The Annals of Probability
1982, Vol. 10, No. 1, 108–123

PROPERTIES OF THE EMPIRICAL DISTRIBUTION FUNCTION FOR INDEPENDENT NON-IDENTICALLY DISTRIBUTED RANDOM VECTORS

BY MARTIEN C. A. VAN ZUIJLEN

Catholic University, Nijmegen

A generalization to the case of independent but not necessarily identically distributed two-dimensional underlying random vectors is obtained of results on univariate empirical df's of van Zuijlen (1976) (linear bounds), Ghosh (1972) and Ruymgaart and van Zuijlen (1978b). No conditions are imposed on the dependence structure of the underlying df's. In the process improvements of van Zuijlen's results concerning linear bounds in the univariate non-i.i.d. case are obtained, whereas also applications of the results on multivariate empirical df's are discussed. Extensions of the two-dimensional results to the k-dimensional case ($k > 2$) are straightforward and therefore omitted.

1. Introduction. For $N \in \mathbb{N}$ let $X_{nN} = (Y_{nN}, Z_{nN})$, $n = 1, 2, \cdots, N$, be N mutually independent two-dimensional random vectors with continuous joint distribution functions (df's)

$$F_{nN}(y, z) = P(Y_{nN} \le y, Z_{nN} \le z), \qquad \text{for} \quad y, z \in \mathbb{R} \tag{1.1}$$

and marginal df's G_{nN} and H_{nN}, i.e.,

$$G_{nN}(y) = P(Y_{nN} \le y); \quad H_{nN}(z) = P(Z_{nN} \le z), \qquad \text{for} \quad y, z \in \mathbb{R}. \tag{1.2}$$

All random vectors are supposed to be defined on a single probability space $(\Omega, \mathscr{A}, P)$. For each N, moreover, let us define the joint empirical df $\mathbb{F}_N$ of $X_{1N}, X_{2N}, \cdots, X_{NN}$ by taking $N\,\mathbb{F}_N(y, z)$ to be the number of elements in the set $\{X_{nN}: Y_{nN} \le y, Z_{nN} \le z, n = 1, 2, \cdots, N\}$ for all $y, z \in \mathbb{R}$, and the averaged df's $\bar{F}_N$, $\bar{G}_N$ and $\bar{H}_N$ as

$$\bar{F}_N(y, z) = N^{-1} \textstyle\sum_{n=1}^{N} F_{nN}(y, z), \tag{1.3}$$

$$\bar{G}_N(y) = N^{-1} \textstyle\sum_{n=1}^{N} G_{nN}(y); \quad \bar{H}_N(z) = N^{-1} \textstyle\sum_{n=1}^{N} H_{nN}(z), \qquad \text{for} \quad y, z \in \mathbb{R}. \tag{1.4}$$

We remark that $\bar{F}_N$ has all the properties of a two-dimensional df and that its marginal df's are $\bar{G}_N$ and $\bar{H}_N$.

It is well-known how several important classes of statistics, such as simple linear rank statistics, certain rank statistics for testing independence and linear combinations of functions of order statistics, can be expressed in terms of the empirical distribution function(s) and how in a Chernoff-Savage approach certain properties of the empirical df can be used to obtain the asymptotic distribution of these statistics and in particular the convergence in probability to zero of the remainder term. In this connection we refer to Bhuchongkul (1964), Govindarajulu, Le Cam and Raghavachari (1967), Koul (1970), Koul and Staudte (1972), Ruymgaart, Shorack and van Zwet (1972), Ruymgaart (1974) and Ruymgaart and van Zuijlen (1978a, 1978b).

Let us consider the following two-dimensional versions of the representations in terms of the empirical df's of linear combinations of functions of order statistics:

$$T_N = \iint J_N(\mathbb{F}_N^*)\psi_N(\bar{F}_N)\, d\mathbb{F}_N, \tag{1.5}$$

where ψ_N and J_N are real-valued functions on $(0, 1)$ and $\mathbb{F}_N^* = N/(N+1)\mathbb{F}_N$. See

Received July 18, 1979; revised December 10, 1980.
AMS 1970 *subject classifications.* Primary 60G17; secondary 62G30.
Key words and phrases. Empirical distribution function, multivariate non-i.i.d. case, linear bounds.

Ruymgaart and van Zuijlen (1978b). Taking $\psi \equiv 1$ in (1.5) we obtain the relevant class of nonlinear rank statistics S_N where

$$S_N = \iint J_N(\mathbb{F}_N^*)\, d\mathbb{F}_N. \tag{1.6}$$

The asymptotic theory of these what we call Kendall-type statistics, has not been developed in the literature. If we take in (1.6) $J_N(s) = (N+1)N^{-1}s$ then

$$S_N = \iint \mathbb{F}_N\, d\mathbb{F}_N,$$

which is a statistic equivalent to Kendall's rank correlation coefficient, as already has been remarked in Ruymgaart (1973).

In order to establish, in a Chernoff-Savage approach, the asymptotic normality of a suitably standardized version of S_N and in particular the convergence to zero of the remainder term in a sense stronger than only convergence in probability, one essentially will need the properties of the multivariate empirical df to be derived in this paper. These theorems on multivariate empirical df's have also applications in sequential statistics and may be of independent interest. No conditions will be imposed on the dependence structure of the underlying df's. The program concerning S_N as described above will be carried out in a forthcoming paper.

In contrast to univariate empirical df's, multivariate empirical df's did not receive much attention in the literature so far: Kiefer and Wolfowitz (1958), Kiefer (1961) and van Zwet (see Ruymgaart (1974)) obtained results in the multivariate i.i.d. case; van Zuijlen (1978) generalized van Zwet's result to the non-i.i.d. case.

In Section 2 we shall derive useful multivariate versions of van Zuijlen's results (1976) concerning linear bounds for the empirical df in the univariate non-i.i.d. case. Compared with the 1-dim. non-i.i.d. case, these generalizations to the multivariate situation will require a different method of proof. In this process we shall also obtain improvements of van Zuijlen's results (1976) in the 1-dim. non-i.i.d. case which are very close to the corresponding well-known results in the 1-dim. i.i.d. case.

Section 3 is devoted to the derivation of multivariate versions of results of Ghosh (1972) and Ruymgaart and van Zuijlen (1978b), whereas in Remark 3.1 of this section some attention will be paid to the concrete applicability of the main results in the theory of rank tests.

For convenience we shall restrict ourselves in this paper to continuous underlying df's. That this restriction is not essential follows from the remarks in Section 3 of van Zuijlen (1978).

2. Linear bounds for the multivariate empirical df. We denote by $Y_{1:N} \le Y_{2:N} \le \cdots \le Y_{N:N}$ the order statistics of the rv's $Y_{1N}, Y_{2N}, \cdots, Y_{NN}$. Let $\mathbb{G}_N$ be the empirical df of these Y's and define in $\mathbb{R}$ the random sets ($n = 1, 2, \cdots, N$)

$$\mathcal{O}_{nN}^1 = \mathcal{O}_{nN}^1(Y_{nN}) = \{y \in \mathbb{R} \mid y \ge Y_{nN},\ \mathbb{G}_N(y) = \mathbb{G}_N(Y_{nN})\}, \tag{2.1}$$

$$\tilde{\mathcal{O}}_{nN}^1 = \tilde{\mathcal{O}}_{nN}^1(Y_{nN}) = \{y \in \mathbb{R} \mid y \le Y_{nN},\ \mathbb{G}_N(y) = \mathbb{G}_N(Y_{nN}) - N^{-1}\} \cup \{Y_{nN}\} \quad \text{if } \mathbb{G}_N(Y_{NN}) > N^{-1}, \tag{2.2}$$

$$= \{\mathrm{Y}_{nN}\} \qquad \text{elsewhere,}$$

$$\mathcal{O}_N^1 = \{y \in \mathbb{R} \mid \mathbb{G}_N(y) = 0\}, \tag{2.3}$$

$$\tilde{\mathcal{O}}_N^1 = \{y \in \mathbb{R} \mid \mathbb{G}_N(y) = 1\}. \tag{2.4}$$

Similarly, in $\mathbb{R}^2$ we define

$$\mathcal{O}_{nN}^2 = \mathcal{O}_{nN}^2(X_{nN}) = \{x \in \mathbb{R}^2 \mid x \ge x_{nN},\ \mathbb{F}_N(x) = \mathbb{F}_N(X_{nN})\}, \tag{2.5}$$

$$\tilde{\mathcal{O}}_{nN}^2 = \tilde{\mathcal{O}}_{nN}^2(X_{nN}) = \{x \in \mathbb{R}^2 \mid x \le X_{nN},\ \mathbb{F}_N(x) = \mathbb{F}_N(X_{nN}) - N^{-1}\} \cup \{X_{nN}\} \quad \text{if } \mathbb{F}_N(X_{nN}) > N^{-1},$$
$$= \{X_{nN}\} \quad \text{elsewhere}, \tag{2.6}$$

$$\mathcal{O}_N^2 = \{x \in \mathbb{R}^2 \mid \mathbb{F}_N(x) = 0\}, \tag{2.7}$$

$$\tilde{\mathcal{O}}_N^2 = \{x \in \mathbb{R}^2 \mid \mathbb{F}_N(x) = 1\}, \tag{2.8}$$

where $x_1 = (y_1, z_1) \ge x_2 = (y_2, z_2)$ stands for $y_1 \ge y_2$ and $z_1 \ge z_2$.

In 1945 Daniels proved an exact result concerning an upper bound for the empirical df in the case of independent and identically distributed one-dimensional random variables (the 1-dim. i.i.d. case). Formulated in terms of the rv's $Y_{1N}, Y_{2N}, \cdots, Y_{NN}$ he showed:

Theorem 2.1 (Daniels). *For $\beta \in (0, 1)$, $N \in \mathbb{N}$ and continuous underlying* df's G_{1N}, $G_{2N}, \cdots, G_{NN}$ *with* $G_{1N} = G_{2N} = \cdots = G_{NN} = \bar{G}_N$, *we have*

$$P_1 \equiv P(\mathbb{G}_N(y) \le \beta^{-1}\bar{G}_N(y), \quad \text{for} \quad y \in \mathbb{R}) = 1 - \beta. \tag{2.9}$$

Proof. An elegant proof of this theorem can be found in Robbins (1954). □

Note that

$$\begin{aligned} P_1 &= P(\mathbb{G}_N(y) \le \beta^{-1}\bar{G}_N(y), && \text{for} \quad y \in ((\cup_{n=1}^N \mathcal{O}_{nN}^1) \cup \mathcal{O}_N^1)) \\ &= P(\mathbb{G}_N(y) \le \beta^{-1}\bar{G}_N(y), && \text{for} \quad y \in \cup_{n=1}^N \mathcal{O}_{nN}^1) \\ &= P(\mathbb{G}_N(y) \le \beta^{-1}\bar{G}_N(y), && \text{for} \quad y \in \{Y_{1N}, Y_{2N}, \cdots, Y_{NN}\}) \\ &= P(\mathbb{G}_N(y) \le \beta^{-1}\bar{G}_N(y), && \text{for} \quad y \in \{Y_{1:N}, Y_{2:N}, \cdots, Y_{N:N}\}). \end{aligned} \tag{2.10}$$

In 1964 Chang proved an exact result concerning a lower bound for the empirical df in the 1-dim. i.i.d. case:

Theorem 2.2 (Chang). *For $N \in \mathbb{N}$ and continuous underlying* df's $G_{1N}, G_{2N}, \cdots$, G_{NN} *with* $G_{1N} = G_{2N} = \cdots = G_{NN} = \bar{G}_N$, *we have*

$$\tilde{P}_1 \equiv P(\mathbb{G}_N(y) > \beta\bar{G}_N(y), \quad \text{for} \quad y \in [Y_{1:N}, \infty)) = 1 - I_N(\beta), \tag{2.11}$$

where

$$I_N(\beta) = \begin{cases} 0 & \text{for} \quad \beta < N^{-1} \\ \left(1 - \dfrac{1}{N\beta}\right)^N \sum_{k=1}^{[N\beta]} \dbinom{N}{k} \dfrac{(k-1)^{k-1}(N\beta - k)^{N-k}}{(N\beta)^N} & \text{for} \quad N^{-1} \le \beta \le 1 \\ 1 & \text{for} \quad \beta > 1. \end{cases}$$

Note that

$$\begin{aligned} \tilde{P}_1 &= P(\mathbb{G}_N(y) \ge \beta\bar{G}_N(y), && \text{for} \quad y \in ((\cup_{n=1}^N \tilde{\mathcal{O}}_{nN}^1) \cup \tilde{\mathcal{O}}_N^1)) \\ &= P(\mathbb{G}_N(y) \ge \beta\bar{G}_N(y), && \text{for} \quad y \in \cup_{n=1}^N \tilde{\mathcal{O}}_{nN}^1) \\ &\le P(\mathbb{G}_N(y) \ge \beta\bar{G}_N(y), && \text{for} \quad y \in \{Y_{1N}, Y_{2N}, \cdots, Y_{NN}\}) \\ &= P(\mathbb{G}_N(y) \ge \beta\bar{G}_N(y), && \text{for} \quad y \in \{Y_{1:N}, Y_{2:N}, \cdots, Y_{N:N}\}). \end{aligned} \tag{2.12}$$

Lower bounds for the probabilities P_1 and $\tilde{P}_1$ in (2.9) and (2.11) respectively, in the case of independent but not necessarily identically distributed random variables (the 1-dim. non-i.i.d. case), are derived in van Zuijlen (1976, 1978) and successfully applied in Ruymgaart and van Zuijlen (1977, 1978), Shorack (1972), Shorack and Wellner (1978) and van Zuijlen (1976).

It is our aim in this section to obtain lower bounds for the probabilities in (2.10) and

(2.12) (part A and part B respectively) in the case of independent but not necessarily identically distributed two-dimensional random vectors (the 2-dim. non-i.i.d. case). A motivation for these multivariate theorems is given in the introduction. See also Remark 3.1.

Part A: *A linear upper bound for the multivariate empirical* df. In van Zuijlen (1976, 1978) a useful lower bound for P_1 in the one-dimensional non-i.i.d. case has been given. When we want to prove an extension of this result to the 2-dim. non-i.i.d. case in the way the theorem has been proved in van Zuijlen (1976) we will meet the problem that for $k > 1$ the order statistics, which play an essential role in that proof, are not defined. However, using the last equality in (2.10) together with a conditioning argument, we are able to prove a sharpening of the result in van Zuijlen (1976) for P_1, in a such a way that the new proof in the 1-dim. non-i.i.d. case will work without difficulties also in the k-dim. non-i.i.d. case.

THEOREM 2.3. *For* $\beta \in (0, 1/2)$, $N \in \mathbb{N}$ *and continuous underlying* df's $G_{1N}, G_{2N}, \cdots, G_{NN}$, *we have*

$$P_1 \geq 1 - \beta - \frac{7\pi^2\beta^2/3}{(1-2\beta)^4}. \tag{2.13}$$

PROOF. We define

$$T_{nN}(y) = \sum_{\substack{j=1 \\ j\neq n}}^{N} Z_j, \tag{2.14}$$

where $Z_j, j \neq n$, are independent Bernoulli (p_j) rv's with $p_j = G_{jN}(y)$ and hence

$$\bar{p} \equiv \bar{p}_n(y) \equiv (N-1)^{-1} \sum_{\substack{j=1 \\ j\neq n}}^{N} p_j = (N-1)^{-1}(N\bar{G}_N(y) - G_{nN}(y)). \tag{2.15}$$

For $N = 1$ the result of the theorem is trivial. For $\beta \in (0, 1/2)$ $N \in \{2, 3, \cdots\}$ and continuous df's we have

$$\begin{aligned} P_1 &= P(\mathbb{G}_N(Y_{nN}) \leq \beta^{-1}\bar{G}_N(Y_{nN}), n = 1, 2, \cdots, N) \\ &\geq 1 - \sum_{n=1}^{N} P(\mathbb{G}_N(Y_{nN}) > \beta^{-1}\bar{G}_N(Y_{nN})) \\ &= 1 - \sum_{n=1}^{N} \int_{-\infty}^{\infty} P(T_{nN}(y) > N\beta^{-1}\bar{G}_N(y) - 1)\, dG_{nN}(y). \end{aligned} \tag{2.16}$$

Next, define for $i = 0, 1, 2, \cdots, N-1$

$$I_i = \left\{y \mid \bar{G}_N(y) \in \left[\frac{i\beta}{N}, \frac{(i+1)\beta}{N}\right)\right\} = \left\{y \mid i - 1 \leq \frac{N}{\beta}\left(\bar{G}_N(y)\right) - 1 < i\right\}, \tag{2.17}$$

$$I_N = \{y \mid \bar{G}_N(y) \geq \beta\} = \left\{y \mid \frac{N}{\beta}(\bar{G}_N(y)) - 1 \geq N - 1\right\}. \tag{2.18}$$

Remark that $\mathbb{R} = \cup_{i=0}^{N} I_i$, $I_i \cap I_j = \varnothing$ for $i \neq j$ and that on I_i $(i = 0, 1, \cdots, N)$ we have

$$P(T_{nN}(y) > N\beta^{-1}\bar{G}_N(y) - 1) = P(T_{nN}(y) \geq i), \tag{2.19}$$

whereas on I_i $(i = 1, 2, \cdots, N)$

$$N\bar{G}_N(y) < (i+1)\beta \leq 2i\beta < i, \tag{2.20}$$

so that

$$i > N\bar{G}_N(y) - G_{nN}(y). \tag{2.21}$$

Using (2.19)-(2.21) and the inequality (1.1) in van Zuijlen (1976) we find (with $\bar{G}_N\{I_i\} \equiv \int_{I_i} d\bar{G}_N$)

$$P_1 \geq 1 - \textstyle\sum_{n=1}^{N} \sum_{i=0}^{N} \displaystyle\int_{I_i} P(T_{nN}(y) \geq i)\, dG_{nN}(y)$$

$$\geq 1 - \textstyle\sum_{n=1}^{N} \displaystyle\Bigg[G_{nN}\{I_0\} + \textstyle\sum_{i=1}^{N-1} \displaystyle\int_{I_i} \frac{3(N\bar{G}_N(y) - G_{nN}(y))^2 + N\bar{G}_N(y) - G_{nN}(y)}{(i - (N\bar{G}_N(y) - G_{nN}(y)))^4}\, dG_{nN}(y) \Bigg]$$

$$\geq 1 - N\bar{G}_N\{I_0\} - \textstyle\sum_{i=1}^{N-1} \sum_{n=1}^{N} \displaystyle\int_{I_i} \frac{3(N\bar{G}_N(y))^2 + N\bar{G}_N(y)}{(i - N\bar{G}_N(y))^4}\, dG_{nN}(y) \tag{2.22}$$

$$\geq 1 - N\bar{G}_N\{I_0\} - \textstyle\sum_{i=1}^{N-1} \sum_{n=1}^{N} \displaystyle\int_{I_i} \frac{12i^2\beta^2 + 2i\beta}{(i - 2i\beta)^4}\, dG_{nN}(y)$$

$$\geq 1 - N\bar{G}_N\{I_0\} - \textstyle\sum_{i=1}^{N-1} \displaystyle\frac{14i^{-2}\beta}{(1 - 2\beta)^4} N\bar{G}_N\{I_i\}$$

$$\geq 1 - N\cdot\frac{\beta}{N} - \frac{14\beta}{(1-2\beta)^4}\cdot\textstyle\sum_{i=1}^{\infty} i^{-2}\cdot N\cdot\displaystyle\frac{\beta}{N} = 1 - \beta - \frac{\frac{7}{3}\pi^2\beta^2}{(1-2\beta)^4}. \ \square$$

Theorem 2.3 is a sharpening of the corresponding result in van Zuijlen (1978) in the sense that for β sufficiently small:

$$1 - \beta - \frac{\frac{7}{3}\pi^2\beta^2}{(1-2\beta)^4} \geq 1 - \frac{\frac{2}{3}\pi^2\beta}{(1-\beta)^4}. \tag{2.23}$$

Comparing Theorem 2.1 with Theorem 2.3 we see that the additional term $7(\pi^2\beta^2/3)\cdot(1-2\beta)^{-4}$ in the lower bound of Theorem 2.3 is apparently the price we pay for allowing the underlying distribution functions to be different.

From (2.10) and the proof of Theorem 2.3 we obtain the following extension of Theorem 2.3 to the two-dimensional non-i.i.d. case:

Theorem 2.4. *For $\beta \in (0, 1/2)$, $N \in \mathbb{N}$ and continuous underlying* df's $F_{1N}, F_{2N}, \cdots, F_{NN}$, *we have*

$$\begin{aligned} P_2 &\equiv P(\mathbb{F}_N(x) \leq \beta^{-1}\bar{F}_N(x), \quad \text{for} \quad x \in (\cup_{n=1}^{N} \mathcal{O}_{nN}^2) \cup \mathcal{O}_N^2) \\ &= P(\mathbb{F}_N(X_{nN}) \leq \beta^{-1}\bar{F}_N(X_{nN}),\ n = 1, 2, \cdots, N) \\ &\geq 1 - NP\left(\bar{F}_N(\bar{X}_N) < \frac{\beta}{N}\right) - \frac{14\beta N}{(1-2\beta)^4} \textstyle\sum_{i=1}^{N-1} i^{-2} P\left(\displaystyle\frac{i\beta}{N} \leq \bar{F}_N(\bar{X}_N) < \frac{(i+1)\beta}{N}\right) \end{aligned} \tag{2.24}$$

where $\bar{X}_N$ is a two-dimensional random vector with df $\bar{F}_N$ *and where the random sets $\mathcal{O}_{nN}^2$ and $\mathcal{O}_N^2$ are defined in* (2.5) *and* (2.7).

Proof. From the proof of Theorem 2.3 it is immediate that for $N \geq 2$

$$P_2 \geq 1 - N\bar{F}_N\{\tilde{I}_0\} - \textstyle\sum_{i=1}^{N-1} \displaystyle\frac{14i^{-2}\beta}{(1-2\beta)^4} N\bar{F}_N\{\tilde{I}_i\}, \tag{2.25}$$

where

$$\tilde{I}_i = \left\{ (y, z) |\ \bar{F}_N(y, z) \in \left[\frac{i\beta}{N}, \frac{(i+1)\beta}{N}\right) \right\}, \tag{2.26}$$

so that (2.24) holds. The result for $N = 1$ is trivial. $\square$

REMARK 2.1. In the special case where $\bar{F}_N(\bar{X}_N)$ is uniformly distributed on (0, 1), which can happen in degenerate cases only, the multidimensional result of Theorem 2.4 reduces to the one-dimensional result of Theorem 2.3.

The lower bound in Theorem 2.4 does not depend on the underlying distribution functions for instance in the special case where the sample elements are independent, identically distributed and also independent component-wise (or more generally where $\bar{F}_N = \bar{G}_N \times \bar{H}_N$):

COROLLARY 2.1. *For $\beta \in (0, 1/2)$, $N \in \mathbb{N}$ and continuous underlying* df's $F_{1N}, F_{2N}, \cdots, F_{NN}$, *with $\bar{F}_N = \bar{G}_N \times \bar{H}_N$, we have*

$$P_2 \geq 1 - \beta - \log \frac{N}{\beta}\left(\beta + \frac{\frac{7}{3}\pi^2\beta^2}{(1-2\beta)^4}\right). \tag{2.27}$$

PROOF. We apply Theorem 2.4 and note that

$$P\left(\bar{F}_N(\bar{X}_N) < \frac{\beta}{N}\right) = P\left(U_2 < \frac{\beta}{N}\right)$$

where U_2 is a random variable which is distributed as the product of two independent random variables, both uniformly distributed on (0, 1). The density function $f(u)$ of U_2 equals

$$f(u) = -1_{(0,1)}(u) \cdot \log u,$$

so that

$$P\left(\bar{F}_N(\bar{X}_N) < \frac{\beta}{N}\right) = \frac{\beta}{N} - \frac{\beta}{N}\log\frac{\beta}{N} \tag{2.28}$$

and for $i = 1, 2, \cdots, N-1$

$$\begin{aligned} P\left(\frac{i\beta}{N} \leq \bar{F}_N(\bar{X}_N) < \frac{(i+1)\beta}{N}\right) &= \frac{\beta}{N}\left(\log\frac{N}{\beta} + i\log i - (i+1)\log(i+1)\right) \\ &\leq \frac{\beta}{N}\log\frac{N}{\beta}. \end{aligned} \tag{2.29}$$

Hence, (2.27) follows from (2.28), (2.29) and Theorem 2.4. □.

Finally, as an example, let us derive a statement like Corollary 2.1 in the case of positive regression dependence (see e.g., Lehmann (1967)).

Let $\mathscr{F}_2$ denote the family of absolutely continuous df's F with either

$$F_1(y, z) \equiv P(Z \leq z \mid Y = y) \text{ is nonincreasing in } y \tag{2.30}$$

or

$$F_2(y, z) \equiv P(Y \leq y \mid Z = z) \text{ is nonincreasing in } z, \tag{2.31}$$

where the 2-dim. random vector (Y, Z) has df F. If (2.30) holds, then Y is said to be positively regression dependent on Z. Similarly, let $\mathscr{G}_2$ be the family of all absolutely continuous df's F with either $F_1(y, z)$ nondecreasing in y or $F_2(y, z)$ nondecreasing in z. If $F_1(y, z)$ is nondecreasing in y then Y is said to be negatively regression dependent on Z.

LEMMA 2.1. *Let X be a 2-dim. random vector with* df F. *For $c \in (0, 1]$ we have*

$$F \in \mathscr{F}_2 \Rightarrow P(F(X) \leq c) \leq P(U_2 \leq c) = c - c\log c \tag{2.32}$$

and

$$F \in \mathscr{G}_2 \Rightarrow P(F(X) \leq c) \geq P(U_2 \leq c) \tag{2.33}$$

where the random variable U_2 is defined above (2.28).

PROOF. Let $X = (Y, Z)$ and let G denote the df of Y. From Rosenblatt (1952) we know that $(G(Y), F_1(Y, Z))$ is uniformly distributed on $[0, 1]^2$. Now, let $F \in \mathscr{F}_2$ and suppose without loss of generality that (2.30) holds. Then

$$F(y, z) = \int_{-\infty}^{y} F_1(u, z)\, dG(u) \geq \int_{-\infty}^{y} F_1(y, z)\, dG(u) = F_1(y, z)\cdot G(y)$$

so that

$$P(F(X) \leq c) \leq P(F_1(Y, Z)\cdot G(Y) \leq c) = P(U_2 \leq c).$$

The proof of the second part of the lemma can be given in a similar way. □

COROLLARY 2.2. *For $\beta \in (0, 1/14)$, $N \in \mathbb{N}$ and continuous underlying* df's $F_{1N}, F_{2N}, \cdots, F_{NN}$ *with $\bar{F}_N \in \mathscr{F}_2$, we have*

$$P_2 \geq 1 - \beta - \log\frac{N}{\beta}\left(\beta + \frac{\frac{7}{3}\pi^2\beta^2}{(1-2\beta)^4}\right). \tag{2.34}$$

PROOF. For brevity let $P_{(i)} \equiv P\left(\bar{F}_N(\bar{X}_N) < \frac{i\beta}{N}\right)$ and let $\bar{P}_{(i)} \equiv P\left(U_2 < \frac{i\beta}{N}\right)$, where U_2 is defined above (2.28). From Lemma 2.1 we have that $P_{(i)} \leq \bar{P}_{(i)}$, for $i = 1, 2, \cdots, N$, so that (2.34) follows from Corollary 2.1 and the following expression for the second part in the lower bound of Theorem 2.4:

$$\begin{aligned}
NP&\left(\bar{F}_N(\bar{X}_N) < \frac{\beta}{N}\right) + \frac{14N\beta}{(1-2\beta)^4}\sum_{i=1}^{N-1} i^{-2}P\left(\frac{i\beta}{N} \leq \bar{F}_N(\bar{X}_N) < \frac{(i+1)\beta}{N}\right)\\
&= NP_{(1)} + \frac{14N\beta}{(1-2\beta)^4}\sum_{i=1}^{N-1} i^{-2}(P_{(i+1)} - P_{(i)})\\
&= N\left(1 - \frac{14\beta}{(1-2\beta)^4}\right)P_{(1)} + \frac{14N\beta}{(1-2\beta)^4}\left\{\left(1 - \frac{1}{4}\right)P_{(2)}\right.\\
&\quad \left. + \left(\frac{1}{4} - \frac{1}{9}\right)P_{(3)} + \cdots + \left(\frac{1}{(N-2)^2} - \frac{1}{(N-1)^2}\right)P_{(N-1)} + \frac{1}{(N-1)^2}P_{(N)}\right\}. \quad \square
\end{aligned}$$

The following corollary has direct applications in the theory of rank statistics.

COROLLARY 2.3. *There exists $N_0 \in \mathbb{N}$ such that for $N \geq N_0$ and continuous underlying* df's $F_{1N}, F_{2N}, \cdots, F_{NN}$, *with $\bar{F}_N \in \mathscr{F}_2$, we have*

$$P(\mathbb{F}_N(x) \leq (\log N)^2\bar{F}_N(x), \quad \text{for} \quad x \in (\cup_{n=1}^{N} \mathcal{O}_{nN}^2) \cup \mathcal{O}_N^2) \geq 1 - \frac{6}{\log N}, \tag{2.35}$$

where the sets $\mathcal{O}_{nN}^2$ and $\mathcal{O}_N^2$ are defined in (2.5) *and* (2.7).

PROOF. Immediate from Corollary 2.2 with $\beta = (\log N)^{-2}$. □

Part B: *A linear lower bound for the multivariate empirical* df. In van Zuijlen (1976) it has been shown that for $\beta \in (0, 1)$, $N \in \mathbb{N}$ and continuous underlying df's which are not necessarily all equal

$$\tilde{P}_1 \geq 1 - \tfrac{2}{3}\pi^2\beta^2(1-\beta)^{-4}. \tag{2.36}$$

REMARK 2.2. In Shorack and Wellner (1978) an exponential upper bound for $1 - \tilde{P}_1$ has been derived from the result of Chang (1964). They showed in the 1-dim. i.i.d. case and for underlying rv's which are uniformly distributed on (0, 1) that

$$\tilde{P}_1 \geq 1 - 16\beta^{-1}\exp(-\beta^{-1}), \qquad \beta \in (0, 1), \qquad N \in \mathbb{N}. \tag{2.37}$$

Moreover, Wellner (1978) proved that the constant 16 can even be replaced by e.

The following substantial improvement of (2.36) gives an exponential upper bound for $1 - \tilde{P}_1$ in the non-i.i.d. case, which has almost the order of the bound (2.37) in the i.i.d. case. The proof of this sharpening is essentially different from the proof of (2.36) in van Zuijlen (1976) and will admit a generalization to the multivariate non-i.i.d. case.

THEOREM 2.5. *For* $\beta \in (0, 1)$, $N \in \mathbb{N}$ *and continuous underlying* df's $G_{1N}, G_{2N}, \cdots, G_{NN}$, *we have*

$$\tilde{P}_1 \geq 1 - 15\beta^{-2}\exp(-\beta^{-1}). \tag{2.38}$$

PROOF. For $N = 1$ the theorem is trivial, so suppose $N \geq 2$. Let $T_{nN}(y)$ be defined as in (2.14). Since

$$\begin{aligned} [\mathbb{G}_N(y) \geq \beta \bar{G}_N(y), y \in [Y_{1:N}, \infty)] &\Leftrightarrow [\cap_{n=2}^N (\mathbb{G}_N(Y_{n:N}-) \geq \beta \bar{G}_N(Y_{n:N}))] \\ &\Leftrightarrow \left[\cap_{n=1}^N \left(\max\left(\frac{1}{N}\# [Y_{jN} < Y_{nN}], \frac{1}{N}\right)\right.\right. \\ &\qquad \left.\left. \geq \beta \bar{G}_N(Y_{nN})\right)\right] \end{aligned} \tag{2.39}$$

we have

$$\begin{aligned} \tilde{P}_1 &= 1 - P(\cup_{n=1}^N (\max(\#[Y_{jN} < Y_{nN}], 1) < N\beta\bar{G}_N(Y_{nN}))) \\ &\geq 1 - \textstyle\sum_{n=1}^N P(\max(\#[Y_{jN} \leq Y_{nN}], 1) < N\beta\bar{G}_N(Y_{nN})) \\ &\qquad\qquad\qquad\qquad\quad {\scriptstyle j \neq n} \\ &= 1 - \textstyle\sum_{n=1}^N \displaystyle\int_{\left\{y | \bar{G}_N(y) > \frac{1}{N\beta}\right\}} P(T_{nN}(y) < N\beta\bar{G}_N(y))\, dG_{nN}(y). \end{aligned} \tag{2.40}$$

Let $i_0 = i_0(N, \beta) \in \mathbb{N} \cup \{0\}$ be such that

$$\frac{i_0}{N\beta} < 1 \quad \text{and} \quad \frac{i_0 + 1}{N\beta} \geq 1 \tag{2.41}$$

and assume without loss of generality that $\beta N > 1$ so that

$$1 \leq i_0. \tag{2.42}$$

Next, define for $i = 1, 2, 3, \cdots, i_0 - 1$

$$\begin{aligned} I_i &= \left\{y \,\middle|\, \frac{i}{N\beta} < \bar{G}_N(y) \leq \frac{i+1}{N}\right\} = \{y \,|\, i < N\beta\bar{G}_N(y) \leq i + 1\} \\ I_{i_0} &= \left\{y \,\middle|\, \frac{i_0}{N\beta} < \bar{G}_N(y) \leq 1\right\} = \{y \,|\, i_0 < N\beta\bar{G}_N(y) \leq N\beta\}. \end{aligned} \tag{2.43}$$

Since $\left\{y \,\middle|\, \bar{G}_N(y) > \frac{1}{N\beta}\right\} = \cup_{i=1}^{i_0} I_i$, $I_i \cap I_j = \emptyset$ for $i \neq j$, and on I_i

$$P(T_{nN}(y) < N\beta\bar{G}_N(y)) = P(T_{nN}(y) \leq i), \tag{2.44}$$

it is clear from (2.40) that

$$1 - \tilde{P}_1 \leq \textstyle\sum_{n=1}^N \sum_{i=1}^{i_0} \displaystyle\int_{I_i} P(T_{nN}(y) \leq i)\, dG_{nN}(y). \tag{2.45}$$

Now, let Z be a binomial $(N - 1, \bar{p})$ rv and let us define

$$h(y) \equiv y(\log y - 1) + 1, \tag{2.46}$$

$$f(y) \equiv yh(y^{-1}) = y + \log(y^{-1}) - 1. \tag{2.47}$$

For $0 < \beta \le 1/3$,

$$(N-1)\bar{p} = N\bar{G}_N(y) - G_{nN}(y) > \frac{i}{\beta} - 1 \ge 3i - 1 \ge i + 1 > i \tag{2.48}$$

on I_i, $i = 1, \cdots, i_0$. Thus, from Hoeffding's (1956) Theorem 4 with $c = i$, we have for $y \in I_i$

$$\begin{aligned} P(T_{nN}(y) \le i) &\le P(Z \le i) \\ &\le P(Z/((N-1)\bar{p}) \le i/((i/\beta) - 1)). \end{aligned} \tag{2.49}$$

Lemma 1(ii) of Wellner (1978) implies that

$$P\left(\frac{Z}{(N-1)\bar{p}} \le \frac{i}{(i/\beta) - 1}\right) \le \exp\left(-(\mathrm{N}-1)\bar{\mathrm{p}}\mathrm{h}\left(\frac{\mathrm{i}}{(1/\beta) - 1}\right)\right) \tag{2.50}$$

where $h(y)$ is defined in (2.46). Hence from (2.48)–(2.50) and with $f(y)$ as defined in (2.47) we find

$$\begin{aligned} P(T_{nN}(y) \le i) &\le \exp\left(-\left(\frac{i}{\beta} - 1\right)h\left(\frac{i}{(i/\beta) - 1}\right)\right) \\ &= \exp(-if(\beta^{-1} - i^{-1})) \\ &= e^1 \cdot \left(1 - \frac{\beta}{i}\right)^i \exp\left(-i\left(\frac{1}{\beta} + \log \beta - 1\right)\right) \\ &\le e^1 \cdot \exp(-if(1/\beta)). \end{aligned} \tag{2.51}$$

Finally, (2.45), (2.51) and the definition of I_i yields for $\beta \in (0, 1/3]$

$$\begin{aligned} 1 - \tilde{P}_1 &\le \textstyle\sum_{n=1}^N \sum_{i=1}^{i_0} e \cdot \exp(-if(1/\beta)) G_{nN}\{I_i\} \\ &\le e \textstyle\sum_{i=1}^{i_0} \exp(-if(1/\beta)) N\bar{G}_N\{I_i\} \\ &\le \frac{e}{\beta} \textstyle\sum_{i=1}^{\infty} \exp(-if(1/\beta)) \\ &= \frac{e}{\beta} \exp(-f(1/\beta)) \cdot (1 - \exp(-f(1/\beta)))^{-1} \\ &\le e^2\beta^{-2} \exp(-1/\beta) \cdot 2 \\ &\le 15\beta^{-2} \exp(-1/\beta). \end{aligned} \tag{2.52}$$

Since $15\beta^{-2}e^{-1/\beta} \ge 1$ for $\beta \in (1/3, 1)$, the inequality holds for all $\beta \in (0, 1)$. □

The author is indebted to Jon A. Wellner, who pointed out how Lemma 1(ii) of Wellner (1978) can be used in the proof of Theorem 2.5 to replace the earlier bound $1 - 9\beta^{-1} \exp(-(32\beta)^{-1})$—which was based on Bernstein's inequality—by the present one.

From (2.12), (2.39) and the proof of Theorem 2.5 it is clear that the following generalization of Theorem 2.5 to the two-dimensional non-i.i.d. case holds true:

THEOREM 2.6. *For $\beta \in (0, 1/3]$, $N \in \mathbb{N}$ and continuous underlying* df's $F_{1N}, F_{2N}, \cdots, F_{NN}$, *we have*

$$\begin{aligned} \tilde{P}_2 &\equiv P(\mathbb{F}_N(x) \ge \beta\bar{F}_N(x), \quad \text{for} \quad x \in (\cup_{n=1}^N \tilde{\mathcal{O}}_{nN}^2) \cup \tilde{\mathcal{O}}_N^2) \\ &= P\left(\cap_{n=1}^N \left[\max\left(\frac{1}{N} \#[X_{jN} < X_{nN}], \frac{1}{N}\right) \ge \beta\bar{F}_N(X_{nN})\right]\right) \\ &\ge 1 - Ne \textstyle\sum_{i=1}^{i_0} q_i \exp\left(-\frac{i}{\beta} - i \log \beta + i\right), \end{aligned} \tag{2.53}$$

where

$$q_i = P\left(\frac{i}{N\beta} < \bar{F}_N(\bar{X}_N) \le \frac{i+1}{N\beta}\right), \qquad i = 1, 2, \cdots, i_0 - 1,$$
$$q_{i_0} = P\left(\frac{i_0}{N\beta} < \bar{F}_N(\bar{X}_N) \le 1\right), \tag{2.54}$$

where the sets $\tilde{\mathcal{O}}^2_{nN}$ *and* $\tilde{\mathcal{O}}^2_N$ *are defined in* (2.6) *and* (2.8), i_0 *is defined in* (2.41) *and* $\bar{X}_N$ *has* df $\bar{F}_N$.

PROOF. The verification of the equality in (2.53) is straightforward from the definition of the sets $\tilde{\mathcal{O}}^2_{nN}$ and $\tilde{\mathcal{O}}^2_N$ and the fact that $\mathbb{F}_N$ is a stepfunction whereas the underlying df's are continuous. The inequality in (2.53) follows from a reasoning similar to the proof of Theorem 2.5. □

The lower bound in (2.53) does not depend on the underlying distribution functions again in the special case where $\bar{F}_N$ is the product of its marginal distribution functions:

COROLLARY 2.4. *For* $\beta \in (0, 1/3]$, $N \in \mathbb{N}$ *and continuous underlying* df's $F_{1N}, F_{2N}, \cdots, F_{NN}$, *with* $\bar{F}_N = \bar{G}_N \times \bar{H}_N$, *we have*

$$\tilde{P}_2 \ge 1 - \log(N\beta)\cdot 15\beta^{-2}\exp(-\beta^{-1}). \tag{2.55}$$

PROOF. Note that in this situation the density of $\bar{F}_N(\bar{X}_N)$ is given above (2.28) so that for $i = 1, 2, \cdots, i_0 - 1$

$$P\left(\frac{i}{N\beta} < \bar{F}_N(\bar{X}_N) \le \frac{i+1}{N\beta}\right) \le P\left(\frac{1}{N\beta} < \bar{F}_N(\bar{X}_N) \le \frac{2}{N\beta}\right) \le \frac{\log(N\beta)}{N\beta}. \tag{2.56}$$

The proof can be completed with the aid of (2.56) and a reasoning as in (2.52). □

COROLLARY 2.5. *For* $\beta \in (0, 1/3]$, $N \in \mathbb{N}$ *and continuous* df's $F_{1N}, F_{2N}, \cdots, F_{NN}$, *we have*

$$\tilde{P}_2 \ge 1 - 15N\beta^{-1}\exp(-\beta^{-1}). \tag{2.57}$$

PROOF. The reasoning of the proof of Corollary 2.4 can be followed, with now for $i = 1, 2, \cdots, i_0$

$$P\left(\frac{i}{N\beta} < \bar{F}_N(\bar{X}_N) \le \frac{i+1}{N\beta}\right) \le 1. \quad \square \tag{2.58}$$

The following corollary is very useful in statistics and demonstrates the strength of Theorem 2.6.

COROLLARY 2.6. *For every* $\delta > 0$ *there exist* $N_0 \in \mathbb{N}$ *and* $K = K(\delta) \in (0, \infty)$ *such that for every array of continuous underlying* df's $F_{1N}, F_{2N}, \cdots, F_{NN}$, $N \ge N_0$, *and for every* $N \ge N_0$ *we have*

$$\begin{aligned} &P(\mathbb{F}_N(x) \ge (K\log N)^{-1}\bar{F}_N(x), \quad \text{for} \quad x \in \{X_{1N}, X_{2N}, \cdots, X_{NN}\}) \\ &\ge P(\mathbb{F}_N(x) \ge (K\log N)^{-1}\bar{F}_N(x), \quad \text{for} \quad x \in (\cup_{n=1}^N \tilde{\mathcal{O}}^2_{nN}) \cup \tilde{\mathcal{O}}^2_N) \\ &\ge 1 - N^{-1-\delta}. \end{aligned} \tag{2.59}$$

PROOF. Immediate from Corollary 2.5 with $\beta = (K\log N)^{-1}$. □

REMARK 2.3. Exploiting the approach of this paper it is possible to also obtain lower bounds for

$$(2.60)\qquad Q_2 \equiv P(\mathbb{F}_N(x) \le \beta^{-1}\bar{F}_N(x), \quad \text{for} \quad x \in \mathbb{R}^2)$$

and

$$(2.61)\qquad \tilde{Q}_2 \equiv P(\mathbb{F}_N(x) \ge \beta\bar{F}_N(x), \quad \text{for} \quad x \in \mathbb{R}^2 \quad \text{with} \quad \mathbb{F}_N(x) \ne 0),$$

i.e., to obtain still further generalizations of the Theorems 2.4 and 2.6. As an example let us mention without proofs the analogues of Theorem 2.6 with its Corollaries 2.5 and 2.6 for Q_2, which the author obtained jointly with J. Beirlant.

THEOREM 2.7. *For $\beta \in (0, \frac{1}{4}]$, $N \in \mathbb{N}$ and continuous underlying* df's $F_{1N}, F_{2N}, \cdots, F_{NN}$, *we have*

$$(2.62)\qquad \tilde{Q}_2 \ge 1 - e^2 \textstyle\sum_{i=1}^{i_0} \tilde{q}_i \exp(-i\beta^{-1} - i\log\beta + i) - 30\beta^{-2}\exp(-\beta^{-1}),$$

where i_0 is defined in (2.41) *and where*

$$(2.63)\qquad \begin{aligned} \tilde{q}_i &= \sum_{\substack{m=1\\ m\ne n}}^N \sum_{n=1}^N P\left(\frac{i}{N\beta} < \bar{F}_N(Y_{nN}, Z_{mN}) \le \frac{i+1}{N\beta}\right), \qquad i = 1, 2, \cdots, i_0 - 1, \\ \tilde{q}_{i_0} &= \sum_{\substack{m=1\\ m\ne n}}^N \sum_{n=1}^N P\left(\frac{i_0}{N\beta} < \bar{F}_N(Y_{nN}, Z_{mN}) \le 1\right). \end{aligned}$$

COROLLARY 2.7. *For $\beta \in (0, \frac{1}{4}]$, $N \in \mathbb{N}$ and continuous underlying* df's $F_{1N}, F_{2N}, \cdots, F_{NN}$, *we have*

$$(2.64)\qquad \tilde{Q}_2 \ge 1 - \{30\beta^{-1} + 41N(N-1)\}\beta^{-1}\exp(-\beta^{-1}).$$

COROLLARY 2.8. *For every $\delta > 0$ there exist $N_0 \in \mathbb{N}$ and $K = K(\delta) \in (0, \infty)$ such that for every array of continuous underlying* df's $F_{1N}, F_{2N}, \cdots, F_{NN}$, $N \ge N_0$ *and for every $N \ge N_0$ we have*

$$(2.65)\qquad P(\mathbb{F}_N(x) \ge (K\log N)^{-1}\bar{F}_N(x), \quad \text{for} \quad x \in \mathbb{R}^2 \quad \text{with} \quad \mathbb{F}_N(x) \ne 0) \ge 1 - N^{-1-\delta}.$$

3. Other theorems on multivariate empirical df's. The motivation for the theorems in this section is given in the introduction. See also Remark 3.1.

The first theorem is an extension of a result of Ghosh (1972) in the 1-dim i.i.d. case to the 2-dim. non-i.i.d. case. We refer to Ruymgaart and van Zuijlen (1978b) for the 1-dim. non-i.i.d. case.

THEOREM 3.1. *For every $\delta \ge 0$ there exist $K = K(\delta) \in (0, \infty)$ and $N_0 = N_0(\delta) \in \mathbb{N}$ such that for every array of continuous underlying* df's $F_{1N}, F_{2N}, \cdots, F_{NN}$, $N = 1, 2, \cdots$ *and for every $N \ge N_0$ we have*

$$(3.1)\qquad P\left(\sup_{\{x\in\mathbb{R}^2 \mid \bar{F}_N(x)\in[3/N, 1-(3/N)]\}} \frac{N^{1/2}\,|\,\mathbb{F}_N(x) - \bar{F}(x)\,|}{(\bar{F}_N(x)(1-\bar{F}_N(x)))^{1/2}} \le K\log N\right) \ge 1 - N^{-1-\delta}.$$

PROOF. Define an inverse of a df G on $(-\infty, \infty)$ by

$$(3.2)\qquad G^{-1}(u) = \inf\{y \mid G(y) \ge u\}, \quad \text{for} \quad 0 < u \le 1,$$

whereas $G^{-1}(0) = -\infty$ and $G^{-1}(a) = \infty$ for $a > 1$. Moreover, let for $i, j \in \{1, 2, \cdots, N+1\}$, $N \in \mathbb{N}$, $x \in \bar{\mathbb{R}}^2$,

$$a_{i,N} = \bar{G}_N^{-1}(iN^{-1}); \qquad b_{i,N} = \bar{H}_N^{-1}(iN^{-1}),$$

$$u_{i,N}^j = (a_{i,N}, b_{j,N}) \in \bar{\mathbb{R}}^2,$$

$$D_{iN} = \left\{ x \in \mathbb{R}^2 \mid \bar{F}_N(x) \in \left[\frac{i}{N}, 1 - \frac{i}{N}\right] \right\}, \tag{3.3}$$
$$R_N = \{u^j_{i,N} \in \bar{\mathbb{R}}^2 \mid i, j = 3, 4, \cdots, N+1\},$$
$$U_N(x) = \frac{\mathbb{F}_N(x) - \bar{F}_N(x)}{(\bar{F}_N(x)(1 - \bar{F}_N(x)))^{1/2}}.$$

For $N \geq 6$, $i, j \in \{4, 5, \cdots, N+1\}$ and $x \in ([a_{i-1,N}, a_{i,N}] \times [b_{j-1,N}, b_{j,N}]) \cap D_{3N}$ we have

$$U_N(x) \leq \frac{\mathbb{F}_N(u^j_{i,N}) - \bar{F}_N(u^{j-1}_{i-1,N})}{(\bar{F}_N(u^{j-1}_{i-1,N})(1 - \bar{F}_N(u^j_{i,N})))^{1/2}}$$
$$= \left(\frac{\bar{F}_N(u^j_{i,N})}{\bar{F}_N(u^{j-1}_{i-1,N})}\right)^{1/2} \frac{\mathbb{F}_N(u^j_{i,N}) - \bar{F}_N(u^j_{i,N})}{(\bar{F}_N(u^j_{i,N})(1 - \bar{F}_N(u^j_{i,N})))^{1/2}} + \frac{\bar{F}_N(u^j_{i,N}) - \bar{F}_N(u^{j-1}_{i-1,N})}{(\bar{F}_N(u^{j-1}_{i-1,N})(1 - \bar{F}_N(u^j_{i,N})))^{1/2}}$$
$$\leq (2+1)^{1/2} U_N(u^j_{i,N}) + 6N^{-1/2}$$

and

$$U_N(x) \geq \frac{\mathbb{F}_N(u^{j-1}_{i-1,N}) - \bar{F}_N(u^j_{i,N})}{(\bar{F}_N(u^j_{i,N})(1 - \bar{F}_N(u^{j-1}_{i-1,N})))^{1/2}}$$
$$= \left(\frac{\bar{F}_N(u^{j-1}_{i-1,N})}{\bar{F}_N(u^j_{i,N})}\right)^{1/2} \frac{\mathbb{F}_N(u^{j-1}_{i-1,N}) - \bar{F}_N(u^{j-1}_{i-1,N})}{(\bar{F}_N(u^{j-1}_{i-1,N})(1 - \bar{F}_N(u^{j-1}_{i-1,N})))^{1/2}} - \frac{\bar{F}_N(u^j_{i,N}) - \bar{F}_N(u^{j-1}_{i-1,N})}{(\bar{F}_N(u^j_{i,N})(1 - \bar{F}_N(u^{j-1}_{i-1,N})))^{1/2}}$$
$$\geq \frac{1}{\sqrt{3}} U_N(u^{j-1}_{i-1,N}) - 4N^{-1/2},$$

so that

$$\sup_{D_{3N}} |U_N(x)| \leq \sqrt{3} \max_{R_N \cap D_{1N}} |U_N(x)| + 6N^{-1/2}. \tag{3.4}$$

Hence, for proving (3.1) it is sufficient to show that for every $\delta \geq 0$ there exist $K = K(\delta) \in (0, \infty)$ and $N_0 = N_0(\delta) \in \mathbb{N}$ such that for every array of continuous underlying df's F_{1N}, $F_{2N}, \cdots, F_{NN}$, $N = 1, 2, \cdots$ and for every $N \geq N_0$ we have

$$P(\max_{R_N \cap D_{1N}} |N^{1/2} U_N(x)| \geq K \log N) \leq N^{-1-\delta}. \tag{3.5}$$

It follows from Bonferroni's inequality and Theorem 5 in Hoeffding (1956) that the probability in (3.5) is bounded above by

$$\textstyle\sum_{R_N \cap D_{1N}} P(|N\mathbb{F}_N(x) - N\bar{F}_N(x)| \geq KN^{1/2} \log N(\bar{F}_N(x)(1 - \bar{F}_N(x)))^{1/2}) \tag{3.6}$$

where $N\mathbb{F}_N(x)$ has a binomial distribution with parameters N and $\bar{F}_N(x) \in \left[\frac{1}{N}, 1 - \frac{1}{N}\right]$. Applying Bernstein's inequality and proceeding as in Ghosh (1972, page 352), with $K_1/\sqrt{2}$ replaced by K, we find for $N \geq N_0(K)$ the following upper bound for (3.6)

$$2 \textstyle\sum_{R_N \cap D_{1N}} \exp(-\frac{1}{8} K\sqrt{2} \log N) \leq 2N^2 \exp(-\frac{1}{8} K \log N) = 2N^{2-K/8}, \tag{3.7}$$

which is less than $N^{-1-\delta}$ for $K \geq 32 + 8\delta$. □

The second result is an extension of the first part of Theorem 3.2 in Ruymgaart and van Zuijlen (1978) to the 2-dim. non-i.i.d. case. See also van Zuijlen (1978) for a weaker, but similar multivariate result.

By an abuse of notation we write $\mathbb{F}_N$ and $\bar{F}_N$ for the measure induced by the df's, thus $\mathbb{F}_N\{B\} = \int_B d\mathbb{F}_N$, $\bar{F}\{B\} = \int_B d\bar{F}_N$ for a Borel set B in $\mathbb{R}^2$. An interval in $\mathbb{R}^2$ is defined as the product set of two intervals, closed, open or half open, bounded or unbounded, on the line.

THEOREM 3.2. *Let I be an interval in $\mathbb{R}^2$ and let $\mathcal{I} = \{I^*: I^*$ is an interval contained in $I\}$. For every $\delta \geq 0$ there exists $K = K(\delta)$ such that for every array of continuous underlying* df's $F_{1N}, F_{2N}, \cdots, F_{NN}$, $N = 3, 4, \cdots$, *for every* $N = 3, 4, \cdots$ *and for every interval I, with* $\bar{F}_N\{I\} \geq N^{-1}$, we have

$$\mathrm{P}\left(\sup_{I^* \in \mathcal{I}} | \mathbb{F}_N\{I^*\} - \bar{F}_N\{I^*\} | \leq K\left(\frac{\bar{F}_N\{I\}}{N}\right)^{1/2} \log N\right) \geq 1 - N^{-1-\delta}. \tag{3.8}$$

PROOF. We follow the lines of the proof of Theorem 2.1 in van Zuijlen (1978). We start by applying Lemma 2.1 in the above mentioned paper with $k = 2$ and obtain in the notation of this lemma for $K \geq 8$, $N \geq 3$

$$\begin{aligned}
&P\left(\sup_{I \in \mathcal{I}} | \mathbb{F}_N\{I^*\} - \bar{F}_N\{I^*\} | \leq K\left(\frac{\bar{F}_N\{I\}}{N}\right)^{1/2} \log N\right) \\
&\geq P\left(\max_{\tilde{I}_N \in \tilde{\mathcal{I}}_N} | \mathbb{F}_N\{\tilde{I}_N\} - \bar{F}_N\{\tilde{I}_N\} | \leq K\left(\frac{\bar{F}_N(I)}{N}\right)^{1/2} \log N - 4\frac{\bar{F}_N\{I\}}{N}\right) \\
&\geq P\left(\max_{\tilde{I}_N \in \tilde{\mathcal{I}}_N} | \mathbb{F}_N\{\tilde{I}_N\} - \bar{F}_N\{\tilde{I}_N\} | \leq \frac{1}{2} K\left(\frac{\bar{F}_N\{I\}}{N}\right)^{1/2} \log N\right) \\
&\geq 1 - \textstyle\sum_{\tilde{I}_N \in \tilde{\mathcal{I}}_N} P(| N\mathbb{F}_N\{\tilde{I}_N\} - N\bar{F}_N\{\tilde{I}_N\} | > \frac{1}{2}KN^{1/2}(\bar{F}_N\{I\})^{1/2} \log N).
\end{aligned} \tag{3.9}$$

Since $\frac{1}{2}KN^{1/2}(\bar{F}_N\{I\})^{1/2} \log N \geq 1$, Theorem 5 in Hoeffding (1956) is applicable, so that we may assume $N\mathbb{F}_N\{\tilde{I}_N\}$ in (3.9) to be a binomial rv with parameters N and $\bar{F}_N\{\tilde{I}_N\}$. From Bernstein's inequality (see, e.g., Bahadur (1966), page 578) it follows that for $K \geq 8$, $N \geq 3$ we have

$$\begin{aligned}
&P(| N\mathbb{F}_N\{\tilde{I}_N\} - N\mathbb{F}_N\{\tilde{I}_N\} | > \tfrac{1}{2}KN^{1/2}(\bar{F}_N\{I\})^{1/2} \log N) \\
&\leq 2 \exp\left(-\frac{\frac{1}{4}K^2 N\bar{F}_N\{I\}(\log N)^2}{2N\bar{F}_N\{\tilde{I}_N\} + \frac{1}{3}KN^{1/2}(\bar{F}_N\{I\})^{1/2} \log N}\right) \\
&\leq 2 \exp\left(-\frac{\frac{1}{4}K^2(\log N)^2}{2 + \frac{1}{3}K \log N(N\bar{F}_N\{I\})^{-1/2}}\right) \\
&\leq 2 \exp\left(-\frac{\frac{1}{4}K^2(\log N)^2}{2 \log N + \frac{1}{3}K \log N}\right) \\
&= 2N^{-(1/4)K^2/(2 + (1/3)K)}.
\end{aligned} \tag{3.10}$$

Noting that the number of elements in $\tilde{\mathcal{I}}_N$ is bounded above by $625N^8$, we obtain from (3.9) and (3.10)

$$P\left(\sup_{I^* \in \mathcal{I}} | \mathbb{F}_N\{I^*\} - \bar{F}_N\{I^*\} | \leq K\left(\frac{\bar{F}_N\{I\}}{N}\right)^{1/2} \log N\right) \geq 1 - 1250N^{8-(1/4)K^2/(2 + K/3)},$$

which completes the proof of the theorem. □

Finally, we shall give a useful theorem which is a direct consequence of Corollary 3.1 in Ruymgaart and van Zuijlen (1978b) (with $t - s$ replaced by $\bar{F}_N(t) - \bar{F}_N(s)$). We define for $0 \leq s < t \leq 1$

$$I_{s,t}^N = \{x \in \mathbb{R}^2 \mid \bar{F}_N(x) \in (s, t]\} \tag{3.11}$$

$$I_t^N = \{x \in \mathbb{R}^2 \mid \bar{F}_N(x) \leq t\} \tag{3.12}$$

$$X_N(t) = N^{1/2}(\mathbb{F}_N\{I_t^N\} - \bar{F}_N\{I_t^N\}). \tag{3.13}$$

THEOREM 3.3. *For every $\delta \geq 0$ there exists $K = K(\delta) > 0$ such that for every array of continuous underlying* df's $F_{1N}, F_{2N}, \cdots, F_{NN}$, $N = 3, 4, \cdots$, *for every $N = 3, 4, \cdots$ and for every $0 < \tau \leq 1$ we have, with s and t restricted to* $[0, 1]$,

$$P(\sup_{\{(s,t) \mid |\bar{F}_N\{I_t^N\} - \bar{F}_N\{I_s^N\}| \leq N^{-\tau}\}} |X_N(t) - X_N(s)| \leq KN^{-\tau/2} \log N) \geq 1 - N^{-1-\delta}. \tag{3.14}$$

PROOF. Let $\hat{W}_N$ be the empirical df based on $W_{1N}, W_{2N}, \cdots, W_{NN}$, where $W_{nN} \equiv \bar{F}_N(X_{nN})$, $n = 2, 1, \cdots, N$. The theorem is immediate from Corollary 3.1 in Ruymgaart and van Zuijlen (1978b) with $\hat{F}_N = \hat{W}_N$ and $\bar{F}_N(t) = 1/N \sum_{n=1}^N P(W_{nN} \leq t)$ for $t \in [0, 1]$ since

$$\bar{F}_N(X_{nN}) \leq t \Leftrightarrow X_{nN} \in I_t^N$$

and hence

$$\hat{W}_N(t) = N^{-1} \# [\bar{F}_N(X_{nN}) \leq t] = \frac{1}{N} \# [X_{nN} \in I_t^N] = \mathbb{F}_N\{I_t^N\},$$

$$N^{-1} \sum_{n=1}^N P(W_{nN} \leq t) = N^{-1} \sum_{n=1}^N P(\bar{F}_N(X_{nN}) \leq t) = N^{-1} \sum_{n=1}^N P(X_{nN} \in I_t^N) = \bar{F}_N\{I_t^N\} \square$$

REMARK 3.1 *Applications.* As an example we shall demonstrate first how Corollary 2.6 can be used in the theory of rank statistics.

For $N \in \mathbb{N}$ let $J_N: (0, 1) \to \mathbb{R}$ be such that

$$|J_N(s)| \leq C(s(1 - s))^{-\alpha} \equiv Cr^\alpha(s)$$

for some $C \in (0, \infty)$, $\alpha \in (0, \frac{1}{4})$ and let

$$\bar{D}_N = \{x \in \mathbb{R}^2 \mid \bar{F}_N(x) \leq 3N^{-1}\}.$$

One of the remainder terms in the decomposition of a suitably standardized Kendall-type statistic as given in (1.6) is B_N, where

$$B_N = N^{1/2} \int \int_{\bar{D}_N} J_N\left(\frac{N}{N + 1} \mathbb{F}_N\right) d\mathbb{F}_N. \tag{3.15}$$

Because of Corollary 2.6 we have on a subset, say $\tilde{\Omega}_N$, of Ω with very high probability

$$\begin{aligned} |B_N| &\leq N^{-1/2} \textstyle\sum_{X_{nN} \in \bar{D}_N} Cr^\alpha\left(\frac{N}{N+1} \mathbb{F}_N(X_{nN})\right) \\ &\leq C'N^{-1/2} \textstyle\sum_{X_{nN} \in \bar{D}_N} (K \log N)^\alpha (\bar{F}_N(X_{nN}))^{-\alpha} \\ &\leq C''N^{1/2} (\log N)^\alpha \int \int_{\bar{D}_N} \bar{F}_N^{-\alpha}\, d\mathbb{F}_N. \end{aligned} \tag{3.16}$$

With Hölder's inequality, taking $q = 1 + (4\alpha)^{-1}$, $p = (1 - q^{-1})^{-1}$, we find on $\tilde{\Omega}_N$ for (3.16) the upper bound

$$C''N^{1/2} (\log N)^\alpha (\mathbb{F}_N\{\bar{D}_n\})^{1/p} \left(\int \int \bar{F}_N^{-q\alpha}\, d\mathbb{F}_N \right)^{1/q}.$$

Since $p < 2$ and $q\alpha < \frac{1}{2}$, (3.15), (3.16) together with Theorem 3 in Sen (1970) yields for $\bar{F}_N \in \mathscr{F}_2$ *on* $\tilde{\Omega}_N$

$$\begin{aligned} |B_N| &\leq C'''N^{1/2} (\log N)^\alpha (\bar{F}_N\{\bar{D}_N\})^{1/p} (\log N)^{1/p} \\ &\leq C'''N^{1/2} (\log N)^{\alpha + 1/p} \left(\frac{3}{N} - \frac{3}{N} \log \frac{3}{N}\right)^{1/p} \to 0, \qquad \text{as} \quad N \to \infty. \end{aligned} \tag{3.17}$$

Note that this example demonstrates that we only have to control the multivariate empirical df in the observations X_{nN}, $n = 1, 2, \cdots, N$.

Finally, we shall give two corollaries of the theorems from this section which can be applied directly when dealing with the other remainder terms in the decomposition of Kendall-type statistics. To avoid technicalities Corollary 3.2 will be formulated under the hypothesis that $\bar{F}_N = \bar{G}_N \times \bar{H}_N$. For the proofs of these corollaries we refer to van Zuijlen (1979).

Let $D_N \equiv \left\{x \in \mathbb{R}^2 \mid \bar{F}_N(x) \in \left[\frac{3}{N}, 1 - \frac{3}{N}\right]\right\}$ and for $N \in \mathbb{N}$ let $\tilde{J}_N: (0, 1) \to \mathbb{R}$ be such that

$$|\tilde{J}_N^{(i)}| \le Mr^{\alpha+i+1}, \qquad i = 0, 1$$

for some $M \in (0, \infty)$, $\alpha \in (0, \frac{1}{4})$, where $\tilde{J}_N^{(0)} \equiv \tilde{J}_N$ and where $\tilde{J}_N^{(1)}$ denotes the derivative of the measurable functions $\tilde{J}_N$.

COROLLARY 3.1. *For every $\delta \ge 0$ there exist $K = K(\delta) \in (0, \infty)$ and $N_0 = N_0(\alpha) \in \mathbb{N}$ such that for every array of continuous underlying* df's $F_{1N}, F_{2N}, \cdots, F_{NN}$, $N \ge N_0$ *and for every $N \ge N_0$ we have*

$$P\left(\int\int_{D_N} (r(\bar{F}_N))^{3/4+\alpha} d\mathbb{F}_N \le \int\int_{D_N} (r(\bar{F}_N))^{3/4+\alpha}\, d\bar{F}_N + 2\right) \ge 1 - N^{-1-\delta}. \tag{3.18}$$

COROLLARY 3.2. *For every $\delta \ge 0$ there exist $N_0 = N_0(\alpha, \delta) \in \mathbb{N}$ and $K_\alpha = K_\alpha(\delta) \in (1, \infty)$ such that for every array of continuous underlying df's $F_{1N}, F_{2N}, \cdots, F_{NN}$, with $\bar{F}_N = \bar{G}_N \times \bar{H}_N$, $N \ge N_0$ and for every $N \ge N_0$ we have*

$$P\left(\left|\int\int_{D_N} V_N \tilde{J}_N(\bar{F}_N)\, d(\mathbb{F}_N - \bar{F}_N)\right| \le K_\alpha M N^{\alpha/4-1/16}\right) \ge 1 - N^{-1-\delta}, \tag{3.19}$$

where

$$V_N(x) = N^{1/2}(\mathbb{F}_N(x) - \bar{F}_N(x)), \qquad x \in \bar{\mathbb{R}}^2. \tag{3.20}$$

Acknowledgement. The author is indebted to W. R. van Zwet for some helpful and stimulating discussions. Also note the acknowledgement of Wellner below the proof of Theorem 2.5.

REFERENCES

[1] BAHADUR, R. R. (1966). A note on quantiles in large samples. *Ann. Math. Statist.* **37** 577–580.
[2] BHUCHONGKUL, S. (1964). A class of nonparametric tests for independence in bivariate populations. *Ann. Math. Statist.* **35** 138–149.
[3] CHANG, LI-CHIEN (1964). On the ratio of the empirical distribution function to the theoretical distribution function. *Selected Transl. Math. Statist. Probability*. Amer. Math. Soc. **4** 17–38 Providence, Rhode Island.
[4] DANIELS, H. E. (1945). The statistical theory of the strength of bundles of threads, I. *Proc. Roy. Soc. London Ser. A* **183** 405–435.
[5] GHOSH, M. (1972). On the representation of linear functions of order statistics. *Sankhyā* **34** 349–356.
[6] GOVINDARAJULU, Z., LE CAM, L. and RAGHAVACHARI, M. (1967). Generalization of theorems of Chernoff-Savage on asymptotic normality of nonparametric test statistics. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* 609–638. Univ. California Press.
[7] HOEFFDING, W. (1956). On the distribution of the number of successes in independent trials. *Ann. Math. Statist.* **27** 713–721.
[8] KIEFER, J. and WOLFOWITZ, J. (1958). On the deviations of the empiric distribution function of vector chance variables. *Trans. Amer. Math. Soc.* **87** 173–186.
[9] KIEFER, J. (1961). On large deviations of the empiric df of vector chance variables and a law of the iterated logarithm. *Pacific J. Math.* **11** 649–660.
[10] KOUL, H. L. (1970). A class of ADF tests for subhypothesis in the multiple linear regression. *Ann. Math. Statist.* **41** 1273–1281.

[11] KOUL, H. L. and STAUDTE, R. G. (1972). Weak convergence of weighted empirical cumulatives based on ranks. *Ann. Math. Statist.* **43** 832–841.

[12] LEHMANN, E. L. (1966). Some concepts of dependence. *Ann. Math. Statist.* **37** 1137–1153.

[13] ROBBINS, H. (1954). A one-sided confidence interval for an unknown distribution function. *Ann. Math. Statist.* **25** 409.

[14] ROSENBLATT, M. (1952). Remarks on a multivariate transformation. *Ann. Math. Statist.* **23** 470–472.

[15] RUYMGAART, F. H., SHORACK, G. R. and VAN ZWET, W. R. (1972). Asymptotic normality of nonparametric tests for independence. *Ann. Math. Statist.* **43** 1122–1135.

[16] RUYMGAART, F. H. (1973). Asymptotic theory of rank tests for independence. *Mathematical Centre Tracts* **43** Amsterdam.

[17] RUYMGAART, F. H. (1974). Asymptotic normality of nonparametric tests for independence. *Ann. Statist.* **5** 892–910.

[18] RUYMGAART, F. H. and VAN ZUIJLEN, M. C. A. (1977). Asymptotic normality of linear combinations of functions of order statistics in the non-i.i.d. case. *Indag. Math.* **80** 432–447.

[19] RUYMGAART, F. H. and VAN ZUIJLEN, M. C. A. (1978a). Asymptotic normality of multivariate linear rank statistics in the non-i.i.d. case. *Ann Statist.* **6** 588–602.

[20] RUYMGAART, F. H. and VAN ZUIJLEN, M. C. A. (1978b). On convergence of the remainder term in linear combinations of functions of order statistics in the non-i.i.d. case. *Sankhyā Ser. A* **40** 369–387.

[21] SEN, P. K. (1970). On some convergence properties of one-sample rank order statistics. *Ann. Math. Statist.* **41** 2140–2143.

[22] SHORACK, G. R. (1972). Functions of order statistics. *Ann. Math. Statist.* **43** 412–427.

[23] SHORACK, G. R. and WELLNER, J. A. (1978). Linear bounds on the empirical distribution function. *Ann. Probability* **6** 349–353.

[24] WELLNER, J. A. (1978). Limit theorems for the ratio of the empirical distribution function to the true distribution function. *Z. Wahrsch. verw. Gebiete* **45** 73–88.

[25] ZUIJLEN, M. C. A. VAN (1976). Some properties of the empirical distribution function in the non-i.i.d. case. *Ann Statist.* **4** 406–408.

[26] ZUIJLEN, M. C. A. VAN (1978). Properties of the empirical distribution function for independent nonidentically distributed random variables. *Ann. Probability* **6** 250–266.

[27] ZUIJLEN, M. C. A. VAN (1979). Fundamental properties of the empirical distribution function in the two-dimensional non-i.i.d. case. Report 7931 Math. Inst. Katholieke Universiteit, Nijmegen.

MATHEMATISCH INSTITUUT
KATH. UNIV. NIJMEGEN
TOERNOOIVELD
6525 ED NIJMEGEN
THE NETHERLANDS