APPLICATIONES MATHEMATICAE
51,2 (2024), pp. 131–146

Agnieszka Siedlaczek (Opole)

NONPARAMETRIC ESTIMATION OF QUANTILE VERSIONS OF THE BONFERRONI CURVE

Abstract. We consider the quantile versions of the Bonferroni curve and of the Bonferroni index. Distribution-free estimators of these quantile versions are proposed. Asymptotic properties of the estimators are proved. In simulations we compare the accuracy of the estimators. To illustrate the estimators under study, we perform real data analysis concerning the remission times of bladder cancer patients.

The Bonferroni curve was introduced by Bonferroni [3]. It is used in examining inequality of distribution of some feature in a population. Although it was introduced for economics, it is currently applied in many fields, such as industry or medicine. In addition to the Bonferroni curve, also other curves were defined to investigate inequality of distribution, such as the Lorenz curve [11], the Zenga curve [16, 17] and the Gastwirth curve [4]. The biggest advantage of the Bonferroni curve is its simple interpretation, which we will discuss later.

Despite the wide range of applications of the Bonferroni curve, we will present interpretations of the concepts introduced in the work using the example of income.

Let $X : \mathcal{X} \to \mathbb{R}$ be a random variable with distribution

(A1) absolutely continuous with respect to the Lebesgue measure, with cumulative distribution function (cdf) F and density f,

(A2) with support $[0, \infty)$.

2020 *Mathematics Subject Classification*: Primary 62G05.
Key words and phrases: nonparametric estimation, concentration, Bonferroni curve, Bonferroni index, quantile versions of the Bonferroni curve, quantile versions of the Bonferroni index.
Received 10 October 2023; revised 11 June 2024.
Published online 25 July 2024.

DOI: 10.4064/am2504-6-2024

Denote

$$Q(p) = F^{-1}(p) = \inf\{x : F(x) \geq t\}.$$

With the above assumptions, the definition of the Bonferroni curve is as follows.

Definition 0.1. The *Bonferroni curve*, corresponding to the distribution of the non-negative random variable X with finite expected value $\mu \neq 0$, with cumulative distribution function F and density f with respect to the Lebesgue measure, is the set of points $(p, B(p; F))$

$$B(p; F) = \frac{1}{p\mu} \int_0^p Q(t)\, dt \qquad \text{for } p \in (0, 1].$$

We will often omit the argument F in $B(p; F)$, as long as it does not lead to a misunderstanding.

The Bonferroni curve has a simple interpretation. For each $p \in (0, 1]$, the value of $B(p; F)$ is the ratio of the mean income of the fraction p of people with the lowest income to the mean income in the entire population, when the distribution of income in the entire population is determined by the distribution function F.

The disadvantage of the Bonferroni curve is that it does not exist for distributions with infinite expected value. A similar problem occurs for the Lorenz curve, which is also a tool to measure inequality of distribution. To deal with this difficulty Prendergast and Staudte [13] proposed quantile versions of the Lorenz curve. We introduce quantile versions of the Bonferroni curve as well.

Definition 0.2. The quantile versions B_1, B_2, B_3 of the Bonferroni curve, corresponding to the cumulative distribution F, are defined as follows:

$$\begin{aligned} B_1(p; F) &= \frac{Q(p/2)}{Q(0.5)}, \\ B_2(p; F) &= \frac{Q(p/2)}{Q(1-p/2)}, \qquad (0.1) \\ B_3(p; F) &= 2\frac{Q(p/2)}{Q(p/2) + Q(1-p/2)} = \frac{2}{1 + 1/B_2(p)} \end{aligned}$$

for $p \in (0, 1)$ and $B_i(0; F) = 0$, $B_i(1; F) = 1$, $i = 1, 2, 3$.

We will omit F in $B_i(p; F)$, just as for the Bonferroni curve.

Notice that since (A1) and (A2) are satisfied, we have $\lim_{p\to 0^+} B_i(p) = 0$.

To our best knowledge, only the second quantile version of the Bonferroni curve was considered by Prendergast and Staudte [14], in the context of inequality measures. It is called the symmetric ratio of quantiles in [14].

With assumptions (A1) and (A2), all the quantile inequality curves $\{(p, B_i(p))\}$ are scale invariant and increasing from $B_i(0) = 0$ to $B_i(p)(1)=1$, but they are not shape invariant.

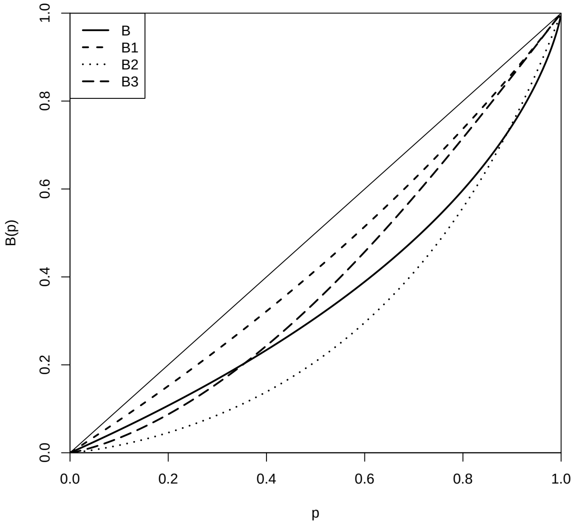

Fig. 1. The Bonferroni curve and its quantile versions for the exponential distribution

Figure 1 presents the Bonferroni curve and its quantile versions for the exponential distribution and Table 1 contains quantile versions of the Bonferroni curve for various distributions.

Table 1. The quantile versions of the Bonferroni curve for various distributions with parameters $\alpha > 0$, $a > 0$, $\lambda > 0$, $k > 0$.

Model	CDF	B_1	B_2	B_3
Pareto Par(a, α) $x > 0$	$1 - \left(\frac{\alpha}{x}\right)^a$	$\sqrt[a]{\frac{1}{2-p}}$	$\sqrt[a]{\frac{p}{2-p}}$	$\frac{2\sqrt[a]{\frac{2}{2-p}}}{\sqrt[a]{\frac{2}{2-p}}+\sqrt[a]{\frac{2}{p}}}$
Exponential Exp(λ) $x > 0$	$1 - e^{-\lambda x}$	$\frac{\ln\left(1-\frac{p}{2}\right)}{(-\ln 2)}$	$\frac{\ln\left(1-\frac{p}{2}\right)}{\ln\left(\frac{p}{2}\right)}$	$\frac{2\ln\left(1-\frac{p}{2}\right)}{\ln\left(\frac{p(2-p)}{4}\right)}$
Weibull Wei(λ, k) $x > 0$	$1 - e^{-\left(\frac{x}{\lambda}\right)^k}$	$\sqrt[k]{\frac{\ln\left(1-\frac{p}{2}\right)}{(-\ln 2)}}$	$\sqrt[k]{\frac{\ln\left(1-\frac{p}{2}\right)}{\ln\left(\frac{p}{2}\right)}}$	$\frac{2\sqrt[k]{-\ln\left(1-\frac{p}{2}\right)}}{\sqrt[k]{-\ln\left(\frac{p}{2}\right)}+\sqrt[k]{-\ln\left(1-\frac{p}{2}\right)}}$
Fréchet Fre(λ, k)	$e^{-\left(\frac{x}{\lambda}\right)^{-k}}$	$\sqrt[k]{\frac{\ln 2}{\ln\left(\frac{2}{p}\right)}}$	$\sqrt[k]{\frac{\ln\left(1-\frac{p}{2}\right)}{\ln\left(\frac{p}{2}\right)}}$	$\frac{2\sqrt[k]{-\ln\left(1-\frac{p}{2}\right)}}{\sqrt[k]{-\ln\left(1-\frac{p}{2}\right)}+\sqrt[k]{-\ln\left(\frac{p}{2}\right)}}$
Power Pow(a) $x \in (0, 1)$	x^a	$\sqrt[a]{p}$	$\sqrt[a]{\frac{p}{2-p}}$	$\frac{2\sqrt[a]{p}}{\sqrt[a]{p}+\sqrt[a]{2-p}}$

Apart from the Bonferroni curve, to measure the inequality of distribution of some feature in the population, various inequality measures are also applied. Some of them are not associated with any curves, such as Atkinson measure [2] and Theil index [15], and others are based on different curves, such as the Lorenz curve [5], the quantile versions of the Lorenz curve [13] or the Zenga curve [16, 17].

Since these indices have different sensitivity to transfers in some parts of the distribution, in applications it is recommended to use a few inequality measures, which supplement each other, to get better information about the distribution [6].

We recall the Bonferroni index, which is based on the Bonferroni curve, and propose its quantile versions.

Definition 0.3. The *Bonferroni index* corresponding to the distribution of the random variable X with finite expected value $\mu \neq 0$ satisfying (A1) and (A2) is given by

$$BI(F) = 1 - \int_0^1 B(p;F)\,dp.$$

Based on this index we propose new inequality measures.

Definition 0.4. The *quantile versions of the Bonferroni index* corresponding to the distribution of the random variable X satisfying (A1) and (A2) are given by

$$BI_1(F) = 1 - \int_0^1 B_1(p;F)\,dp,$$

$$BI_2(F) = 1 - \int_0^1 B_2(p;F)\,dp,$$

$$BI_3(F) = 1 - \int_0^1 B_3(p;F)\,dp.$$

We will often omit F in $BI(F)$ and $BI_i(F)$.

1. Estimators of the quantile versions of the Bonferroni curve and Bonferroni index and their asymptotic properties. We use the nonparametric method of estimation of the quantile versions of the Bonferroni curve. Using the plug-in method we can obtain estimators of these quantile versions by substituting various estimators of the quantile function (see for example [8]).

Let $\mathbf{X}_n = (X_1, \ldots, X_n)$ be a random sample from the cdf F satisfying (A1) and (A2). Denote by $\hat{F}_n^E$ the standard empirical estimator of F based on $\mathbf{X}_n$ defined as $\hat{F}_n^E(t;\mathbf{X}_n) := \frac{1}{n}\sum_{i=1}^n \mathbb{I}_{(-\infty,t]}(X_i)$ where $\mathbb{I}_A(x) = 1$ if $x \in A$, and 0 otherwise.

Let $\tilde{F}_n$ be an estimator of F such that with probability 1,

$$\sup_{t\in\mathbb{R}} |\tilde{F}_n(t) - \hat{F}_n^E(t)| \le 1/n. \tag{1.1}$$

Denote the estimator of the pth quantile based on $\tilde{F}_n$ by

$$\tilde{x}_{p,n} := \tilde{Q}_n(p) = \tilde{F}_n^{-1}(p). \tag{1.2}$$

We will consider estimators satisfying the above property. We obtain estimators of the quantile versions of the Bonferroni curve by substituting $\tilde{Q}_n(p)$ with proper order to formulas (0.1). We will denote the estimators of $B_i(p)$ based on $\tilde{Q}_n(p)$ by $\tilde{B}_{i,n}(p;\mathbf{X})$. They are given by

$$\begin{aligned}
\tilde{B}_{1,n}(p;\mathbf{X}) &= \frac{\tilde{Q}_n(p/2)}{\tilde{Q}_n(0.5)},\\
\tilde{B}_{2,n}(p;\mathbf{X}) &= \frac{\tilde{Q}_n(p/2)}{\tilde{Q}_n(1-p/2)},\\
\tilde{B}_{3,n}(p;\mathbf{X}) &= 2\frac{\tilde{Q}_n(p/2)}{\tilde{Q}_n(p/2)+\tilde{Q}_n(1-p/2)}
\end{aligned} \tag{1.3}$$

for $p \in (0,1]$.

For all the estimators $\tilde{B}_{i,n}(p;\mathbf{X})$ the following theorems hold.

THEOREM 1.1. *Let $X_1,\dots,X_n$ be an independent and identically distributed random variables satisfying* (A1) *and* (A2) *with F. Let $\tilde{F}_n$ be an estimator of F satisfying* (1.1), *$\tilde{Q}_n$ the estimator of the pth quantile based on $\tilde{F}_n$, and $\tilde{B}_{i,n}$ the estimators of the quantile versions of the Bonferroni curve based on $\tilde{Q}_n(p)$, $i = 1,2,3$. Fix $p \in (0,1)$. Then*

$$\tilde{B}_{i,n}(p;\mathbf{X}) \xrightarrow{P1} B_i(p) \quad \text{as } n\to\infty \text{ for } i=1,2,3.$$

Proof. We will prove the convergence of $\tilde{B}_{1,n}(p;\mathbf{X})$. The remaining convergences can be shown analogously.

Since the distribution of $X_1,\dots,X_n$ satisfies (A1) and (A2), we have

$$F(x_{p/2}-\epsilon) < p/2 < F(x_{p/2}+\epsilon) \quad \text{and} \quad F(x_{1/2}-\epsilon) < 1/2 < F(x_{1/2}+\epsilon).$$

Due to [10, Theorem 3.2], under our assumptions, the estimators $\tilde{x}_{p,n}$ are strongly consistent for x_p. Hence $\tilde{x}_{p/2,n} \xrightarrow{P1} x_{p/2}$ and $\tilde{x}_{1/2,n} \xrightarrow{P1} x_{1/2}$. Moreover, $P(\tilde{x}_{p/2,n} \neq 0) = 1$ and $x_{1/2} \neq 0$. Therefore

$$\tilde{B}_{1,n}(p;\mathbf{X}) = \frac{\tilde{x}_{p/2,n}}{\tilde{x}_{1/2,n}} \xrightarrow{P1} \frac{x_{1/2,n}}{x_{1/2,n}} = B_1(p),$$

from the definition of convergence with probability 1, and the proof is complete. ∎

THEOREM 1.2. *Let $X_1, \dots, X_n$ be independent and identically distributed random variables satisfying* (A1) *and* (A2) *with cdf F. Then*

$$\sup_{p\in(0,1)} |\tilde{B}_{i,n}(p;\mathbf{X}) - B_i(p)| \xrightarrow{P1} 0 \quad \textit{for } i = 1,2,3.$$

Proof. We will prove the convergence for $\tilde{B}_{1,n}(p;\mathbf{X})$. The remaining convergences can be shown analogously.

Since B_1 is continuous on $[0,1]$, it is uniformly continuous. Hence for all $\epsilon > 0$ there exists a finite partition $0 = p_1 < \dots < p_r = 1$ such that

$$B_1(p_{l+1}) - B_1(p_l) < \epsilon/2 \quad \text{for } l = 1,\dots,r-1. \tag{1.4}$$

By Theorem 1.1, $\tilde{B}_{1,n}(p_l;\mathbf{X})$ is a strongly consistent estimator of $B_1(p_l)$. Therefore there exist $N_l(\epsilon) > 0$ such that, for all $n_l \geq N_l(\epsilon)$, with probability 1,

$$|\tilde{B}_{1,n_l}(p_l;\mathbf{X}) - B_1(p_l)| < \epsilon/2 \quad \text{for } l = 1,\dots,r.$$

Denote $N(\epsilon) = \max\{N_1(\epsilon),\dots,N_r(\epsilon)\}$. Then for all $n \geq N(\epsilon)$, with probability 1,

$$|\tilde{B}_{1,n_l}(p_l;\mathbf{X}) - B_1(p_l)| < \epsilon/2 \quad \text{for } l = 1,\dots,r. \tag{1.5}$$

Let $p \in (0,1)$. Then $p_l \leq p < p_{l+1}$ for some $l \in \{1,\dots,r\}$. Moreover, $n \geq N(\epsilon)$. Then by (1.4) and (1.5) and the continuity of B_1, with probability 1, we have

$$\begin{aligned} B_1(p_l) - \epsilon/2 < \tilde{B}_{1,n}(p_l;\mathbf{X}) &\leq \tilde{B}_{1,n}(p;\mathbf{X}) \leq \tilde{B}_{1,n}(p_{l+1};\mathbf{X}) \\ &< B_1(p_l) + \epsilon/2 \leq B_1(p) + \epsilon \\ &\leq B_1(p_{l+1}) + \epsilon. \end{aligned} \tag{1.6}$$

Hence, with probability 1,

$$0 \leq B_1(p) - \tilde{B}_{1,n}(p;\mathbf{X}) + \epsilon \leq B_1(p_{l+1}) + \epsilon - B_1(p_l) + \epsilon/2 \leq 2\epsilon.$$

Therefore for all $n \geq N(\epsilon)$, with probability 1,

$$|\tilde{B}_{1,n}(p;\mathbf{X}) - B_1(p)| < \epsilon \quad \text{for all } p \in (0,1),$$

and the theorem is proved. ■

Using the estimators of the quantile versions of the Bonferroni curve from (1.3), we obtain estimators of the quantile versions of the Bonferroni index $\widetilde{BI}_{i,n}(\mathbf{X})$ for $i = 1,2,3$. The following theorem holds for them.

THEOREM 1.3. *Let $X_1, \dots, X_n$ be independent and identically distributed random variables satisfying* (A1) *and* (A2), *with cdf F. Then*

$$\widetilde{BI}_{i,n}(\mathbf{X}) \xrightarrow{P1} BI_i \quad \textit{as } n \to \infty \textit{ for } i = 1,2,3.$$

Proof. We have to show that

$$P\Big(\lim_{n\to\infty} \widetilde{BI}_{i,n}(\mathbf{X}) = BI_i\Big) = 1,$$

which is equivalent to

$$P\Big(\lim_{n\to\infty}(\widetilde{BI}_{i,n}(\mathbf{X}) - BI_i) = 0\Big) = 1,$$

for $i = 1, 2, 3$. From the definition of BI_i and $\widetilde{BI}_{i,n}(\mathbf{X})$ we have

$$\widetilde{BI}_{i,n}(\mathbf{X}) - BI_i = \int_0^1 (B_i(p) - \tilde{B}_{i,n}(p;\mathbf{X}))\,dp.$$

By using the fact that

$$\begin{aligned}(1.7)\qquad & -\sup_{p\in(0,1)} |B_i(p) - \tilde{B}_{i,n}(p;\mathbf{X})| \\ & \le \int_0^1 (B_i(p) - \tilde{B}_{i,n}(p;\mathbf{X}))\,dp \le \sup_{p\in(0,1)} |B_i(p) - \tilde{B}_{i,n}(p;\mathbf{X})|,\end{aligned}$$

from Theorem 1.2 we obtain the assertion. ∎

Estimators of the quantile function are discussed in detail in [10]. Here we will recall six such estimators, based on $\tilde{F}_n$, which we will consider in our simulation study.

Denote by $X_{1:n}, \dots, X_{n:n}$ the order statistics of the sample $\mathbf{X}_n$. A natural estimator of $Q(p)$ is the empirical quantile function

$$\hat{Q}_n^E(p) := X_{([np]+1):n},$$

where $[x]$ denotes the greatest integer not greater than x.

The next three estimators are based on quantile function estimators $\hat{Q}_n$ constructed by linearly interpolating between so called plotting positions, i.e., the points p_k, $k = 1, \dots, n$, for which $\hat{Q}_n(p_k) = X_{k:n}$. These are the following:

- the estimator of Hazen [7], $\hat{Q}_n^H(p)$, with plotting positions
$$p_k^H = \frac{k - 1/2}{n},$$
- the estimator of Hyndman and Fan [8], $\hat{Q}_n^{HF}(p)$, with plotting positions
$$p_k^{HF} = \frac{k - 1/3}{n + 1/3},$$
- the estimator of Makkonen and Pajari [12], $\hat{Q}_n^{MP}(p)$, with plotting positions
$$p_k^{MP} = \frac{k}{n+1},$$

where $k = 1, \dots, n$.

Assume $X_{0:n} = 0$ and

$$X_{(n+1):n} = X_{n:n} + \frac{X_{n:n} - X_{(n-1):n}}{2} = \frac{3}{2}X_{n:n} - \frac{1}{2}X_{(n-1):n}.$$

The estimator of Jokiel-Rokita and Pulit [9] is

$$\hat{Q}_n^{JP}(p) = \frac{n(X_{(k+1):n} - X_{(k-1):n})}{2}p - (k-1)\frac{X_{(k+1):n} - X_{(k-1):n}}{2} + \frac{X_{(k-1):n} + X_{k:n}}{2}$$

if $(k-1)/n < p \le k/n$ for $k = 1, \dots, n$, and $\hat{Q}_n^{JP}(0) = 0$.

The quantile function estimator $\hat{Q}_n^M(p)$, a modification of $\hat{Q}_n^{JP}$ proposed in [10], is

$$\hat{Q}_n^M(p) = \begin{cases} \dfrac{X_{1:n}}{\hat{F}_n^M(X_{1:n})}p & \text{for } p \in [0, \hat{F}_n^M(X_{1:n})], \\[2ex] X_{k:n} + \dfrac{D_{k+1}[p - \hat{F}_n^M(X_{k:n})]}{2[k/n - \hat{F}_n^M(X_{k:n})]} & \text{for } p \in [\hat{F}_n^M(X_{k:n}), k/n], \\ & k = 1, \dots, n-1, \\[2ex] \dfrac{X_{k:n}+X_{(k+1):n}}{2} + \dfrac{D_{k+1}(p - k/n)}{2[\hat{F}_n^M(X_{(k+1):n}) - k/n]} & \text{for } p \in [k/n, \hat{F}_n^M(X_{(k+1):n})], \\ & k = 1, \dots, n-1, \\[2ex] X_{n:n} + \dfrac{\frac{1}{2}D_n[p - \hat{F}_n^M(X_{n:n})]}{1 - \hat{F}_n^M(X_{n:n})} & \text{for } p \in [\hat{F}_n^M(X_{n:n}), 1], \end{cases}$$

where $D_i = X_{i:n} - X_{(i-1):n}$, $i = 1, \dots, n+1$, and

$$\hat{F}_n^M(X_{k:n}) = \frac{1}{2}\left[\frac{X_{k:n} - X_{(k-1):n}}{n(X_{(k+1):n} - X_{(k-1):n})} + 1 - \frac{X_{(n-k+1):n} - X_{(n-k):n}}{n(X_{(n-k+2):n} - X_{(n-k):n})} + \frac{n-1}{n}\right]$$

for $k = 1, \dots, n$.

All of the estimators $\hat{Q}_n^E$, $\hat{Q}_n^H(p)$, $\hat{Q}_n^{HF}(p)$, $\hat{Q}_n^{MP}(p)$, $\hat{Q}_n^{JP}$ and $\hat{Q}_n^M$ satisfy (1.2), i.e. they are examples of estimators $\tilde{Q}_n(p)$.

2. Simulation study. In this section we present results of computer simulations. The purpose of the simulations was to compare the accuracy of plug-in estimators of the quantile versions of the Bonferroni curve and the Bonferroni index obtained by substituting nonparametric estimators of the quantile function. There is a wide selection of estimators of quantiles. Most of them are based on quantile function estimators constructed by linearly interpolating between so called plotting positions. There are also many estimators of the quantile function based on a distribution function estimator (e.g. empirical estimator, level crossing empirical estimator, kernel estimator, kernel estimator with random bandwidth). Based on [10], for simulations we chose the following estimators: $\hat{Q}_n^E$, $\hat{Q}_n^H$, $\hat{Q}_n^{HF}$, $\hat{Q}_n^{MP}$, $\hat{Q}_n^{JP}$ and $\hat{Q}_n^M$, mentioned in Section 1. For calculations we used the statistical package *R 4.3.1*.

We generated samples from the generalized Pareto distribution $\mathcal{GP}(\xi, \sigma)$. This distribution has cumulative distribution function

$$F_{\xi,\sigma}(t) = \begin{cases} [1 - (1 + \xi t/\sigma)^{-1/\xi}] I_{[0,\infty)}(t) & \text{when } \xi > 0,\ \sigma > 0, \\ [1 - (1 + \xi t/\sigma)^{-1/\xi}] I_{[0,-\sigma/\xi]}(t) & \text{when } \xi < 0,\ \sigma > 0, \\ [1 - \exp(-t/\sigma)] I_{[0,\infty)}(t) & \text{when } \xi = 0,\ \sigma > 0. \end{cases}$$

It is a two-parameter distribution with scale parameter σ ($\sigma > 0$) and shape parameter ξ ($\xi \in \mathbb{R}$). When $\xi \geq 1$ the expected value of a random variable with generalized Pareto distribution does not exist and the Bonferroni curve and the Bonferroni index cannot be determined.

The quantile function of the generalized Pareto distribution $\mathcal{GP}(\xi, \sigma)$ is

$$Q_{\xi,\sigma}(p) = \begin{cases} \frac{\sigma}{\xi}[(1-p)^{-\xi} - 1] & \text{when } \xi \neq 0,\ \sigma > 0, \\ -\sigma \ln(1-p) & \text{when } \xi = 0,\ \sigma > 0. \end{cases}$$

The quantile versions of the Bonferroni curve corresponding to the generalized Pareto distribution $\mathcal{GP}(\xi, \sigma)$ depend only on the shape parameter ξ and they are expressed by

$$B_1(p;\xi) = \begin{cases} \frac{(1-p/2)^{-\xi}-1}{(1/2)^{-\xi}-1} & \text{when } \xi \neq 0, \\ \frac{-\ln(1-p/2)}{\ln 2} & \text{when } \xi = 0, \end{cases}$$

$$B_2(p;\xi) = \begin{cases} \frac{(1-p/2)^{-\xi}-1}{(p/2)^{-\xi}-1} & \text{when } \xi \neq 0, \\ \frac{\ln(1-p/2)}{\ln p/2} & \text{when } \xi = 0, \end{cases}$$

$$B_3(p;\xi) = \begin{cases} 2\frac{(1-p/2)^{-\xi}-1}{(1-p/2)^{-\xi}+(p/2)^{-\xi}-2} & \text{when } \xi \neq 0, \\ 2\frac{\ln(1-p/2)}{\ln(p/2-p^2/4)} & \text{when } \xi = 0. \end{cases}$$

Using the *rgpd* function from the *eva* library, we generated samples of sizes $n = 50$ and $n = 100$ from the generalized Pareto distribution $\mathcal{GP}(\xi, \sigma)$ with $\xi \in \left\{\frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1, 1\frac{1}{2}, 2\right\}$ and $\sigma = 1$. Figure 2 presents the quantile versions of the Bonferroni curve corresponding to the generalized Pareto distribution with the shape parameter values selected for simulations. The function *quantile* from the *stats* package in *R* was used to calculate quantile estimates based on plotting positions, i.e. $\hat{x}_{p,n}^H$, $\hat{x}_{p,n}^{HF}$, $\hat{x}_{p,n}^{MP}$. We implemented our own functions to calculate the remaining quantile estimates, i.e. $\hat{x}_{p,n}^E$, $\hat{x}_{p,n}^{JP}$, $\hat{x}_{p,n}^M$.

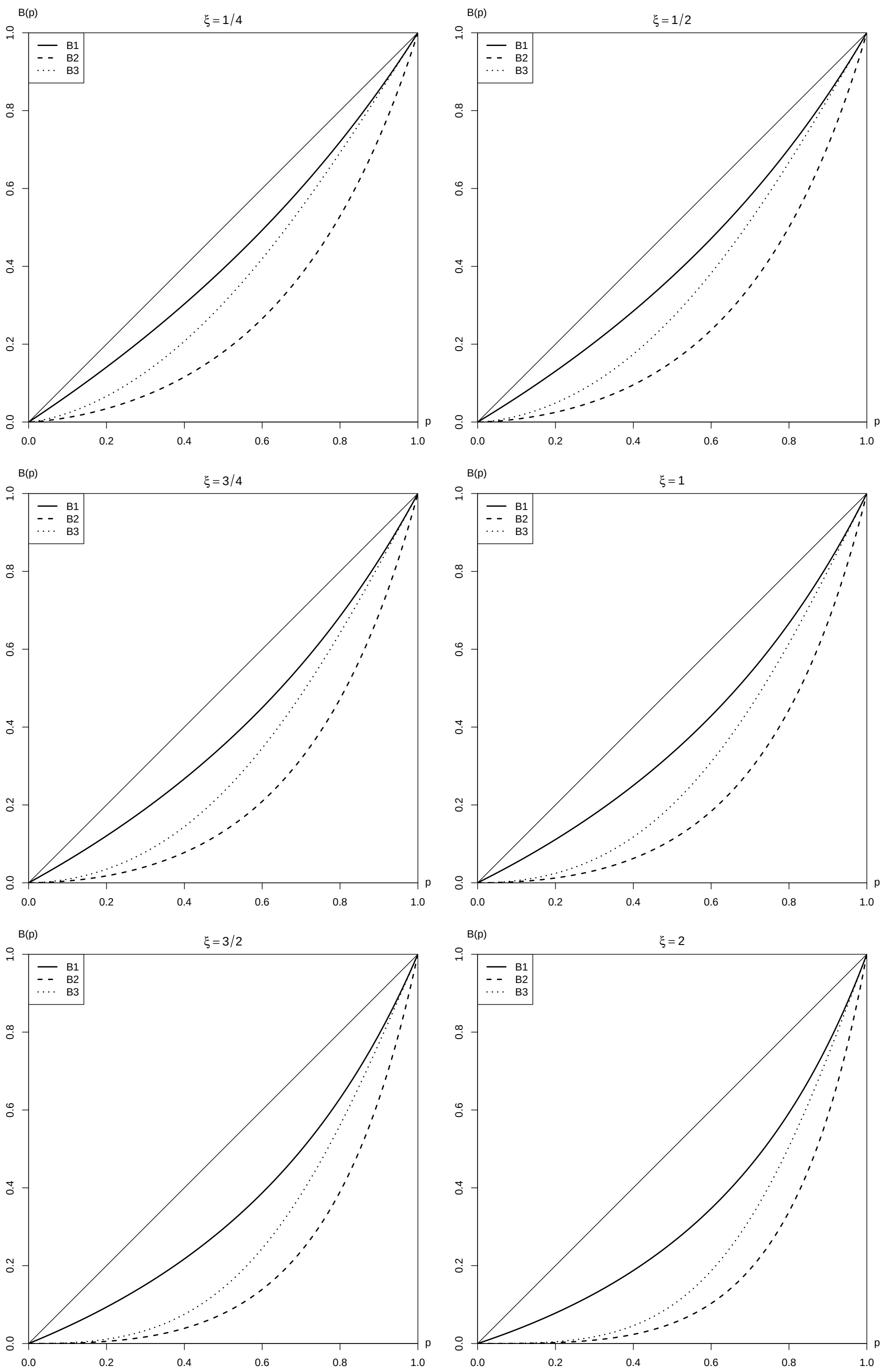

Fig. 2. The quantile versions of the Bonferroni curve for the generalized Pareto distribution for different values of ξ

Table 2. The values of the quantile versions of the Bonferroni index corresponding to the generalized Pareto distribution for the selected values of ξ

	$\xi = 1/4$	$\xi = 1/2$	$\xi = 3/4$	$\xi = 1$	$\xi = 3/2$	$\xi = 2$
BI_1	0.5716	0.5858	0.5998	0.6137	0.6408	0.6667
BI_2	0.7213	0.7397	0.7568	0.7726	0.8005	0.8240
BI_3	0.6278	0.6510	0.6728	0.6931	0.7295	0.7603

Table 2 contains values of the quantile versions of the Bonferroni index corresponding to the distribution $\mathcal{GP}(\xi, 1)$ for the values of ξ considered in the simulations.

We will denote by $\hat{B}_i^E(p)$, $\hat{B}_i^H(p)$, $\hat{B}_i^{HF}(p)$, $\hat{B}_i^{MP}(p)$, $\hat{B}_i^{JP}(p)$ and $\hat{B}_i^M(p)$, $i = 1, 2, 3$, the plug-in estimators obtained by replacing the unknown quantiles by their estimators $\hat{Q}_n^E$, $\hat{Q}_n^H$, $\hat{Q}_n^{HF}$, $\hat{Q}_n^{MP}$, $\hat{Q}_n^{JP}$ and $\hat{Q}_n^M$, respectively. The error of the estimators $\hat{B}_{i,n}$, $i = 1, 2, 3$, based on the sample $\mathbf{X} = (X_1, \ldots, X_n)$ from the distribution F, was measured by the integrated mean square error (MISE), i.e.

$$\mathrm{MISE}(\hat{B}_{i,n}) = \mathrm{E}_F\Big\{\int_0^1 [\hat{B}_{i,n}(p; \mathbf{X}) - B_i(p)]^2 \, dp\Big\}.$$

We estimated MISE by

$$\widehat{\mathrm{MISE}}(\hat{B}_{i,n}) = \frac{1}{S}\sum_{j=1}^{S}\int_0^1 [\hat{B}_{i,n}^{(j)}(p) - B_i(p)]^2 \, dp,$$

where S is the number of random samples generated, whereas $\hat{B}_{i,n}^{(j)}(p)$, $i = 1, 2, 3$, $j = 1, \ldots, S$, is the estimate of the ith quantile version of the Bonferroni curve, based on the implementation of the jth sample.

We also considered the accuracy of estimators of the quantile versions of the Bonferroni index, measured in terms of bias and mean square error given by

$$\mathrm{BIAS}(\widehat{BI}_i) = \mathrm{E}(\widehat{BI}_i(\mathbf{X}) - BI_i) \quad \text{and} \quad \mathrm{MSE}(\widehat{BI}_i) = \mathrm{E}(\widehat{BI}_i(\mathbf{X}) - BI_i)^2,$$

respectively. We determined the estimates of the above values from the formulas

$$\widehat{\mathrm{BIAS}}(\widehat{BI}_i) = \frac{1}{S}\sum_{j=1}^{S}(\widehat{BI}_i^{(j)} - BI_i) \quad \text{and} \quad \widehat{\mathrm{MSE}}(\widehat{BI}_i) = \frac{1}{S}\sum_{j=1}^{S}(\widehat{BI}_i^{(j)} - BI_i)^2,$$

where S is the number of random samples generated, whereas $\widehat{BI}_i^{(j)}$ is the estimate of the ith quantile version of the Bonferroni index, based on the implementation of the jth sample.

The smallest values of estimates of integrated mean square errors, bias estimates and estimates of mean square errors, for each value of ξ and n, are shown in boldface.

Table 3. Simulation results concerning the quantile versions of the Bonferroni curve and the Bonferroni index for $n = 50$ (multiplied by 100)

	B_1			B_2			B_3		
	MISE	BIAS	MSE	MISE	BIAS	MSE	MISE	BIAS	MSE
				$\xi = 1/4$					
E	0.6720	-0.3088	0.3330	0.3889	-0.9575	0.1630	0.7543	-0.8085	0.1970
H	0.6158	-1.0975	0.3229	0.3605	-0.8795	0.1607	0.7223	-0.7187	0.1952
HF	0.6055	-0.8288	0.3142	0.3550	-0.6929	0.1557	0.6933	-0.4732	0.1898
MP	0.5931	**-0.2987**	**0.3014**	0.3475	**-0.3294**	**0.1482**	**0.6418**	**0.0058**	**0.1829**
JP	**0.5889**	-1.0953	0.3211	**0.3403**	-0.8501	0.1592	0.7049	-0.6942	0.1938
M	0.5988	-1.0541	0.3196	0.3429	-0.8570	0.1592	0.7071	-0.6923	0.1937
				$\xi = 1/2$					
E	0.6793	**-0.2414**	0.3357	0.3911	-0.9449	0.1552	0.7195	-0.7890	0.1894
H	0.6160	-1.1021	0.3222	0.3610	-0.8802	0.1530	0.6882	-0.7164	0.1877
HF	0.6059	-0.8411	0.3136	0.3555	-0.7046	0.1483	0.6618	-0.4840	0.1825
MP	0.5933	-0.3260	**0.3006**	0.3480	**-0.3634**	**0.1409**	**0.6148**	**-0.0320**	**0.1756**
JP	**0.5890**	-1.1062	0.3205	**0.3396**	-0.8472	0.1513	0.6698	-0.6873	0.1862
M	0.5991	-1.0617	0.3190	0.3422	-0.8581	0.1514	0.6723	-0.6911	0.1862
				$\xi = 3/4$					
E	0.7011	**-0.0803**	0.3482	0.3932	-0.8987	0.1491	0.6896	-0.7260	0.1858
H	0.6391	-0.9616	0.3371	0.3614	-0.8418	0.1473	0.6584	-0.6648	0.1844
HF	0.6297	-0.7096	0.3290	0.3562	-0.6771	0.1429	0.6343	-0.4458	0.1796
MP	0.6183	-0.2120	**0.3169**	0.3490	**-0.3577**	**0.1360**	**0.5918**	**-0.0212**	**0.1730**
JP	**0.6116**	-0.9700	0.3356	**0.3390**	-0.8028	0.1457	0.6389	-0.6300	0.1830
M	0.6220	-0.9226	0.3340	0.3417	-0.8170	0.1458	0.6417	-0.6379	0.1831
				$\xi = 1$					
E	0.7110	**-0.1604**	0.3534	0.3943	-0.9410	0.1432	0.6715	-0.7789	0.1790
H	0.6420	-1.1101	0.3389	0.3604	-0.8950	0.1412	0.6400	-0.7298	0.1776
HF	0.6322	-0.8651	0.3302	0.3551	-0.7400	0.1369	0.6174	-0.5231	0.1727
MP	0.6193	-0.3813	**0.3167**	0.3473	**-0.4399**	**0.1300**	**0.5774**	**-0.1230**	**0.1657**
JP	**0.6141**	-1.1242	0.3374	**0.3365**	-0.8525	0.1393	0.6192	-0.6910	0.1758
M	0.6248	-1.0734	0.3357	0.3394	-0.8696	0.1395	0.6223	-0.7025	0.1760
				$\xi = 3/2$					
E	0.7435	**-0.1262**	0.3611	0.4137	-0.9643	0.1381	0.6608	-0.8135	0.1785
H	0.6657	-1.1845	0.3465	0.3752	-0.9298	0.1363	0.6272	-0.7791	0.1770
HF	0.6557	-0.9548	0.3375	0.3697	-0.7925	0.1322	0.6071	-0.5952	0.1721
MP	0.6415	-0.5011	**0.3231**	0.3615	**-0.5276**	**0.1254**	**0.5716**	**-0.2409**	**0.1648**
JP	**0.6359**	-1.2093	0.3451	**0.3479**	-0.8793	0.1341	0.6034	-0.7322	0.1751
M	0.6470	-1.1529	0.3431	0.3510	-0.9020	0.1344	0.6071	-0.7502	0.1754
				$\xi = 2$					
E	0.7850	**-0.0751**	0.3837	0.4308	-1.0526	0.1348	0.6678	-0.9147	0.1773
H	0.7002	-1.2243	0.3685	0.3883	-1.0243	0.1329	0.6316	-0.8877	0.1757
HF	0.6901	-1.0097	0.3594	0.3824	-0.9016	0.1289	0.6130	-0.7231	0.1707
MP	0.6752	-0.5855	**0.3443**	0.3732	**-0.6656**	**0.1220**	**0.5799**	**-0.4071**	**0.1629**
JP	**0.6689**	-1.2599	0.3672	**0.3583**	-0.9680	0.1304	0.6050	-0.8345	0.1735
M	0.6806	-1.1984	0.3650	0.3614	-0.9951	0.1307	0.6093	-0.8572	0.1738

Table 4. Simulation results concerning the quantile versions of the Bonferroni curve and the Bonferroni index for $n = 100$ (multiplied by 100)

	B_1			B_2			B_3		
	MISE	BIAS	MSE	MISE	BIAS	MSE	MISE	BIAS	MSE
				$\xi = 1/4$					
E	0.3433	-0.4684	0.1722	0.1911	-0.6369	0.0802	0.5209	-0.5172	0.0966
H	0.3302	-0.6531	0.1691	0.1864	-0.5759	0.0794	0.5094	-0.4435	0.0960
HF	0.3266	-0.5158	0.1665	0.1847	-0.4825	0.0779	0.4960	-0.3197	0.0944
MP	**0.3223**	**-0.2412**	**0.1624**	0.1822	**-0.2988**	**0.0755**	**0.4710**	**-0.0765**	**0.0922**
JP	0.3225	-0.6509	0.1688	**0.1805**	-0.5678	0.0792	0.5046	-0.4368	0.0957
M	0.3252	-0.6384	0.1685	0.1812	-0.5693	0.0792	0.5052	-0.4360	0.0957
				$\xi = 1/2$					
E	0.3483	-0.3840	0.1750	0.1909	-0.6142	0.0760	0.4834	-0.4751	0.0926
H	0.3342	-0.5909	0.1712	0.1861	-0.5688	0.0754	0.4739	-0.4231	0.0922
HF	0.3310	-0.4581	0.1687	0.1845	-0.4811	0.0739	0.4619	-0.3063	0.0907
MP	0.3270	**-0.1926**	**0.1649**	0.1821	**-0.3092**	**0.0716**	**0.4394**	**-0.0777**	**0.0887**
JP	**0.3264**	-0.5903	0.1709	**0.1799**	-0.5594	0.0751	0.4688	-0.4151	0.0920
M	0.3292	-0.5773	0.1706	0.1806	-0.5619	0.0751	0.4694	-0.4158	0.0920
				$\xi = 3/4$					
E	0.3488	-0.3821	0.1739	0.1948	-0.5880	0.0748	0.4576	-0.4623	0.0926
H	0.3334	-0.6385	0.1699	0.1896	-0.5562	0.0743	0.4493	-0.4277	0.0922
HF	0.3301	-0.5099	0.1674	0.1881	-0.4740	0.0730	0.4383	-0.3177	0.0908
MP	0.3259	**-0.2527**	**0.1634**	0.1857	**-0.3131**	**0.0707**	**0.4177**	**-0.1029**	**0.0887**
JP	**0.3255**	-0.6396	0.1697	**0.1830**	-0.5461	0.0740	0.4438	-0.4186	0.0920
M	0.3283	-0.6261	0.1694	0.1838	-0.5494	0.0741	0.4445	-0.4204	0.0920
				$\xi = 1$					
E	0.3527	-0.3613	0.1747	0.1984	-0.6403	0.0730	0.4444	-0.5201	0.0914
H	0.3362	-0.6323	0.1705	0.1930	-0.6175	0.0726	0.4369	-0.4980	0.0911
HF	0.3330	-0.5080	0.1680	0.1914	-0.5402	0.0712	0.4265	-0.3943	0.0896
MP	0.3288	**-0.2595**	**0.1641**	0.1888	**-0.3891**	**0.0689**	**0.4072**	**-0.1921**	**0.0872**
JP	**0.3281**	-0.6347	0.1703	**0.1860**	-0.6059	0.0723	0.4310	-0.4877	0.0909
M	0.3311	-0.6207	0.1700	0.1868	-0.6098	0.0723	0.4318	-0.4902	0.0909
				$\xi = 3/2$					
E	0.3734	-0.3652	0.1852	0.2037	-0.6589	0.0688	0.4176	-0.5391	0.0893
H	0.3550	-0.6899	0.1805	0.1977	-0.6508	0.0686	0.4112	-0.5353	0.0891
HF	0.3519	-0.5738	0.1780	0.1960	-0.5825	0.0673	0.4022	-0.4433	0.0876
MP	0.3473	**-0.3418**	**0.1736**	0.1933	**-0.4492**	**0.0650**	**0.3853**	**-0.2642**	**0.0852**
JP	**0.3464**	-0.6952	0.1803	**0.1896**	-0.6370	0.0682	0.4044	-0.5228	0.0888
M	0.3495	-0.6802	0.1800	0.1905	-0.6422	0.0683	0.4053	-0.5268	0.0889
				$\xi = 2$					
E	0.4002	-0.3675	0.1979	0.2095	-0.6631	0.0660	0.4014	-0.5660	0.0881
H	0.3793	-0.7508	0.1931	0.2031	-0.6640	0.0660	0.3956	-0.5724	0.0881
HF	0.3760	-0.6428	0.1904	0.2015	-0.6034	0.0648	0.3876	-0.4905	0.0867
MP	0.3710	**-0.4267**	**0.1857**	0.1987	**-0.4851**	**0.0627**	**0.3725**	**-0.3310**	**0.0842**
JP	**0.3699**	-0.7591	0.1929	**0.1941**	-0.6485	0.0656	0.3880	-0.5584	0.0878
M	0.3733	-0.7431	0.1926	0.1951	-0.6547	0.0656	0.3891	-0.5634	0.0878

Taking into account the MISE, while estimating $\hat{B}_1$ and $\hat{B}_2$, the $\hat{B}^{JP}$ estimator has the least $\widehat{\mathrm{MISE}}(\hat{B}_{i,n})$ (except the case of estimating B_1 with $\xi = 1/4$ and $n = 100$), and for $\hat{B}_3$ the best estimator is $\hat{B}^{MP}$.

In the sense of MSE, the best estimator of BI for $i = 1, 2, 3$ is $\widehat{BI}^{MP}$. Considering BIAS, in estimation of BI_i the smallest $\widehat{\mathrm{BIAS}}(\widehat{BI}_i)$ occurs for

- $\widehat{BI}^{MP}$, for $i = 1$ with $n = 100$ and for $i = 2, 3$ with $n = 50$,
- $\widehat{BI}^{E}$, for $i = 1$ and $n = 50$ (except $\xi = 1/4$).

3. Real data analysis. To illustrate all the nonparametric estimators under study, we apply them to the data representing the remission times (in months) of a random sample of 128 bladder cancer patients. The data come from [1] and are contained in Table 5.

Table 5. Remission times (in months) of a random sample of 128 bladder cancer patients

0.08	0.20	0.40	0.50	0.51	0.81	0.90	1.05	1.19	1.26
1.35	1.40	1.46	1.76	2.02	2.02	2.07	2.09	2.23	2.26
2.46	2.54	2.62	2.64	2.69	2.69	2.75	2.83	2.87	3.02
3.25	3.31	3.36	3.36	3.48	3.52	3.57	3.64	3.70	3.82
3.88	4.18	4.23	4.26	4.33	4.34	4.40	4.50	4.51	4.87
4.98	5.06	5.09	5.17	5.32	5.32	5.34	5.41	5.41	5.49
5.62	5.71	5.85	6.25	6.54	6.76	6.93	6.94	6.97	7.09
7.26	7.28	7.32	7.39	7.59	7.62	7.63	7.66	7.87	7.93
8.26	8.37	8.53	8.65	8.66	9.02	9.22	9.47	9.74	10.06
10.34	10.66	10.75	11.25	11.64	11.79	11.98	12.02	12.03	12.07
12.63	13.11	13.29	13.80	14.24	14.76	14.77	14.83	15.96	16.62
17.12	17.14	17.36	18.10	19.13	20.28	21.73	22.69	23.63	25.74
25.82	26.31	32.15	34.26	36.66	43.01	46.12	79.05		

Table 6. Estimates of the quantile versions of the Bonferroni index for the data set considered

	BI_1	BI_2	BI_3
E	0.5049	0.6535	0.5425
H	0.4937	0.6534	0.5425
HF	0.4949	0.6543	0.5437
MP	0.4975	0.6561	0.5460
JP	0.4936	0.6533	0.5425
M	0.4937	0.6534	0.5425

Table 6 contains the estimates of the quantile versions of the Bonferroni index. Figure 3 shows the plots of the plug-in estimators of the quantile versions of the Bonferroni curve for the data set considered. Only the empirical estimator of the first version of the Bonferroni curve differs visibly

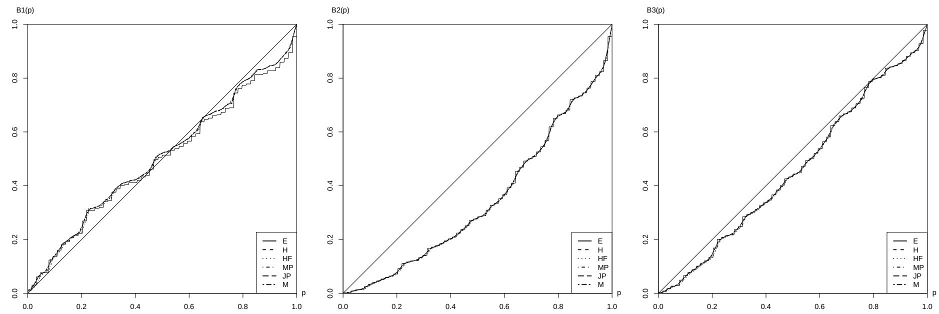

Fig. 3. Plots of estimates of the plug-in estimators of the quantile versions of the Bonferroni curve for the data set considered

in Figure 3. This is confirmed by the value of the empirical estimator of the first version of the Bonferroni index, which differs from the remaining estimators by more than 0.01, while for other estimators, and other versions of the index, these differences are less than 0.01.

References

[1] M. Aldeni, C. Lee and F. Famoye, *Families of distributions arising from the quantile of generalized lambda distribution*, J. Statist. Distributions Appl. 4 (2017), art 25. 8 pp.

[2] A. B. Atkinson, *On the measurement of inequality*, J. Economic Theory 2 (1970), 244–263.

[3] C. E. Bonferroni, *Elementi di Statistica Generale*, Libreria Seeber, Firenze, 1930.

[4] J. L. Gastwirth, *Measures of economic inequality focusing on the status of the lower and middle income groups*, Statistics and Public Policy 3 (2016), 1–9.

[5] C. Gini, *Variabilità e mutabilità: contributo allo studio delle distribuzioni e delle relazioni statistiche [Fasc. I]*, Tipogr. di P. Cuppini, Bologna, 1912.

[6] G. M. Giorgi and S. Nadarajah, *Bonferroni and Gini indices for various parametric families of distributions*, Metron 68 (2010), 23–46.

[7] A. Hazen, *Storage to be provided in impounding municipal water supply*, Trans. Amer. Soc. Civil Engineers 77 (1914), 1539–1640.

[8] R. J. Hyndman and Y. Fan, *Sample quantiles in statistical packages*, Amer. Statistician 50 (1996), 361–365.

[9] A. Jokiel-Rokita and M. Pulit, *Nonparametric estimation of the ROC curve based on smoothed empirical distribution function*, Statistics and Computing 23 (2013), 703–712.

[10] A. Jokiel-Rokita and A. Siedlaczek, *Quantile estimation via distribution fitting*, Appl. Math. (Warsaw) 46 (2019), 283–301.

[11] M. O. Lorenz, *Methods of measuring the concentration of wealth*, Publ. Amer. Statist. Assoc. 9 (1905), 209–219

[12] L. Makkonen and M. Pajari, *Defining sample quantiles by the true rank probability*, J. Probab. Statist. 2014, art. 326579, 6 pp.
[13] L. A. Prendergast and R. G. Staudte, *Quantile versions of the Lorenz curve*, Electron. J. Statist. 10 (2016), 1896–1926.
[14] L. A. Prendergast and R. G. Staudte, *A simple and effective inequality measure*, Amer. Statistician 72 (2018), 328–343.
[15] H. Theil, *Economics and Information Theory*, North-Holland, 1967.
[16] M. Zenga, *Proposta per un indice di concentrazione basato sui rapporti fra quantili di popolazione e quantili di reddito*, Giornale degli Economisti e Annali di Economia 43 (1984), 301–326.
[17] M. Zenga, *Inequality curve and inequality index based on the ratios between lower and upper arithmetic means*, Statistica e Applicazioni 5 (2007), 3–27.

Agnieszka Siedlaczek
Faculty of Mathematics, Physics and Computer Science
University of Opole
45-040 Opole, Poland
E-mail: asiedlaczek@uni.opole.pl