The Annals of Probability
1994, Vol. 22, No. 4, 1725–1744

U-STATISTIC PROCESSES: A MARTINGALE APPROACH[1]

BY WINFRIED STUTE

Justus-Liebig University

For i.i.d. data $X_1, \dots, X_n$ and a kernel h, the associated U-statistic process is defined as

$$U_n(u,v) = \frac{1}{n(n-1)} \sum_{1 \le i \ne j \le n} h(X_i, X_j) 1_{\{X_i \le u, X_j \le v\}}.$$

Variants of these processes occur, for example, in the representation of the product-limit estimator of a lifetime distribution for censored/truncated data or in trimmed U-statistics. We derive an almost sure representation of U_n under weak moment assumptions on h. Proofs rely on a proper decomposition of the remainder term into strong two-parameter martingales.

1. Introduction and main results. Assume that $X_1, \dots, X_n$ is a (finite) sequence of independent identically distributed (i.i.d.) random variables with distribution function (d.f.) F, defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Let h be a function (the kernel) on m-dimensional Euclidean space and set (for $n \ge m$)

$$U_n = \frac{(n-m)!}{n!} \sum_{\pi} h(X_{i_1}, \dots, X_{i_m}),$$

where π extends over all multiindices $\pi = (i_1, \dots, i_m)$ of pairwise distinct $1 \le i_j \le n$, $1 \le j \le m$. Commonly U_n is called a U-statistic. U-statistics were introduced by Hoeffding (1948). They have been extensively investigated over the last 40 years. Most of the basic theory is contained in Serfling (1980), Denker (1985) and Lee (1990). See also Randles and Wolfe (1979).

More recently much attention has been given to what has been called a U (-statistic) process. For ease of representation we shall restrict ourselves to degree $m = 2$. The U-statistic process is then defined as follows: for real u and v set

$$U_n(u,v) = \frac{1}{n(n-1)} \sum_{1 \le i \ne j \le n} h(X_i, X_j) 1_{\{X_i \le u, X_j \le v\}}.$$

Write

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} 1_{\{X_i \le x\}}, \qquad x \in \mathbb{R},$$

Received January 1993; revised February 1994.
[1]Work partially supported by the Deutsche Forschungsgemeinschaft.
AMS 1991 *subject classifications*. Primary 62G30, 60G42; secondary 62G05.
Key words and phrases. U-statistic processes, maximal inequalities, martingale decomposition, Hájek projection.

the empirical d.f. of the sample. Then $U_n(u,v)$ becomes (assuming no ties)

$$U_n(u,v) = \frac{n}{n-1}\int_{-\infty}^{u}\int_{-\infty}^{v} h(x,y)1_{\{x\neq y\}}F_n(dx)F_n(dy).$$

Now, a standard method to analyze the (large sample) distributional behavior of U_n is to write U_n as

$$U_n = \widehat{U}_n + R_n,$$

in which $\widehat{U}_n$ is the Hájek projection of U_n and the remainder $R_n = U_n - \widehat{U}_n$ is a degenerate U-statistic that is asymptotically negligible when compared with $\widehat{U}_n$. In fact, provided that h has a finite pth moment and zero mean,

$$n^{1/2}\widehat{U}_n \to \mathcal{N}(0,\sigma^2) \quad \text{in distribution}$$

and

$$\mathbb{E}|R_n|^p \leq Cn^{-p}.$$

See Serfling [(1980), page 188]. It follows that also

$$n^{1/2}U_n \to \mathcal{N}(0,\sigma^2) \quad \text{in distribution.}$$

Of course, this approach immediately applies to $U_n(u,v)$ for each (u,v) fixed. Just replace $h(x,y)$ by $h(x,y)1_{\{x\leq u, y\leq v\}}$. Unfortunately, this is insufficient for handling $U_n(u,v)$ [resp. $R_n(u,v)$] as a process in (u,v). Particularly, the (pointwise) Hájek approach does not yield bounds for $\sup_{u,v}|R_n(u,v)|$. Such bounds are, however, extremely useful in applications. For example, in survival analysis, U-statistic processes or variants of them appear in the context of estimating the lifetime distribution F and the cumulative hazard function Λ when the data are censored or truncated [cf. Lo and Singh (1986), Lo, Mack and Wang (1989), Major and Rejtő (1988) and Chao and Lo (1988)]. In Lo and Singh (1986) the analysis of the remainder term incorporated known global and local properties of empirical processes. In Lo, Mack and Wang (1989) the error bounds were improved by applying a sharp moment bound for degenerate U-statistics due to Dehling, Denker and Philipp (1987). In Major and Rejtő (1988) a bound for $\sup_u|R_n(u,1)|$ of large deviation type due to Major (1988) was applied, which required h to be bounded. In all these papers estimation of F (resp. Λ) could only be carried through on intervals strictly contained in the support of the distribution of the observed data; similarly in Chao and Lo (1988) for truncated data situations. This general drawback mainly arose because of lack of a sharp bound for $\sup_{u,v}|R_n(u,v)|$ when the kernel h is not necessarily bounded.

Classes of degenerate U-statistics also have been studied, from a different point of view, by Nolan and Pollard (1987). In their Theorem 6 they derive an upper bound for the mean of the supremum by first decoupling the U-process of interest and then using a chaining argument conditionally on the observations. Now, by Hölder, a more efficient inequality would be one relating the pth order mean of the supremum to the pth order mean of the envelope function, $p \geq 2$.

At least this is a typical feature of many other maximal inequalities. We also refer to de la Peña (1992) and the literature cited there. In these papers the main emphasis is on relating the maximum of interest to the maximum of a decoupled process. No explicit bounds for a degenerate U-statistic process are derived that are comparable to ours. Note, however, that in applications the leading (Hájek) part is well understood and it is the degenerate part that creates the more serious problems.

In this paper we shall employ martingale methods to provide a maximal bound satisfying the above requirements. As a consequence we would be able to improve the a.s. representations of the product-limit estimators of F for censored and truncated data as discussed above; see Stute (1993, 1994).

Denote by $\widehat{U}_n(u,v)$ the Hájek projection of $U_n(u,v)$. As for proofs, unfortunately, as a process in (u,v),

$$U_n(u,v) - \widehat{U}_n(u,v)$$

does not enjoy any particular properties, so that standard maximal inequalities could be applied. As another possibility, assume for a moment that h is nonnegative. Then $U_n(u,v)$ is nondecreasing in (u,v) and adapted to the filtration

$$\mathcal{F}_{u,v} = \sigma\left(1_{\{X_i \le x\}},\ x \le \max(u,v)\right).$$

Let $C_n(u,v)$ denote the compensator in the Doob–Meyer decomposition of $U_n(u,v)$; see, for example, Dozzi (1981). At first sight one might expect

$$U_n(u,v) - C_n(u,v)$$

to be a two-parameter martingale to which standard maximal bounds could be applied; see Cairoli and Walsh (1975). A serious drawback of this approach is that with this choice of $\mathcal{F}$:

1. The process $(U_n(\cdot) - C_n(\cdot),\ \mathcal{F})$ does not satisfy the fundamental conditional independence property (F4) in Cairoli and Walsh (1975).
2. The compensator $C_n(\cdot)$ is still a U-statistic process rather than a sum of i.i.d. processes.
3. $U_n(\cdot) - C_n(\cdot)$ turns out *not* to be a degenerate U-statistic.

The last comments were meant only to express the author's difficulties when writing the paper, in finding a proper decomposition of $U_n(u,v)$, in which the remainder term (at least the most interesting part of it) is both a two-parameter (strong) martingale in (u,v) and a degenerate U-statistic for (u,v) fixed. Given such a decomposition we could then apply standard maximal inequalities for (strong) two-parameter martingales. Having thus replaced $\sup_{u,v}|R_n(u,v)|$ by a single $R_n(u,v)$, $\mathbb{E}|R_n(u,v)|^p$ could be further dealt with by applying Burkholder's inequality.

Furthermore, in our analysis, the Doob–Meyer decomposition of the process

$$\sum_j h(s,X_j)1_{\{X_j \le v\}} \quad (s \text{ fixed})$$

will be employed. Finally, some global bounds for empirical d.f.'s *á la* Dvoretzky–Kiefer–Wolfowitz (1956) will be required.

Now, the process $U_n(u,v)$ may be written as

$$\begin{aligned} n(n-1)U_n(u,v) &= \sum_{1\le i<j\le n} h(X_i, X_j)1_{\{X_i\le u, X_j\le v\}} \\ &\quad + \sum_{1\le j<i\le n} h(X_i, X_j)1_{\{X_i\le u, X_j\le v\}} \\ &\equiv I_n(u,v) + II_n(u,v). \end{aligned}$$

The following theorem contains the key representation of $I_n(u,v)$ in terms of a sum of independent random processes.

THEOREM 1.1. *Assume $h \in \mathcal{L}_p(F\otimes F)$, with $p\ge 2$. Then we have*

$$\begin{aligned} I_n(u,v) = \sum_{1\le i<j\le n} \Bigg[\int_{-\infty}^{u} h(x,X_j)1_{\{X_j\le v\}}F(dx) + \int_{-\infty}^{v} h(X_i,y)1_{\{X_i\le u\}}F(dy) \\ - \int_{-\infty}^{u}\int_{-\infty}^{v} h(x,y)F(dx)F(dy)\Bigg] + R_n(u,v), \end{aligned}$$

where for each u_0, v_0,

$$\mathbb{E}\Bigg[\sup_{u\le u_0,\, v\le v_0} |R_n(u,v)|^p\Bigg] \le C^p n^p. \tag{1.1}$$

The constant C satisfies

$$C \le \widetilde{C}\Bigg[\int_{-\infty}^{u_0}\int_{-\infty}^{v_0} |h(x,y)|^p F(dx)F(dy)\Bigg]^{1/p}$$

with $\widetilde{C}$ depending only on p.

A similar representation also holds for $II_n(u,v)$. Putting these together we get the following corollary.

COROLLARY 1.1. *Under the assumptions of Theorem* 1.1,

$$\begin{aligned} n(n-1)U_n(u,v) = n(n-1)\Bigg[&\int_{-\infty}^{u}\int_{-\infty}^{v} h(x,y)F_n(dx)F(dy) \\ &+ \int_{-\infty}^{u}\int_{-\infty}^{v} h(x,y)F(dx)F_n(dy) \\ &- \int_{-\infty}^{u}\int_{-\infty}^{v} h(x,y)F(dx)F(dy)\Bigg] + R_n(u,v). \end{aligned}$$

where the remainder satisfies (1.1), *with $\widetilde{C}$ replaced by $2\widetilde{C}$.*

Since (assuming no ties)

$$n(n-1)U_n(u,v) = n^2 \int_{-\infty}^{u}\int_{-\infty}^{v} h(x,y)1_{\{x \neq y\}}F_n(dx)F_n(dy),$$

we may write the equation in Corollary 1.1 as

$$\begin{aligned}\frac{n}{n-1}\int_{-\infty}^{u}\int_{-\infty}^{v} h(x,y)1_{\{x \neq y\}}F_n(dx)F_n(dy) &= U_n(u,v)\\ &= \int_{-\infty}^{u}\int_{-\infty}^{v} h(x,y)F_n(dx)F(dy)\\ &\quad + \int_{-\infty}^{u}\int_{-\infty}^{v} h(x,y)F(dx)F_n(dy)\\ &\quad - \int_{-\infty}^{u}\int_{-\infty}^{v} h(x,y)F(dx)F(dy) + R_n(u,v)/n(n-1).\end{aligned}$$

Inequality (1.1) together with the Markov inequality yield, with probability 1,

$$\sup_{\substack{u \le u_0\\ v \le v_0}} |R_n(u,v)| = o\big(n^{1+1/p}(\ln n)^{\delta}\big), \tag{1.2}$$

whenever δ satisfies $p\delta > 1$. Furthermore, if h is bounded, (1.1) may be applied for each $p \ge 2$.

So far we kept u_0 and v_0 fixed. In such a situation integrability of h^p only up to (u_0, v_0) is sufficient. Actually, it may happen that $(u_0, v_0) = (u_n, v_n)$ varies with n in such a way that h^p is not integrable over the whole plane, but $\int_{-\infty}^{u_n}\int_{-\infty}^{v_n} |h|^p\, dF\, dF \to \infty$ at a prescribed rate. Theorem 1.1 is particularly useful also in this case. On the other hand, if either u_n or v_n becomes small as $n \to \infty$ (such situations occur quite often in nonparametric curve estimation), then the integral

$$\int_{-\infty}^{u_n}\int_{-\infty}^{v_n} |h(x,y)|^p F(dx)F(dy)$$

also becomes small, to the effect that the bound in (1.2) may be replaced by smaller ones. The last remarks also apply to the results that follow.

Interestingly enough (1.2) may be improved a lot. This is due to the fact that according to Berk (1966) a sequence of normalized U-statistics is a reverse-time martingale. Utilizing this, we get the following result.

THEOREM 1.2. *Under the assumptions of Theorem* 1.1, *with probability* 1,

$$\sup_{\substack{u \le u_0\\ v \le v_0}} |R_n(u,v)| = o\big(n(\ln n)^{\delta}\big)$$

whenever $p\delta > 1$. *For bounded h's, we may therefore take any* $\delta > 0$.

With some extra work the logarithmic factor may be pushed down so as to get a bounded LIL. The necessary methodology may be found, for a fixed U-statistic rather than a process, in a notable paper by Dehling, Denker and Philipp (1986). After truncation, they applied their moment inequality, at stage n, with a $p = p_n$ depending on n such that $p_n \to \infty$ slowly, to the effect that for a bounded LIL the moment inequality "serves the same purpose as an exponential bound" (personal communication by M. Denker). Since this method is well established now, we need not dwell on this here again.

In the next theorem we are concerned with a two-sample situation. Let $X_1, \dots, X_n$ be i.i.d. with common d.f. F and let, independently of the X's, $Y_1, \dots, Y_m$ be another i.i.d. sequence with common d.f. G. We shall derive a representation of the process

$$nmU_{nm}(u,v) = \sum_{i=1}^{n}\sum_{j=1}^{m} h(X_i, Y_j)1_{\{X_i \le u, Y_j \le v\}}.$$

THEOREM 1.3. *Assume $h \in \mathcal{L}_p(F \otimes G)$, with $p \geq 2$. Then we have*

$$\begin{aligned} nmU_{nm}(u,v) = \sum_{i=1}^{n}\sum_{j=1}^{m} \Bigg[& \int_{-\infty}^{u} h(x, Y_j)1_{\{Y_j \le v\}}F(dx) \\ & + \int_{-\infty}^{v} h(X_i, y)1_{\{X_i \le u\}}G(dy) \\ & - \int_{-\infty}^{u}\int_{-\infty}^{v} h(x,y)F(dx)G(dy)\Bigg] + R_{nm}(u,v), \end{aligned}$$

where for each u_0, v_0,

$$\mathbb{E}\left[\sup_{\substack{u \le u_0 \\ v \le v_0}} |R_{nm}(u,v)|^p\right] \le \left[C^2 nm\right]^{p/2}.$$

The constant C satisfies

$$C \le \widetilde{C}\left[\int_{-\infty}^{u_0}\int_{-\infty}^{v_0} |h(x,y)|^p F(dx)G(dy)\right]^{1/p}.$$

The analogue of Theorem 1.2 is only formulated for $m = n$.

THEOREM 1.4. *Under the assumptions of Theorem* 1.3, *with probability* 1 *as* $n \to \infty$,

$$\sup_{\substack{u \le u_0 \\ v \le v_0}} |R_{nn}(u,v)| = o\big(n(\ln n)^{\delta}\big)$$

whenever $p\delta > 1$.

Another variant (resp. extension) of Theorem 1.1, which is extremely useful in applications, comes up when, in addition to X_i, there is a Y_i paired with X_i.

Typically X_i is correlated with Y_i. We may then form

$$I_n(u,v) = \sum_{1 \le i < j \le n} h(X_i, Y_j) 1_{\{X_i \le u, Y_j \le v\}}.$$

Clearly, this I_n equals the I_n from Theorem 1.1 if $X_i = Y_i$; similarly, for $II_n(u,v)$. Theorem 1.5 is an extension of Theorem 1.1 to paired observations.

THEOREM 1.5. *Assume that* (X_i, Y_i), $1 \le i \le n$, *is an* i.i.d. *sample from some bivariate* d.f. H *with marginals* F *and* G. *Assume* $h \in \mathcal{L}_p(F \otimes G)$ *with* $p \ge 2$. *Then we have*

$$I_n(u,v) = \sum_{1 \le i < j \le n} \left[\int_{-\infty}^{u} h(x, Y_j) 1_{\{Y_j \le v\}} F(dx) + \int_{-\infty}^{v} h(X_i, y) 1_{\{X_i \le u\}} G(dy) - \int_{-\infty}^{u} \int_{-\infty}^{v} h(x,y) F(dx) G(dy) \right] + R_n(u,v),$$

where R_n *satisfies* (1.1) *and the h-integral in the bound is taken w.r.t.* $F \otimes G$. *The assertion of Theorem* 1.2 *also extends to the present case.*

REMARK. The results of this section may be extended to U-statistic processes of degree $m > 2$, but proofs become more complicated and the notation even more cumbersome. As far as applications are concerned, however, the case $m = 2$ is by far the most important one.

We end this section by presenting five examples to which the theorems may be applied. For these we remark that in the formulation of the previous results, the point infinity could also be included in the parameter set. What matters is that the parameter sets of the coordinate spaces need to be linearly ordered.

EXAMPLE 1.1 (Censored data). In the random censorship model the actually observed data are $Z_i = \min(X_i, C_i)$ and $\delta_i = 1_{\{X_i \le C_i\}}$, where X_i is the variable of interest (the lifetime), which is at risk of being censored by C_i, the censoring variable. For estimation of the cumulative hazard function of X, a crucial role is played by the (one-parameter) process

$$\sum_{1 \le i < j \le n} \frac{\delta_i 1_{\{Z_j > Z_i\}}}{(1-H)^2(Z_i)} 1_{\{Z_i \le u\}} = I_n(u), \qquad u \in \mathbb{R}.$$

Here H is the d.f. of Z_i. If we introduce

$$Y_i = \begin{cases} Z_i, & \text{if } \delta_i = 1, \\ \infty, & \text{if } \delta_i = 0, \end{cases}$$

then

$$\delta_i 1_{\{Z_i \le u\}} = 1_{\{Y_i \le u\}}$$

and, therefore,

$$I_n(u) = \sum_{1 \le i < j \le n} \frac{1_{\{Z_j > Y_i\}}}{(1-H)^2(Y_i)} 1_{\{Y_i \le u\}} = I_n(u, \infty)$$

for an appropriate kernel h. The fact that Y_i is an extended random variable is of no importance to us. The theorems have been formulated for real variables just for the sake of convenience, but may be generalized easily to the foregoing setup. This example is discussed in greater detail in Stute (1994).

EXAMPLE 1.2 (Truncated data). Here one observes (X_i, Y_i) only if $Y_i \le X_i$. Though originally X_i is assumed independent of Y_i, the actually observed pair has dependent components. For estimation of the cumulative hazard function the following process constitutes a crucial part in the analysis:

$$I_n(u) = \sum_{1 \le i < j \le n} \frac{1_{\{Y_j \le X_i \le X_j\}}}{C^2(X_i)} 1_{\{X_i \le u\}}$$

for some particular function C. Obviously I_n may be decomposed into two parts, each of which is of the type as discussed in Theorem 1.5, with $v = \infty$. See Stute (1993) for a thorough discussion of this example.

EXAMPLE 1.3 (Two samples). In the situation of Theorem 1.3, the Wilcoxon two-sample rank test for H_0: $F = G$ versus H_1: $F = G(\cdot - \Delta)$ is based on the U-statistic

$$T_{nm} = \sum_{i=1}^{n} \sum_{j=1}^{m} h(X_i, Y_j),$$

with $h(x,y) = 1_{\{x \le y\}}$. In our previous notation,

$$T_{nm} = nmU_{nm}(\infty, \infty).$$

We may now consider the associated process U_{nm} in order to construct tests for H_0 versus H_1, which are based on the whole of U_{nm} rather than only T_{nm}. It would be interesting to compare the power of these tests with that of the standard Wilcoxon test.

EXAMPLE 1.4 (Trimmed U-statistics). Gijbels, Janssen and Veraverbeke (1988) investigated so-called trimmed U-statistics,

$$n(n-1)U_n^0(\alpha, \beta) = \sum_{i=1}^{[n\alpha]} \sum_{j=1}^{[n\beta]} h(X_{i:n}, X_{j:n}),$$

where $X_{1:n} \le \cdots \le X_{n:n}$ denotes the ordered sample.

Since

$$n(n-1)U_n^0(\alpha,\beta) = \sum_{i=1}^{n}\sum_{j=1}^{n} h(X_i, X_j)1_{\{X_i \le F_n^{-1}(\alpha), X_j \le F_n^{-1}(\beta)\}},$$

we see that $U_n^0(\alpha,\beta)$ is related to $U_n(F_n^{-1}(\alpha), F_n^{-1}(\beta))$, neglecting the sum over $i = j$ for a moment. Observe that $u = F_n^{-1}(\alpha)$ and $v = F_n^{-1}(\beta)$ are random in this case. The (uniform) representation of U_n in terms of a (simple) sum of independent random processes together with their tightness in the two-dimensional Skorokhod space allows for a simple analysis of $U_n^0(\alpha,\beta)$, not just for a fixed (α,β) [as done in Gijbels, Janssen and Veraverbeke (1988)], but as a process in (α,β). Details are omitted.

U-statistic processes also occur in the analysis of linear rank statistics. We only mention the possibility of representing a linear signed rank statistic (up to an error term) as a sum of i.i.d. random processes [cf. Sen (1981), Theorem 5.4.2].

EXAMPLE 1.5 (Linear signed rank statistics). For a sample $X_1,\dots,X_n$ and a proper score function φ, it is required to represent the double sum

$$\sum_{1\le i \ne j \le n} \varphi(X_i)1_{\{|X_j| \le X_i\}}1_{\{0 \le X_i \le x\}}.$$

We see that Theorem 1.5 applies with $Y_j = |X_j|$.

2. Proofs. We may and do assume without loss of generality that the X's are uniformly distributed on the unit interval. Otherwise, use the representation $X_i = F^{-1}(U_i)$ in terms of a uniform U_i and the quantile function of F, upon replacing u and v by $F(u)$ and $F(v)$, respectively, and $h(x,y)$ by the transformed kernel $h(F^{-1}(x), F^{-1}(y))$.

In the following we shall always write U_i instead of X_i. In order to prove Theorem 1.1, we start from the decomposition

$$\begin{aligned} I_n(u,v) &= \sum_{1\le i<j\le n} h(U_i, U_j)1_{\{U_i \le u, U_j \le v\}} \\ &= \sum_{1\le i<j\le n}\biggl[\int_0^u h(s, U_j)1_{\{U_j \le v\}}\,ds + \int_0^v h(U_i, t)1_{\{U_i \le u\}}\,dt \\ &\qquad\qquad - \int_0^u\int_0^v h(s,t)\,ds\,dt\biggr] \\ &\quad + \beta_n(u,v) - \gamma_n(u,v) - \delta_n(u,v) + \varepsilon_n(u,v), \end{aligned}$$

where

$$\beta_n(u,v) = \sum_{1\le i<j\le n} \beta_{ij}(u,v),$$
$$\gamma_n(u,v) = \sum_{1\le i<j\le n} \gamma_{ij}(u,v),$$
$$\delta_n(u,v) = \sum_{1\le i<j\le n} \delta_{ij}(u,v),$$
$$\varepsilon_n(u,v) = \sum_{1\le i<j\le n} \varepsilon_{ij}(u,v),$$

with

$$\begin{aligned}\beta_{ij}(u,v) &= h(U_i,U_j)1_{\{U_i\le u, U_j\le v\}} - \int_0^u \frac{1_{\{U_i\ge s\}}}{1-s} h(s,U_j)1_{\{U_j\le v\}}\,ds \\ &\quad - \int_0^v \frac{1_{\{U_j\ge t\}}}{1-t} h(U_i,t)1_{\{U_i\le u\}}\,dt + \int_0^u\int_0^v \frac{1_{\{U_i\ge s\}}}{1-s}\frac{1_{\{U_j\ge t\}}}{1-t} h(s,t)\,ds\,dt, \\ \gamma_{ij}(u,v) &= \int_0^u \frac{1_{\{U_i\le s\}}-s}{1-s}\left[h(s,U_j)1_{\{U_j\le v\}} - \int_0^v \frac{1_{\{U_j\ge t\}}}{1-t} h(s,t)\,dt\right] ds, \\ \delta_{ij}(u,v) &= \int_0^v \frac{1_{\{U_j\le t\}}-t}{1-t}\left[h(U_i,t)1_{\{U_i\le u\}} - \int_0^u \frac{1_{\{U_i\ge s\}}}{1-s} h(s,t)\,ds\right] dt\end{aligned}$$

and

$$\varepsilon_{ij}(u,v) = \int_0^u\int_0^v \frac{1_{\{U_i\le s\}}-s}{1-s}\frac{1_{\{U_j\le t\}}-t}{1-t} h(s,t)\,ds\,dt.$$

Hence, the remainder term in Theorem 1.1 is given by

$$R_n(u,v) = \beta_n(u,v) - \gamma_n(u,v) - \delta_n(u,v) + \varepsilon_n(u,v).$$

As is well known from univariate empirical process theory, the bracketed quantity in γ_{ij} constitutes, for fixed s, the martingale part in the Doob–Meyer decomposition of $h(s,U_j)1_{\{U_j\le v\}}$; similarly for δ_{ij}. This observation led us to consider β_{ij} as the crucial martingale part in the decomposition of the two-parameter process $h(U_i,U_j)1_{\{U_i\le u, U_j\le v\}}$.

Furthermore, to cope with the dependence of the summands of β_n, we consider the associated "sliced processes"

$$\beta_n^k(u,v) \equiv \sum_{1\le i<k\le j\le n} \beta_{ij}(u,v), \qquad 2\le k\le n.$$

Sliced processes have the advantage that the variables appearing in the first coordinate are independent of those in the second component. This fact will guarantee that for a proper filtration to be introduced, the crucial conditional independence property (F4) of Cairoli and Walsh (1975) will always be satisfied.

For this, let

$$\mathcal{F}_{u,v}^k = \sigma\big(1_{\{U_i\le s\}},\ 1_{\{U_j\le t\}}, 1\le i<k, k\le j\le n, s\le u, t\le v\big).$$

Then β_n^k is adapted to the filtration $\mathcal{F}^k$. Moreover, we have the following lemma.

LEMMA 2.1. *For each* $2 \le k \le n$:

(a) β_n^k *is a (centered strong) martingale w.r.t. the filtration* $\mathcal{F}^k$.
(b) *For each* (u,v), $\beta_n^k(u,v)$ *is a degenerate (zero mean) U-statistic.*

PROOF. For (a) we have to verify that the conditional expectation of an increment over a rectangle given the past equals zero. By independence, we may restrict ourselves to a particular pair $i < k \le j$. For $u < u'$ and $v < v'$, write $R = (u,u'] \times (v,v']$. Check that, by independence of U_i and U_j,

$$\begin{aligned}
&\mathbb{E}\big[h(U_i,U_j)1_{\{(U_i,U_j)\in R\}} \mid U_i \le u, U_j\big] \\
&= \mathbb{E}\bigg[\int_{(u,u']} \frac{1_{\{U_i \ge s\}}}{1-s} h(s,U_j) 1_{\{v < U_j \le v'\}}\, ds \mid U_i \le u, U_j\bigg] \\
&= \frac{1_{\{v < U_j \le v', U_i > u\}}}{1-u} \int_{(u,u']} h(s,U_j)\, ds
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{E}\bigg[\int_{(v,v']} \frac{1_{\{U_j \ge t\}}}{1-t} h(U_i,t) 1_{\{u < U_i \le u'\}}\, dt \mid U_i \le u, U_j\bigg] \\
&= \mathbb{E}\bigg[\int\!\!\int_R \frac{1_{\{U_i \ge s\}}}{1-s} \frac{1_{\{U_j \ge t\}}}{1-t} h(s,t)\, ds\, dt \mid U_i \le u, U_j\bigg] \\
&= \frac{1_{\{U_i > u\}}}{1-u} \int\!\!\int_R h(s,t) \frac{1_{\{U_j \ge t\}}}{1-t}\, ds\, dt.
\end{aligned}$$

Conditional expectations w.r.t. $\sigma(U_j \le v, U_i)$ are dealt with similarly. A Markov-argument completes the proof of (a); (b) is straightforward. □

LEMMA 2.2. *For any* $0 \le u_0, v_0 < 1$ *and* $p > 1$,

$$\mathbb{E}\Bigg[\sup_{\substack{0 \le u \le u_0 \\ 0 \le v \le v_0}} |\beta_n^k(u,v)|^p\Bigg] \le \left(\frac{p}{p-1}\right)^{2p} \sup_{\substack{0 \le u \le u_0 \\ 0 \le v \le v_0}} \mathbb{E}\Big[|\beta_n^k(u,v)|^p\Big].$$

PROOF. Follows from Lemma 2.1(a) and, for example, Theorem 1.2(b) of Cairoli and Walsh (1975). □

We now bound the right-hand side in Lemma 2.2. Fix some (u,v) and put, for $k \le r \le n$,

$$S_r = \sum_{\substack{1 \le i < k \\ k \le j \le r}} \beta_{ij}(u,v), \qquad \mathcal{F}_r = \sigma(U_1, \dots, U_{k-1}, \dots, U_r).$$

Clearly, S_r is adapted to $\mathcal{F}_r$. Moreover,

$$\begin{aligned}\mathbb{E}\left[S_r \mid \mathcal{F}_{r-1}\right] &= S_{r-1} + \sum_{1 \le i < k} \mathbb{E}\left[\beta_{ir}(u,v) \mid \mathcal{F}_{r-1}\right] \\ &= S_{r-1} \quad \text{by Lemma 2.1(b).}\end{aligned}$$

Hence $(S_r, \mathcal{F}_r)_{k \le r \le n}$ is a martingale, with $S_n = \beta_n^k(u,v)$. Set

$$D_r = \sum_{1 \le i < k} \beta_{ir}(u,v), \qquad k \le r \le n.$$

From Burkholder's inequality [cf. Chow and Teicher (1978), Theorem 1, page 384],

$$\begin{aligned}\mathbb{E}\left[\left|\beta_n^k(u,v)\right|^p\right] &\le B_p^p\, \mathbb{E}\left[\left|\sum_{r=k}^n D_r^2\right|^{p/2}\right] \\ &\le B_p^p (n-k+1)^{p/2-1} \sum_{r=k}^n \mathbb{E}|D_r|^p \quad \text{if } p \ge 2,\end{aligned}$$

where

$$B_p = 18 p^{3/2}/(p-1)^{1/2}.$$

A similar argument now yields an upper bound also for $\mathbb{E}|D_r|^p$. Set, for $1 \le i < k$ and $k \le r \le n$,

$$T_i = \sum_{j=1}^i \beta_{jr}(u,v), \qquad \mathcal{G}_i = \sigma(U_1, \dots, U_i, U_r).$$

Then T_i is adapted to $\mathcal{G}_i$. Because of

$$\mathbb{E}\left[\beta_{ir}(u,v) \mid \mathcal{G}_{i-1}\right] = 0 \quad \text{[cf. Lemma 2.1(b)]},$$

$(T_i, \mathcal{G}_i)_{1 \le i < k}$ is a martingale. Putting $E_i = \beta_{ir}(u,v)$ we get as before, since $T_{k-1} = D_r$,

$$\mathbb{E}|D_r|^p \le B_p^p (k-1)^{p/2-1} \sum_{i=1}^{k-1} \mathbb{E}|E_i|^p \quad \text{if } p \ge 2.$$

In summary, we therefore have, if $p \ge 2$,

$$\mathbb{E}\left[\left|\beta_n^k(u,v)\right|^p\right] \le B_p^{2p} (n-k+1)^{p/2-1} (k-1)^{p/2-1} \sum_{r=k}^n \sum_{i=1}^{k-1} \mathbb{E}\left[|\beta_{ir}(u,v)|^p\right].$$

Because the U's are i.i.d. we arrive at the following lemma.

LEMMA 2.3. *For* $p \geq 2$ *and* $0 \leq u_0, v_0 < 1$,

$$\mathbb{E}\left[\sup_{\substack{0 \leq u \leq u_0 \\ 0 \leq v \leq v_0}} \left|\beta_n^k(u,v)\right|^p\right] \leq C_1^p (n-k+1)^{p/2}(k-1)^{p/2}$$

with

$$C_1^p = \left(\frac{p}{p-1}\right)^{2p} B_p^{2p} \sup_{\substack{0 \leq u \leq u_0 \\ 0 \leq v \leq v_0}} \mathbb{E}\left[|\beta_{12}(u,v)|^p\right].$$

Lemma 2.3 will now be used to bound the supremum of

$$\beta_n(u,v) = \sum_{1 \leq i < j \leq n} \beta_{ij}(u,v).$$

For this note that the function $u \to u(1-u)$ on $[0,1]$ attains its maximum at $u = 1/2$. It follows that

$$(n-k+1)^{p/2}(k-1)^{p/2} \leq (n/2)^p, \qquad 2 \leq k \leq n.$$

Lemma 2.3 implies

$$\left\| \sup_{\substack{0 \leq u \leq u_0 \\ 0 \leq v \leq v_0}} \left|\beta_n^k(u,v)\right| \right\|_p \leq C_1 n/2,$$

where $\|\cdot\|_p$ denotes the norm in pth mean. To bound the sup of $|\beta_n|$, we first consider the case $n = 2^m$ with some integer m. Write

$$\beta_n(u,v) = \sum_{l=1}^{m} B_l(u,v)$$

with

$$B_l(u,v) = \sum_{k=1}^{2^{m-l}} B_{kl}(u,v)$$

and

$$B_{kl}(u,v) = \sum_{(k-1)2^l < i \leq (2k-1)2^{l-1} < j \leq k2^l} \beta_{ij}(u,v).$$

Note that since the B_{kl}'s are sums over pairwise disjoint index sets, each B_l is also a strong martingale. The Cairoli–Walsh (1975) inequality is therefore also applicable to B_l:

$$\mathbb{E}\left[\sup_{\substack{0 \leq u \leq u_0 \\ 0 \leq v \leq v_0}} |B_l(u,v)|^p\right] \leq \left[\frac{p}{p-1}\right]^{2p} \sup_{\substack{0 \leq u \leq u_0 \\ 0 \leq v \leq v_0}} \mathbb{E}\left[|B_l(u,v)|^p\right]. \tag{2.1}$$

To bound the right-hand side we shall use the fact that for some $\varepsilon = \varepsilon(p) > 0$,

$$\mathbb{E}|\xi + \eta|^p \le (1-\varepsilon)2^p \max\big(\mathbb{E}|\xi|^p, \mathbb{E}|\eta|^p\big),$$

if ξ and η are independent with $\mathbb{E}\xi = 0 = \mathbb{E}\eta$. The preceding inequality is trivial if p is an even integer and may be readily extended to a general $p \ge 2$. Iterative application to B_l together with (2.1) yields

$$\mathbb{E}\left[\sup_{\substack{0\le u\le u_0\\ 0\le v\le v_0}} |B_l(u,v)|^p\right] \le (1-\varepsilon)^{m-l} 2^{p(m-l)} C_1^p 2^{pl} = (1-\varepsilon)^{m-l} C_1^p n^p.$$

Conclude

$$\left\|\sup_{u,v} |\beta_n(u,v)|\right\|_p \le C_1 n \sum_{l=1}^{m} (1-\varepsilon)^{(m-l)/p} \le K C_1 n,$$

where $1 \le K = K(p) < \infty$ is a constant depending only on p, but not on h. In the following we verify the last inequality for a general n, with $2K$ rather than K.

LEMMA 2.4. *For $p \ge 2$ and $0 \le u_0, v_0 < 1$,*

$$\left\|\sup_{\substack{0\le u\le u_0\\ 0\le v\le v_0}} |\beta_n(u,v)|\right\|_p \le 2KC_1 n. \tag{2.2}$$

PROOF. We shall show (2.2) by induction on n. It is already known that the left-hand side of (2.2) is less than or equal to $KC_1 n$ if $n = 2^m$. In particular, (2.2) holds for $n = 2$. For a general n, denote with $\underline{n}$ the largest power of 2 contained in n. Since

$$\beta_n = \sum_{1\le i<\underline{n}\le j\le n} \beta_{ij} + \sum_{1\le i<j\le \underline{n}} \beta_{ij} + \sum_{\underline{n}\le i<j\le n} \beta_{ij},$$

Lemmas 2.2–2.4 together with the induction hypothesis imply

$$\left\|\sup_{u,v} |\beta_n(u,v)|\right\|_p \le C_1 \frac{n}{2} + KC_1\underline{n} + 2KC_1(n-\underline{n}). \tag{2.3}$$

By construction, $n \le 2\underline{n}$. Hence the right-hand side of (2.3) is less than or equal to $2KC_1 n$. The proof is complete. □

NOTE 2.1. From Hölder's inequality we obtain

$$C_1 \le \left(\frac{p}{p-1}\right)^2 B_p^2 \left[\int_0^{u_0}\int_0^{v_0} |h(s,t)|^p \,ds\,dt\right]^{1/p} \times \left[1 + \frac{1}{1-u_0} + \frac{1}{1-v_0} + \frac{1}{(1-u_0)(1-v_0)}\right].$$

Next we introduce the process

$$\gamma_n^k(u,v) = \sum_{1 \le i < k \le j \le n} \gamma_{ij}(u,v),$$

where as before

$$\gamma_{ij}(u,v) = \int_0^u \frac{1_{\{U_i \le s\}} - s}{1-s} \left[h(s, U_j) 1_{\{U_j \le v\}} - \int_0^v \frac{1_{\{U_j \ge t\}}}{1-t} h(s,t)\, dt \right] ds.$$

Set

$$\mathcal{F}_v^k = \sigma\big(U_1, \dots, U_{k-1}, 1_{\{U_j \le t\}}, k \le j \le n, t \le v\big).$$

For each u, $\gamma_n^k(u, \cdot)$ is a martingale w.r.t. the filtration $\mathcal{F}^k$. Consequently,

$$\sup_{0 \le u \le u_0} \left|\gamma_n^k(u, \cdot)\right|$$

is a (nonnegative) submartingale. From Doob's inequality,

$$\mathbb{E}\left[\sup_{\substack{0 \le u \le u_0 \\ 0 \le v \le v_0}} \left|\gamma_n^k(u,v)\right|^p \right] \le \left(\frac{p}{p-1}\right)^p \mathbb{E}\left[\sup_{0 \le u \le u_0} \left|\gamma_n^k(u,v_0)\right|^p \right].$$

Now for $0 \le u \le u_0 < 1$,

$$\begin{aligned} |\gamma_n^k(u,v_0)| \le{}& \frac{(k-1)\|F_{k-1} - Id\|_\infty}{1-u_0} \\ &\times \int_0^{u_0} \left| \sum_{k \le j \le n} \left[h(s,U_j) 1_{\{U_j \le v_0\}} - \int_0^{v_0} \frac{1_{\{U_j \ge t\}}}{1-t} h(s,t)\, dt \right] \right| ds, \end{aligned}$$

with F_{k-1} denoting the empirical d.f. of $U_1, \dots, U_{k-1}$. From the Dvoretzky–Kiefer–Wolfowitz (1956) bound,

$$\mathbb{E}\left[\|F_{k-1} - Id\|_\infty^p\right] \le C_0^p (k-1)^{-p/2},$$

where C_0 is a universal constant. Since F_{k-1} is independent of $\sigma(U_k, \dots, U_n)$, an application of Hölder's inequality in connection with Burkholder's inequality (the Marcinkiewicz–Zygmund version for independent summands suffices here) yields, for $p \ge 2$,

$$\begin{aligned} &\mathbb{E}\left[\sup_{0 \le u \le u_0} \left|\gamma_n^k(u,v_0)\right|^p \right] \\ &\quad \le \frac{C_0^p B_p^p}{(1-u_0)^p} (k-1)^{p/2} (n-k+1)^{p/2} \\ &\quad\quad \times \int_0^{u_0} \int_0^1 \left| h(s,w) 1_{\{w \le v_0\}} - \int_0^{v_0} \frac{1_{\{w \ge t\}}}{1-t} h(s,t)\, dt \right|^p dw\, ds. \end{aligned}$$

This bound is similar to the one in Lemma 2.3. Also the arguments leading to Lemma 2.4 may be carried over so that in summary we get the next lemma.

LEMMA 2.5. *For*

$$\gamma_n(u,v) = \sum_{1\leq i<j\leq n} \gamma_{ij}(u,v)$$

and $p \geq 2$,

$$\left\| \sup_{\substack{0\leq u\leq u_0 \\ 0\leq v\leq v_0}} |\gamma_n(u,v)| \right\|_p \leq 2KC_2 n.$$

NOTE 2.2. Again, by Hölder, we obtain

$$C_2 \leq \frac{pC_0B_p}{(p-1)(1-u_0)}\left(\int_0^{u_0}\int_0^{v_0} |h(s,t)|^p\, ds\, dt\right)^{1/p}\left(1+\frac{1}{1-v_0}\right).$$

The proof of Lemma 2.6 follows the same pattern as that of Lemma 2.5.

LEMMA 2.6. *Recall*

$$\delta_n(u,v) = \sum_{1\leq i<j\leq n} \delta_{ij}(u,v),$$

with

$$\delta_{ij}(u,v) = \int_0^v \frac{1_{\{U_j\leq t\}}-t}{1-t}\left[h(U_i,t)1_{\{U_i\leq u\}} - \int_0^u \frac{1_{\{U_i\geq s\}}}{1-s}h(s,t)\,ds\right]dt.$$

Then, for $p \geq 2$,

$$\left\| \sup_{\substack{0\leq u\leq u_0 \\ 0\leq v\leq v_0}} |\delta_n(u,v)| \right\|_p \leq 2KC_3 n.$$

NOTE 2.3. The constant C_3 satisfies

$$C_3 \leq \frac{pC_0B_p}{(p-1)(1-v_0)}\left(\int_0^{u_0}\int_0^{v_0} |h(s,t)|^p\, ds\, dt\right)^{1/p}\left(1+\frac{1}{1-u_0}\right).$$

Our final result is concerned with the process

$$\varepsilon_n(u,v) = \sum_{1\leq i<j\leq n} \varepsilon_{ij}(u,v),$$

where as before

$$\varepsilon_{ij}(u,v) = \int_0^u \int_0^v \frac{1_{\{U_i \le s\}} - s}{1-s} \frac{1_{\{U_j \le t\}} - t}{1-t} h(s,t)\,ds\,dt.$$

Denote

$$\varepsilon_n^k(u,v) = \sum_{1 \le i < k \le j \le n} \varepsilon_{ij}(u,v).$$

As in the proof of Lemma 2.5, we may apply the Dvoretzky–Kiefer–Wolfowitz bound to get for $0 \le u_0, v_0 < 1$ and $p \ge 2$,

$$\mathbb{E}\left[\sup_{\substack{0 \le u \le u_0 \\ 0 \le v \le v_0}} \left|\varepsilon_n^k(u,v)\right|^p\right] \le C_4^p (k-1)^{p/2}(n-k+1)^{p/2}.$$

LEMMA 2.7. *For $0 \le u_0, v_0 < 1$ and $p \ge 2$,*

$$\left\| \sup_{\substack{0 \le u \le u_0 \\ 0 \le v \le v_0}} |\varepsilon_n(u,v)| \right\|_p \le 2KC_4 n.$$

NOTE 2.4. The constant C_4 satisfies

$$C_4 \le \frac{C_0^2}{(1-u_0)(1-v_0)} \left(\int_0^{u_0} \int_0^{v_0} |h(s,t)|^p \,ds\,dt \right)^{1/p}.$$

We are now in the position to give the proof.

PROOF OF THEOREM 1. We have already seen that

$$\begin{aligned} &\sum_{1 \le i < j \le n} h(U_i, U_j) 1_{\{U_i \le u, U_j \le v\}} \\ &\quad = \sum_{1 \le i < j \le n} \left[\int_0^u h(s, U_j) 1_{\{U_j \le v\}}\,ds + \int_0^v h(U_i, t) 1_{\{U_i \le u\}}\,dt \right. \\ &\qquad\qquad \left. - \int_0^u \int_0^v h(s,t)\,ds\,dt \right] \\ &\qquad + \beta_n(u,v) - \gamma_n(u,v) - \delta_n(u,v) + \varepsilon_n(u,v). \end{aligned} \tag{2.4}$$

When both u_0 and v_0 are less than 1/2, the assertion of the theorem is an immediate consequence of Lemmas 2.4–2.7 and Notes 2.1–2.4. If at least one of them is larger than 1/2, the indicator $1_{\{U_i \le u, U_j \le v\}}$ on the left-hand side of (2.4) needs to be decomposed into (when both $u, v \ge 1/2$)

$$\begin{aligned} &1_{\{U_i \le 1/2, U_j \le 1/2\}} + 1_{\{1/2 < U_i \le u, 1/2 < U_j \le v\}} \\ &\quad + 1_{\{U_i \le 1/2, 1/2 < U_j \le v\}} + 1_{\{1/2 < U_i \le u, U_j \le 1/2\}}. \end{aligned}$$

The theorem has been proved already for each of the four U-statistic processes pertaining to the above indicators (after a possible time change), upon observing that the intervals $[0, 1/2), (1/2, u]$ and $(1/2, v]$ all have length less than or equal to 1/2. □

PROOF OF THEOREM 1.2. Since for each $(u,v), R_n(u,v)/(n(n-1))$ is a sequence of normalized U-statistics, it constitutes a reverse-time martingale w.r.t. the decreasing sequence of σ-fields

$$\mathcal{F}_n = \sigma(X_{i:n}: 1 \le i \le n, X_{n+1}, \ldots), \qquad n \ge 2.$$

Hence

$$\sup_{\substack{u \le u_0 \\ v \le v_0}} \left[|R_n(u,v)|/\big(n(n-1)\big)\right]^p$$

is a nonnegative reverse-time submartingale. Set $\varepsilon_n = n(\ln n)^{-\delta}$. By Borel–Cantelli it suffices to show that for each $\varepsilon > 0$,

$$\sum_{k=0}^{\infty} \mathbb{P}\left(\varepsilon_{2^{k+1}} \max_{2^k \le n \le 2^{k+1}} \sup_{\substack{u \le u_0 \\ v \le v_0}} |R_n(u,v)|/n(n-1) \ge \varepsilon\right) < \infty. \tag{2.5}$$

By Doob's maximal inequality the last probability is less than or equal to

$$\big(2^k(2^k-1)\big)^{-p} \varepsilon^{-p} \varepsilon_{2^{k+1}}^p \mathbb{E}\left[\sup_{\substack{u \le u_0 \\ v \le v_0}} |R_{2^k}(u,v)|^p\right] = O[k^{-p\delta}],$$

in view of Theorem 1.1. Since $p\delta > 1$, the proof of (2.5) is complete. □

NOTE 2.5. The last proof utilized ideas from Serfling [(1980), Section 5.3.3].

PROOF OF THEOREM 1.3. The arguments are almost the same as for Theorem 1.1. In fact, some simplifications are possible. Because the two samples are independent, introduction of $\beta_n^k, \gamma_n^k, \ldots$ is superfluous. Rather, we may deal with $\beta_n, \gamma_n, \ldots$ directly. For example, after a reduction to the uniform case,

$$\beta_n(u,v) \equiv \sum_{i=1}^{n} \sum_{j=1}^{m} \beta_{ij}(u,v)$$

with

$$\begin{aligned} \beta_{ij}(u,v) = {} & h(U_i, V_j) 1_{\{U_i \le u, V_j \le v\}} \\ & - \int_0^u \frac{1_{\{U_i \ge s\}}}{1-s} h(s, V_j) 1_{\{V_j \le v\}}\, ds \\ & - \int_0^v \frac{1_{\{V_j \ge t\}}}{1-t} h(U_i, t) 1_{\{U_i \le u\}}\, dt \\ & + \int_0^u \int_0^v \frac{1_{\{U_i \ge s\}}}{1-s} \frac{1_{\{V_j \ge t\}}}{1-t} h(s,t)\, ds\, dt \end{aligned}$$

is a strong martingale w.r.t.

$$\mathcal{F}_{u,v} = \sigma\big(1_{\{U_i \le s\}}, 1_{\{V_j \le t\}},\ 1 \le i \le n,\ 1 \le j \le m, s \le u, t \le v\big).$$

So a direct application of the maximal inequality for two-parameter strong martingales and the Burkholder inequality is possible, to the effect that the arguments leading to Lemma 2.4 are superfluous. Similarly, for γ_n, δ_n and ε_n. □

PROOF OF THEOREM 1.4. With only minor modifications the proof is the same as for Theorem 1.2. □

PROOF OF THEOREM 1.5. This follows almost verbatim the arguments for Theorem 1.1. The reduction to the uniform case incorporates a sequence (U_i, V_i) of i.i.d. random vectors in the unit square with uniform marginals such that

$$X_i = F^{-1}(U_i), \qquad Y_i = G^{-1}(V_i)$$

for which the joint distribution of each (U_i, V_i) is the copula C of H, that is,

$$H = C(F, G).$$

The processes $\beta_{ij}, \gamma_{ij}, \dots$ are defined similarly, with U_j replaced by V_j. Because of possible dependencies between U_i and V_i, β_n, for example, has to be sliced again by introducing the β_n^k's. □

Acknowledgment. The constructive comments of a referee, which led to an improvement of the original Lemma 2.4, are highly appreciated.

REFERENCES

BERK, R. H. (1966). Limiting behavior of posterior distributions when the model is incorrect. *Ann. Math. Statist.* **37** 51–58.

CAIROLI, R. and WALSH, J. B. (1975). Stochastic integrals in the plane. *Acta Math.* **134** 111–183.

CHAO, M. T. and LO, S. H. (1988). Some representations of the nonparametric maximum likelihood estimators with truncated data. *Ann. Statist.* **16** 661–668.

CHOW, Y. S. and TEICHER, H. (1978). Probability theory. In *Independence, Interchangeability, Martingales.* Springer, Berlin.

DEHLING, H., DENKER, M. and PHILIPP, W. (1986). A bounded law of the iterated logarithm for Hilbert space valued martingales and its application to *U*-statistics. *Probab. Theory Related Fields* **72** 111–131.

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* **23** 121–134.

DE LA PEÑA, V. H. (1992). Decoupling and Khintchine's inequalities for *U*-statistics. *Ann. Probab.* **20** 1877–1892.

DENKER, M. (1985). *Asymptotic Distribution Theory in Nonparametric Statistics.* Vieweg, Braunschweig-Wiesbaden.

DOZZI, M. (1981). On the decomposition and integration of two-parameter stochastic processes. *Processus Aléatoires à Deux Indicies. Lecture Notes in Math.* **863** 162–171. Springer, Berlin.

DVORETZKY, A., KIEFER, J. and WOLFOWITZ, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Statist.* **27** 642–669.

GIJBELS, I., JANSSEN, P. and VERAVERBEKE, N. (1988). Weak and strong representations for trimmed U-statistics. *Probab. Theory Related Fields* **77** 179–194.

HOEFFDING, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* **19** 293–325.

LEE, A. J. (1990). *U-Statistics. Theory and Practice.* Dekker, New York.

LO, S. H., MACK, Y. P. and WANG J. L. (1989). Density and hazard rate estimation for censored data via strong representation of the Kaplan–Meier estimator. *Probab. Theory Related Fields* **80** 461–473.

LO, S. H. and SINGH, K. (1986). The product-limit estimator and the bootstrap: Some asymptotic representations. *Probab. Theory Related Fields* **71** 419–435.

MAJOR, P. (1988). On the tail behaviour of the distribution function of multiple stochastic integrals. *Probab. Theory Related Fields* **78** 419–435.

MAJOR, P. and REJTŐ, L. (1988). Strong embedding of the estimator of the distribution function under random censorship. *Ann. Statist.* **16** 1113–1132.

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

RANDLES, R. H. and WOLFE, D. A. (1979). *Introduction to the Theory of Nonparametric Statistics.* Wiley, New York.

SEN, P. K. (1981). *Sequential Nonparametrics.* Wiley, New York.

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

STUTE, W. (1993). Almost sure representations of the product-limit estimator for truncated data. *Ann. Statist.* **21** 146–156.

STUTE, W. (1994). Strong and weak representations of cumulative hazard function and Kaplan–Meier estimators on increasing sets. *J. Statist. Plann. Inference* **42**.

MATHEMATICS INSTITUTE
UNIVERSITY OF GIESSEN
ARNDTSTRASSE 2
D-35392 GIESSEN
GERMANY