_m) d\mu_k^*(x_1, \ldots, x_m) = \int_0^1 f_t(x) dB_k(x).$$

The covariance structure of the process $\left\{ \int_0^1 f_t(x) dB_k(x), t \in \mathbb{R} \right\}$ can be computed as follows:

$$E\left\{ \int_0^1 f_s(x) dB_1(x) \int_0^1 f_t(x) dB_1(x) \right\} = \int_0^1 (f_s(x) - Ef_s(X_1))(f_t(x) - Ef_t(X_1)) dx.$$

But this is just the covariance of $\{ f_t(X_1), t \in \mathbb{R} \}$ and hence of $\{ Y_1(t), t \in \mathbb{R} \}$. Hence the two Gaussian processes $\left\{ \int_0^1 f_t(x) dB_1(x), t \in \mathbb{R} \right\}$ and $\{ Y_1(t), t \in \mathbb{R} \}$ have the same distribution and using (5.4) and (5.5) we find

$$(5.6)\qquad \sum_{j \leq n} Y_j(t) = \int_0^1 \cdots \int_0^1 h_t(x_1, \ldots, x_m) d\gamma_n(x_1, \ldots, x_m).$$

Using Lemma 2.11 of Dudley and Philipp we may actually assume that both sides of (5.6) are identical a. e.

To finish the proof of Theorem 1 we apply Theorem 4 to the terms with $|J| \leq m - 2$ on the right-hand side of (2.6). This and the above remarks together with (5.2) yield Theorem 1.

6. PROOF OF THEOREM 6

As is shown in Serfling [7] $T(\mu_n) - T(\mu)$ can be decomposed into a linear part and a small remainder in the following way:

$$(6.1)\qquad T(\mu_n) - T(\mu) = d_1(\mu, \mu_n - \mu) + \Delta_{1,n} + \Delta_{2,n}$$

where

$$(6.2)\qquad \Delta_{1,n} = - \int_{-\infty}^{\infty} W_{\mu_n,\mu}(y)(\mu_n(y) - \mu(y)) dy$$

$$(6.3)\qquad \Delta_{2,n} = \sum_{j=1}^{d} a_j \left(\mu_n^{-1}(p_j) - \mu^{-1}(p_j) - \frac{p_j - \mu_n(p_j)}{m(\mu^{-1}(p_j))} \right)$$

and

$$W_{F,G}(y) = \frac{K(F(y)) - K(G(y))}{F(y) - G(y)} - J(F(y)) \tag{6.4}$$

$$K(x) = \int_0^x J(u)du .$$

By the law of the iterated logarithm for non-degenerate U-statistics it suffices to show

$$\frac{\sqrt{n}}{\sqrt{\log \log n}} \Delta_{1,n} = 0(1) \quad \text{a. s.} \quad \text{and} \tag{6.5}$$

$$\frac{\sqrt{n}}{\sqrt{\log \log n}} \Delta_{2,n} = 0(1) \quad \text{a. s.} \tag{6.6}$$

(6.5) can be proved as in Serfling [7], replacing his Lemma 3.3 by our Corollary 2, which shows that $\| \mu_n - \mu \|_\infty = O(n^{-1/2} (\log \log n)^{1/2})$. The individual summands in (6.3) are the remainder terms in a generalized Bahadur representation for the quantiles of the e. d. f. of U-statistics structure. Then (6.6) is a consequence of the following proposition.

PROPOSITION 9. — Let $0 < p < 1$ and suppose that μ is twice differentiable at $\mu^{-1}(p)$ with $\mu'(\mu^{-1}(p)) > 0$. Then

$$\mu_n^{-1}(p) = \mu^{-1}(p) + \frac{p - \mu_n(\mu^{-1}(p))}{\mu'(\mu^{-1}(p))} + R_n$$

where

$$R_n = O(n^{-3/4} (\log n)^{3/4}) \quad \text{a. s.} \tag{6.7}$$

The proof of this Proposition is analogous to the proof of the classical Bahadur representation, as given in Serfling [6], e. g. The only difference is that Hoeffding's and Bernstein's inequalities have to be replaced by their counterparts for U-statistics (Serfling [6], p. 201).

REFERENCES

[1] G. BENNETT, Probability inequalities for the sum of independent random variables. *JASA*, t. **57**, 1962, p. 33-45.

[2] S. CSÖRGÖ, L. HORVATH, R. SERFLING, *An approximation for the empirical process* of U-statistic structure. Technical report, Johns Hopkins University, 1983.

[3] R. M. DUDLEY, W. PHILIPP, Invariance principles for sums of Banach space valued random elements and empirical processes. *Zeitschrift Wahrscheinlichkeitstheorie verw. Gebiete*, t. **62**, 1983, p. 509-552.
[4] R. HELMERS, P. JANSSEN, R. SERFLING, *Asymptotic theory for the e. d.f. of U-statistics and applications to GL-statistics and the bootstrap*. Preprint, 1984-1985.
[5] P. JANSSEN, R. SERFLING, N. VERAVERBEKE, Asymptotic normality for a general class of statistical functions and applications to measures of spread. *Ann. Stat.*, t. **12**, 1984, p. 1369-1379.
[6] R. SERFLING, *Approximation Theorems of Mathematical Statistics*. John Wiley and Sons, New York, 1980.
[7] R. SERFLING, Generalized L-, M- and R-statistics. *Ann. Stat.*, t. **12**, 1984, p. 76-86.
[8] B. W. SILVERMAN, Limit theorems for dissociated random variables. *Adv. Appl. Prob.*, t. **8**, 1976, p. 806-819.
[9] B. W. SILVERMAN, Convergence of a class of empirical distribution functions of dependent random variables. *Ann. Prob.*, t. *11*, 1983, p. 745-751.

(Manuscrit reçu le 25 avril 1985)

(Révisé le 13 août 1986)