PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 14, Fasc. 1 (1993), pp. 143–155

U-FUNCTIONS OF CONCOMITANTS OF ORDER STATISTICS

BY

WINFRIED STUTE (GIESSEN)

Abstract. Let (X_i, Y_i), $1 \leqslant i \leqslant n$, be i.i.d. $\boldsymbol{R}^{1+d}$-valued random vectors. Denote by $Y_{[i:n]}$ the Y-value associated with the i-th order statistic $X_{i:n}$. Concomitants of order statistics may be used to exhibit special features of the dependence structure between X_i and Y_i. We prove various distributional limit theorems for so-called U-functions (of degree two) of concomitants. The method of proof is based on a new conditional projection lemma.

1. Introduction and main results. The main subject of this paper* is to provide new results for so-called U-functional of concomitants of order statistics. To be precise, assume that (X_i, Y_i), $1 \leqslant i \leqslant n$, is a sequence of independent identically distributed $\boldsymbol{R}^{1+d}$-valued random vectors on some probability space $(\Omega, \mathscr{A}, \boldsymbol{P})$. Denote by $X_{1:n} \leqslant \ldots \leqslant X_{n:n}$ the order statistics of the X-sample. The Y-vector $Y_{[i:n]}$ pertaining to the i-th order statistic is called the *i-th concomitant*. Concomitants of order statistics rather than the Y's themselves play an important role, e.g., when the X-random variables are type-II censored, i.e., when the X's are time-sequentially observed up to $X_{\langle nt \rangle:n}$, where $0 < t < 1$, and $\langle \cdot \rangle$ denotes the integer part of $\cdot$. In this case, $Y_1, \ldots, Y_n$ are not all available, and statistical inference about the Y's may be only based on $Y_{[1:n]}, \ldots, Y_{[\langle nt \rangle:n]}$. What is more, even if all pairs $(X_i, Y_i)_i$ are observed, grouping the X's and analyzing the within-group Y's amounts to studying certain (functions of) concomitants (see, e.g., [10]). The most familiar theoretical function describing mean outputs of Y given some (quantile-) side condition on X is the so-called *Lorenz curve*, as well as the closely related total time on test transform (see, e.g., [4]). A general account of the distributional properties of concomitants (of order statistics) was given by Yang [12]. L-statistics of concomitants were studied by Sandstroem [8] and Yang [13]. An interesting invariance principle for the partial sum process of concomitants was derived by Bhattacharya [1]. Applications to testing about a regression function are due to Bhattacharya [2]; see also [3] for a comprehensive review of results available so far.

* This work has been partially supported by the "Deutsche Forschungsgemeinschaft".

In this paper we extend Bhattacharya's [1] result to U-functions of concomitants. For this, let h be any symmetric U-kernel (of degree two), and set, for $0 \leqslant t \leqslant 1$ and $n \geqslant 2$,

$$Y_n(t) = \frac{1}{n(n-1)} \sum_{1 \leqslant i \neq j \leqslant \langle nt \rangle} h(Y_{[i:n]}, Y_{[j:n]}),$$

for the partial sum process of U-type based on the (dependent) concomitants. For $t = 1$, $Y_n(1)$ becomes a familiar U-statistic of degree two, based on all of the Y's (see, e.g., [9]). For $t < 1$, $Y_n(t)$ is an estimator of

$$E[h(Y_1, Y_2) 1_{\{X_1 \leqslant F^{-1}(t), X_2 \leqslant F^{-1}(t)\}}],$$

where F denotes the distribution function of the X's and

$$F^{-1}(u) = \inf\{x \in \mathbf{R}: F(x) \geqslant u\}, \quad 0 < u < 1,$$

is its left-continuous inverse. In other words, the parameter of interest is the same as for classical U-statistics, up to the fact that we are only interested in the mean of $h(Y_1, Y_2)$ given that the pertaining X's fall below the t-quantile. Examples will be postponed to the end of this section.

Let $m(dy|x)$ denote a (regular) conditional distribution of Y given $X = x$. We know from [12] that conditionally on $X_{1:n}, \ldots, X_{n:n}$ the concomitants are independent and

$$\mathscr{L}(Y_{[i:n]}|X_{i:n} = x) = m(dy|x)$$

(see also [11]). Write, for $i \neq j$,

$$E_{ij} = \int\int h(x, y) m(dx|X_{i:n}) m(dy|X_{j:n}).$$

Then $h(Y_{[i:n]}, Y_{[j:n]}) - E_{ij}$, $i \neq j$, are centered conditionally on $\mathscr{F} \equiv \sigma(X_{r:n}: 1 \leqslant r \leqslant n)$. Consider, for $n \geqslant 2$, the process

$$S_n(t) = \frac{1}{n(n-1)} \sum_{1 \leqslant i \neq j \leqslant \langle nt \rangle} [h(Y_{[i:n]}, Y_{[j:n]}) - E_{ij}], \quad 0 \leqslant t \leqslant 1.$$

Theorem 1.1 below yields the asymptotic normality of $\sqrt{n} S_n(t)$ for $0 < t \leqslant 1$ fixed. The invariance principle is stated in Theorem 1.2 under some regularity assumptions on h. Our method of proof is different from that of Bhattacharya [1], who utilized a strong embedding argument. In contrast, we shall apply a conditional projection lemma (Section 3), which may be interesting in itself. Analyzing the projection $\hat{S}_n$ of S_n requires some consistency results for U-type Lorenz curves (Section 2). Asymptotic normality and the invariance principle for $\hat{S}_n$ are proved in Section 5, while the proofs of Theorems 1.1 and 1.2 are presented in Section 6.

We shall prove Theorem 1.1 under the assumption

$$Eh^2(Y_1, Y_2) < \infty. \tag{1.1}$$

It follows from (1.1) that E_{ij} as well as the functions f_1 and f_2 to be introduced now are well defined (almost surely):

$$f_1(a, b, c) = \int\int\int h(x, y)h(x, z)m(dx|a)m(dy|b)m(dz|c),$$
$$f_2(a, b, c) = \int\int h(x, y)m(dx|a)m(dy|b)\int\int h(x, y)m(dx|a)m(dy|c).$$

Also, let

$$g_1(a, b, c, d, e)$$
$$= \int\int\int\int\int h(x, y)h(x, z)h(x, u)h(x, v)m(dx|a)m(dy|b)m(dz|c)m(du|d)m(dv|e),$$

$a, b, c, d, e \in \mathbf{R}$.

1.1. THEOREM. *Under* (1.1), *for each* $0 \leqslant t \leqslant 1$,

$$\sqrt{n}\, S_n(t) \to \mathcal{N}(0, 4V(t)) \text{ in distribution,}$$

where $V(t) = V_1(t) - V_2(t)$ *and*

$$V_i(t) = \int_0^t\int_0^t\int_0^t f_i(F^{-1}(u), F^{-1}(v), F^{-1}(w))dudvdw, \quad i = 1, 2.$$

An alternative representation of V will be given in the lines following Lemma 4.1. As for an invariance principle, for $s \leqslant t$ set

$$K(s, t) = 4\int_0^s\int_0^s\int_0^t (f_1 - f_2)(F^{-1}(u), F^{-1}(v), F^{-1}(w))dudvdw.$$

1.2. THEOREM. *Assume that*

(1.2) f_1, f_2 *and* g_1 *are bounded.*

Then in distribution

$$\{\sqrt{n}\, S_n(t)\colon 0 \leqslant t \leqslant 1\} \to \{B(t)\colon 0 \leqslant t \leqslant 1\}$$

in the space $D[0, 1]$. *Here* B *is a continuous zero means Gaussian process with covariance function* K.

Condition (1.2) is satisfied if

(i) h is bounded or

(ii) the conditional distributions $m(dx\,|\,a)$ are dominated by some measure ν with Radon–Nikodym derivatives $f(x, a)$ such that (as, e.g., for f_1) the functions

$$(x, y, z) \to h(x, y)h(x, z)f(x, a)f(y, b)f(z, c)$$

are bounded in $L_1(\nu\otimes\nu\otimes\nu)$.

We only mention here that Theorem 1.2 also admits a bootstrap version. This will be needed if, for a particular h, the distribution (of a functional) of B is difficult to compute.

In the examples below, Y is assumed to be real valued.

1.3. EXAMPLE. If $h(x, y) = \frac{1}{2}(x-y)^2$, then $Y_n(t)/t^2$ is an estimator of the conditional variance $\mathrm{Var}(Y \mid X \leqslant F^{-1}(t))$ (provided that $F \circ F^{-1}(t) = t$). For multivariate Y's, a slight modification of this example yields an estimator of conditional covariances.

1.4. EXAMPLE. Put $h(x, y) = 1_{\{x+y>0\}}$. In classical nonparametrics this h is related to the Wilcoxon one-sample signed rank statistic designed for testing symmetry at zero. In the present (conditional) setup

$$\int\int h(x, y) m(dx \mid a) m(dy \mid b) = 1/2$$

under symmetry (and continuity). Suppose we want to test the hypothesis

$$\mathrm{H}_0\colon\ m(\cdot \mid a) \text{ is symmetric at zero on } [F^{-1}(t_1), F^{-1}(t_2)].$$

A test of H_0 may then be based on $S_n(t) - S_n(t_1)$, $t_1 \leqslant t \leqslant t_2$, with E_{ij} replaced by 1/2.

1.5. EXAMPLE. For bivariate $Y = (Y^1, Y^2)$, the expression

$$h(Y_i, Y_j) = \mathrm{sgn}[(Y_i^1 - Y_j^1)(Y_i^2 - Y_j^2)]$$

leads to a conditional version of Kendall's tau. This may be used to test the independence of Y^1 and Y^2 given $X \leqslant F^{-1}(t)$.

2. U-type Lorenz curves: consistency. For a distribution function (d.f.) F on the real line with existing nonvanishing expectation $\mu = \int xF(dx)$, the (theoretical) Lorenz curve is defined as

$$L(t) = \mu^{-1} \int_0^t F^{-1}(u)\,du.$$

In economics, when F may be interpreted as the income distribution of an individual from a given population, $L(t)$ represents the (normalized) mean income of an individual belonging to the lowest t-th fraction of income possessors. An empirical analogue of L is given by

$$L_n(t) = \mu_n^{-1} \int_0^t F_n^{-1}(u)\,du,$$

where F_n is the empirical d.f. of the observed data, and μ_n is the sample mean. A detailed study of L_n may be found [5] and [4]. Since F_n^{-1} admits a representation

$$F_n^{-1}(u) = F^{-1}(\bar{F}_n^{-1}(u)), \tag{2.1}$$

in which $\bar{F}_n$ is the empirical d.f. of a uniform sample, we may write

$$L_n(t) = \mu_n^{-1} \int_0^t h(\bar{F}_n^{-1}(u))\,du$$

with $h = F^{-1}$. For the purpose of this paper, we need to generalize L, resp. L_n, in two different directions. Firstly, more general (not necessarily monotone) h's are required. Secondly, functions h of k ($k \geqslant 2$) variables need to be considered. In view of (2.1), we may restrict ourselves to a uniform sample. So, let h be a measurable function defined on the (open) unit cube I^k satisfying

$$\int_{I^k} |h(\boldsymbol{u})| d\boldsymbol{u} < \infty, \quad \boldsymbol{u} = (u_1, \ldots, u_k). \tag{2.2}$$

Write

$$\mu = \int_{I^k} h(\boldsymbol{u}) d\boldsymbol{u}$$

and set (assuming $\mu \neq 0$)

$$L(t) = \mu^{-1} \int_0^t \ldots \int_0^t h(\boldsymbol{u}) d\boldsymbol{u}.$$

An empirical analogue of L is given by

$$L_n(t) = R_n(t)/R_n(1),$$

where

$$R_n(t) = n^{-k} \sum_{\substack{i_1 \neq \ldots \neq i_k \\ i_j \leqslant \langle nt \rangle}} h(X_{i_1:n}, \ldots, X_{i_k:n}).$$

Note that $R_n(1)$ is (up to a slight difference in the normalizing factor) a classical U-statistic. Also,

$$L_n(t) = R_n^{-1}(1) \int_0^t \ldots \int_0^t h \circ \bar{F}_n^{-1}(\boldsymbol{u}) 1_{A_n}(\boldsymbol{u}) d\boldsymbol{u}.$$

The set A_n is such that its complement has Lebesgue measure $O(1/n)$, and

$$h \circ \bar{F}_n^{-1}(\boldsymbol{u}) \equiv h(\bar{F}_n^{-1}(u_1), \ldots, \bar{F}_n^{-1}(u_k))$$

for short.

2.1. Lemma. *Under* (2.2), *with probability one*

$$\sup_{0 \leqslant t \leqslant 1} |L_n(t) - L(t)| \to 0.$$

Proof. We may assume without loss of generality that h is nonnegative, otherwise decompose h into its positive and negative parts. For $h \geqslant 0$, L_n and L are nondecreasing and continuous. By a usual uniformity argument (introducing appropriate grids), we only need to prove pointwise consistency. So, fix $0 < t < 1$ ($t = 0$ and $t = 1$ are trivial). Since $R_n(1) \to \mu$ with probability one, by the SLLN for U-statistics (cf. [9]) it suffices to show

$$\int_0^t \ldots \int_0^t [h(\boldsymbol{u}) - h \circ \bar{F}_n^{-1}(\boldsymbol{u})] 1_{A_n}(\boldsymbol{u}) d\boldsymbol{u} \to 0 \quad \boldsymbol{P}\text{-a.s.} \tag{2.3}$$

Since $\|\bar{F}_n^{-1} - \mathrm{Id}\| \to 0$ P-a.s., (2.3) is immediate for a uniformly continuous h. For a general h, choose a uniformly continuous function g such that, for given $\varepsilon > 0$,

$$\int_{I^k} |g-h|(u)du < \varepsilon,$$

which is possible by Lusin's theorem. Apply (2.3) to g. On the other hand,

$$\left|\int_0^t \cdots \int_0^t [g \circ \bar{F}_n^{-1}(u) - h \circ \bar{F}_n^{-1}(u)] 1_{A_n}(u)du\right| \leqslant \int_{I^k} |g-h| \circ \bar{F}_n^{-1}(u) 1_{A_n}(u)du$$
$$\to \int_{I^k} |g-h|(u)du < \varepsilon$$

by the SLLN for U-statistics. Since $\varepsilon > 0$ was arbitrary, this completes the proof. ∎

2.2. Remark. The results of this section may be easily extended to functions L_n of k variables, i.e., for which integration is taken over $[0, t_1] \times \ldots \times [0, t_k]$ with not necessarily equal $t_1, \ldots, t_k$. Also we have formulated Lemma 2.1 for the normalized Lorenz curve, though we shall only consider the nonnormalized estimators.

3. A conditional projection lemma. Let $Y_1, \ldots, Y_n$ be arbitrary random vectors and let S be any square-integrable statistic, i.e., a measurable function of the Y's. Also, let $\mathscr{F}$ be any sub-σ-field of the basic σ-field $\mathscr{A}$. We seek for a random variable L of the form

$$L = \sum_{i=1}^{n} Z_i, \tag{3.1}$$

where Z_i is $\sigma(Y_i, \mathscr{F})$-measurable, such that L approximates S well within the class of statistics satisfying (3.1). When the Y's are independent (and if formally we set $\mathscr{F} = \{\emptyset, \Omega\}$), Hájek [6] showed that the function L minimizing the L^2-distance to S is of the form

$$\hat{S} = \sum_{i=1}^{n} E(S \mid Y_i) - (n-1)E(S). \tag{3.2}$$

To motivate our conditional projection lemma, note that in our situation S will be a function of the concomitants, which are typically dependent. On the other hand, we know that the concomitants are conditionally independent given the order statistics. Consequently, it is likely that a proper approximation of S by functions L should allow for summands Z_i which are measurable Y_i, enlarged by $\mathscr{F} = \sigma(X_{j:n}, 1 \leqslant j \leqslant n)$.

A basic assumption throughout this section will be

$$E[E(S \mid Y_i, \mathscr{F}) \mid Y_j, \mathscr{F}] = E(S \mid \mathscr{F}) \quad \text{for } i \neq j. \tag{3.3}$$

3.1. LEMMA. *Under* $ES^2 < \infty$ *and* (3.3), *let*

$$\hat{S} = \sum_{i=1}^{n} E(S \mid Y_i, \mathscr{F}) - (n-1)E(S \mid \mathscr{F}).$$

Then the following holds:

(i) $E(\hat{S} \mid \mathscr{F}) = E(S \mid \mathscr{F})$;

(ii) $E[(S-\hat{S})^2 \mid \mathscr{F}] = \mathrm{Var}(S \mid \mathscr{F}) - \mathrm{Var}(\hat{S} \mid \mathscr{F})$;

(iii) *for any* L *of the form* (3.1),

$$E[(S-L)^2 \mid \mathscr{F}] = E[(S-\hat{S})^2 \mid \mathscr{F}] + E[(\hat{S}-L)^2 \mid \mathscr{F}],$$

i.e., $\hat{S}$ *minimizes the left-hand side.*

3.2. Remark. Recall that for Lemma 3.1 no independence assumption was required. On the other hand, if the Y's are independent and if we set $\mathscr{F} = \{\emptyset, \Omega\}$, then (3.3) is easily verified, and $\hat{S}$ reduces to (3.2).

Proof of Lemma 3.1. The proof is similar to that of Hájek [6], appropriately modified to meet the conditional setup. First, needless to say that $\hat{S}$ is of the form (3.1). Equality (i) is trivial, since $\mathscr{F} \subset \sigma(Y_i, \mathscr{F})$. Relation (ii) follows from (iii) if we set $L = E(S \mid \mathscr{F}) = E(\hat{S} \mid \mathscr{F})$. For (iii), assume $E(S \mid \mathscr{F}) = 0 = E(\hat{S} \mid \mathscr{F})$ w.l.o.g. We then have

$$E[(S-\hat{S})(\hat{S}-L) \mid \mathscr{F}] = \sum_{i=1}^{n} E\{(S-\hat{S})(E(S \mid Y_i, \mathscr{F}) - Z_i) \mid \mathscr{F}\}$$

$$= \sum_{i=1}^{n} E\{[E(S \mid Y_i, \mathscr{F}) - Z_i] E[S-\hat{S} \mid Y_i, \mathscr{F}] \mid \mathscr{F}\}.$$

From (3.3) we obtain

$$E[E(S \mid Y_i, \mathscr{F}) \mid Y_j, \mathscr{F}] = \begin{cases} E(S \mid \mathscr{F}) & \text{for } i \neq j, \\ E(S \mid Y_i, \mathscr{F}) & \text{for } i = j. \end{cases}$$

It follows that $E(\hat{S} \mid Y_i, \mathscr{F}) = E(S \mid Y_i, \mathscr{F})$, whence

$$E[(S-\hat{S})(\hat{S}-L) \mid \mathscr{F}] = 0,$$

and therefore we get (iii). ∎

3.3. Remark. Equality (ii) will be applied in the following way. Assume that as $n \to \infty$ the right-hand side converges to zero in probability. Then so does the left-hand side. By a conditional Chebyshev inequality (neglecting the dependence on n) for each $\varepsilon > 0$ we have

$$P(|S-\hat{S}| \geqslant \varepsilon \mid \mathscr{F}) \to 0 \text{ in probability.}$$

After integrating we get

$$P(|S-\hat{S}| \geqslant \varepsilon) \to 0 \quad \text{for each } \varepsilon > 0,$$

i.e.,

$$S-\hat{S}\to 0 \text{ in probability.}$$

Apart from the applications we have in mind in this paper, conditioning on $\mathscr{F}$ is always useful in other situations, when S contains awkward $\mathscr{F}$-measurable components.

4. U-functions of concomitants: variance and projection. In this section we compute the asymptotic variance of a standardized U-function of concomitants. So, let h be a symmetric U-kernel of degree two. Recall $S_n(t)$. Clearly,

$$(4.1)\qquad n\operatorname{Var}(S_n(t)\mid X_{r:n},\ 1\leqslant r\leqslant n)$$

$$=\frac{1}{n(n-1)^2}\sum_{\substack{i\neq j\\ k\neq m}} E\big[\big(h(Y_{[i:n]},Y_{[j:n]})-E_{ij}\big)\big(h(Y_{[k:n]},Y_{[m:n]})-E_{km}\big)\mid X_{r:n},\ 1\leqslant r\leqslant n\big],$$

where the summation always extends from 1 to $\langle nt\rangle$. Since each summand of $S_n(t)$ is conditionally centered and the concomitants are independent conditionally of the order statistics, the summands in (4.1) vanish for pairwise distinct indices. For $i=k\neq j=m$, the conditional expectation is less than or equal to

$$E[h^2(Y_{[i:n]},Y_{[j:n]})\mid X_{r:n},\ 1\leqslant r\leqslant n]=\int\int h^2(x,y)m(dx\mid X_{i:n})m(dy\mid X_{j:n})$$
$$\equiv: g(X_{i:n},X_{j:n}).$$

But with probability one, by Lemma 2.1,

$$n^{-2}\sum_{1\leqslant i\neq j\leqslant\langle nt\rangle} g(X_{i:n},X_{j:n})\to\int_0^t\int_0^t g(F^{-1}(u),F^{-1}(v))\,du\,dv.$$

By symmetry of h, similar arguments hold for $i=m$ and $j=k$. Since the factor in (4.1) is of order n^{-3}, the contribution of these index combinations is asymptotically zero. So it remains to study the index combinations for which, say, $i=k$ but $i\neq j\neq m$. Recall the definition of f_1, f_2 and V_1, V_2, respectively. As in Lemma 2.1,

$$\frac{1}{n(n-1)^2}\sum_{\substack{i\neq j\neq m\\ \leqslant\langle nt\rangle}} E[h(Y_{[i:n]},Y_{[j:n]})h(Y_{[i:n]},Y_{[m:n]})\mid X_{r:n},\ 1\leqslant r\leqslant n]$$
$$=\frac{1}{n(n-1)^2}\sum_{\substack{i\neq j\neq m\\ \leqslant\langle nt\rangle}} f_1(X_{i:n},X_{j:n},X_{m:n})\to V_1(t)$$

and

$$\frac{1}{n(n-1)^2}\sum_{\substack{i\neq j\neq m\\ \leqslant\langle nt\rangle}} E_{ij}E_{im}\to V_2(t).$$

By symmetry of h, we thus get the following

4.1. Lemma. *With probability one,*

$$n\operatorname{Var}(S_n(t) \mid X_{r:n},\ 1 \leqslant r \leqslant n) \to 4[V_1(t)-V_2(t)].$$

Set

$$\gamma(x, t) = \int_{-\infty}^{F^{-1}(t)} \int h(x, y)m(dy \mid v)F(dv) = E[h(x, Y)1_{\{X \leqslant F^{-1}(t)\}}].$$

Then (provided F is continuous)

$$V_1(t) = \int_{-\infty}^{F^{-1}(t)} \int \gamma^2(x, t)m(dx \mid u)F(du) = E[\gamma^2(Y, t)1_{\{X \leqslant F^{-1}(t)\}}]$$

and

$$V_2(t) = \int_{-\infty}^{F^{-1}(t)} \left[\int \gamma(x, t)m(dx \mid u)\right]^2 F(du).$$

In other words,

$$V_1(t)-V_2(t) = \int_{-\infty}^{F^{-1}(t)} \operatorname{Var}(\gamma(Y, t) \mid X = u)F(du).$$

In the following we shall derive, with $\mathscr{F} = \sigma(X_{r:n},\ 1 \leqslant r \leqslant n)$, the conditional projection of $S_n(t)$. Use conditional independence to verify (3.3). Set

$$\tilde{h}(y, X_{j:n}) = \int h(y, z)m\,(dz \mid X_{j:n}).$$

Then

$$\hat{S}_n(t) = \frac{1}{n(n-1)} \sum_{1 \leqslant i \neq j \leqslant \langle nt \rangle} [\tilde{h}(Y_{[i:n]}, X_{j:n}) + \tilde{h}(Y_{[j:n]}, X_{i:n}) - 2E_{ij}]$$

$$= \frac{2}{n} \sum_{i=1}^{\langle nt \rangle} \left\{ \frac{1}{n-1} \sum_{\substack{j=1 \\ j \neq i}}^{\langle nt \rangle} [\tilde{h}(Y_{[i:n]}, X_{j:n}) - E_{ij}] \right\} \equiv \frac{2}{n} \sum_{i=1}^{\langle nt \rangle} h_i'(Y_{[i:n]})$$

for short, after noticing that $h_i'(Y_{[i:n]})$ is also a function of the first $\langle nt \rangle$ order statistics. The $h_i'(Y_{[i:n]})$ variables are conditionally independent and centered. To compute the conditional variance of $\hat{S}_n(t)$, note that for $1 \leqslant i \leqslant \langle nt \rangle$, and $1 \leqslant j, m \leqslant \langle nt \rangle$ distinct from i

$$E[\tilde{h}(Y_{[i:n]}, X_{j:n})\tilde{h}(Y_{[i:n]}, X_{m:n}) \mid \mathscr{F}]$$
$$= \int\int\int h(y, z)h(y, x)m(dz \mid X_{j:n})m(dx \mid X_{m:n})m(dy \mid X_{i:n}),$$

which in terms of f_1 equals $f_1(X_{i:n}, X_{j:n}, X_{m:n})$. As for Lemma 4.1 we obtain

4.2. Lemma. *With probability one*

$$n\operatorname{Var}(\hat{S}_n(t) \mid X_{r:n},\ 1 \leqslant r \leqslant n) \to 4[V_1(t)-V_2(t)].$$

Refer to Remark 3.3 and recall Lemma 4.1 to get

$$\sqrt{n}(\hat{S}_n(t)-S_n(t)) \to 0 \text{ in probability.} \tag{4.2}$$

We want to show that (4.2) holds uniformly in $0 \leqslant t \leqslant 1$. For this, the following lemma will be crucial.

4.3. LEMMA. *For each* $n \geqslant 1$, *the process*

$$D_n(t) \equiv S_n(t)-\hat{S}_n(t) = \frac{1}{n(n-1)} \sum_{\substack{i,j=1 \\ i\neq j}}^{\langle nt\rangle} [h(Y_{[i:n]}, Y_{[j:n]})+E_{ij}-2\tilde{h}(Y_{[i:n]}, X_{j:n})]$$

is a martingale in $0 \leqslant t \leqslant 1$.

The proof is standard. ∎

We may also consider the process D_n as defined on a basic sample space, in which the $X_{i:n}$'s take on given values and the concomitants are independent with d.f.'s specified by the values taken on by the order statistics. Then D_n is also a martingale in this conditional setup. As before denote by $\mathscr{F}$ the σ-field generated by the order statistics. Kolmogorov's maximal inequality implies

$$P\big(\sqrt{n} \sup_{0\leqslant t\leqslant 1} |D_n(t)| \geqslant \varepsilon \mid \mathscr{F}\big) \leqslant n\varepsilon^{-2} E[D_n^2(1) \mid \mathscr{F}].$$

Conclude from Lemmas 4.1, 4.2 and the conditional projection lemma that the right-hand side converges to zero with probability one. So does the left-hand side. Integrating we obtain

$$P\big(\sqrt{n} \sup_{0\leqslant t\leqslant 1} |D_n(t)| \geqslant \varepsilon\big) \to 0 \text{ for each } \varepsilon > 0,$$

i.e.,

$$\sqrt{n} \sup_{0\leqslant t\leqslant 1} |S_n(t)-\hat{S}_n(t)| \to 0 \text{ in probability,} \tag{4.3}$$

which enables us to restrict ourselves to the process $\hat{S}_n$. This will be done in the next section.

5. The projected process: an invariance principle. We first compute the limit covariance structure of $\hat{S}_n$. Fix $s \leqslant t$. By conditional independence we obtain

$$\begin{aligned} & nE[\hat{S}_n(s)\hat{S}_n(t) \mid \mathscr{F}] \\ &= \frac{4}{n}\sum_{i=1}^{\langle nt\rangle} E\left[\left\{\frac{1}{n-1}\sum_{\substack{j=1 \\ j\neq i}}^{\langle ns\rangle} [\tilde{h}(Y_{[i:n]}, X_{j:n})-E_{ij}]\right\}\left\{\frac{1}{n-1}\sum_{\substack{j=1 \\ j\neq i}}^{\langle nt\rangle} [\tilde{h}(Y_{[i:n]}, X_{j:n})-E_{ij}]\right\} \,\middle|\, \mathscr{F}\right] \\ &= \frac{4}{n(n-1)^2} \sum_{i=1}^{\langle ns\rangle} \sum_{\substack{j=1 \\ j\neq i}}^{\langle ns\rangle} \sum_{\substack{m=1 \\ m\neq i}}^{\langle nt\rangle} [f_1(X_{i:n}, X_{j:n}, X_{m:n})-E_{ij}E_{im}] \to K(s, t) \end{aligned}$$

by Remark 2.2. Note that $K(t, t) = 4[V_1(t)-V_2(t)]$.

5.1. LEMMA. *Under* (1.1), *for* $0 \leqslant t \leqslant 1$,

$$\hat{S}_n(t) \to \mathcal{N}(0, K(t, t)) \text{ in distribution.}$$

Proof. Since, conditionally on $\mathcal{F}$, $\hat{S}_n(t)$ is a sum of independent centered random variables with existing finite second moments for which the limit variance exists, we only need to verify Lindeberg's condition. Now, for $\delta > 0$ fixed, consider

$$L_n = L_n(\delta; X_{r:n}, 1 \leqslant r \leqslant n) = \frac{4}{n} \sum_{i=1}^{\langle nt \rangle} \int_{\{|h_i'(y)| \geqslant \delta n^{1/2}\}} [h_i'(y)]^2 m(dy \mid X_{i:n}).$$

If h is bounded, so are the h_i'-functions. Since $\delta n^{1/2} \to \infty$, the $\{\ldots\}$ sets are empty from one n on. So, $L_n \to 0$ with probability one. For an arbitrary h, we know that the conditional variances of $\hat{S}_n(t)$ converge to

$$4[V_1(t) - V_2(t)] \leqslant 4E\gamma^2(Y, t) \leqslant 4Eh^2(Y_1, Y_2)$$

by the Cauchy–Schwarz inequality. Conclude that the limit variance is small whenever h is small in L^2. Thus, for a given $\varepsilon > 0$, choosing a bounded kernel g such that

$$E[(h-g)^2(Y_1, Y_2)] \leqslant \varepsilon,$$

we see that we may approximate $\hat{S}_n = \hat{S}_n^h$ by some $\hat{S}_n^g$ for which the Lindeberg condition holds, and such that

$$\limsup_{n \to \infty} E[(\hat{S}_n^h - \hat{S}_n^g)^2 \mid X_{r:n}, 1 \leqslant r \leqslant n] \leqslant \varepsilon.$$

From the CLT we get P-a.s.

$$P(\hat{S}_n(t) \leqslant x \mid X_{r:n}, 1 \leqslant r \leqslant n) \to P(\xi \leqslant x), \qquad \xi \sim \mathcal{N}(0, K(t, t)).$$

Integrating, we get $\hat{S}_n(t) \to \mathcal{N}(0, K(t, t))$. ∎

In the following lemma we prove the invariance principle for $\{\hat{S}_n\colon n \geqslant 1\}$.

5.2. LEMMA. *Under* (1.2)

$$\{\hat{S}_n(t)\colon 0 \leqslant t \leqslant 1\} \to \{B(t)\colon 0 \leqslant t \leqslant 1\}$$

in distribution in the space $D[0, 1]$. *Here* B *is a continuous zero means Gaussian process with covariance function* K.

Proof. Convergence of the finite-dimensional distributions follows similarly to Lemma 5.1, by the Cramér-Wold device. For tightness, since $\hat{S}_n(0) = 0$ is uniformly (stochastically) bounded, it suffices to prove, for $0 \leqslant s \leqslant t \leqslant u \leqslant 1$,

$$E\big[(\hat{S}_n(t) - \hat{S}_n(s))^2 (\hat{S}_n(u) - \hat{S}_n(t))^2\big] \leqslant \text{const}\,(u-s)^2. \tag{5.1}$$

We shall prove a conditional version of the last inequality. Integrating then yields (5.1). Fix $0 \leqslant s \leqslant t \leqslant u \leqslant 1$. Setting

$$Z_{ij} \equiv \tilde{h}(Y_{[i:n]}, X_{j:n}) - E_{ij},$$

we have

$$\hat{S}_n(t) - \hat{S}_n(s) = \frac{2}{n^{1/2}(n-1)} \Big[\sum_{i=\langle ns\rangle+1}^{\langle nt\rangle} \sum_{j=1}^{\langle ns\rangle} Z_{ij} + \sum_{i=1}^{\langle nt\rangle} \sum_{j=\langle ns\rangle+1}^{\langle nt\rangle} Z_{ij} \Big]$$

$$\equiv \mathrm{I}(s, t) + \mathrm{II}(s, t),$$

say. We proceed similarly for (t, u). Observe that $\mathrm{I}(s, t)$ and $\mathrm{I}(t, u)$ are conditionally independent. Use $(a+b)^2 \leqslant 2(a^2+b^2)$ and the Cauchy–Schwarz inequality to get

$$E\big[(\hat{S}_n(u) - \hat{S}_n(t))^2 (\hat{S}_n(t) - \hat{S}_n(s))^2 \mid \mathscr{F}\big]$$
$$\leqslant 4\{E(\mathrm{I}^2(t, u) + \mathrm{II}^2(t, u))(\mathrm{I}^2(s, t) + \mathrm{II}^2(s, t)) \mid \mathscr{F}\}$$
$$\leqslant 4\{E[\mathrm{I}^2(t, u) \mid \mathscr{F}]E[\mathrm{I}^2(s, t) \mid \mathscr{F}] + \sqrt{E[\mathrm{I}^4(t, u) \mid \mathscr{F}]E[\mathrm{II}^4(s, t) \mid \mathscr{F}]}$$
$$+ \sqrt{E[\mathrm{II}^4(t, u) \mid \mathscr{F}]E[\mathrm{I}^4(s, t) \mid \mathscr{F}]} + \sqrt{E[\mathrm{II}^4(t, u) \mid \mathscr{F}]E[\mathrm{II}^4(s, t) \mid \mathscr{F}]}\}.$$

By the assumed boundedness of f_1, we obtain

$$E[\mathrm{I}^2(s, t) \mid \mathscr{F}] \leqslant \frac{4}{n(n-1)^2} \sum_{i=\langle ns\rangle+1}^{\langle nt\rangle} \sum_{j,m=1}^{n} |f_1(X_{i:n}, X_{j:n}, X_{m:n})|$$
$$\leqslant \mathrm{const} \frac{\langle nt\rangle - \langle ns\rangle}{n}.$$

Similarly, applying the Zygmund–Marcinkiewicz inequality (cf., e.g., [7], p. 186) for 4-th moments, we get by boundedness of g_1

$$E[\mathrm{I}^4(s, t) \mid \mathscr{F}] \leqslant \mathrm{const} \left(\frac{\langle nt\rangle - \langle ns\rangle}{n} \right)^2.$$

Finally, again by the Zygmund–Marcinkiewicz inequality and boundedness of g_1,

$$E[\mathrm{II}^4(s, t) \mid \mathscr{F}] \leqslant \mathrm{const} \left(\frac{\langle nt\rangle - \langle ns\rangle}{n} \right)^4.$$

Similar bounds, of course, hold for (t, u). Let us integrate to get (5.1) whenever $u - s \geqslant 1/n$. For $u - s < 1/n$, the left-hand side of (5.1) is zero. We also have

$$E[B(t) - B(s)]^2 \leqslant \mathrm{const}\, |t - s|.$$

Since B is Gaussian, this yields continuity of B. The proof of the lemma is complete. ∎

6. Proofs of Theorems 1.1 and 1.2. Theorem 1.1 follows from (4.2) and Lemma 5.1, while Theorem 1.2 is immediate from (4.3) and Lemma 5.2.

REFERENCES

[1] P. K. Bhattacharya, *Convergence of sample paths of normalized sums of induced order statistics*, Ann. Statist. 2 (1974), pp. 1034–1039.

[2] — *An invariance principle in regression analysis*, ibidem 4 (1976), pp. 621–624.

[3] — *Induced order statistics: theory and applications*, in: P. R. Krishnaiah and P. K. Sen (Eds.), *Handbook of Statistics*, Vol. 4, 1984, pp. 383–403.

[4] M. Csörgö, S. Csörgö and L. Horváth, *An asymptotic theory for empirical reliability and concentration processes*, Lecture Notes in Statist. 33, Springer, New York–Heidelberg–Berlin 1986.

[5] C. M. Goldie, *Convergence theorems for empirical Lorenz curves and their inverses*, Adv. in Appl. Probab. 9 (1977), pp. 765–791.

[6] J. Hájek, *Asymptotic normality of simple linear rank statistics under alternatives*, Ann. Math. Statist. 39 (1968), pp. 325–346.

[7] I. A. Ibragimov and R. Z. Has'minskii, *Statistical Estimation*, Springer, New York–Heidelberg–Berlin 1981.

[8] A. Sandstroem, *Asymptotic normality of linear functions of concomitants of order statistics*, Metrika 34 (1987), pp. 129–142.

[9] R. J. Serfling, *Approximation Theorems of Mathematical Statistics*, Wiley, New York 1980.

[10] N. L. Spruill and J. L. Gastwirth, *On the estimation of the correlation coefficient from grouped data*, J. Amer. Statist. Assoc. 77 (1982), pp. 614–620.

[11] W. Stute and L. J. Wang, *The strong law under random censorship*, Ann. Statist. (to appear).

[12] S. S. Yang, *General distribution theory of the concomitants of order statistics*, ibidem 5 (1977), pp. 996–1002.

[13] — *Linear combination of concomitants of order statistics with application to testing and estimation*, Ann. Inst. Statist. Math. 33 (1981), pp. 464–470.

Mathematical Institute
University of Giessen
Arndtstr. 2
D-35392 Giessen, Germany

Received on 22.5.1992;
revised version on 25.2.1993