stichting
mathematisch
centrum

MC

AFDELING MATHEMATISCHE STATISTIEK SW 37/75 OKTOBER

M.C.A. VAN ZUYLEN

SOME PROPERTIES OF THE EMPIRICAL DISTRIBUTION FUNCTION IN THE NON-I.I.D. CASE (PART II)

Prepublication

2e boerhaavestraat 49 amsterdam

Printed at the Mathematical Centre, 49, 2e Boerhaavestraat, Amsterdam.

The Mathematical Centre, founded the 11-th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications. It is sponsored by the Netherlands Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O), by the Municipality of Amsterdam, by the University of Amsterdam, by the Free University at Amsterdam, and by industries.

AMS(MOS) subject classification scheme (1970): 60G17, 62G30

SOME PROPERTIES OF THE EMPIRICAL DISTRIBUTION FUNCTION IN THE NON-I.I.D. CASE*) (part II)

Abbreviated title:
ON EMPIRICAL DISTRIBUTION FUNCTIONS

by

M.C.A. van Zuylen
Mathematisch Centrum, Amsterdam

ABSTRACT

Two theorems are proved on empirical df's in the non-i.i.d. case, where moreover the underlying df's need not to be continuous.

These theorems are useful for proving asymptotic normality of rank statistics in the case where the multivariate sample elements are allowed to have different df's and where the scores generating functions are allowed to have a finite number of discontinuities of the first kind. The theorems may also be of interest in their own right.

KEY WORDS & PHRASES: *Empirical distribution, empirical process.*

---

*) This paper is not for review; it is meant for publication in a journal.

1. INTRODUCTION

Let k be a fixed positive integer and for each N = 1,2,..., let $X_{nN} = (X_{1nN}, X_{2nN}, \ldots, X_{knN})$, n = 1,2,...,N, be N mutually independent k-dimensional random vectors with joint distribution function (df) F_{nN} and marginal df's $F_{1nN}, F_{2nN}, \ldots, F_{knN}$. For each N, moreover, let $\mathbb{F}_N$ be the joint empirical df based on the N random vectors $X_{1N}, X_{2N}, \ldots, X_{NN}$. All random vectors are supposed to be defined on a single probability space (Ω, A, P).

We shall also use the notation $\bar{F}_N = N^{-1} \sum_{n=1}^{N} F_{nN}$ and $\bar{F}_{iN} = N^{-1} \sum_{n=1}^{N} F_{inN}$ for i = 1,2,...,k. Let us observe that $\bar{F}_N$ has all the properties of a k-variate df and that its marginal df's are the $\bar{F}_{iN}$. Finally we define, for i = 1,2,...,k, the inverse $\bar{F}_{iN}^{-1}$ by $\bar{F}_{iN}^{-1}(s) = \inf \{x \mid \bar{F}_{iN}(x) \geq s\}$, for $0 \leq s \leq 1$. Note that in the case where $\bar{F}_N$ is continuous we have $\bar{F}_{iN}(\bar{F}_{iN}^{-1}(s)) = s$ for $s \in [0,1]$.

This paper contains two theorems which are useful for proving asymptotic normality of rank statistics in the case where the multivariate sample elements are allowed to have different df's and where the scores generating functions are allowed to have a finite number of discontinuities of the first kind. The theorems may also be of interest in their own right.

The first theorem is a generalisation to the non-i.i.d. case of a slightly weaker version of a theorem, due to VAN ZWET (lemma 4.4 in [4]). See also [1]. In fact, VAN ZWET proved that in the i.i.d. case theorem 1 holds true, without the factor $(\log (N+1))^{\frac{1}{2}}$ in (1.1). However, we remark that our theorem 1 still is much stronger than, for instance, a generalisation of the Glivenko-Cantelli theorem to the multivariate non-i.i.d. case would be. Theorem 1 makes it possible to handle problems, connected with discontinuities in the scores generating functions of rank statistics.

For any Borel set B in $\mathbb{R}^k$ we shall write $\int_B d\bar{F}_N = \bar{F}_N(B)$ and $\int_B d\mathbb{F}_N = \mathbb{F}_N(B)$. By an interval in $\mathbb{R}^k$ the product set of k half open (closed on the right) intervals on the line will be meant.

THEOREM 1. *Let* $I_1, I_2, \ldots$ *be a sequence of intervals in* $\mathbb{R}^k$ *and let* $\mathcal{I}_N = \{I_N^*\colon I_N^*$ *is an interval contained in* $I_N\}$, $N = 1,2,\ldots$. *For every* $\varepsilon > 0$ *and every positive integer* k, *there exists* $M = M(\varepsilon,k)$, *such that for every array of* k*-variate* df's $F_{1N}, F_{2N}, \ldots, F_{NN}$, $N = 1,2,\ldots$, *every sequence* $I_1, I_2, \ldots$ *and every* $N = 1,2,\ldots$,

$$(1.1) \qquad P\left(\left\{\sup_{I_N^* \in \mathcal{I}_N} |\mathbb{F}_N\{I_N^*\} - \bar{F}_N\{I_N^*\}| > M\left(\frac{\log(N+1)\,\bar{F}_N\{I_N\}}{N}\right)^{\frac{1}{2}}\right\}\right) \le \varepsilon.$$

The second theorem is a result for the one-dimensional non-i.i.d. case, already given by SEN [5] for continuous underlying df's. However, it is clear from SHORACK [6] that the proof given by SEN is incorrect. An entirely different proof is provided here.

THEOREM 2. *For* $k = 1$ *and every* $\varepsilon > 0$ *and* $\delta \in (0,\frac{1}{2}]$, *there exists* $M = M(\varepsilon,\delta)$, *such that for every array of univariate* df's $F_{1N}, F_{2N}, \ldots, F_{NN}$, $N = 1,2,\ldots$, *and every* $N = 1,2,\ldots$,

$$P\left(\left\{\sup_{x \in \mathbb{R}} \frac{N^{\frac{1}{2}}\,|\mathbb{F}_N(x) - \bar{F}_N(x)|}{(\bar{F}_N(x)(1-\bar{F}_N(x)))^{\frac{1}{2}-\delta}} \ge M\right\}\right) \le \varepsilon.$$

Our basic tools are two related results of HOEFFDING [3]. Suppose that Z_n, $1 \le n \le N$, are independent random variables with $P(Z_n=1) = 1 - P(Z_n=0) = p_n$, and suppose that

$$0 < N^{-1} \sum_{n=1}^{N} p_n = \bar{p} < 1.$$

LEMMA 1 (HOEFFDING). *If* f *is a strictly convex function defined on* $[0,\infty)$ *then*

$$E\left(f\left(\sum_{n=1}^{N} Z_n\right)\right) \le \sum_{k=0}^{N} f(k)\binom{N}{k}\bar{p}^k(1-\bar{p})^{N-k},$$

where equality holds if and only if $p_1 = p_2 = \ldots = p_N = \bar{p}$.

LEMMA 2 (HOEFFDING). *Let* b *and* c *be two integers such that*

$$0 \le b \le N\bar{p} \le c \le N.$$

Then

$$\sum_{n=b}^{c} \binom{N}{n} \bar{p}^n (1-\bar{p})^{N-n} \le P\left(b \le \sum_{n=1}^{N} Z_n \le c\right) \le 1.$$

Both bounds are attained. The lower bound is attained only if $p_1 = p_2 = \ldots$ $\ldots = p_N = \bar{p}$ *unless* $b = 0$ *and* $c = N$.

In section 2 some preliminary results, supplying upper bounds for $\sup |\mathbb{F}_N - \bar{F}_N|$ and for central moments of $\sum_{n=1}^{N} Z_n$, are proved. These results are then used in section 3 to prove the theorems in the case of continuous underlying df's. The fact that the theorems also hold in the case of possibly discontinuous df's is immediate from lemma 9, given at the end of section 3.

2. SOME LEMMAS

By [a] we denote the greatest integer in the number a.

LEMMA 3. *Let for* $N = 1,2,\ldots$ *the* k*-dimensional* df's $F_{1N}, F_{2N}, \ldots, F_{NN}$ *be continuous and let* $I_1, I_2, \ldots$ *be a sequence of intervals in* $\mathbb{R}^k$ *with* $\bar{F}_N\{I_N\} > 0$, *for* $N = 1,2,\ldots$. *Define* $\bar{F}_{iN}^{-1}(1+a) = \infty$, *for* $a > 0$, $i = 1,2,\ldots,k$ *and let*

$$\mathcal{I}_N = \{I_N^*: I_N^* \textit{ is an interval contained in } I_N\}$$

and

$$\tilde{\mathcal{I}}_N = \left\{\tilde{I}_N: \tilde{I}_N = I_N \cap \prod_{i=1}^{k} \left(\bar{F}_{iN}^{-1}\left(\frac{n_{i1}}{N}\bar{F}_N\{I_N\}\right), \bar{F}_{iN}^{-1}\left(\frac{n_{i2}}{N}\bar{F}_N\{I_N\}\right)\right],\right. \tag{2.1}$$

for k *pairs of integers* (n_{i1}, n_{i2}), *with* $n_{i1} < n_{i2}$ *and*

$$\left. n_{ij} \in \left\{0,1,2,\ldots,\left[\frac{N}{\bar{F}_N\{I_N\}}\right]+1\right\}, \textit{ for } i = 1,2,\ldots,k,\ j = 1,2\right\}.$$

Then, for each $\omega \in \Omega$, $N = 1,2,\ldots$, $k = 1,2,\ldots$ *we have*

$$\sup_{I_N^* \in \mathcal{I}_N} |\mathbb{F}_N\{I_N^*\} - \bar{F}_N\{I_N^*\}| \le \max_{\tilde{I}_N \in \tilde{\mathcal{I}}_N} |\mathbb{F}_N\{\tilde{I}_N\} - \bar{F}_N\{\tilde{I}_N\}| + 2kN^{-1}\bar{F}_N\{I_N\}. \tag{2.2}$$

PROOF. Let I_N^* be an arbitrary interval in I_N. Define

$$\bar{I}_N^* = \bigcap_{\substack{\tilde{I}_N \in \tilde{\mathcal{I}}_N \\ \tilde{I}_N \supseteq I_N^*}} \tilde{I}_N, \qquad \underline{I}_N^* = \bigcup_{\substack{\tilde{I}_N \in \tilde{\mathcal{I}}_N \\ \tilde{I}_N \subset I_N^*}} \tilde{I}_N.$$

Note that $\bar{I}_N^*$ and $\underline{I}_N^*$ are elements of $\tilde{\mathcal{I}}_N \cup \emptyset$ and that

$$\bar{F}_N\{\bar{I}_N^*\} - \bar{F}_N\{\underline{I}_N^*\} \le 2kN^{-1}\bar{F}_N\{I_N\}. \tag{2.3}$$

If I_N^* is such that $\mathbb{F}_N\{I_N^*\} - \bar{F}_N\{I_N^*\} \ge 0$, we have using (2.3)

$$\begin{aligned} |\mathbb{F}_N\{I_N^*\} - \bar{F}_N\{I_N^*\}| &= \mathbb{F}_N\{I_N^*\} - \bar{F}_N\{I_N^*\} \le \mathbb{F}_N\{\bar{I}_N^*\} - \bar{F}_N\{\underline{I}_N^*\} \le \\ &\le \mathbb{F}_N\{\bar{I}_N^*\} - \bar{F}_N\{\bar{I}_N^*\} + 2kN^{-1}\bar{F}_N\{I_N\} \le \\ &\le |\mathbb{F}_N\{\bar{I}_N^*\} - \bar{F}_N\{\bar{I}_N^*\}| + 2kN^{-1}\bar{F}_N\{I_N\} \end{aligned}$$

and if $\mathbb{F}_N\{I_N^*\} - \bar{F}_N\{I_N^*\} < 0$, we have

$$\begin{aligned} |\mathbb{F}_N\{I_N^*\} - \bar{F}_N\{I_N^*\}| &= \bar{F}_N\{I_N^*\} - \mathbb{F}_N\{I_N^*\} \le \bar{F}_N\{\bar{I}_N^*\} - \mathbb{F}_N\{\underline{I}_N^*\} \le \\ &\le \bar{F}_N\{\underline{I}_N^*\} - \mathbb{F}_N\{\underline{I}_N^*\} + 2kN^{-1}\bar{F}_N\{I_N\} \le \\ &\le |\mathbb{F}_N\{\underline{I}_N^*\} - F_N\{\underline{I}_N^*\}| + 2kN^{-1}\bar{F}_N\{I_N\}. \qquad \square \end{aligned}$$

The following lemma gives upper bounds for the central moments of $\sum_{n=1}^N Z_n$, where Z_n, $1 \le n \le N$, are independent Bernoulli (p_n) random variables.

LEMMA 4. *For every* $\alpha > \frac{1}{2}$, *there exists* $M_\alpha \in (0,\infty)$, *such that for* $N = 1,2,\ldots,$

$$E\,\Big|\sum_{n=1}^N Z_n - N\bar{p}\Big|^{2\alpha} \le M_\alpha (N\bar{p})^\alpha, \qquad \textit{for } \bar{p} = N^{-1}\sum_{n=1}^N p_n \ge N^{-1} \tag{2.4}$$

and

$$E\,\Big|\sum_{n=1}^N Z_n - N\bar{p}\Big|^{2\alpha} \le M_\alpha N\bar{p}, \qquad \textit{for } \bar{p} = N^{-1}\sum_{n=1}^N p_n \le N^{-1}. \tag{2.5}$$

PROOF. Since $\alpha > \frac{1}{2}$, lemma 1 ensures that it is sufficient to prove lemma 4 in the case where $p_1 = p_2 = \ldots = p_N = \bar{p}$. First let us prove (2.4). For

$N = 1$, we have $\bar{p} = 1$, so that the left-hand side of (2.4) is zero. Let $F_{N\bar{p}}(y)$ be the distribution function of $|\sum_{n=1}^{N} Z_n - N\bar{p}|(N\bar{p}(1-\bar{p}))^{-\frac{1}{2}}$, then using an inequality due to S.N. BERNSTEIN [1, p.578], we have for $y > 0$ that

$$1 - F_{N\bar{p}}(y) = P\left(|\sum_{n=1}^{N} Z_n - N\bar{p}| > y\sqrt{N\bar{p}(1-\bar{p})}\right) \leq 2\exp\left(\frac{-y^2}{2 + \frac{2y}{3\sqrt{N\bar{p}(1-\bar{p})}}}\right).$$

Moreover, for $y \geq 1$ and $\bar{p} \geq N^{-1}$, $N = 2,3,\ldots$, we have

$$(N\bar{p}(1-\bar{p}))^{-\frac{1}{2}} \leq (N/(N-1))^{\frac{1}{2}} \leq 2^{\frac{1}{2}},$$

so that then

$$1 - F_{N\bar{p}}(y) \leq 2\exp\left(\frac{-y^2}{4y}\right) = 2\exp\left(-\frac{y}{4}\right).$$

Hence, for $\bar{p} \geq N^{-1}$, $N = 2,3,\ldots$,

$$E\left|\frac{\sum Z_n - N\bar{p}}{\sqrt{N\bar{p}(1-\bar{p})}}\right|^{2\alpha} = \int_0^\infty y^{2\alpha} dF_{N\bar{p}}(y) = 2\alpha\int_0^\infty y^{2\alpha-1}(1-F_{N\bar{p}}(y))dy \leq$$

$$\leq 2\alpha\int_0^1 dy + 4\alpha\int_1^\infty y^{2\alpha-1}\exp\left(-\frac{y}{4}\right)dy,$$

so that (2.4) is proved.

Let us next concentrate on the proof of (2.5). For $k = 1,2,\ldots,N$, we have $P(\sum_{n=1}^{N} Z_n = k) \leq \frac{(N\bar{p})^k}{k!}$ and $1 + k \leq e^k$, so that for $\alpha > \frac{1}{2}$, $\bar{p} \leq N^{-1}$, $N = 1,2,\ldots$,

$$E\left|\sum_{n=1}^{N} Z_n - N\bar{p}\right|^{2\alpha} = |0 - N\bar{p}|^{2\alpha} P\left(\sum_{n=1}^{N} Z_n = 0\right) + \sum_{k=1}^{N} |k - N\bar{p}|^{2\alpha} P\left(\sum_{n=1}^{N} Z_n = k\right) \leq$$

$$\leq (N\bar{p})^{2\alpha} + \sum_{k=1}^{\infty} (k-N\bar{p})^{2\alpha}\frac{(N\bar{p})^k}{k!} \leq$$

$$\leq N\bar{p} + \sum_{k=0}^{\infty} (k+1-N\bar{p})^{2\alpha}\frac{(N\bar{p})^{k+1}}{(k+1)!} \leq$$

$$\leq N\bar{p} + N\bar{p}\sum_{k=0}^{\infty} (k+1)^{2\alpha}\frac{(N\bar{p})^k}{(k+1)!} \leq N\bar{p} + N\bar{p}\sum_{k=0}^{\infty}\frac{(e^{2\alpha-1})^k}{k!} =$$

$$= N\bar{p}(1 + \exp(e^{2\alpha-1})). \qquad \square$$

3. PROOFS OF THE THEOREMS

PROOF OF THEOREM 1 IN THE CASE OF CONTINUOUS UNDERLYING df's.

If $\bar{F}_N\{I_N\} = 0$, the theorem follows immediately. It proves to be convenient to consider the cases $0 < \bar{F}_N\{I_N\} \le \frac{8 \log (N+1)}{\varepsilon N}$ and $\bar{F}_N\{I_N\} > \frac{8 \log (N+1)}{\varepsilon N}$, for fixed $0 < \varepsilon < 1$, separately.

First suppose that $0 < \bar{F}_N\{I_N\} \le \frac{8 \log (N+1)}{\varepsilon N}$, and choose $M = M_1(\varepsilon) = (2/\varepsilon)^{3/2}$, so that

$$(3.1) \qquad M\left(\frac{\log(N+1)\bar{F}_N\{I_N\}}{N}\right)^{\frac{1}{2}} \ge M\left(\frac{\varepsilon(\bar{F}_N\{I_N\})^2}{8}\right)^{\frac{1}{2}} \ge \frac{\bar{F}_N\{I_N\}}{\varepsilon}.$$

Moreover, since

$$(3.2) \qquad \sup_{I_N^* \in \mathcal{I}_N} |\mathbb{F}_N\{I_N^*\} - \bar{F}_N\{I_N^*\}| \le \max(\mathbb{F}_N\{I_N\}, \bar{F}_N\{I_N\}),$$

we have from (3.1), (3.2) and Markov's inequality that the left-hand side of (1.1) is bounded above by

$$P(\{\max(\mathbb{F}_N\{I_N\},\bar{F}_N\{I_N\}) \ge \bar{F}_N\{I_N\}/\varepsilon\}) = P(\{\mathbb{F}_N\{I_N\} \ge \bar{F}_N\{I_N\}/\varepsilon\}) \le \varepsilon.$$

Next we suppose that $\bar{F}_N\{I_N\} > \frac{8 \log (N+1)}{\varepsilon N}$, whereas the underlying df's are assumed to be continuous. Application of lemma 3 shows that for $M > 0$ and $N = 1,2,\ldots$, the left-hand side of (1.1) is bounded above by

$$(3.3) \qquad P\left(\left\{\max_{\tilde{I}_N \in \tilde{\mathcal{I}}_N} |\mathbb{F}_N\{\tilde{I}_N\} - \bar{F}_N\{\tilde{I}_N\}| > M\left(\frac{\log(N+1)\bar{F}_N\{I_N\}}{N}\right)^{\frac{1}{2}} - 2kN^{-1}\bar{F}_N\{I_N\}\right\}\right) \le$$

$$\le P\left(\left\{\max_{\tilde{I}_N \in \tilde{\mathcal{I}}_N} |\mathbb{F}_N\{\tilde{I}_N\} - \bar{F}_N\{\tilde{I}_N\}| > M\left(\frac{\log(N+1)\bar{F}_N\{I_N\}}{N}\right)^{\frac{1}{2}} - 2k\left(\frac{\bar{F}_N\{I_N\}}{N}\right)^{\frac{1}{2}}\right\}\right),$$

which, for $M > M_2(k) = 4k(\log 2)^{-\frac{1}{2}}$ and $N = 1,2,\ldots$, is bounded above by

$$(3.4) \qquad P\left(\left\{\max_{\tilde{I}_N \in \tilde{\mathcal{I}}_N} |\mathbb{F}_N\{\tilde{I}_N\} - \bar{F}_N\{\tilde{I}_N\}| > \tfrac{1}{2}M\left(\frac{\log(N+1)\bar{F}_N\{I_N\}}{N}\right)^{\frac{1}{2}}\right\}\right) \le$$

$$\le \sum_{\tilde{I}_N \in \tilde{\mathcal{I}}_N} P\left(|\mathbb{F}_N\{\tilde{I}_N\} - \bar{F}_N\{\tilde{I}_N\}| > \tfrac{1}{2}M\left(\frac{\log(N+1)\bar{F}_N\{I_N\}}{N}\right)^{\frac{1}{2}}\right).$$

Since $\frac{1}{2}M(N \log (N+1)\ \bar{F}_N\{I_N\})^{\frac{1}{2}} \geq 1$, for $M \geq M_3 = (\sqrt{2} \log 2)^{-1}$, lemma 2 is applicable, so that we may assume that $N\ \mathbb{F}_N\{\tilde{I}_N\}$ in (3.4) is a binomial rv with parameters N and $\bar{F}_N\{\tilde{I}_N\}$.

Applying Bernstein's inequality [1, p.578] and using $\max(\bar{F}_N\{\tilde{I}_N\}, 1-\bar{F}_N\{\tilde{I}_N\}) \leq 1$, we find for $M > 0$,

$$(3.5) \qquad P\left(|\mathbb{F}_N\{\tilde{I}_N\} - \bar{F}_N\{\tilde{I}_N\}| \geq \tfrac{1}{2}M\left(\frac{\log(N+1)\bar{F}_N\{I_N\}}{N}\right)^{\frac{1}{2}}\right) \leq$$

$$\leq 2 \exp\left(-\frac{\frac{1}{4}M^2 N \log (N+1)\ \bar{F}_N\{I_N\}}{2N\ \bar{F}_N\{\tilde{I}_N\} + \frac{1}{3}M(N \log (N+1)\ \bar{F}_N\{I_N\})^{\frac{1}{2}}}\right).$$

Moreover, since $\bar{F}_N\{I_N\} > \frac{8 \log (N+1)}{\varepsilon N} > \frac{8 \log (N+1)}{N}$ and $\frac{\bar{F}_N\{\tilde{I}_N\}}{\bar{F}_N\{I_N\}} \leq 1$, we obtain for (3.5) the following upper bound

$$(3.6) \qquad 2 \exp\left(-\frac{\frac{3}{2} M^2\sqrt{2} \log (N+1)}{12\sqrt{2} + M}\right) \leq 2 \exp(-\tfrac{3}{4}M \log (N+1)),$$

$$\text{for } M \geq M_4 = 12\sqrt{2}.$$

Noting that the number of elements in $\tilde{I}_N$ is bounded above by

$$\left(\frac{N}{\bar{F}_N\{I_N\}} + 4\right)^{2k} \leq \left(\frac{N^2\varepsilon}{8 \log (N+1)} + 4\right)^{2k} \leq (5N^2)^{2k},$$

we obtain from (3.3) - (3.6) for $M \geq \max(M_1,M_2,M_3,M_4) = M_5$ the following upper bound for the left-hand side of (1.1),

$$(5N^2)^{2k} 2(N+1)^{-\frac{3}{4}M} \leq 2.5^{2k}(N+1)^{4k-\frac{3}{4}M} \leq 2.5^{2k} \,.\, 2^{4k-\frac{3}{4}M},$$

$$\text{for } M > \max(M_5, 5\tfrac{1}{3} k). \qquad \square$$

Before presenting the proof of theorem 2, in the case of continuous underlying df's, we introduce some more notation. Moreover, from now on it will be assumed that $k = 1$. Define, for $n = 1,2,\ldots,N$,

$$Y_{nN} = \bar{F}_N(X_{nN}) \qquad \text{and} \qquad G_{nN}(t) = F_{nN}(\bar{F}_N^{-1}(t)).$$

Note that, because of the assumptions, Y_{nN} has continuous df G_{nN} on [0,1] and also

$$N^{-1} \sum_{n=1}^{N} G_{nN}(t) = t, \qquad \text{for } 0 \le t \le 1.$$

Let $\mathbb{G}_N$ denote the empirical df of $Y_{1N}, Y_{2N}, \ldots, Y_{NN}$; following SHORACK [6] we shall call $\mathbb{G}_N$ the reduced empirical df of $X_{1N}, X_{2N}, \ldots, X_{NN}$. We remark that it suffices to prove theorem 2 for the empirical df $\mathbb{G}_N$. Furthermore, we define the reduced empirical process X_N by

$$X_N(t) = N^{\frac{1}{2}}(\mathbb{G}_N(t) - t), \qquad \text{for } 0 \le t \le 1,$$

and the process S_N by setting $S_N(\frac{i}{N}) = X_N(\frac{i}{N})$, for $i = 0,1,\ldots,N$, and by letting S_N be linear on the intervals $[\frac{i-1}{N}, \frac{i}{N}]$. Finally, we define on $[0,1]$ the stochastic processes $X_{N\delta}(t)$ and $S_{N\delta}(t)$ by $X_{N\delta}(t) = X_N(t)/q_\delta(t)$ and $S_{N\delta}(t) = S_N(t)/q_\delta(t)$, where for $0 < \delta \le \frac{1}{2}$,

$$q_\delta(t) = (t(1-t))^{\frac{1}{2}-\delta}, \qquad \text{for } 0 < t < 1,$$

and

$$q_\delta(1) = q_\delta(0) = 1.$$

The following lemmas are needed for the proof of theorem 2:

<u>LEMMA 5</u>. *For* $k = 1$ *and every* $\alpha > \frac{1}{2}$ *there exists* $M_\alpha \in (0,\infty)$ *such that for every* $0 \le s, t \le 1$, *every* $N = 1,2,\ldots$ *and every array of continuous univariate* df's $F_{1N}, F_{2N}, \ldots, F_{NN}$, $N = 1,2,\ldots,$

$$E\,|S_N(t) - S_N(s)|^{2\alpha} \le M_\alpha\,|t-s|^{\alpha}.$$

<u>PROOF</u>. (The proof follows the pattern of SHORACK [6], where lemma 5 is proved for $\alpha = 2$.)

Let $N \ge 1$, $0 \le s \le t \le 1$, $\alpha > \frac{1}{2}$ and the continuous univariate df's $F_{1N}, F_{2N}, \ldots, F_{NN}$ be arbitrary, but fixed. Choose integers i and j so that

$$\frac{i-1}{N} \le s \le \frac{i}{N} \qquad \text{and} \qquad \frac{j-1}{N} \le t \le \frac{j}{N}.$$

Let $\Delta_{km} = |S_N(\frac{m}{N}) - S_N(\frac{k}{N})|$ for integers k and m, with $m > k$. Then from lemma 4 we have for some number $M_\alpha \in (0,\infty)$,

$$E\Delta_{km}^{2\alpha} = E\left|N^{\frac{1}{2}}\left(\left(\mathbb{G}_N\left(\frac{m}{N}\right) - \mathbb{G}_N\left(\frac{k}{N}\right)\right) - \frac{m-k}{N}\right)\right|^{2\alpha} =$$

$$= N^{-\alpha}E\left|N\left(\mathbb{G}_N\left(\frac{m}{N}\right) - \mathbb{G}_N\left(\frac{k}{N}\right)\right) - N\left(\frac{m-k}{N}\right)\right|^{2\alpha} \le M_\alpha\left(\frac{m-k}{N}\right)^\alpha =$$

$$= e_{km}(\text{say}).$$

We also let $\delta_{uv} = |S_N(v) - S_N(u)|$.

Case 1. $i < j - 1$. Then

$$\delta_{st} \le \Delta_{i,j-1} \vee \Delta_{i,j} \vee \Delta_{i-1,j-1} \vee \Delta_{i-1,j},$$

so that

$$E\delta_{st}^{2\alpha} \le e_{i,j-1} + e_{i,j} + e_{i-1,j-1} + e_{i-1,j} \le 4e_{i-1,j} =$$

$$= 4M_\alpha\left(\frac{j-(i-1)}{N}\right)^\alpha \le 4M_\alpha 3^\alpha(t-s)^\alpha.$$

Case 2. $i = j$. Since the change of a linear function on an interval of length $t - s$ equals the slope times $t - s$, we have

$$\delta_{st} \le N\Delta_{i-1,i}(t-s),$$

so that

$$E\delta_{st}^{2\alpha} \le N^{2\alpha}(t-s)^{2\alpha}e_{i-1,i} \le M_\alpha N^{2\alpha}(t-s)^\alpha(t-s)^\alpha N^{-\alpha} \le M_\alpha(t-s)^\alpha.$$

Case 3. $i = j - 1$. Then

$$\delta_{st} \le \delta_{s,\frac{i}{N}} + \delta_{\frac{i}{N},t} \le 2(\delta_{s,\frac{i}{N}} \vee \delta_{\frac{i}{N},t}),$$

so that by Case 2 we have

$$E\delta_{st}^{2\alpha} \le 2^{2\alpha}\left(E\delta_{s,\frac{i}{N}}^{2\alpha} + E\delta_{\frac{i}{N},t}^{2\alpha}\right) \le 2^{2\alpha}\left(M_\alpha\left(\frac{i}{N} - s\right)^\alpha + M_\alpha\left(t - \frac{i}{N}\right)^\alpha\right) \le$$

$$\le 2^{2\alpha+1}M_\alpha\left\{\left(\frac{i}{N} - s\right) + \left(t - \frac{i}{N}\right)\right\}^\alpha = 2^{2\alpha+1}M_\alpha(t-s)^\alpha. \qquad \square$$

LEMMA 6. *For* $k = 1$ *and every* $\alpha > \frac{1}{2}$ *there exists* $M_\alpha \in (0,\infty)$ *such that for every* $0 \le s, t \le 1$, *every* $N = 1,2,\ldots$, *every* $\delta \in (0,\frac{1}{2}]$ *and every array of continuous univariate* df's $F_{1N},F_{2N},\ldots,F_{NN}$, $N = 1,2,\ldots$,

$$(3.7) \qquad E\{|S_{N\delta}(t) - S_{N\delta}(s)|^{2\alpha}\} \le M_\alpha |t - s|^{2\alpha\delta}.$$

PROOF. Because of symmetry in s and t of (3.7) we suppose $t \ge s$, so that we only have to prove the lemma in the three cases $0 \le s \le t \le \frac{1}{2}$, $0 \le s \le \frac{1}{2} \le t \le 1$ and $\frac{1}{2} \le s \le t \le 1$.

Since for the second case we have (c_r-inequality)

$$E|S_{N\delta}(t) - S_{N\delta}(s)|^{2\alpha} \le c_\alpha E|S_{N\delta}(t) - S_{N\delta}(\tfrac{1}{2})|^{2\alpha} + c_\alpha E|S_{N\delta}(\tfrac{1}{2}) - S_{N\delta}(s)|^{2\alpha},$$

where c_α is a number, depending on α only, it suffices to give a proof for the first and third case. Moreover, the lemma is trivially true for s is zero.

Let $\alpha > \frac{1}{2}$, $0 \le s \le t \le \frac{1}{2}$, $N \ge 1$, $\delta \in (0,\frac{1}{2}]$ and the continuous univariate df's $F_{1N},F_{2N},\ldots,F_{NN}$ be arbitrary, but fixed and let M_i, $i = 1,2,\ldots,6$ be numbers, only depending on α. Because of lemma 5,

$$E\left|\frac{S_N(t)}{q_\delta(t)} - \frac{S_N(s)}{q_\delta(s)}\right|^{2\alpha} \le M_1 E\left|\frac{S_N(t) - S_N(s)}{q_\delta(t)}\right|^{2\alpha} + M_1 E\left|\left(\frac{1}{q_\delta(t)} - \frac{1}{q_\delta(s)}\right)S_N(s)\right|^{2\alpha} \le$$

$$\le M_2(t-s)^{2\alpha\delta} + M_3\left(\frac{1}{q_\delta(s)} - \frac{1}{q_\delta(t)}\right)^{2\alpha} s^\alpha.$$

Moreover, for $s \le \frac{1}{2}t$, we have

$$s^{\frac{1}{2}}\left(\frac{1}{q_\delta(s)} - \frac{1}{q_\delta(t)}\right) \le M_4 \frac{1}{s^{\frac{1}{2}-\delta}} s^{\frac{1}{2}} = M_4 s^\delta \le M_4(t-s)^\delta,$$

whereas for $s > \frac{1}{2}t$, we find

$$s^{\frac{1}{2}}\left(\frac{1}{q_\delta(s)} - \frac{1}{q_\delta(t)}\right) \le M_5 s^{\frac{1}{2}}\left(\frac{1}{s^{\frac{1}{2}-\delta}} - \frac{1}{t^{\frac{1}{2}-\delta}}\right) = M_5 s^{\frac{1}{2}} \int_s^t \frac{du}{u^{3/2-\delta}} \le$$

$$\le M_5 \frac{s^{\frac{1}{2}}(t-s)}{s^{3/2-\delta}} \le M_6 \frac{(t-s)}{t^{1-\delta}} \le M_6(t-s)^\delta.$$

The proof for the case $\frac{1}{2} \le s \le t \le 1$ goes analogously. □

LEMMA 7. *For* k = 1 *and every* $\varepsilon > 0$ *and* $\delta \in (0,\frac{1}{2}]$, *there exists* $M = M(\varepsilon,\delta)$, *such that for every array of continuous univariate* df's $F_{1N}, F_{2N}, \ldots, F_{NN}$, $N = 1,2,\ldots$, *and every* $N = 1,2,\ldots$,

$$P\left(\left\{\sup_{0\le t\le 1} |S_{N\delta}(t)| \ge M(\varepsilon,\delta)\right\}\right) \le \varepsilon.$$

PROOF. Choose arbitrary, but fixed $\varepsilon > 0$, $\delta \in (0,\frac{1}{2}]$, $N \in \{1,2,\ldots\}$ and continuous univariate df's $F_{1N}, F_{2N}, \ldots, F_{NN}$. Because of lemma 6 we have for every $0 \le s, t \le 1$,

$$E\{|S_{N\delta}(t) - S_{N\delta}(s)|^{[\delta^{-1}]+2}\} \le M_\delta |t-s|^{\delta([\delta^{-1}]+2)}, \tag{3.8}$$

where $M_\delta \in (0,\infty)$ only depends on δ. Remark that $\delta([\delta^{-1}]+2) > 1$.

Consider for fixed positive integer m, the random variables

$$\xi_{im} = S_{N\delta}\left(\frac{i}{m}\right) - S_{N\delta}\left(\frac{i-1}{m}\right) \qquad i = 1,2,\ldots,m.$$

Obviously, $S_{N\delta}(\frac{k}{m}) = \xi_{1m} + \ldots + \xi_{km} (S_{N\delta}(0)=0)$.

Because of (3.8) we now have for $0 \le i \le j \le m$, $M \in (0,\infty)$ that

$$P\left(\left|S_{N\delta}\left(\frac{j}{m}\right) - S_{N\delta}\left(\frac{i}{m}\right)\right| \ge M\right) \le M^{-[\delta^{-1}]-2} E\left\{\left|S_{N\delta}\left(\frac{j}{m}\right) - S_{N\delta}\left(\frac{i}{m}\right)\right|^{[\delta^{-1}]+2}\right\} \le$$

$$\le M^{-[\delta^{-1}]-2} M_\delta \left(\frac{j}{m} - \frac{i}{m}\right)^{\delta([\delta^{-1}]+2)} =$$

$$= M^{-[\delta^{-1}]-2}\left(\sum_{i<\ell\le j} u_\ell\right)^{\delta([\delta^{-1}]+2)},$$

where $u_\ell = (M_\delta)^{(\delta([\delta^{-1}]+2))^{-1}} \frac{1}{m}$, for $\ell = 1,2,\ldots,m$.

Applying theorem 12.2 of [2, p.94] we obtain for some $K_\delta \in (0,\infty)$, depending on δ only, that

$$P\left(\left\{\max_{0\le i\le m} \left|S_{N\delta}\left(\frac{i}{m}\right)\right| \ge M\right\}\right) \le K_\delta M^{-[\delta^{-1}]-2}(u_1+u_2+\ldots+u_m)^{\delta([\delta^{-1}]+2)} =$$

$$= K_\delta M^{-[\delta^{-1}]-2} M_\delta.$$

Since, for each $\omega \in \Omega$, $S_{N\delta}$ is a continuous function on [0,1], letting $m \to \infty$ leads to

$$P\left(\left\{\sup_{0\le t\le 1} |S_{N\delta}(t)| \ge M\right\}\right) \le K_\delta M^{-[\delta^{-1}]-2} M_\delta,$$

from which the lemma follows. □

LEMMA 8. *For* k = 1 *and every* $\varepsilon > 0$ *and* $\delta \in (0,\frac{1}{2}]$, *there exists* $M = M(\varepsilon,\delta)$, *such that for every array of continuous univariate* df*'s* $F_{1N}, F_{2N}, \ldots, F_{NN}$, $N = 1,2,\ldots$, *and every* $N = 1,2,\ldots$,

$$P\left(\left\{\sup_{0\le t\le 1} |X_{N\delta}(t) - S_{N\delta}(t)| \ge M(\varepsilon,\delta)\right\}\right) \le \varepsilon.$$

PROOF. We shall only give the proof of the lemma in the case where we restrict t in the supremum to the interval $[0,\frac{1}{2}]$ and $N = 2,3,\ldots$.

We remark that for $M > 0$ and $N = 2,3,\ldots$ and $\delta \in (0,\frac{1}{2}]$,

$$(3.9)\qquad P\left(\left\{\sup_{0\le t\le \frac{1}{2}} |X_{N\delta}(t) - S_{N\delta}(t)| \ge 4M\right\}\right) \le P\left(\left\{\sup_{N^{-1}\le t\le \frac{1}{2}} |X_{N\delta}(t) - S_{N\delta}(t)| \ge M\right\}\right) + $$
$$+ P\left(\left\{\sup_{0\le t\le N^{-1}} |S_{N\delta}(t)| \ge M\right\}\right) + $$
$$+ P\left(\left\{\sup_{0\le t\le N^{-1}} \frac{N^{\frac{1}{2}}\mathbb{G}_N(t)}{q_\delta(t)} \ge M\right\}\right) + $$
$$+ P\left(\left\{\sup_{0\le t\le N^{-1}} \frac{N^{\frac{1}{2}}t}{q_\delta(t)} \ge M\right\}\right).$$

Let us now derive upper bounds for these terms seaprately, assuming throughout continuous univariate df's $F_{1N}, F_{2N}, \ldots, F_{NN}$.

From the fact that $|X_N(t) - S_N(t)| \le |X_N(t) - S_N(\frac{i}{N})| \vee |X_N(t) - S_N(\frac{i-1}{N})|$ for $\frac{i-1}{N} \le t \le \frac{i}{N}$, it follows for $M > 0$, $N = 2,3,\ldots$ and $\delta \in (0,\frac{1}{2}]$ that

$$(3.10)\qquad P\left(\left\{\sup_{N^{-1}\le t\le \frac{1}{2}} |X_{N\delta}(t) - S_{N\delta}(t)| \ge M\right\}\right) \le$$

$$\le \sum_{i=2}^{\left[\frac{N+1}{2}\right]} P\left(N\mathbb{G}_N\left(\frac{i}{N}\right) - N\mathbb{G}_N\left(\frac{i-1}{N}\right) - 1 \ge N^{\frac{1}{2}} M q_\delta\left(\frac{i-1}{N}\right) - 3\right).$$

Moreover, from Chebyshev's inequality, (3.10) and lemma 4 we find that for $\delta \in (0,\frac{1}{2}]$, $\beta = \beta(\delta) = (2-2\delta)/(1-2\delta) > 2$, $M \ge 6\sqrt{2}$ and $N = 2,3,\ldots$, (3.11) $$P\left(\left\{\sup_{N^{-1}\le t\le \frac{1}{2}} |X_{N\delta}(t) - S_{N\delta}(t)| \ge M\right\}\right) \le$$

$$\le \sum_{i=2}^{\left[\frac{N+1}{2}\right]} P\left(\left|N\mathbb{G}_N\left(\frac{i}{N}\right) - N\mathbb{G}_N\left(\frac{i-1}{N}\right) - NN^{-1}\right| \ge \tfrac{1}{2}MN^{\frac{1}{2}}q_\delta\left(\frac{i-1}{N}\right)\right) \le$$

$$\le \sum_{i=2}^{\left[\frac{N+1}{2}\right]} (\tfrac{1}{2}M)^{-\beta}q_\delta^{-\beta}\left(\frac{i-1}{N}\right)N^{-\frac{1}{2}\beta}E\left\{\left|N\mathbb{G}_N\left(\frac{i}{N}\right) - N\mathbb{G}_N\left(\frac{i-1}{N}\right) - NN^{-1}\right|^{\beta}\right\} \le$$

$$\le M_{\frac{1}{2}\beta}(\tfrac{1}{2}M)^{-\beta}N^{1-\frac{1}{2}\beta}\,\frac{1}{N}\sum_{i=2}^{\left[\frac{N+1}{2}\right]} q_\delta^{-\beta}\left(\frac{i-1}{N}\right) \le M_{\frac{1}{2}\beta}(\tfrac{1}{2}M)^{-\beta}2^{1-\frac{1}{2}\beta}K_\delta,$$

where $K_\delta \in (0,\infty)$ is a number only depending on δ.

Next, since $S_N(t) = NtS_N(N^{-1})$ for $0 \le t \le N^{-1}$, we have for $M > 0$, $\delta \in (0,\frac{1}{2}]$ and $N = 2,3,\ldots$, that (Chebyshev's inequality and lemma 4)

(3.12) $$P\left(\left\{\sup_{0\le t\le \frac{1}{N}} \frac{|S_N(t)|}{q_\delta(t)} \ge M\right\}\right) \le P\left(\frac{|S_N(N^{-1})|}{q_\delta(N^{-1})} \ge M\right) =$$

$$= P(|N\mathbb{G}_N(N^{-1}) - NN^{-1}| \ge Mq_\delta(N^{-1})N^{\frac{1}{2}}) \le$$

$$\le \frac{M_1}{M^2Nq_\delta^2\left(\frac{1}{N}\right)} \le K'_\delta M^{-2}2^{-2\delta},$$

where K'_δ is a number only depending on δ.

Furthermore we have for $M \ge 2^{\frac{1}{2}}$, $N = 2,3,\ldots$ and $\delta \in (0,\frac{1}{2}]$,

(3.13) $$P\left(\left\{\sup_{0\le t\le N^{-1}} \frac{N^{\frac{1}{2}}t}{q_\delta(t)} \ge M\right\}\right) \le P\left(\left\{\sup_{0\le t\le \frac{1}{N}} \frac{N^{\frac{1}{2}}t}{t^{\frac{1}{2}-\delta}} \ge M2^{\delta-\frac{1}{2}}\right\}\right) =$$

$$= P(N^{-\delta} \ge M2^{\delta-\frac{1}{2}}) \le P(2^{-\delta} \ge M2^{\delta-\frac{1}{2}}) = 0.$$

Finally, we have for $0 \le t \le N^{-1}$, $M > 0$, $N = 2,3,\ldots$ and $\delta \in (0,\frac{1}{2}]$ that

$$\frac{\mathbb{G}_N(t)}{q_\delta(t)} = \frac{\mathbb{G}_N(t)}{t}\,\frac{t^{\frac{1}{2}+\delta}}{(1-t)^{\frac{1}{2}-\delta}} \le \frac{\mathbb{G}_N(t)}{t}\,2^{\frac{1}{2}-\delta}\left(\frac{1}{N}\right)^{\frac{1}{2}+\delta}$$

and hence

$$(3.14)\qquad P\left(\left\{\sup_{0\le t\le \frac{1}{N}} \frac{N^{\frac{1}{2}}\mathbb{G}_N(t)}{q_\delta(t)} \le M\right\}\right) \ge P\left(\left\{\sup_{0\le t\le 1} \frac{\mathbb{G}_N(t)}{t} \le \frac{M2^{\delta}}{2^{\frac{1}{2}-\delta}}\right\}\right) .$$

In [7]*) it is proved that for each $\eta > 0$, there exists a α, $0 < \alpha < 1$, independent of N and of the continuous df's $F_{1N}, F_{2N}, \ldots, F_{NN}$, such that

$$(3.15)\qquad P\left(\left\{\sup_{0\le t\le 1} \frac{\mathbb{G}_N(t)}{t} \le \alpha^{-1}\right\}\right) \ge 1 - \eta .$$

Hence, the lemma follows after combining (3.9), (3.11), (3.12), (3.13) and (3.14), together with the remark above. □

PROOF OF THEOREM 2 IN THE CASE OF CONTINUOUS UNDERLYING df's

Remark that it suffices to prove theorem 2 for the empirical df $\mathbb{G}_N$. Then the theorem follows immediately from lemma 7, lemma 8 and the fact that for $M > 0$,

$$P\left(\left\{\sup_{0\le t\le 1} |X_{N\delta}(t)| \ge M\right\}\right) \le P\left(\left\{\sup_{0\le t\le 1} |S_{N\delta}(t)| \ge \frac{M}{2}\right\}\right) + P\left(\left\{\sup_{0\le t\le 1} |X_{N\delta}(t) - S_{N\delta}(t)| \ge \frac{M}{2}\right\}\right) . \quad \square$$

The following lemma and its proof make it clear that the theorems also hold in the case where the underlying df's are allowed to be discontinuous.

<u>LEMMA 9</u>. *Let* k *be a fixed positive integer and let* $\mathbb{F}_N$ *be the empirical df based on* N k*-variate sample elements* $X_n = (X_{1n}, X_{2n}, \ldots, X_{kn})$, $n = 1,2,\ldots,N$, *where the* X_n *are distributed independently according to given, possibly discontinuous* df*'s* F_n. *There exist* N k*-variate random vectors* $Y_n = (Y_{1n}, Y_{2n}, \ldots, Y_{kn})$, $n = 1,2,\ldots,N$, *where the* Y_n *are distributed independently according to continuous* df*'s* G_n, *such that with probability one*

---

*) The proof of (3.15) can also be found in Report SW 34/75 of the Department of Mathematical Statistics, Mathematisch Centrum, Amsterdam.

$$(3.16)\qquad \sup_{-\infty<x_1,x_2,\dots,x_k<\infty} |\mathbb{G}_N(x_1,x_2,\dots,x_k) - \bar{G}_N(x_1,x_2,\dots,x_k)| \geq$$
$$\geq \sup_{-\infty<x_1,x_2,\dots,x_k<\infty} |\mathbb{F}_N(x_1,\dots,x_k) - \bar{F}_N(x_1,x_2,\dots,x_k)|,$$

where $\mathbb{G}_N$ *is the* ***empirical*** df *based on the* Y_n, $n = 1,2,\dots,N$ *and* $\bar{F}_N = N^{-1}\sum_{n=1}^N F_n$, $\bar{G}_N = N^{-1}\sum_{n=1}^N G_n$.

PROOF. For $i = 1,2,\dots,k$ let us denote by F_{in} the i^{th} marginal df of F_n, let $\bar{F}_{iN} = N^{-1}\sum_{n=1}^N F_{in}$ and let $\{\xi_\nu^{(i)}, \nu = 1,2,\dots\}$ be the countable set of discontinuity points of $\bar{F}_{iN}$. This set contains the discontinuity points of each F_{in}, $n = 1,2,\dots,N$. Moreover, let $p_\nu^{(i)}$ be the jump at $\xi_\nu^{(i)}$ of $\bar{F}_{iN}$ and let $\{U_\nu^{(i)}, \nu = 1,2,\dots\}$ be a set of uniform (0,1) distributed random variables, mutually independent and also independent of the random vectors X_n, $n = 1,2,\dots,N$.

Since $\sum_\nu p_\nu^{(i)} \leq 1$ for $i = 1,2,\dots,k$, we can define for $n = 1,2,\dots,N$ the random vector $Y_n = (Y_{1n},Y_{2n},\dots,Y_{kn})$ as follows ($\mathbb{1}$ denotes the indicator function):

$$(3.17)\qquad Y_{in} = X_{in} + \sum_\nu \mathbb{1}_{X_{in}>\xi_\nu^{(i)}}\, p_\nu^{(i)} + \sum_\nu \mathbb{1}_{X_{in}=\xi_\nu^{(i)}}\, p_\nu^{(i)}\, U_\nu^{(i)},$$
$$\text{for } i = 1,2,\dots,k,$$

so that X_n is transformed stochastically to Y_n.
Let G_n be the df of Y_n and let $\mathbb{G}_N$ be the empirical df based on $Y_1,Y_2,\dots,Y_N$. It is clear that all the marginal df's of G_n are continuous and hence G_n is continuous.
From definition (3.17) it is immediate that for $n = 1,2,\dots,N$

$$(3.18)\qquad [X_{in} \leq x_i, \text{ for } i = 1,2,\dots,k] \Longleftrightarrow [Y_{in} \leq x_i + \sum_\nu \mathbb{1}_{x_i\geq\xi_\nu^{(i)}}\, p_\nu^{(i)},$$
$$\text{for } i = 1,2,\dots,k].$$

Moreover, from (3.18) it is obvious that (with $\bar{G}_N = N^{-1}\sum_{n=1}^N G_n$)

$$\bar{F}_N(x_1,x_2,\dots,x_k) = \bar{G}_N\Big(x_1+\sum_\nu \mathbb{1}_{x_1\geq\xi_\nu^{(1)}}\, p_\nu^{(1)},\dots,x_k + \sum_\nu \mathbb{1}_{x_k\geq\xi_\nu^{(k)}} p_\nu^{(k)}\Big)$$

and with probability one

$$\mathbb{F}_N(x_1,x_2,\ldots,x_k) = \mathbb{G}_N\left(x_1+\sum_\nu 1_{x_1\geq\xi_\nu^{(1)}}p_\nu^{(1)},\ldots,x_k + \sum_\nu 1_{x_k\geq\xi_\nu^{(k)}}p_\nu^{(k)}\right).$$

so that (3.16) holds. □

ACKNOWLEDGEMENT. I wish to thank Professor W.R. VAN ZWET for his careful reading of the manuscript and his valuable comments.

REFERENCES

[1] BAHADUR, R.R. (1966), *A note on quantiles in large samples*, Ann. Math. Statist. 37, 577-580.

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

[3] HOEFFDING, W. (1956), *On the distribution of the number of successes in independent trials*, Ann. Math. Statist. 27, 713-721.

[4] RUYMGAART, F.H. (1974), *Asymptotic normality of nonparametric tests for independence*, Ann. Statist. 2, 892-910.

[5] SEN, P. (1970), *On the distribution of one-sample rank order statistics*, Nonparametric Techniques in Statistical Inference, ed. M. Puri, Cambridge Univ. Press.

[6] SHORACK, G.R. (1973), *Convergence of reduced empirical and quantile processes with application to functions of order statistics in the non-i.i.d. case*, Ann. Statist. 1, 146-152.

[7] VAN ZUYLEN, M.C.A. (1975), *Some properties of the empirical distribution function in the non-i.i.d. case*, To appear in Ann. Statist.