Z. Wahrscheinlichkeitstheorie verw. Gebiete
61, 417-430 (1982)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete

On Functions Bounding the Empirical Distribution of Uniform Spacings

J. Beirlant[1], E.C. van der Meulen[1], F.H. Ruymgaart[2] and M.C.A. van Zuijlen[2]

[1] Department of Mathematics, Katholieke Universiteit Leuven, Belgium
[2] Department of Mathematics, Katholieke Universiteit Nijmegen, The Netherlands

Summary. In two different ways a result is proved on the inclusion between a pair of non-random functions of the empirical distribution function based on uniform spacings. Applications in nonparametric statistics are discussed.

1. Introduction

It is the purpose of this paper to prove a result on the inclusion of the empirical distribution function (df) based on uniform spacings, between a pair of non-random functions by two different methods. The first method is directly based on the well-known property that normalized spacings formed from ordered uniform spacings are themselves distributed as uniform spacings. It subsequently enables us to convert the problem into a similar one for the ordinary empirical df, based on the original independent uniform random variables (rv's), so that the sharp results in Robbins (1954) and Wellner (1978) can be applied.

The main tool in the second approach is an upper bound for the moment-generating function (mgf) of a certain sum of dependent zero-one rv's in terms of the mgf of the corresponding sum of independent zero-one rv's, which is based on the negative dependence structure of uniform spacings. This method is self-contained and is of interest in its own right since it can be used in other situations as well.

The results we obtain with these two methods described above are comparable and close to the classical results in case of an empirical df based on independent and identically distributed rv's.

To be more precise, let $X_1, X_2, \ldots, X_{N-1}$ $(N \geqq 2)$ be a random sample from the uniform $(0,1)$ distribution and let

$$0 \equiv X_{0:N-1} \leqq X_{1:N-1} \leqq \cdots \leqq X_{N-1:N-1} \leqq X_{N:N-1} \equiv 1$$

0044-3719/82/0061/0417/$02.80

be the ordered sample. The uniform spacings are then defined by

$$D_{n,N} = X_{n:N-1} - X_{n-1:N-1} \qquad (n = 1, 2, \ldots, N)$$

whereas for the ordered spacings we use the notation

$$D_{1:N} \leqq D_{2:N} \leqq \ldots \leqq D_{N:N}.$$

It is well-known that the uniform spacings are exchangeable rv's (see Pyke (1965)). The common marginal df is

$$F_N(s) = P(D_{1,N} \leqq s) = P(X_{1:N-1} \leqq s) = 1 - (1-s)^{N-1}, \qquad s \in [0, 1] \tag{1.1}$$

and the empirical df based on the uniform spacings is as usual defined by

$$\hat{F}_N(s) = N^{-1} \sum_{n=1}^{N} 1_{[0,s]}(D_{n,N}), \qquad s \in [0, 1]$$

where $1_S(\cdot)$ denotes the indicator function of a set S.

We shall develop the two approaches described above in the Sects. 2 and 3 in order to obtain lower bounds for the probabilities P_1 and P_2 defined by

$$P_1 = P(\hat{F}_N(s) \leqq \beta^{-1} F_N(s), \quad \textit{for } s \in [0, 1]),$$
$$P_2 = P(\hat{F}_N(s) \geqq \beta F_N(s), \quad \textit{for } s \in [D_{1:N}, 1]), \ \beta \in (0, 1),$$

whereas in Sect. 4 we shall discuss the applicability of these results towards statistics.

2. The First Approach

Throughout the paper the symbol $\mathscr{L}$ will be used to denote the law of a random variable or vector.

The main tool in this section is the well-known relation

$$\mathscr{L}(D_{1,N}, D_{2,N}, \ldots, D_{N,N}) = \mathscr{L}(N D_{1:N}, (N-1)(D_{2:N} - D_{1:N}), \ldots, D_{N:N} - D_{N-1:N}), \tag{2.1}$$

see e.g. Karlin (1966, p. 264) or Pyke (1965).

It is immediate from (2.1) that

$$\mathscr{L}(D_{1:N}, D_{2:N}, \ldots, D_{N:N}) = \mathscr{L}(T_1, T_2, \ldots, T_N) \tag{2.2}$$

where

$$T_n = \sum_{m=1}^{n} (X_{m:N-1} - X_{m-1:N-1})/(N-m+1), \qquad n = 1, 2, \ldots, N.$$

It follows by simple algebra that

$$X_{n:N-1}/N \leqq T_n \leqq X_{n:N-1}/(N-n+1); \qquad n = 1, 2, \ldots, N. \tag{2.3}$$

It is already clear from relation (2.3) that *with probability one the empirical df* $\hat{F}_N$ *lies above the polygonal line connecting the points* $(1/N, 0)$, $(1/(N-1), 1/N)$, $\ldots$, $(1, (N-1)/N)$.

For the evaluation of lower bounds for P_1 and P_2 we also need an approximation of the inverse

$$F_N^{-1}(s) = 1-(1-s)^{1/(N-1)}, \qquad \text{for } s \in [0,1); \; F_N^{-1}(1) = 1,$$

of F_N in (1.1). Using the expansion for the binomial series we find that

$$1-(1-s)^{1/(N-1)} = s/(N-1) + \rho_N(s), \qquad \text{for } s<1,$$

where, provided $N \geqq 2$, the remainder term satisfies

$$0 \leqq \rho_N(s) \leqq s^2/((N-1)(1-s)), \qquad \text{for } 0<s<1. \tag{2.4}$$

Furthermore we use some results on the ordinary empirical df of the uniform rv's $\hat{\Gamma}_{N-1}$, say, viz.

$$P(\hat{\Gamma}_{N-1}(s) < (1/\gamma)s, \quad \textit{for } s \in [0,1]) = 1-\gamma \tag{2.5}$$

for every $\gamma \in (0,1)$ and $N \geqq 2$ (see Robbins (1954)), and

$$\begin{aligned} &P(\hat{\Gamma}_{N-1}(s) > \gamma s, \quad \textit{for } s \in [X_{1:N-1}, 1]) \\ &\qquad \geqq 1 - e(1/\gamma)\exp(-1/\gamma)) \end{aligned} \tag{2.6}$$

for every $\gamma \in (0,1)$ and $N \geqq 2$ (see Wellner (1978)).

We can then state the following theorems.

Theorem 2.1. *For* $\beta \in (0, \frac{1}{4})$, $N \geqq 2$ *we have*

$$P_1 \geqq 1 - \beta/(1-\beta).$$

Proof. We shall use (2.2) along with the first inequality in (2.3) and the second inequality in (2.4). Note that in particular $X_{N:N-1} = 1 > NF_N^{-1}(\beta)$ for $\beta \in (0, \frac{1}{4})$ and $N \geqq 2$. It follows that

$$\begin{aligned} P_1 &= P(F_N(D_{n:N}) \geqq \beta n/N, \quad \textit{for } n=1,2,\ldots,N) \\ &\geqq P(X_{n:N-1} \geqq NF_N^{-1}(\beta n/N), \quad \textit{for } n=1,2,\ldots,N-1) \\ &\geqq P(X_{n:N-1} \geqq (\beta/(1-\beta))(n/(N-1)), \quad \textit{for } n=1,2,\ldots,N-1) \\ &= P(\hat{\Gamma}_{N-1}(s) \leqq ((1-\beta)/\beta)s, \quad \textit{for } s \in [0,1]) \\ &= 1-\beta/(1-\beta), \end{aligned}$$

where in the last step we use (2.5) with $\gamma = \beta/(1-\beta)$. □

Theorem 2.2. *For* $\beta \in (0,1)$ *and* $N \geqq 2$ *we have*

$$P_2 \geqq 1 - e((1-\beta)/\beta)\exp(-(1-\beta)/\beta).$$

Proof. Again (2.2) will be used, but this time together with the second inequality in (2.3) and the first inequality in (2.4). We see that

$$\begin{aligned}P_2 &= P(F_N(D_{n:N}) \leqq (n-1)/(\beta N), \quad \textit{for } n=2,3,\ldots,[\beta N]+1) \\ &\geqq P(X_{n:N-1} \leqq (N-n+1)F_N^{-1}((n-1)/(\beta N)), \quad \textit{for } n=2,3,\ldots,[\beta N]+1) \\ &\geqq P(X_{n:N-1} \leqq ((1-\beta)/\beta)((n-1)/(N-1)), \quad \textit{for } n=2,3,\ldots,[\beta N]+1) \\ &= P(\hat{\Gamma}_{N-1}(s) \geqq (\beta/(1-\beta))s, \quad \textit{for } s \in [X_{1:N-1}, 1]) \\ &\geqq 1-e((1-\beta)/\beta)\exp(-(1-\beta)/\beta),\end{aligned}$$

where in the last step we use (2.6) with $\gamma=\beta/(1-\beta)$. □

3. The Second Approach

For $k \in \{1,2,\ldots,N\}$ we denote the k-dimensional df of a subset of k uniform spacings out of $D_{1,N},\ldots,D_{N,N}$ by

$$F_{k,N}(s_1,\ldots,s_k) = P(D_{1,N} \leqq s_1, D_{2,N} \leqq s_2, \ldots, D_{k,N} \leqq s_k), \quad \text{for all } (s_1,s_2,\ldots,s_k) \in [0,1]^k.$$

In particular remark that for $k=1$, $F_{1,N} \equiv F_N$ as defined in (1.1). For $k \in \{1,2,\ldots,N\}$ let

$$\bar{F}_{k,N}(s_1,\ldots,s_k) = P(D_{1,N} > s_1, D_{2,N} > s_2, \ldots, D_{k,N} > s_k), \quad \text{for all } (s_1,\ldots,s_k) \in [0,1]^k.$$

For notational convenience we will write $(x,x,\ldots,x)=x^{(i)}$ for $x^{(i)} \in \mathbb{R}^i$.

Let Dir $(\alpha_1,\alpha_2,\ldots,\alpha_m;\alpha_{m+1})$ denote a Dirichlet distribution with parameters $\alpha_1,\alpha_2,\ldots,\alpha_m;\alpha_{m+1}$. Recall that if $\mathscr{L}(Y_1,Y_2,\ldots,Y_m)=\mathrm{Dir}(\alpha_1,\alpha_2,\ldots,\alpha_m;\alpha_{m+1})$ then we have

$$\mathscr{L}\left(\frac{Y_2}{1-Y_1},\frac{Y_3}{1-Y_1},\ldots,\frac{Y_m}{1-Y_1}\,\middle|\,Y_1\right)=\mathrm{Dir}(\alpha_2,\alpha_3,\ldots,\alpha_m;\alpha_{m+1})$$

(see e.g. Wilks (1962, p. 180)). Moreover, from Wilks (1962, p. 179) it is also clear that

$$\mathscr{L}(D_{1,N},D_{2,N},\ldots,D_{N-1,N})=\mathrm{Dir}(1,1,\ldots,1;1).$$

Hence we have the following Lemma.

Lemma 3.1.

$$\mathscr{L}\left(\frac{D_{2,N}}{1-D_{1,N}},\frac{D_{3,N}}{1-D_{1,N}},\ldots,\frac{D_{N,N}}{1-D_{1,N}}\,\middle|\,D_{1,N}\right)=\mathscr{L}(D_{1,N-1},D_{2,N-1},\ldots,D_{N-1,N-1}),$$

where $(D_{1,N-1},D_{2,N-1},\ldots,D_{N-1,N-1})$ *is a vector of* $(N-1)$ *uniform spacings of a sample of size* $N-2$ *from the uniform* (0,1) *distribution.*

The main tool in this section is Corollary 3.1, which is based on the negative orthant dependence structure of uniform spacings. Following Block et al. (1980) a random vector $(Y_1, \ldots, Y_m)$ is said to be *negatively lower orthant dependent* (NLOD) if for every $(y_1, y_2, \ldots, y_m) \in \mathbb{R}^m$

$$P(Y_1 \leqq y_1, Y_2 \leqq y_2, \ldots, Y_m \leqq y_m) \leqq \prod_{i=1}^{m} P(Y_i \leqq y_i),$$

and is said to be *negatively upper orthant dependent* (NUOD) if for every $(y_2, y_2, \ldots, y_m) \in \mathbb{R}^m$

$$P(Y_1 > y_1, Y_2 > y_2, \ldots, Y_m > y_m) \leqq \prod_{i=1}^{m} P(Y_i > y_i).$$

Lemma 3.2. *The vector of uniform spacings* $(D_{1,N}, D_{2,N}, \ldots, D_{N,N})$ *is both* NLOD *and* NUOD, *which implies that for every* $k \in \{1, 2, \ldots, N\}$ *and every* $(s_1, \ldots, s_k) \in [0, 1]^k$ *we have*

$$F_{k,N}(s_1, \ldots, s_k) \leqq \prod_{i=1}^{k} F_{1,N}(s_i)$$

and

$$\bar{F}_{k,N}(s_1, \ldots, s_k) \leqq \prod_{i=1}^{k} \bar{F}_{1,N}(s_i).$$

Proof. This fact can easily be derived from Block et al. (1980, p. 32) since $(D_{1,N}, D_{2,N}, \ldots, D_{N,N})$ is uniformly distributed on the (negatively tilted) area

$$\left((s_1, s_2, \ldots, s_N) \geqq (0, \ldots, 0) \,\middle|\, \sum_{n=1}^{N} s_n = 1\right). \quad \square$$

Corollary 3.1. *For* $N \geqq 2$ *let* S_N *be a sum of* N *exchangeable zero-one rv's* $Z_1, Z_2, \ldots, Z_N$ *with* $E(Z_1) = p = F_N(s)$ *for certain* $s \in (0, 1)$. *Moreover suppose that*

$$E\left(\prod_{i=1}^{k} Z_i\right) = F_{k,N}(s^{(k)}), \qquad \textit{for } k = 2, 3, \ldots, N.$$

Then for $h \in \mathbb{R}$ *we have*

$$E(e^{hS_N}) \leqq E(e^{h\tilde{S}_N}),$$

where $\tilde{S}_N$ *is a Binomial* (N, p) *rv.*

Proof. The statement follows from Lemma 3.2 in view of Proposition 3.1 of Block et al. (1980) which says that a vector $(Y_1, \ldots, Y_m)$ is NLOD (resp. NUOD) if and only if

$$E\left(\prod_{i=1}^{m} \phi_i(Y_i)\right) \leqq \prod_{i=1}^{m} E(\phi_i(y_i))$$

whenever all ϕ_i are nonnegative and decreasing (resp. increasing). The lemma was also established independently in Beirlant et al. (1981). Their proof hinges on the expansion of the mgf's of S_N and $\tilde{S}_N$ in powers of h. $\square$

We next state the first main theorem of this section which is the analogue of Theorem 2.1 but slightly different and proved by a different method.

Theorem 3.1. *For* $\beta\in(0,\frac{1}{4})$, $N\geqq 2$ *we have*

$$P_1\geqq 1-\beta-(4e\beta^2/(1-4\beta\exp(1-4\beta))).$$

Proof. Following the approach of van Zuijlen (1982), we have

$$\begin{aligned}
P_1&=P(\hat{F}_N(D_{n,N})\leqq\beta^{-1}F_N(D_{n,N}))\\
&\geqq 1-\sum_{n=1}^{N} P(\hat{F}_N(D_{n,N})>\beta^{-1}F_N(D_{n,N}))\\
&=1-\sum_{n=1}^{N} P(\#[D_{\substack{i,N\\ i\neq n}}\leqq D_{n,N}]>N\beta^{-1}F_N(D_{n,N})-1)\\
&=1-\sum_{n=1}^{N}\int_0^1 P(\#[D_{\substack{i,N\\ i\neq n}}\leqq s]>N\beta^{-1}F_N(s)-1\,|\,D_{n,N}=s)\,dF_N(s)\\
&=1-N\int_0^1 P(\#[D_{\substack{i,N\\ 2\leqq i\leqq N}}\leqq s]>N\beta^{-1}F_N(s)-1\,|\,D_{1,N}=s)\,dF_N(s).
\end{aligned}\tag{3.1}$$

For $s\in(\frac{1}{2},1)$ we have by Lemma 3.1 that with probability one, given $D_{1,N}=s$, the number of $D_{i,N}$, $2\leqq i\leqq N$, not greater than s is equal to $N-1$. Moreover, $\beta\leqq F_N(s)$ for $\beta<\frac{1}{4}$ and $d\geqq\frac{1}{2}$, so that $N\beta^{-1}F_N(s)-1\geqq N-1$. So

$$P(\#[D_{\substack{i,N\\ 2\leqq i\leqq N}}\leqq s]>N\beta^{-1}F_N(s)-1\,|\,D_{1,N}=s)=0,\qquad\text{for } s\in(\tfrac{1}{2},1).$$

Hence, (3.1) and Lemma 3.1 imply that

$$P_1\geqq 1-N\int_0^{\frac{1}{2}} P(S_{N-1}>N\beta^{-1}F_N(s)-1)\,dF_N(s)\tag{3.2}$$

where S_{N-1} is a sum of $N-1$ exchangeable zero-one rv's $Z_1,\ldots,Z_{N-1}$ with

$$E\left(\prod_{i=1}^{k} Z_i\right)=F_{k,N-1}\left(\left(\frac{s}{1-s}\right)^{(k)}\right),\qquad k=1,2,\ldots,N-1.$$

Next, define for $j=0,1,2,\ldots,N-1$

$$\begin{aligned}
I_j&=\left\{s\in(0,\tfrac{1}{2})\,|\,F_N(s)\in\left[\frac{j\beta}{N},\frac{(j+1)\beta}{N}\right)\right\}\\
&=\{s\in(0,\tfrac{1}{2})\,|\,j-1\leqq N\beta^{-1}F_N(s)-1<j\},
\end{aligned}$$

and

$$I_N=\{s\in(0,\tfrac{1}{2})\,|\,F_N(s)\geqq\beta\}=\{s\in(0,\tfrac{1}{2})\,|\,N\beta^{-1}F_N(s)-1\geqq N-1\}.$$

On I_j, $j=0,1,\ldots,N$, we have

$$P(S_{N-1}>N\beta^{-1}F_N(s)-1)=P(S_{N-1}\geqq j).\tag{3.3}$$

Since it can be seen easily that for $s\in(0,\frac{1}{2})$

$$F_{N-1}(2s)\leqq F_N(2s)\leqq 2F_N(s),$$

we have on I_j, $j=1, 2, \ldots, N$, for $\beta<\frac{1}{4}$

$$(N-1)F_{N-1}\left(\frac{s}{1-s}\right)\leqq 2NF_N(s)<2N\frac{(j+1)\beta}{N}\leqq 4j\beta. \tag{3.4}$$

Now, using Markov's inequality and Corollary 3.1 we have on I_j, $j=1,2,\ldots,N$, that for every $h>0$

$$\begin{aligned}P(S_{N-1}\geqq j)&\leqq e^{-hj}E(e^{hS_{N-1}})\\&\leqq e^{-hj}E(e^{h\bar{S}_{N-1}})\\&=e^{-hj}\left[1-F_{N-1}\left(\frac{s}{1-s}\right)(1-e^h)\right]^{N-1}\\&\leqq\exp\left(-hj-(N-1)F_{N-1}\left(\frac{s}{1-s}\right)(1-e^h)\right).\end{aligned}$$

Since $0<\beta<\frac{1}{4}$, we can take $h=-\log(4\beta)$, and using (3.4) we have on I_j, $j=1,2,\ldots,N$,

$$P(S_{N-1}\geqq j)\leqq\exp(-j(-\log(4\beta)+4\beta-1)). \tag{3.5}$$

Writing for convenience $F_N(I_j)$ for the measure induced by F_N assigned to I_j, it follows from (3.2), (3.3), and (3.5) that for $\beta\in(0,\frac{1}{4})$,

$$\begin{aligned}P_1&\geqq 1-N\sum_{j=0}^{N}\int_{I_j}P(S_{N-1}\geqq j)\,dF_N(s)\\&\geqq 1-N\left\{F_N(I_0)+\sum_{j=1}^{N-1}\exp(-j(\log(4\beta)+4\beta-1))\,F_N(I_j)\right\}\\&\geqq 1-\beta-\beta\sum_{j=1}^{+\infty}\exp(-j(\log(4\beta)+4\beta-1))\\&=1-\beta-\frac{\beta\exp(\log(4\beta)-4\beta+1)}{(1-\exp(\log(4\beta)-4\beta+1))}\\&\geqq 1-\beta-(4e\beta^2/(1-4\beta\exp(1-4\beta))).\quad\square\end{aligned}$$

The section's analogue to Theorem 2.2 is the following

Theorem 3.2. *For $\beta\in(0,1)$, $N\geqq 2$, we have*

$$P_2\geqq 1-\frac{e\beta^{-1}\exp(-\beta^{-1})}{1-e\beta^{-1}\exp(-\beta^{-1})}.$$

Proof. Following the approach taken in van Zuijlen (1976) we have

$$\begin{aligned}P_2&=1-P\left(\bigcup_{n=2}^{N}\left(\frac{n-1}{\beta N}<F_N(D_{n:N})\right)\right)\\&\geqq 1-\sum_{n=2}^{N}P\left(F_N(D_{n:N})>\frac{n-1}{\beta N}\right)\\&=1-\sum_{n=2}^{[\beta N]+1}P\left(F_N(D_{n:N})>\frac{n-1}{\beta N}\right).\end{aligned}$$

Now, for $n=2, 3, \ldots, [\beta N]+1$

$$P\left(F_N(D_{n:N})>\frac{n-1}{\beta N}\right)=P(S_N\leqq n-1)$$

where S_N is a sum of N exchangeable zero-one rv's $Z_1, Z_2, \ldots, Z_N$ with $E(Z_1)=p=\frac{n-1}{\beta N}$ and for every $k\in\{2, 3, \ldots, N\}$

$$E\left(\prod_{i=1}^{k} Z_i\right)=F_{k,N}\left(\left(F_N^{-1}\left(\frac{n-1}{\beta N}\right)\right)^{(k)}\right).$$

From Markov's inequality and Corollary 3.1 it follows, for $h>0$ and $n=2, 3, \ldots, [\beta N]+1$, that

$$\begin{aligned} P(S_N\leqq n-1) &= P(e^{-hS_N}\geqq e^{-h(n-1)}) \\ &\leqq e^{h(n-1)}E(e^{-hS_N}) \\ &\leqq e^{h(n-1)}(1-p+pe^{-h})^N \\ &\leqq e^{h(n-1)-Np(1-e^{-h})} \end{aligned}$$

since $1-x\leqq e^{-x}$. Choosing $h=-\log\beta$ yields for $n=2, 3, \ldots, [\beta N]+1$

$$P(S_N\leqq n-1)\leqq e^{-n(\log\beta+\beta^{-1}-1)},$$

where $\log\beta+\beta^{-1}-1>0$ for $\beta\in(0, 1)$.

Hence

$$P_2\geqq 1-\sum_{n=1}^{+\infty} e^{-n(\log\beta+\beta^{-1}-1)}=1-\frac{e\beta^{-1}\exp(-\beta^{-1})}{1-e\beta^{-1}\exp(-\beta^{-1})}. \quad \square$$

Remark 3.1. In order to give applications in nonparametric statistics (as done in Sect. 4) it is also necessary to establish lower bounds for

$$\tilde{P}_1=P(1-\beta^{-1}(1-F_N(s))\leqq\hat{F}_N(s), \textit{ for } s\in[0, 1])$$

and

$$\tilde{P}_2=P(\hat{F}_N(s)\leqq 1-\beta(1-F_N(s)), \textit{ for } s\in[0, D_{N:N}]), \qquad \beta\in(0, 1).$$

This can be done by looking at the rv's $D'_{n,N}=1-D_{n,N}$, $n=1, 2, \ldots, N$ with df F'_N and empirical df $\hat{F}'_N$, say. Clearly one has the following relations

$$\tilde{P}_1=P(\hat{F}'_N(s)\leqq\beta^{-1}F'_N(s), \textit{ for } s\in[0, 1])$$

and

$$\tilde{P}_2=P(\hat{F}'_N(s)\geqq\beta F'_N(s), \textit{ for } s\in[D'_{1:N}, 1]),$$

where $D'_{1:N}$ is of course the smallest of the $D'_{n,N}$, $n=1, 2, \ldots, N$. Since $(D'_{1,N}, D'_{2,N}, \ldots, D'_{N,N})$ is also both NLOD and NUOD, lower bounds for $\tilde{P}_1$ and $\tilde{P}_2$ can be derived following the approach of this section.

Paralleling the proofs of the Theorems 3.1 and 3.2 one obtains that for some universal number $N_0\in\mathbb{N}$,

$$\tilde{P}_1 \geqq 1-\beta-(2e\beta^2/(1-2\beta\exp(1-2\beta)), \quad \text{for } \beta\in(0,\tfrac{1}{2}) \text{ and } N\geqq N_0$$

and

$$\tilde{P}_2 \geqq 1-\frac{e\beta^{-1}\exp(-\beta^{-1})}{1-e\beta^{-1}\exp(-\beta^{-1})}, \quad \text{for } \beta\in(0,1) \text{ and } N\geqq 2.$$

The method developed in Sect. 2 doesn't seem to yield a good result in this case. This can be intuitively understood as the probabilities $\tilde{P}_1$ and $\tilde{P}_2$ are to a high degree determinated by the behaviour in $D'_{1:N}$ resp. $D'_{2:N}$, i.e. in $D_{N:N}$ resp. $D_{N-1:N}$. The inequalities (2.3), however, become less sharp for larger values of n.

4. Applications

In this section we give two applications of the linear bounds for $\hat{F}_N$ obtained in the Sect. 2 and 3. It will be convenient to work with the rescaled spacings $ND_{n,N}=D^*_{n,N}$. Remark from (1.1) that the common marginal df of the $D^*_{n,N}$ is

$$F^*_N(x)=1-\left(1-\frac{x}{N}\right)^{N-1}, \quad x\in[0,N].$$

For the ordered $D^*_{n,N}$ we use the notation $D^*_{1:N}\leqq D^*_{2:N}\leqq\ldots\leqq D^*_{N:N}$, whereas the empirical df of the rescaled spacings $D^*_{n,N}$ will be denoted by $\hat{F}^*_N$. Note that

$$\begin{aligned}[\hat{F}^*_N(x)\leqq\beta^{-1}F^*_N(x), \textit{ for } x\in[0,N]] \\ =[\hat{F}_N(s)\leqq\beta^{-1}F_N(s), \textit{ for } s\in[0,1]]\end{aligned} \tag{4.1}$$

and

$$\begin{aligned}[\hat{F}^*_N(x)\geqq\beta F^*_N(x), \textit{ for } x\in[D^*_{1:N},N]] \\ =[\hat{F}_N(s)\geqq\beta F_N(s), \textit{ for } s\in[D_{1:N},1]],\end{aligned} \tag{4.2}$$

so that the theorems of Sects. 2 and 3 can be restated in terms of $\hat{F}^*_N$.

We next define the *empirical spacings process* X^*_N by

$$X^*_N(x)=N^{\frac{1}{2}}(\hat{F}^*_N(x)-F^*_N(x)), \quad x\in[0,N]$$

and the *reduced empirical spacings process* $\bar{X}_N$ by

$$\bar{X}_N(s)=N^{\frac{1}{2}}(\hat{F}_N(F_N^{-1}(s))-s), \quad s\in[0,1].$$

Moreover, let $\{V(s)|0\leqq s\leqq 1\}$ be a Gaussian stochastic process (defined on the probability space of X^*_N) with mean zero and covariance function $E(V(s)V(t))=s(1-t)-v(s)v(t)$ for all $0\leqq s\leqq t\leqq 1$ where $v(s)=-(1-s)\ln(1-s)$.

Finally, we define $R:(0,1)\to\mathbb{R}$ to be the function

$$R(s)=[s(1-s)]^{-1},$$

and write $\|f\|_a^b$ for the supremum over $[a,b]\subseteq\mathbb{R}$ of the absolute value of any $f:[a,b]\to\mathbb{R}$.

Application 1

In a separate paper Beirlant and van Zuijlen (1982) show that for $\delta\in(0,\frac{1}{2}]$,

$$\|(X_N^*-V(F_N^*))R^{\frac{1}{2}-\delta}(F_N^*)\|_0^N=\|(\bar{X}_N-V)R^{\frac{1}{2}-\delta}\|_0^1\xrightarrow[P]{}0, \tag{4.3}$$

a result which implies the weak convergence of the weighted empirical spacings process. For $\delta=\frac{1}{2}$ (the unweighted case) (4.3) has been established in Shorack (1972a). A critical part of the proof of (4.3) consists in showing that

$$\|(\bar{X}_N-V)R^{\frac{1}{2}-\delta}\|_0^{1/N}\xrightarrow[P]{}0 \quad\text{and}\quad \|(\bar{X}_N-V)R^{\frac{1}{2}-\delta}\|_{1-\frac{1}{N}}^1\xrightarrow[P]{}0. \tag{4.4}$$

To show (4.4) we need the following lemma in the proof of which we make an essential use of the linear bounds.

Lemma 4.1. *There exists $K\in(0,+\infty)$ such that for every $N=2,3,\dots$, every $\delta\in(0,\frac{1}{2}]$, and every $c>0$*

$$P\Big(\sup_{\substack{0<s\leq\frac{1}{N}\\1-\frac{1}{N}\leq s<1}}|\bar{X}_N(s)|\,R^{\frac{1}{2}-\delta}(s)\geq c\Big)\leq K(cN^{\delta})^{-1}.$$

Proof. The proof is patterned on that of Lemma 1.4 in van Zuijlen (1978). We note from Theorem 2.1 or 3.1 that for $0\leq\beta\leq\frac{1}{2}$, $N\geq 2$ there exists $C>1$ with

$$P\left(\sup_{0<s\leq\frac{1}{N}}\frac{|\bar{X}_N(s)|}{s}\geq(\beta^{-1}-1)\sqrt{N}\right)\leq C\beta,$$

so that for every $\delta\in(0,\frac{1}{2}]$

$$P\Big(\sup_{0<s\leq\frac{1}{N}}|\bar{X}_N(s)|\,R^{\frac{1}{2}-\delta}(s)>\sqrt{2}(\beta^{-1}-1)N^{-\delta}\Big)\leq C\beta$$

which proves the lemma for $c>\sqrt{2}N^{-\delta}$ and hence for all $c>0$. The part of the lemma where we consider $s\geq 1-\dfrac{1}{N}$ follows by a similar argument. □

Moreover we mention a Lemma, also proved in Beirlant and van Zuijlen (1982) which together with (4.3) implies that for every $\varepsilon>0$ and $\delta\in(0,\frac{1}{2}]$ there exists $M(\varepsilon,\delta)$ such that for every $N\in\mathbb{N}$

$$P(\|\bar{X}_N R^{\frac{1}{2}-\delta}\|_0^1\geq M)\leq\varepsilon \tag{4.5}$$

which means that $\|\bar{X}_N R^{\frac{1}{2}-\delta}\|_0^1=\|X_N^* R^{\frac{1}{2}-\delta}(F_N^*)\|_0^N$ is bounded in probability.

Lemma 4.2. *There exists $K\in(0,+\infty)$ such that for every $\delta\in(0,\frac{1}{2}]$, every $0<\theta<\frac{1}{2}$, and every $c>0$*

$$P(|V(s)|\leq cR^{-\frac{1}{2}+\delta}(s),\ \textit{for all } s\in[0,\theta]\cup[1-\theta,1])$$
$$\geq 1-Kc^{-4}\left(\int_0^\theta R^{1-2\delta}(s)\,ds\right)^2.$$

Application 2

We next give an application of the linear bounds to linear combinations of (functions of) ordered spacings. Such statistics are useful in a number of practical problems, such as problems of fit; see Pyke (1965) for an extensive survey. It is interesting to include linear combinations with relatively high weights for the smaller and larger spacings as e.g. in Jackson (1967). See also Beirlant and van Zuijlen (1982). More precisely, let us consider statistics of the form

$$S_N = \sum_{n=1}^{N} c_{nN} \Psi(D^*_{n:N})$$

where $c_{1N}, c_{2N}, \ldots, c_{NN}$ are given numbers (the weights) and $\Psi: (0, +\infty) \to \mathbb{R}$ is a given function.

We assume that the weights are generated by a given function $J: (0,1) \to \mathbb{R}$ according to

$$N c_{nN} = J(n/(N+1)).$$

Writing $J(sN/(N+1)) = J_N(s)$ for $s \in (0, 1]$ we may represent S_N by

$$S_N = \int_0^N J_N(\hat{F}^*_N(x)\, \Psi(x)\, d\hat{F}^*_N(x). \tag{4.6}$$

We are interested in the asymptotic distribution of statistics of this type, in particular when we allow $|J|$ to tend to ∞ as $s \downarrow 0$ or $s \uparrow 1$, which corresponds to relatively high weights for the smaller and larger spacings.

This way of representing the statistics S_N in the form (4.6) almost naturally leads to a Chernoff-Savage (1958) approach for the asymptotic distribution theory; see Ruymgaart and van Zuijlen (1977; 1978) where the same approach has been used for linear combinations of (functions of) ordinary order statistics. In the proof a number of properties of empirical df's like the one in Theorem 2.2 or 3.2, that might be of independent interest, appears to be needed.

Writing

$$\mu_N = \int_0^N J_N(F^*_N(x))\, \Psi(x)\, dF^*_N(x),$$

a proper standardization of S_N turns out to be

$$N^{\frac{1}{2}}(S_N - \mu_N) = A_{0N} + A_{1N} + B_N$$

where

$$A_{0N} = N^{\frac{1}{2}} \int_0^N J_N(F^*_N(x))\, \Psi(x)\, d[\hat{F}^*_N(x) - F^*_N(x)],$$

$$A_{1N} = \int_0^N X^*_N(x) J^{(1)}_N(F^*_N(x))\, \Psi(x)\, dF^*_N(x),$$

$$B_N = N^{\frac{1}{2}} \int_0^N [J_N(\hat{F}^*_N(x)) - J_N(F^*_N(x))]\, \Psi(x)\, d\hat{F}^*_N(x) - A_{1N}.$$

As far as the A-terms are concerned, notice that $A_{0N}+A_{1N}$ is a sum function of *unordered* spacings, to which a version of central limit theorems established by Holst (1979), Beirlant et al. (1982) can be applied for proving the asymptotic normality.

For the asymptotic normality of the A-terms some conditions are needed, in particular in order to guarantee the existence of moments. These conditions will also be needed to prove the asymptotic negligibility of B_N which is in fact the only real problem. *These conditions are that Ψ is measurable, J has a continuous first derivative $J^{(1)}$ throughout* $(0, 1)$, *and that there are numbers* $c\in(0, +\infty)$ *and* $a, b\in(0, \frac{1}{2})$ *with* $a+b<\frac{1}{2}$ *such that*

$$|\Psi(F_N^{*-1}(s))| \leqq c R^b(s), \quad s\in(0,1), \quad \textit{uniformly in } N, \tag{4.7}$$

$$|J^{(i)}(s)| \leqq c R^{a+i}(s), \quad s\in(0,1), \quad i\in\{0,1\}. \tag{4.8}$$

It will be convenient to briefly write B_N as

$$B_N = N^{\frac{1}{2}} \int_0^N (\ldots)\, d\hat{F}_N^* - N^{\frac{1}{2}} \int_0^N (\ldots)\, dF_N^*.$$

The proof that $B_N \xrightarrow[P]{} 0$, as $N \to +\infty$, will follow at once if we can show that each of the integrals

$$B_{1N} = N^{\frac{1}{2}} \int_0^{F_N^{*-1}(\gamma)} (\ldots)\, d\hat{F}_N^*, \qquad B_{2N} = N^{\frac{1}{2}} \int_{F_N^{*-1}(1-\gamma)}^{N} (\ldots)\, d\hat{F}_N^*, \tag{4.9}$$

$$B_{3N} = N^{\frac{1}{2}} \int_0^{F_N^{*-1}(\gamma)} (\ldots)\, dF_N^*, \qquad B_{4N} = N^{\frac{1}{2}} \int_{F_N^{*-1}(1-\gamma)}^{N} (\ldots)\, dF_N^* \xrightarrow[P]{} 0, \tag{4.10}$$

as $\gamma \downarrow 0$ and $N \to +\infty$, along with

$$N^{\frac{1}{2}} \int_{F_N^{*-1}(\gamma)}^{F_N^{*-1}(1-\gamma)} (\ldots)\, d\hat{F}_N^* - N^{\frac{1}{2}} \int_{F_N^{*-1}(\gamma)}^{F_N^{*-1}(1-\gamma)} (\ldots)\, dF_N^* \xrightarrow[P]{} 0 \tag{4.11}$$

as $N \to +\infty$, for all $\gamma\in(0, \frac{1}{2})$.

In dealing with the integrals in (4.9) the importance of Theorem 2.2 or Theorem 3.2 will become apparent. We restrict ourselves to a full proof of the asymptotic negligibility of B_{1N}. The second integral B_{2N} can be handled by similar methods using the results stated in Remark 3.1.

Let us notice that the random measure $d\hat{F}_N^*$ restricts integration to the random interval

$$\Delta_N = [D_{1:N}^*, D_{N:N}^*],$$

and that $N/(N+1)\,\hat{F}_N^*(x) \in \left[\frac{1}{N+1}, 1-\frac{1}{N+1}\right] \subset (0,1)$ for $x\in\Delta_N$. Hence the mean value theorem applies to the integrand and yields (for $x\in\Delta_N$)

$$N^{\frac{1}{2}}(J_N(\hat{F}_N^*(x)) - J_N(F_N^*(x))) = N^{\frac{1}{2}}(\hat{F}_N^*(x) - F_N^*(x))\, J_N^{(1)}(\hat{G}_N^*(x)) = X_N^*(x)\, J_N^{(1)}(\hat{G}_N^*(x)),$$

where $\hat{G}_N^*(x)$ is a random point strictly between $\hat{F}_N^*(x)$ and $F_N^*(x)$. The point $\hat{G}_N^*(x)$ satisfies

$$\hat{F}_N^*(x) \wedge F_N^*(x) \leqq \hat{G}_N^*(x) \leqq F_N^*(x) + \|\hat{F}_N^* - F_N^*\|_0^N. \tag{4.13}$$

We next introduce the subsets

$$\Omega_{N,\beta}^{(1)} = \{\omega \in \Omega \mid \hat{F}_N^*(x) \geqq \beta F_N^*(x),\ x \in \Delta_N\},$$
$$\Omega_{N,C}^{(2)} = \{\omega \in \Omega \mid \|X_N^* R^{\frac{1}{2}-\delta}(F_N^*)\|_0^N \leqq C\}$$

for some $\delta \in (0, \frac{1}{2} - a - b)$. From now on let us fix an arbitrary $\varepsilon > 0$. As a result of relations (4.1) and (4.2), Theorem 2.2 (or 3.2), and the boundedness in probability of $\|X_N^* R^{\frac{1}{2}-\delta}(F_N^*)\|_0^N$ (see 4.5)) we may claim that there exist β_ε and C_ε such that

$$P(\Omega_{N,\beta}^{(1)} \cap \Omega_{N,C}^{(2)}) \geqq 1 - \varepsilon \quad \text{for all } N \geqq 2 \tag{4.14}$$

provided $\beta \leqq \beta_\varepsilon$ and $C \geqq C_\varepsilon$. From now on we shall also take β and C fixed such that (4.14) is satisfied.

For any $\gamma \in (0, \frac{1}{2})$ we have on $\Omega_{N,C}^{(2)}$ that

$$F_N^*(x) + \|\hat{F}_N^* - F_N^*\|_0^N \leqq F_N^*(x) + C N^{-\frac{1}{2}} \leqq \tfrac{1}{2}, \quad \text{for } x \in [0, F_N^{*^{-1}}(\gamma)] \tag{4.15}$$

and provided N is chosen sufficiently large. Combining (4.13) and (4.15) we see that on $\Omega_{N,C}^{(2)}$ we have

$$\hat{F}_N^*(x) \wedge F_N^*(x) \leqq \hat{G}_N^*(x) \leqq \tfrac{1}{2}, \quad \text{for } x \in \Delta_N \cap [0, F_N^{*^{-1}}(\gamma)], \tag{4.16}$$

provided N is sufficiently large.

The condition (4.8) on J and (4.16) entail that on $\Omega_{N,\beta}^{(1)} \cap \Omega_{N,C}^{(2)}$ we find the bound

$$\begin{aligned} |J_N^{(1)}(\hat{G}_N^*(x))| &\leqq c R^{a+1}(\hat{G}_N^*(x)) \\ &\leqq c R^{a+1}(\hat{F}_N^*(x) \wedge F_N^*(x)) \\ &\leqq c R^{a+1}(\beta F_N^*(x)) \\ &\leqq c (1/\beta)^{a+1} R^{a+1}(F_N^*(x)), \end{aligned} \tag{4.17}$$

for $x \in \Delta_N \cap [0, F_N^{*^{-1}}(\gamma)]$.

Finally, it follows from (4.7), (4.12), and (4.17) that on $\Omega_{N,\beta}^{(1)} \cap \Omega_{N,C}^{(2)}$ the integrand (including $N^{\frac{1}{2}}$) of the integral B_{1N} is bounded by

$$\begin{aligned} & N^{\frac{1}{2}} |J_N(\hat{F}_N^*(x)) - J_N(F_N^*(x))|\,|\Psi(x)| \\ &\quad \leqq c^2 C (1/\beta)^{a+1} R^{a+1}(F_N^*(x)) R^b(F_N^*(x)) R^{-\frac{1}{2}+\delta}(F_N^*(x)) \\ &\quad \leqq c^2 C (1/\beta)^{a+1} R^{a+b+\frac{1}{2}+\delta}(F_N^*(x)), \qquad x \in \Delta_N \cap [0, F_N^*(\gamma)]. \end{aligned}$$

So for each $\omega \in \Omega_{N,\beta}^{(1)} \cap \Omega_{N,c}^{(2)}$, B_{1N} itself is bounded by

$$c^2 C (1/\beta)^{a+1} \int_0^{F_N^{*-1}(\gamma)} R^{a+b+\frac{1}{2}+\delta}(F_N^*(x))\, d\hat{F}_N^*(x)$$

and hence

$$E(1_{(\Omega_{N,\beta}^{(1)}\cap\Omega_{N,C}^{(2)})}|B_{1N}|)\leq c^2 C(1/\beta)^{a+1}\int_0^{F_N^{*-1}(\gamma)} R^{a+b+\frac{1}{2}+\delta}(F_N^*(x))\,dF_N^*(x)$$

$$=c^2 C(1/\beta)^{a+1}\int_0^{\gamma} R^{a+b+\frac{1}{2}+\delta}(s)\,ds,$$

$$\to 0, \quad \text{as} \quad \gamma\downarrow 0, \tag{4.18}$$

because $a+b+\frac{1}{2}+\delta<1$.

It is now obvious from (4.14) and (4.18) that B_{1N} converges to zero in probability as $N\to+\infty$ and $\gamma\downarrow 0$.

References

1. Beirlant, J., Janssen, P., Veraverbeke, N.: An elementary proof for the asymptotic normality of functions of uniform spacings. Report LUC, Belgium (1982)
2. Beirlant, J., van der Meulen, E.C., Zuijlen, M.C.A. van: Bounds on the empirical distribution function for uniform spacings. Mededeling uit het Wiskundig Instituut KUL, no 132 (1981)
3. Beirlant, J., Zuijlen, M.C.A. van: Empirical distribution functions and functions of order statistics for uniform spacings. Mededeling uit het Wiskundig Instituut KUL (1982)
4. Block, H., Savits, T., Shaked, M.: Some concepts of negative dependence. Technical Report, Univ. of Pittsburgh (1980)
5. Chernoff, H., Savage, I.R.: Asymptotic normality and efficiency of certain nonparametric test statistics. Ann. Math. Statist. **38**, 52–72 (1958)
6. Holst, L.: Asymptotic normality of sum-functions of spacings. Ann. Probab. **7**, 1066–1072 (1979)
7. Jackson, O.: An analysis of departures from the exponential distribution. J. Roy. Statist. Soc., Ser. B. **29**, 540–549 (1967)
8. Karlin, S.: A first course in stochastic processes. New York: Academic Press 1966
9. Pyke, R.: Spacings. J. Roy. Statist. Soc., Ser. B. **27**, 359–449 (1965)
10. Robbins, H.: A one-sided confidence interval for an unknown distribution function. Ann. Math. Statist. **25**, 409 (1954)
11. Ruymgaart, F.H.: A note on bounds for the empirical df of uniform spacings. Report 8107 Math. Instituut Kath. Univ. Nijmegen (1981)
12. Ruymgaart, F.H., Zuijlen, M.C.A. van: Asymptotic normality of linear combinations of functions of order statistics in the non-i.i.d. case. Indag. Math. **80**, 432–447 (1977)
13. Ruymgaart, F.H., Zuijlen, M.C.A. van: 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 (1978)
14. Shorack, G.R.: Convergence of quantile and spacings processes with applications. Ann. Math. Statist. **43**, 1400–1411 (1972)
15. Wellner, J.: Limit theorems for the ratio of the empirical distribution function to the true distribution function. Z. Wahrscheinlichkeitstheorie Verw. Gebiete **45**, 73–88 (1978)
16. Wilks, S.S.: Mathematical Statistics. New York: Wiley 1962
17. Zuijlen, M.C.A. van: Some properties of the empirical distribution function in the non-i.i.d. case. Ann. Statist. **4**, 406–408 (1976)
18. Zuijlen, M.C.A. van: Properties of the empirical distribution function for independent non-identically distributed random variables. Ann. Probab. **6**, 250–266 (1978)
19. Zuijlen, M.C.A. van: Properties of the empirical distribution function for independent non-identically distributed random vectors. Ann. Probab. **10**, 108–123 (1982)

Received March 13, 1981; in revised form June 25, 1982