The Annals of Probability
1993, Vol. 21, No. 2, 1170–1184

WEAK CONVERGENCE AND GLIVENKO–CANTELLI RESULTS FOR WEIGHTED EMPIRICAL U-PROCESSES

BY WILHELM SCHNEEMEIER

University of Munich

Empirical processes of U-statistic structure were introduced by Serfling and studied in detail by Silverman, who proved weak convergence of weighted versions in the i.i.d. case. Our main theorem shows that this result can be generalized in two directions: First, the i.i.d. assumption can be omitted, and second, our proof holds for a richer class of weight functions. In addition, we obtain almost sure convergence of weighted U-processes in the i.i.d. case which improves the results of Helmers, Janssen and Serfling, Aerts, Janssen and Mason and (in the special situation of the real line) Nolan and Pollard.

1. Introduction. Following Silverman (1983), let $(\xi_i)_{i \in \mathbb{N}}$ be a sequence of i.i.d. random variables on some probability space $(\Omega, \mathscr{F}, \mathbb{P})$, and let $m \in \mathbb{N}$ be fixed. Consider some Borel-measurable kernel function $h\colon \mathbb{R}^m \to \mathbb{R}$ and define the so-called U-process U_n for all $n \geq m$ in the following way:

$$E_n := \left\{(i_1, \dots, i_m) \in \{1, \dots, n\}^m : i_j \neq i_k \text{ for } j \neq k\right\},$$

$$H_n(t) := \operatorname{card}(E_n)^{-1} \sum_{(i_1, \dots, i_m) \in E_n} 1\left(h(\xi_{i_1}, \dots, \xi_{i_m}) \leq t\right),$$

$$\overline{H}(t) := \mathbb{E}(H_n(t)),$$

$$U_n(t) := n^{1/2}\left[H_n(t) - \overline{H}(t)\right] \quad \text{for all } t \in \mathbb{R}.$$

The main result of Silverman (1983) is the weak convergence of $U_n \cdot q^{-1}(\overline{H})$ w.r.t. the supremum norm, where $q\colon [0,1] \to \mathbb{R}_+$ fulfills the following conditions [with $v(t) := t^{1/2}q^{-1}(t)$]:

(Q)(i) q and v are increasing on $[0, 1/2]$, q is symmetric about $1/2$ and continuous.
(ii) $\int_0^{1/2}[\log 1/x]^{1/2} v(dx) < \infty$ and $v(t) \downarrow 0$ as $t \downarrow 0$.

Received August 1987; revised September 1991.
AMS 1991 *subject classification*. Primary 60F17.
Key words and phrases. U-process, weight function, $\mathscr{L}_b$-convergence, empirical process, pseudometric.

Condition (Q)(ii) is slightly stronger than square integrability of q^{-1} because partial integration at once implies:

$$\text{(i)}\quad \int_0^{1/2}\left[\log\frac{1}{x}\right]^{1/2} v(dx) \geq \int_0^{\varepsilon} q^{-2}(x)\,dx \quad \text{for } \varepsilon > 0 \text{ small enough.}$$

$$\text{(ii)}\quad \int_0^{\varepsilon}\left[\log\frac{1}{x}\right]^{1/2} v(dx) = \infty \quad\text{and}\quad \int_0^{\varepsilon} q^{-2}(x)\,dx < \infty$$

$$\text{for } q(x) := \left(x\log\frac{1}{x}\right)^{1/2}\left(\log\log\frac{1}{x}\right)^{3/4} \quad \text{and} \quad \varepsilon > 0 \text{ small enough.}$$

In Section 3 we will show that Silverman's result continues to hold for all weight functions q having the following property (Q_{k_0}) for some $k_0 \in \mathbb{N}$:

(Q_{k_0})(i) q is increasing on $[0, 1/2]$, symmetric about $1/2$ and continuous.
(ii) $\int_0^{1/2} q^{-2k_0}(x)x^{k_0-1}\,dx < \infty$.

Besides the fact that monotonicity of v can be omitted, a simple calculation shows that (Q_k) is strictly stronger than (Q_{k+1}) for all $k \in \mathbb{N}$ [note that (Q_1) means square integrability of q^{-1}]. This implies that the class of all q fulfilling (Q_k) for some $k \in \mathbb{N}$ is larger than that of all q fulfilling (Q).

Moreover, we prove not only weak convergence for $U_n \cdot q^{-1}(\overline{H})$ but also for a truncated version of $U_n \cdot q^{-1}(V_n)$, where the functions $V_n\colon \mathbb{R} \to [0,1]$ are close to the variance structure of U_n [for this result the condition (Q_k) has to be strengthened, which might be compensated by a suitable choice of V_n].

This type of argument function is quite new and gives better results in all cases where the variance of $U_n(t)$ is $o(\overline{H}(t))$; see Remark 3.9.2.

Nolan and Pollard (1988) proved a different generalization of the Silverman result by studying function-indexed U-processes, but restricted to the case of the real line their class of weight functions is smaller than ours.

All our results on weak convergence are stated (for triangular arrays in the non-i.i.d. case) in Section 3, while in Section 4 almost sure convergence for weighted U-processes in the i.i.d. case (in order to apply martingale arguments) is investigated. Proofs of the results are given in Section 5.

2. Basic definitions and characterization of weak convergence. Let $(\xi_{ni})_{i \leq i(n),\ n \in \mathbb{N}}$ be a triangular array of rowwise independent random elements taking values in some measurable space $(Z, \mathscr{D})$, which are all defined on the same probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Throughout this paper let $m \in \mathbb{N}$ be fixed and w.l.o.g. let $i(n) \geq m$ for all $n \in \mathbb{N}$. Consider some (not necessarily symmetric) measurable kernel function $h\colon Z^m \to \mathbb{R}$ and define for all $n \in \mathbb{N}$

and $t \in \mathbb{R}$ (with $E_{i(n)}$ as in Section 1):

$$H_n(t) := \operatorname{card}(E_{i(n)})^{-1} \sum_{(i_1,\dots,i_m)\in E_{i(n)}} 1\big(h(\xi_{ni_1},\dots,\xi_{ni_m}) \le t\big),$$

$$\overline{H}_n(t) := \mathbb{E}(H_n(t)),$$

$$U_n(t) := i(n)^{1/2}\big[H_n(t) - \overline{H}_n(t)\big].$$

Now we have to state some regularity conditions for the variance structure of U_n. We make the general assumption that there exists some $K \ge 0$ and for all $n \in \mathbb{N}$ an increasing right-continuous function V_n: $\mathbb{R} \to [0,1]$ such that for all $-\infty < r < t < \infty$:

$$\mathbb{E}\big([U_n(t) - U_n(r)]^2\big) \le K\big[V_n(t) - V_n(r) + i(n)^{-1}\big].$$

Note that this condition is especially fulfilled by the choice of $V_n = \overline{H}_n$. Our main result will be the weak convergence of the process

$$W_n(t) := \begin{cases} U_n(t)q^{-1}(V_n(t)), & \text{for } i(n)^{-1} \le V_n(t) < 1 - i(n)^{-1}, \\ 0, & \text{elsewhere,} \end{cases}$$

where q denotes a weight function satisfying, for some integer $k \ge 3/2$, the condition:

($\tilde{Q}_k$)(i) q is increasing on $[0, 1/2]$, symmetric about $1/2$ and continuous.
(ii) $\int_0^{1/2} q^{-2k}(x)x^{k-3/2}\,dx < \infty$.

In the case of $V_n = \overline{H}_n$, we will even be able to prove a nontruncated result if q only fulfills (Q_k) of Section 1. But in many cases of interest (as pointed out in Section 1), the ($\tilde{Q}_k$) proposition together with suitable V_n gives better results.

Now we will give a short abstract of the type of weak convergence used throughout this paper, the so-called $\mathscr{L}_b$-convergence; for further details see Gaenssler (1983).

For some totally bounded pseudometric on $\mathbb{R}$, we define the continuity modulus

$$\omega_f^s(\delta) := \sup\{|f(t) - f(r)|: s(r,t) \le \delta\}, \qquad \delta > 0,\ f \in l^\infty(\mathbb{R}).$$

Then the following characterization is crucial [cf. Gaenssler and Schneemeier (1986)].

THEOREM 2.1. *$(W_n)_{n\in\mathbb{N}}$ is relatively $\mathscr{L}_b$-sequentially compact iff*

$$\text{(A)} \qquad \lim_{\delta\downarrow 0}\limsup_{n\to\infty} \mathbb{P}^*\big(\omega_{W_n}^s(\delta) \ge \varepsilon\big) = 0 \quad \textit{for all } \varepsilon > 0,$$

$$\text{(B)} \qquad \lim_{M\uparrow\infty}\limsup_{n\to\infty} \mathbb{P}\big(|W_n(t)| \ge M\big) = 0 \quad \textit{for all } t \in \mathbb{R}.$$

REMARK 2.2. If (A) and (B) hold, we call $(W_n)_{n\in\mathbb{N}}$ $\mathscr{L}_b$-convergent iff the finite-dimensional distributions of $(W_n)_{n\in\mathbb{N}}$ converge. Thus the main problem is to prove the "tightness" of $(W_n)_{n\in\mathbb{N}}$ by Theorem 2.1.

3. Weak convergence of the weighted U-process. The following inequality is important for all proofs [for a similar inequality in the degenerate i.i.d. case, see Lemma 7 of Dehling, Denker and Philipp (1987)].

PROPOSITION 3.1. *For all $k \in \mathbb{N}$ there exists $C = C(k, m) \geq 0$ with*

$$\mathbb{E}\Big(\big[U_n(t) - U_n(r)\big]^{2k}\Big) \leq C\Big[\mathbb{E}\Big(\big[U_n(t) - U_n(r)\big]^2\Big)^k + i(n)^{-k}\Big]$$

for all $r, t \in \mathbb{R}$, $n \in \mathbb{N}$.

SHORT OUTLINE OF THE PROOF. In case $m = 1$, U_n reduces to the empirical process and so Proposition 3.1 is an immediate consequence of the Bernstein inequality. For general m, we make use of an approximation of U_n by a sum of m empirical processes.

Now, our aim is to establish conditions (A) and (B) of Theorem 2.1 for the process W_n. The following inequalities are crucial.

LEMMA 3.2. *If $(\tilde{Q}_{k_0})$ holds for some $k_0 \geq 3/2$, then there exists $C = C(k_0, m) \geq 0$ such that for all $n \in \mathbb{N}$, $0 < \delta \leq 1/2$ and $\varepsilon > 0$:*

$$\mathbb{P}\big(\sup\{|W_n(t)|: V_n(t) \leq \delta\} \geq \varepsilon\big) \leq C\varepsilon^{-4k_0}\left[\int_0^\delta q^{-2k_0}(x)x^{k_0-3/2}\,dx\right]^2.$$

LEMMA 3.3. *There exists $C \geq 0$ such that for all $\varepsilon > 0$, $0 < \delta \leq 1$ and $i(n) \geq \delta^{-1}$:*

$$\mathbb{P}\big(\sup\{|U_n(t) - U_n(r)|: |V_n(t) - V_n(r)| \leq \delta; r, t \in \mathbb{R}\} \geq \varepsilon\big) \leq C\varepsilon^{-6}\delta.$$

These two inequalities are the key for controlling the modulus of continuity of W_n. It is obvious that we have to impose some additional asymptotic properties of $(V_n)_{n \in \mathbb{N}}$, which are of the following form:

(V) There exists some pseudometric s on $\mathbb{R}$ with:

(i) $\lim_{\delta \downarrow 0} \limsup_{n \to \infty} \sup\{|V_n(t) - V_n(r)|: r, t \in \mathbb{R}; s(r, t) \leq \delta\} = 0$.
(ii) $(\mathbb{R}, s)$ is totally bounded.

Now we are able to prove one of the main results.

THEOREM 3.4. *Assume that $(\tilde{Q}_{k_0})$ holds for some $k_0 \geq 3/2$ and that (V) is fulfilled together with $i(n) \to \infty$ for $n \to \infty$. Then $(W_n)_{n \in \mathbb{N}}$ is relatively $\mathscr{L}_b$-sequentially compact.*

In view of Remark 2.2 this implies the following theorem.

THEOREM 3.5. *Under the assumptions of Theorem 3.4, the following statements are equivalent:*

(i) *$(W_n)_{n \in \mathbb{N}}$ is $\mathscr{L}_b$-convergent to some zero-mean Gaussian process G.*
(ii) *$(\mathbb{E}(W_n(r)W_n(t)))_{n \in \mathbb{N}}$ converges for all $r, t \in \mathbb{R}$.*

Now the question arises which functions $(V_n)_{n\in\mathbb{N}}$ are possible and whether the results of Silverman (1983) follow from Theorem 3.5. Note in this context that $V_n = \overline{H}_n$, $n \in \mathbb{N}$, is suitable; this follows by a simple calculation. Then the following corollaries are consequences of Lemma 3.2 and Theorems 3.4 and 3.5.

COROLLARY 3.6. *If (Q_{k_0}) holds for some $k_0 \in \mathbb{N}$, then there exists $C \geq 0$ with*

$$\mathbb{P}\big(\sup\{|U_n(t)|q^{-1}(\overline{H}_n(t)) \colon \overline{H}_n(t) \leq \delta\} \geq \varepsilon\big)$$
$$\leq C\varepsilon^{-4k_0}\left[\int_0^{\delta} q^{-2k_0}(x)x^{k_0-1}\,dx\right]^2 + \varepsilon^{-1}\left[\int_0^{1/i(n)} q^{-2k_0}(x)x^{k_0-1}\,dx\right]^{1/(2k_0)}$$

for all $\varepsilon > 0$, $0 < \delta \leq 1/2$ and $n \in \mathbb{N}$.

COROLLARY 3.7. *Assume that (Q_{k_0}) holds for some $k_0 \in \mathbb{N}$ and that (V) is satisfied for $V_n = \overline{H}_n$, $n \in \mathbb{N}$. Then, if $i(n) \to \infty$ for $n \to \infty$, $(U_n q^{-1}(\overline{H}_n))_{n\in\mathbb{N}}$ is relatively $\mathscr{L}_b$-sequentially compact.*

COROLLARY 3.8. *Under the assumptions of Corollary 3.7, $(U_n q^{-1}(\overline{H}_n))_{n\in\mathbb{N}}$ is $\mathscr{L}_b$-convergent to some zero-mean Gaussian process iff*

$$\Big(\mathbb{E}\big(\big[U_n(t)q^{-1}(\overline{H}_n(t))\big]\big[U_n(r)q^{-1}(\overline{H}_n(r))\big]\big)\Big)_{n\in\mathbb{N}}$$

is convergent for all $r, t \in \mathbb{R}$.

REMARK 3.9.

1. Suppose that there exists an increasing right-continuous function H with
$$\limsup_{n\to\infty}\,[V_n(t) - V_n(r)] \leq H(t) - H(r) \text{ for all } -\infty < r < t < \infty.$$
Then it is obvious that the pseudometric $s(r,t) := |H(t) - H(r)|$ satisfies (V)(i) and (V)(ii). This is true in particular for the i.i.d. case so that the result of Silverman (1983) is an immediate consequence of Corollary 3.8 (note also that we have given an elementary proof, whereas Silverman used the KMT approximation).
2. Another interesting aspect is that the condition on the weight function in Corollary 3.7 might be weakened by using Theorem 3.4, though the integral condition is stronger. We may compensate for this by a suitable choice of $(V_n)_{n\in\mathbb{N}}$ (but note that Theorem 3.4 gives a "truncated" result). For an example, set $\mathscr{L}(\xi_{ni}) = U(0,1)$ for all $i \leq i(n) = n$, $n \in \mathbb{N}$, $m = 2$ and $h(x,y) := \max(x,y)$. Then an easy calculation shows that there exists $C \geq 0$ with $\mathbb{E}([U_n(t) - U_n(r)]^2) \leq C[t^3 - r^3 + n^{-1}]$ for all $0 \leq r \leq t \leq 1$, $n \in \mathbb{N}$ [note also that $\overline{H}_n(t) = t^2$ for all $0 \leq t \leq 1$ and $n \in \mathbb{N}$]. By using $q(t) := [t(1-t)]^{1/3}$ and $V_n(t) := t^3$ for $0 \leq t \leq 1$, $n \in \mathbb{N}$, Theorem 3.5 proves $\mathscr{L}_b$-convergence of $(W_n)_{n\in\mathbb{N}}$, that is, the weight function is of the form

$q^{-1}(V_n(t)) \geq \frac{1}{2}t^{-1}$ near 0. This result cannot be achieved by Corollary 3.8 because in order to establish this, a weight function $q^*(t) \leq 2t^{1/2}$ near 0 would be necessary; this function does not satisfy (Q_k) for any $k \in \mathbb{N}$.

4. Rates for almost sure convergence of the weighted U-process in the i.i.d. case. In this section let $(\xi_i)_{i \in \mathbb{N}}$ be a sequence of independent and identically distributed (i.i.d.) random elements in $(Z, \mathscr{D})$ defined on $(\Omega, \mathscr{F}, \mathbb{P})$, $m \in \mathbb{N}$, and let $h: Z^m \to \mathbb{R}$ be such that $h(\xi_{i_1}, \dots, \xi_{i_m})$ is measurable for all $(i_1, \dots, i_m) \in E_n$, $n \geq m$. Furthermore, we will use the definitions of Section 1. Then we consider a weight function q: $[0,1] \to \mathbb{R}_+$ with the following property:

> q is increasing on $[0, 1/2]$, symmetric about $1/2$ and continuous; $\int_0^1 q^{-1}(x)\,dx < \infty$.

Our aim is to study the almost sure convergence of $(a_n D_n^q)_{n \geq m}$, where $(a_n)_{n \geq m}$ is an increasing sequence of positive real numbers and D_n^q is defined for $n \geq m$ by

$$D_n^q := \sup\left\{q^{-1}\left(\overline{H}(t)\right)\left|H_n(t) - \overline{H}(t)\right| : 0 < \overline{H}(t) < 1\right\}.$$

The next proposition is an immediate consequence of the fact that $H_n = (n!)^{-1}\sum_{i=1}^{n!} H_n^i$, where for $1 \leq i \leq n!$ and $n \geq m$, H_n^i is the empirical measure of $[[n/m]]$ independent random variables all having the distribution of $h(\xi_1, \dots, \xi_m)$ (here $[[\cdot]]$ denotes the greatest-integer function); this representation is due to Hoeffding (1963).

PROPOSITION 4.1. *Let for all $i \in \mathbb{N}$, $\eta_1, \dots, \eta_i$ be independent $U(0,1)$-distributed random variables,*

$$\alpha_i(t) := i^{-1/2} \sum_{j=1}^{i} \left(1(n_j \leq t) - t\right), \qquad 0 \leq t \leq 1,$$

and

$$\overline{D}_i^q := \sup\left\{i^{-1/2} q^{-1}(t)|\alpha_i(t)| : 0 < t < 1\right\}.$$

Then we have for all $n \geq m$, $\mathbb{E}(D_n^q) \leq \mathbb{E}(\overline{D}_{[[n/m]]}^q)$.

Based on these definitions we prove the following lemma.

LEMMA 4.2. *If $\int_0^1 q^{-1-c}(x)\,dx < \infty$ for some $0 \leq c \leq 1$, then there exist $C \geq 0$ and a sequence $(\gamma_n)_{n \in \mathbb{N}}$ with $\gamma_n \downarrow 0$ such that*

$$\mathbb{E}\left(\overline{D}_n^q\right) \leq C n^{-c/(1+c)} (\gamma_n)^{1-c} \quad \text{for all } n \in \mathbb{N}.$$

THEOREM 4.3. *Assume that* $\int_0^1 q^{-1-c}(x)\,dx < \infty$ *for some* $0 \le c \le 1$. *For all increasing sequences* $(a_n)_{n \ge m} \subset \mathbb{R}_+$ *which satisfy*

$$\sum_{n=m+1}^{\infty} (a_n - a_{n-1}) n^{-c/(1+c)} < \infty$$

and either (i) $c < 1$ *and* $a_n = o(n^{c/(1+c)})$ *or* (ii) $c = 1$ *and* $a_n = o(n^{1/2})$, *we have*: $a_n D_n^q \to 0$ $\mathbb{P}$-*a.s.*

COROLLARY 4.4. *Provided that* $\int_0^1 q^{-1-c}(x)\,dx < \infty$ *for some* $0 \le c \le 1$, *we have for all decreasing* $f\colon (0,1] \to \mathbb{R}_+$ *with* $\int_0^1 f(x)\,dx < \infty$: $f(n^{-c/(1+c)}) D_n^q \to 0$ $\mathbb{P}$-*a.s.*

COROLLARY 4.5.

(i) *For* $\int_0^1 q^{-1}(x)\,dx < \infty$, *we have*: $D_n^q \to 0$ $\mathbb{P}$-*a.s.*
(ii) *For* $\int_0^1 q^{-2}(x)\,dx < \infty$, *we have*: $n^{1/2}[\log n (\log\log n)^{1+\varepsilon}]^{-1} D_n^q \to 0$ $\mathbb{P}$-*a.s. for all* $\varepsilon > 0$.

REMARK 4.6.

1. Corollary 4.5(i) is Theorem 2.2 of Helmers, Janssen and Serfling (1985) and Theorem 1(i) of Aerts, Janssen and Mason (1986).
2. The rates in Theorem 4.3 may be compared with those in Theorem (9) of Nolan and Pollard (1987), who study the more general case of function-indexed U-processes. Applied to our special situation for many kernel functions, their theorem gives only bounded weights $q^{-1}(\nu)$ (because the function ν used there is bounded away from 0) in contrast to Theorem 4.3.

We conclude this section by giving an example which shows that for special kernel functions h, Corollary 4.5(i) is the best possible sufficient condition (note that this might not be the case for "all" h).

EXAMPLE. Let $\mathscr{L}(\xi_1) = U(0,1)$ and $h(x_1, \dots, x_m) := \max(x_1, \dots, x_m)$. Then we have for all $n \ge m$:

$$(+) \qquad H_n(t) = R_n \prod_{i=0}^{m-1} \left[F_n(t) - \frac{i}{n} \right] \quad \text{for } 0 \le t \le 1,$$

where

$$F_n(t) := n^{-1} \sum_{i=1}^{n} 1(\xi_i \le t) \quad \text{and} \quad R_n = n^m \frac{(n-m)!}{n!} \ge 1.$$

We suppose $\int_0^1 q^{-1}(x)\,dx = \infty$. For all $M \geq q^{-1}(1/2)$, we define $0 < a_n \leq 1/2$ for $n \in \mathbb{N}$ by $q^{-1}(a_n^m) = Mn^m$. Then $(a_n)_{n \in \mathbb{N}}$ is decreasing and it follows that

$$\sum_{n=1}^{\infty} a_n^m n^{m-1} \geq 2^{-m} M^{-1} \int_0^{a_1^m} q^{-1}(x)\,dx = \infty.$$

An application of Mori (1976), Lemma 3, implies:

$$\mathbb{P}(\xi_{n:m} \leq a_n \quad \text{for infinitely many } n \in \mathbb{N}) = 1,$$

where $\xi_{n:m}$ denotes the mth-order statistic of $\xi_1, \dots, \xi_n$ for all $n \in \mathbb{N}$. Noting that $\xi_{n:m} \leq a_n$ gives $q^{-1}(\overline{H}(a_n))H_n(a_n) \geq M$ [by $(+)$], we get

$$\mathbb{P}\Big(\limsup_{n \to \infty} \sup\big\{q^{-1}\big(\overline{H}(t)\big)H_n(t) \colon 0 < t \leq 1/2\big\} \geq M\Big) = 1.$$

$M \uparrow \infty$ immediately implies: $\limsup_{n \to \infty} D_n^q = \infty$ $\mathbb{P}$-a.s.

So this example is an improvement of Theorem 1(ii) of Aerts, Janssen and Mason (1986) if we restrict ourselves to this special situation.

5. Proofs.

Proof of Proposition 3.1. The proof is given in two steps. Suppose w.l.o.g. $m \geq 2$.

Part 1. For all $n \in \mathbb{N}$ we consider the following processes, indexed by $t \in \mathbb{R}$:

$$L_n(t) := \sum_{j=1}^{m} \left[\mathop{\times}_{1}^{j-1} \nu_n \times \beta_n \times \mathop{\times}_{j+1}^{m} \nu_n \right](\{h \leq t\}),$$

$$\overline{L}_n(t) := \sum_{j=1}^{m} \left[\mathop{\times}_{1}^{j-1} \nu_n \times \beta_n \times \mathop{\times}_{j+1}^{m} \mu_n \right](\{h \leq t\}),$$

where for all $D \in \mathscr{D}$,

$$\mu_n(D) := i(n)^{-1} \sum_{i=1}^{i(n)} 1_D(\xi_{ni}),$$

$$\nu_n(D) := i(n)^{-1} \sum_{i=1}^{i(n)} \nu_{ni}(D) \quad \text{with } \nu_{ni} := \mathscr{L}(\xi_{ni}),$$

$$\beta_n(D) := i(n)^{1/2}\big[\mu_n(D) - \nu_n(D)\big].$$

We prove (1): For all $k \in \mathbb{N}$ there exists a constant $C = C(k, m) \geq 0$ with $\mathbb{E}([L_n(t) - U_n(t)]^{2k}) \leq Ci(n)^{-k}$ for all $t \in \mathbb{R}$, $n \in \mathbb{N}$.

It is obvious that $\overline{L}_n = i(n)^{1/2}[\times_1^m \mu_n - \times_1^m \nu_n]$, which immediately implies

$$|U_n - \overline{L}_n| \leq 2\left[1 - \prod_{i=0}^{m-1} \frac{i(n) - i}{i(n)}\right] i(n)^{1/2} \leq 2m^3 i(n)^{-1/2}.$$

It only remains to show that (1) holds with U_n replaced by $\overline{L}_n$, which is a consequence of the following statement:

(2) For all $k \in \mathbb{N}$ and $r \in \mathbb{N}$ there exists $C = C(k, r) > 0$ with

$$\mathbb{E}\left(\left[\beta_n \times \left[\bigtimes_1^r \mu_n - \bigtimes_1^r \nu_n\right](B)\right]^{2k}\right) \leq Ci(n)^{-k} \quad \text{for all } n \in \mathbb{N},\ B \in \bigotimes_1^{r+1} \mathscr{D}.$$

The proof of (2) proceeds by induction on $r \in \mathbb{N}$. First, we treat the case $r = 1$: For all $B \in \otimes_1^2 \mathscr{D}$ and $n \in \mathbb{N}$, it follows with $M_n := \{1, \dots, i(n)\}^{4k}$:

$$\begin{aligned}
&\mathbb{E}\left([\beta_n \times (\mu_n - \nu_n)(B)]^{2k}\right)\\
&\quad = i(n)^{-3k} \sum_{\substack{(i_1, \dots, i_{4k}) \in M_n \\ \operatorname{card}(\{i_1, \dots, i_{4k}\}) \leq 2k}} \mathbb{E}\left(\left[\bigtimes_{j=1}^{4k} \left(1_{\xi_{ni_j}} - \nu_{ni_j}\right)\right]\left(\bigtimes_1^{2k} B\right)\right)\\
&\quad \leq i(n)^{-3k} \sum_{j=1}^{2k} \binom{4k}{j} i(n)^j j^{4k-j} 2^{2k} \leq C(k) i(n)^{-k}.
\end{aligned}$$

Now suppose that (2) holds for $r \in \mathbb{N}$. Then we obtain for all $n \in \mathbb{N}$ and $B \in \otimes_1^{r+2} \mathscr{D}$:

$$\begin{aligned}
&\mathbb{E}\left(\beta_n \times \left[\bigtimes_1^{r+1} \mu_n - \bigtimes_1^{r+1} \nu_n\right](B)^{2k}\right)\\
&\quad \leq 2^{2k}\Bigg[i(n)^{-1} \sum_{i=1}^{i(n)} \mathbb{E}\left(\beta_n \times \left[\bigtimes_1^r \mu_n - \bigtimes_1^r \nu_n\right] \times 1_{\xi_{ni}}(B)^{2k}\right)\\
&\qquad\qquad + \mathbb{E}\left(\beta_n \times \bigtimes_1^r \nu_n \times (\mu_n - \nu_n)(B)^{2k}\right)\Bigg].
\end{aligned}$$

The first part of the sum on the right-hand side can be estimated by an application of the induction hypothesis, while the second may be reduced to the case $r = 2$ by Jensen's inequality. This proves (2).

Part 2. We show (3): For all $k \in \mathbb{N}$ there is a $C = C(k, m) > 0$ with

$$\mathbb{E}\left([L_n(t) - L_n(r)]^{2k}\right) \leq C\left(\mathbb{E}\left([L_n(t) - L_n(r)]^2\right)^k + i(n)^{-k}\right)$$
$$\text{for all } r, t \in \mathbb{R} \text{ and } n \in \mathbb{N}.$$

Combining (1) and (3) finishes the proof of Proposition 3.1. But (3) is an immediate consequence of the well-known Bernstein inequality [cf. Gaenssler

(1983), page 9]: Let $n \in \mathbb{N}$ and $-\infty < r < t < \infty$ be fixed. It follows with $v := \mathbb{E}([L_n(t) - L_n(r)]^2)$ and $u := [3vi(n)^{1/2}/(2m)]^{2k}$:

$$
\begin{aligned}
&\mathbb{E}\big([L_n(t) - L_n(r)]^{2k}\big)\\
&= \int_0^\infty \mathbb{P}\big(|L_n(t) - L_n(r)| \geq \varepsilon^{1/(2k)}\big)\,d\varepsilon\\
&\leq \int_0^\infty 2\exp\left(-\frac{3\varepsilon^{1/k}}{6v + 4\varepsilon^{1/(2k)}mi(n)^{-1/2}}\right)d\varepsilon\\
&\leq \int_0^u 2\exp\left(-\frac{\varepsilon^{1/k}}{4v}\right)d\varepsilon + \int_u^\infty 2\exp\left(-\frac{3\varepsilon^{1/(2k)}i(n)^{1/2}}{8m}\right)d\varepsilon\\
&= 2\int_0^{uv^{-k}} v^k \exp\left(-\frac{1}{4}\varepsilon^{1/k}\right)d\varepsilon + 2\int_{ui(n)^k}^\infty i(n)^{-k}\exp\left(-\frac{3}{8m}\varepsilon^{1/(2k)}\right)d\varepsilon\\
&\leq 2\int_0^\infty \left[\exp\left(-\frac{1}{4}\varepsilon^{1/k}\right) + \exp\left(-\frac{3}{8m}\varepsilon^{1/(2k)}\right)\right]d\varepsilon\big[v^k + i(n)^{-k}\big]. \qquad \square
\end{aligned}
$$

PROOF OF LEMMA 3.2. For all $n \in \mathbb{N}$ we define a sequence $(t_i^n)_{i \in \mathbb{N}}$ as follows (with $t_0^n := -\infty$ and $\inf\varnothing := \infty$): $t_{i+1}^n := \inf\{t > t_i^n\colon V_n(t) - V_n(t_i^n) \geq i(n)^{-1}\}$ for $i \in \mathbb{N}$. Then it is obvious that there exists some $l_n \in \mathbb{N}$ with

$$-\infty = t_0^n < t_1^n < \cdots < t_{l_n-1}^n < t_{l_n}^n = \infty.$$

This implies for all $n \in \mathbb{N}$, $0 < \delta \leq 1/2$ and $1 \leq i < j \leq l_n$ with $V_n(t_j^n) \leq \delta$ (where C is a variable constant only depending on k_0 and m):

$$
\begin{aligned}
\mathbb{E}\big([W_n(t_j^n) - W_n(t_i^n)]^{4k_0}\big) &\leq C\Big[q^{-4k_0}(V_n(t_j^n))\mathbb{E}\big([U_n(t_j^n) - U_n(t_i^n)]^{4k_0}\big)\\
&\qquad + \mathbb{E}\big(U_n(t_i^n)^{4k_0}\big)\big[q^{-1}(V_n(t_i^n)) - q^{-1}(V_n(t_j^n))\big]^{4k_0}\Big]\\
&\overset{(3.1)}{\leq} C\Big[q^{-4k_0}(V_n(t_j^n))\big[V_n(t_j^n) - V_n(t_i^n)\big]^{2k_0}\\
&\qquad + V_n(t_i^n)^{2k_0}\big[q^{-1}(V_n(t_i^n)) - q^{-1}(V_n(t_j^n))\big]^{4k_0}\Big]\\
&\leq C\left[\sum_{r=i+1}^{j}\big(\tilde{I}(V_n(t_r^n)) - \tilde{I}(V_n(t_{r-1}^n))\big)\right]^2,
\end{aligned}
$$

where

$$
\begin{aligned}
\tilde{I}(t) &:= \int_0^t \big[q^{-2k_0}(x)\,d(x^{k_0}) + x^{k_0}\,d(-q^{-2k_0}(x))\big]\\
&\leq 2k_0\int_0^t q^{-2k_0}(x)x^{k_0-1}\,dx \quad \text{for all } 0 \leq t \leq 1/2.
\end{aligned}
$$

Theorem 12.2 in Billingsley (1968) immediately gives for all $\varepsilon > 0$:

$$\mathbb{P}(T_n \geq \varepsilon) \leq C\varepsilon^{-4k_0}\left[\int_0^\delta q^{-2k_0}(x)x^{k_0-1}\,dx\right]^2 \tag{4}$$

$$\text{with } T_n = \sup\{|W_n(t_i^n)|: V_n(t_i^n) \leq \delta\}.$$

To complete the proof of Lemma 3.2, we make use of the following simple inequality (with $M_n := \{i \geq 1: V_n(t_i^n) \leq \delta\}$):

$$\sup\{|W_n(t)|: V_n(t) \leq \delta\} \leq T_n + \max\{q^{-1}(V_n(t_i^n))T_i^n: i \in M_n\},$$

where

$$T_i^n := \sup\{|U_n(t) - U_n(t_i^n)|: t_i^n \leq t < t_{i+1}^n\} \quad \text{for } i \in M_n.$$

Now we show:

There exists $C > 0$ with $\mathbb{P}(T_i^n \geq \varepsilon) \leq C\varepsilon^{-4k_0}i(n)^{1-2k_0}$ for all $\varepsilon > 0$, $i \in M_n$, $n \in \mathbb{N}$, which immediately implies for $\varepsilon > 0$ and $n \in \mathbb{N}$:

$$\begin{aligned}
&\mathbb{P}\left(\max\{q^{-1}(V_n(t_i^n))T_i^n: i \in M_n\} \geq \varepsilon\right)\\
&\quad\leq \sum_{i \in M_n} C\varepsilon^{-4k_0}q^{-4k_0}(V_n(t_i^n))i(n)^{1-2k_0}\\
&\quad\leq \sum_{i \in M_n} C\varepsilon^{-4k_0}q^{-4k_0}(V_n(t_i^n))[V_n(t_i^n) - V_n(t_{i-1}^n)]^{2k_0-1}\\
&\quad\leq Ck_0^2\varepsilon^{-4k_0}\left[\int_0^\delta q^{-2k_0}(x)x^{k_0-3/2}\,dx\right]^2.
\end{aligned} \tag{5}$$

Together with (4) this finishes the proof of Lemma 3.2 so that only (5) remains to be shown: For this let $n \in \mathbb{N}$ and $i \in M_n$ be fixed. Then we define some sequence $(s_j^n)_{j\in\mathbb{N}}$ as follows (with $s_0^n := t_i^n$ and $\inf\varnothing := \infty$):

$$s_{j+1}^n := \inf\left\{s > s_j^n: \overline{H}_n(s) - \overline{H}_n(s_j^n) \geq i(n)^{-1}\right\} \quad \text{for } j \in \mathbb{N}.$$

Then it is obvious that there exists some $r_n \in \mathbb{N}$ with $s_{r_n}^n < t_{i+1}^n \leq s_{r_n+1}^n$ and $r_n \leq i(n)$ such that [$U_n(t-)$ denoting left-hand limit]:

$$\begin{aligned}
\mathbb{P}(T_i^n \geq 2\varepsilon) &\leq \sum_{j=0}^{r_n}\Big[\mathbb{P}\left(\sup\{|U_n(t) - U_n(s_j^n)|: s_j^n \leq t < s_{j+1}^n\} \geq \varepsilon\right)\\
&\qquad\qquad + \mathbb{P}\left(|U_n(s_j^n) - U_n(s_0^n)| \geq \varepsilon\right)\Big]\\
&\leq \sum_{j=0}^{r_n}\Big[\mathbb{P}\left(|U_n(s_{j+1}^n -) - U_n(s_j^n)| + 2i(n)^{-1/2} \geq \varepsilon\right)\\
&\qquad\qquad + \mathbb{P}\left(|U_n(s_j^n) - U_n(s_0^n)| \geq \varepsilon\right)\Big]\\
&\overset{(3.1)}{\leq} C\varepsilon^{-4k_0}i(n)^{-2k_0+1}.
\end{aligned}$$

□

PROOF OF LEMMA 3.3. Let $0 \le a \le b \le 1$ and $n \in \mathbb{N}$ be fixed. Similar to the proof of Lemma 3.2 it follows: There exists $C > 0$ with

$$\mathbb{P}(\sup\{|U_n(t) - U_n(r)|: a \le V_n(r) \le V_n(t) \le b\} \ge \varepsilon)$$
$$\le C\varepsilon^{-6}\left[b - a + i(n)^{-1}\right]^2 \quad \text{for all } \varepsilon > 0 \text{ and } n \in \mathbb{N}$$

[compare with the case $q \equiv 1$ and $k_0 = 3/2$; the term $i(n)^{-1}$ results from the fact that in order to use Billingsley (1968), 12.2, we partitioned $\mathbb{R}$ into at most $i(n)$ intervals of at least "V_n"-length $i(n)^{-1}$]. Now dividing $\mathbb{R}$ into at most δ^{-1} intervals of at least "V_n"-length δ completes the proof. □

PROOF OF THEOREM 3.4. Let $\omega^s_{W_n}(\cdot)$ be defined as in Section 2. Then we obtain for all $n \in \mathbb{N}$, $0 < 2\delta \le \tilde{\delta} \le 1/2$ with

$$\omega_n(\delta) := \sup\{|W_n(t) - W_n(r)|: |V_n(t) - V_n(r)| \le \delta\}:$$

$$\begin{aligned}\omega_n(\delta) \le\ & 2\sup\{|W_n(t)|: V_n(t) \le \tilde{\delta}\} + 2\sup\{|W_n(t)|: V_n(t) \ge 1 - \tilde{\delta}\}\\ &+ q^{-1}\left(\frac{\tilde{\delta}}{2}\right)\sup\{|U_n(t) - U_n(r)|: |V_n(t) - V_n(r)| \le \delta\}\\ &+ \sup\{|U_n(t)|: t \in \mathbb{R}\}\sup\left\{|q^{-1}(a) - q^{-1}(b)|:\right.\\ &\qquad\left.\frac{\tilde{\delta}}{2} \le a \le b \le 1 - \frac{\tilde{\delta}}{2};\, b - a \le \delta\right\}.\end{aligned}$$

Application of Lemmas 3.2 and 3.3 immediately implies (note that q is continuous and that a result similar to Lemma 3.2 holds for W_n if V_n is "near 1"; this case can be reduced to Lemma 3.2 by using the kernel function $-h$):

$$\lim_{\delta \downarrow 0} \limsup_{n \to \infty} \mathbb{P}(\omega_n(\delta) \ge \varepsilon) \le C\varepsilon^{-4k_0}\left(\int_0^{\tilde{\delta}} q^{-2k_0}(x)^{k_0 - 3/2}\right)^2 \quad \text{for all } \varepsilon > 0.$$

Then $\tilde{\delta} \downarrow 0$ and use of (V)(i) show that Theorem 2.1(A) holds. Theorem 2.1(B) is a trivial consequence of Proposition 3.1 because we have for all $t \in \mathbb{R}$, $n \in \mathbb{N}$:

$$\mathbb{E}\left(W_n(t)^4\right) \le C\frac{V_n(t)^2 + i(n)^{-2}}{q^4\left(\max\left(V_n(t), i(n)^{-1}\right)\right)} \le 2C\left[k_0\int_0^1 q^{-2k_0}(x)x^{k_0-1}\,dx\right]^{2/k_0}$$

with C not depending on n and t. □

PROOF OF THEOREM 3.5. The last inequality in the proof of Theorem 3.4 also proves the implication (i) ⇒ (ii). It remains to show that (ii) ⇒ (i). It is obvious that for $t \in \mathbb{R}$ one of the following statements must hold:

1. $(V_n(t))_{n \in \mathbb{N}}$ is bounded away from 0 and 1.
2. $W_n(t) \to_{\mathbb{P}} 0$.

[Note that if there is some subsequence (n') with $V_{n'}(t) \to 0$ or 1, we obtain $\mathbb{E}(W_{n'}(t)^2) \to 0$, that is, (ii) implies $\mathbb{E}(W_n(t)^2) \to 0$.] So we are able to restrict ourselves to case (1) if we prove convergence of the finite-dimensional distributions. Considering now the auxiliary process L_n of the proof of Proposition 3.1, we have for all $t \in \mathbb{R}$: $\mathbb{E}([q^{-1}(V_n(t))L_n(t)1_{[i(n)^{-1}, 1-i(n)^{-1})}(V_n(t)) - W_n(t)]^2) \to 0$ as $n \to \infty$, so that only the convergence of the finite-dimensional distributions of $q^{-1}(V_n(t))L_n(t)$ to normal distributions has to be shown [note that we are in case (1)]. This is an immediate consequence of the Cramér–Wold device and the central limit theorem. □

Proof of Corollary 3.6. This follows by noticing that in the case $V_n = \overline{H}_n$, $n \in \mathbb{N}$, we are able to improve relation (5) in the proof of Lemma 3.2 as follows:

There exists $C > 0$ with

$$\mathbb{P}(T_i^n \geq \varepsilon) \leq C\varepsilon^{-4k_0} i(n)^{-2k_0} \quad \text{for all } \varepsilon > 0,\ n \in \mathbb{N}. \tag{6}$$

This is an easy consequence of the fact that (using the notation of the proof of Lemma 3.2) $T_i^n \leq |U_n(t_{i+1}^n -) - U_n(t_i^n)| + 2i(n)^{-1/2}$. It only has to be shown (note that there is no truncation) that

$$S_n := \sup\left\{q^{-1}\left(\overline{H}_n(t)\right)|U_n(t)|\colon \overline{H}_n(t) \leq i(n)^{-1}\right\}$$

behaves well:

$$\begin{aligned}\mathbb{E}(S_n) &\leq 2i(n)^{1/2} \int_0^{i(n)^{-1}} q^{-1}(x)\,dx \\ &\leq 2\left[\frac{2k_0 - 1}{k_0}\right]^{(2k_0-1)/(2k_0)} \left[\int_0^{i(n)^{-1}} q^{-2k_0}(x)x^{k_0-1}\,dx\right]^{1/(2k_0)}. \qquad \square\end{aligned}$$

Proof of Lemma 4.2. Using the notation of Lemma 4.1, we obtain by an application of Chow's inequality [cf. Gaenssler and Stute (1977), 6.6.1] to the martingale $(\alpha_n(t)/(1-t))_{0 \leq t < 1}$:

$$\mathbb{P}(R_n^q \geq \varepsilon) \leq 36\varepsilon^{-2} \int_{1/n}^{1/2} q^{-2}(x)\,dx \quad \text{for all } n \geq 3 \text{ and } \varepsilon > 0,$$

where

$$R_n^q := \sup\left\{q^{-1}(t)\left|\alpha_n(t) - \alpha_n\left(\frac{1}{2}\right)\right| : \frac{1}{n} \leq t \leq \frac{1}{2}\right\}.$$

Integration over all $\varepsilon > 0$ gives for $n \geq 3$

$$\mathbb{E}(R_n^q) \leq \int_0^\infty \mathbb{P}(R_n^q \geq \varepsilon)\,d\varepsilon \leq \left[\int_0^\infty \min(1, 36u^{-2})\,du\right]\left[\int_{1/n}^{1/2} q^{-2}(x)\,dx\right]^{1/2}.$$

With $I(s) := \int_0^s q^{-1-c}(x)\,dx$ for $0 \leq s \leq 1/2$ and $C := \int_0^\infty \min(1, 36u^{-2})\,du$ this implies for all $n \geq 3$:

$$\mathbb{E}(n^{-1/2}R_n^q) \leq Cn^{-1/2}\left[q^{c-1}\left(\frac{1}{n}\right)\int_{1/n}^{1/2} q^{-1-c}(x)\,dx\right]^{1/2}$$

$$= Cn^{-c/(1+c)}\left[n^{-1}q^{-1-c}\left(\frac{1}{n}\right)\right]^{(1-c)/(2+2c)}\left[I\left(\frac{1}{2}\right) - I\left(\frac{1}{n}\right)\right]^{1/2}$$

$$\leq C_1 n^{-c/(1+c)} I\left(\frac{1}{n}\right)^{(1-c)/(2+2c)} I\left(\frac{1}{2}\right)^{1/2}.$$

Similarly to the end of the proof of Corollary 3.6, we obtain for $S_n^q := \sup\{|\alpha_n(t)|: t \leq 1/n\}$:

$$\mathbb{E}(n^{-1/2}S_n^q) \leq 2\int_0^{1/n} q^{-1}(x)\,dx \leq 2n^{-c/(1+c)} I\left(\frac{1}{n}\right)^{1/(1+c)} \quad \text{for all } n \geq 2.$$

This completes the proof because we have for all $n \geq 2$:

$$\mathbb{E}(\overline{D}_n^q) \leq 2\left[\mathbb{E}(n^{-1/2}S_n^q) + \mathbb{E}(n^{-1/2}R_n^q)\right]. \qquad \square$$

PROOF OF THEOREM 4.3. It is obvious that $(D_n^q)_{n \geq m}$ defines a reversed submartingale [compare Gaenssler (1983), Lemma 5]. Then Theorem 4.3 is an immediate consequence of Lemma 4.2 and Chow's inequality. $\square$

PROOF OF COROLLARY 4.4. This follows with $a_n = f(n^{-c/(1+c)})$, $n \geq m$. $\square$

PROOF OF COROLLARY 4.5. Apply Corollary 4.4: For (i) we use $f \equiv 1$, $c = 0$ and for (ii) we use

$$c = 1 \quad \text{and} \quad f(x) = \left[x \log\frac{1}{x}\left(\log\log\frac{1}{x}\right)^{1+\varepsilon}\right]^{-1}$$

for small $x > 0$, $\varepsilon > 0$. $\square$

Acknowledgment. I am especially grateful to Professor Dr. Peter Gaenssler for his support during the preparation of my thesis, which the present paper is part of.

REFERENCES

AERTS, M., JANSSEN, P. and MASON, D. (1986). Glivenko-Cantelli convergence for weighted empirical and quantile processes of U-statistic structure. *Probab. Theory Math. Statist.* **1** 23–30.

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

DEHLING, H., DENKER, M. and PHILIPP, W. (1987). The almost sure invariance principle for the empirical process of U-statistic structure. *Ann. Inst. H. Poincaré Sect. B (N.S.)*. **23** 121–134.

GAENSSLER, P. (1983). *Empirical Processes*. IMS, Hayward, CA.

GAENSSLER, P. and SCHNEEMEIER, W. (1986). On functional limit theorems for a class of stochastic processes indexed by pseudo-metric parameter spaces (with applications to empirical processes). In *Probability in Banach Spaces 6. Proc. Sixth Internat. Conf., Denmark, 1986.* [*Progress in Probability* **20** (1990).] Birkhäuser, Boston.

GAENSSLER, P. and STUTE, W. (1977). *Wahrscheinlichkeitstheorie*. Springer, Berlin.

HELMERS, R., JANSSEN, P. and SERFLING, R. (1985). Glivenko–Cantelli properties of some generalized empirical distribution functions and strong convergence of generalized L-Statistics. Preprint, Dept. Math. Sci., Johns Hopkins Univ.

HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.

MORI, T. (1976). The strong law of large numbers when extreme terms are excluded from sums. *Z. Wahrsch. Verw. Gebiete* **36** 189–194.

NOLAN, D. and POLLARD, D. (1987). U-Processes: Rates of convergence. *Ann. Statist.* **15** 780–799.

NOLAN, D. and POLLARD, D. (1988). Functional limit theorems for U-processes. *Ann. Probab.* **16** 1291–1298.

SILVERMAN, B. W. (1983). Convergence of a class of empirical distribution functions of dependent random variables. *Ann. Probab.* **11** 745–761.

MATHEMATICAL INSTITUTE
UNIVERSITY OF MUNICH
THERESIENSTRASSE 39
D-8000 MUNICH 2
GERMANY