APPLICATIONES MATHEMATICAE
46,2 (2019), pp. 283–301

Alicja Jokiel-Rokita (Wrocław)
Agnieszka Siedlaczek (Opole)

QUANTILE ESTIMATION VIA DISTRIBUTION FITTING

Abstract. This paper focuses on nonparametric estimation of quantiles, based on estimators of the distribution function. We review some known and recommended quantile estimators and propose a new one, which has all the desired properties of quantile estimators. The consistency and asymptotic normality of the estimators is proved. The estimators considered are compared in a small simulation study.

1. Introduction. Let X be a real-valued random variable with cumulative distribution function (c.d.f.) $F(t) = P(X \leq t)$. This article deals with estimating quantiles of X of level $p \in (0,1)$, i.e.,

$$x_p = \inf\{t \in \mathbb{R} : F(t) \geq p\} =: Q(p).$$

The function Q is called the *quantile function*. Throughout this paper we assume that $F \in \mathcal{F}$, where $\mathcal{F}$ is the family of all continuous and strictly increasing distribution functions on the real line, i.e., $F \in \mathcal{F}$ if and only if $F(a) = 0$, $F(b) = 1$, and F is strictly increasing on (a,b) for some $-\infty \leq a < b \leq +\infty$, where a and b are unknown. Under this assumption $x_p = F^{-1}(p)$, where F^{-1} is the inverse of F in the usual sense.

Given a random sample $X_n = (X_1, \ldots, X_n)$ from the distribution F we are interested in estimation of x_p. A natural estimator of the x_p is the value of the *empirical quantile function* (EQF)

$$\hat{x}^E_{p,n} := \hat{Q}^E_n(p; X_n) := \inf\{t : \hat{F}^E_n(t) \geq p\},$$

2010 *Mathematics Subject Classification*: Primary 62G05; Secondary 62G30.
Key words and phrases: nonparametric estimation, quantile function, level crossing empirical distribution function, kernel estimator.
Received 13 December 2018.
Published online 23 July 2019.

where

$$\hat{F}_n^E(t) := \frac{1}{n}\sum_{i=1}^{n} I_{(-\infty,t]}(X_i) \tag{1.1}$$

is the *empirical distribution function* (EDF), and $I_A(x) = 1$ if $x \in A$ and 0 otherwise. The estimators $\hat{x}_{p,n}^E$ are called *sample quantiles*. In the literature and statistical software there are a large number of different definitions of sample quantiles. In a widely cited article [11], the authors analysed nine different definitions. Most of them are based on quantile function estimators $\hat{Q}_n$ constructed by linearly interpolating between so called *plotting positions*, i.e., the points p_k, $k = 1, \ldots, n$, for which $\hat{Q}_n(p_k) = X_{k:n}$, where $X_{1:n}, \ldots, X_{n:n}$ denote the order statistics of the sample X_n. In [11] the authors compared sample quantiles by describing their motivation and analyzing whether or not they enjoy the following six properties:

P1: $\hat{Q}_n(p)$ is a continuous function of p for each realization $x_n = (x_1, \ldots, x_n)$ of the random sample X_n,

P2: for each realization x_n of X_n, and for each $p \in (0, 1)$,

$$\sum_{i=1}^{n} I_{(-\infty,\hat{Q}_n(p)]}(x_i) \geq np,$$

P3: for each realization x_n of X_n, and for each $p \in (0, 1)$,

$$\sum_{i=1}^{n} I_{(-\infty,\hat{Q}_n(p)]}(x_i) = \sum_{i=1}^{n} I_{[\hat{Q}_n(1-p),\infty)}(x_i),$$

P4: where $\hat{Q}_n^{-1}(x)$ is uniquely defined,

$$\hat{Q}_n^{-1}(X_{k:n}) + \hat{Q}_n^{-1}(X_{(n-k+1):n}) = 1 \quad \text{for } k = 1, \ldots, n,$$

P5: where $\hat{Q}_n^{-1}(x)$ is uniquely defined,

$$\hat{Q}_n^{-1}(X_{1:n}) > 0 \quad \text{and} \quad \hat{Q}_n^{-1}(X_{n:n}) < 1,$$

P6: $\hat{Q}_n(0.5)$ is equal to the sample median defined by

$$\begin{cases} (X_{l:n} + X_{(l+1):n})/2 & \text{if } n = 2l, \\ X_{(l+1):n} & \text{if } n = 2l + 1. \end{cases}$$

Among the estimators compared in [11], only the estimator proposed by Hazen [7], which is based on the plotting positions

$$p_k^H = \frac{k - 1/2}{n}, \qquad k = 1, \ldots, n, \tag{1.2}$$

has all six properties. However, Hyndman and Fan [11] recommend the estimator based on the plotting positions

$$p_k^{HF} = \frac{k - 1/3}{n + 1/3}, \qquad k = 1, \ldots, n, \tag{1.3}$$

because it gives approximately median-unbiased estimates of $Q(p)$ regardless of the distribution (it fails to satisfy P3). The need for a standard definition of sample quantiles was also discussed by Langford [13] who identified twelve different definitions used in statistical software. In [15] the sample quantiles proposed by Weibull [21] and Gumbel [6], based on plotting positions

$$p_k^{WG} = \frac{k}{n + 1}, \qquad k = 1, \ldots, n, \tag{1.4}$$

are recommended.

In quantile estimation, smoothed versions of the empirical distribution function $\hat{F}_n^E$ have also been considered. As a result one obtains order statistics or their linear combinations (see e.g. [2]), or some more sophisticated estimators for more advanced smoothing techniques ([1], [3], [4], [5], [23]).

A different approach was presented in a series of Zieliński's papers (e.g., [24], [25], [26], [28]), where the best equivariant nonparametric estimators of quantiles are derived under various criterions.

Let us notice that if an estimator $\hat{Q}_n$ is based on an estimator $\hat{F}_n$ of the distribution function, i.e. $\hat{Q}_n(p) = \hat{F}_n^{-1}(p)$, then it satisfies P1–P6 if for each realization x_n of X_n the estimator $\hat{F}_n$ has the following properties:

PF1: $\hat{F}_n(t)$ is a continuous and strictly increasing function of t,
PF2: $\hat{F}_n(x_{k:n}) \le k/n$ for $k = 1, \ldots, n$,
PF3: $\sum_{i=1}^n I_{(-\infty,p]}(\hat{F}_n(x_i)) = \sum_{i=1}^n I_{[1-p,\infty)}(\hat{F}_n(x_i))$ for $p \in (0, 1)$,
PF4: $\hat{F}_n(x_{k:n}) + \hat{F}_n(x_{(n-k+1):n}) = 1$,
PF5: $\hat{F}_n(x_{1:n}) > 0$ and $\hat{F}_n(x_{n:n}) < 1$,
PF6: $\hat{F}_n((x_{l:n} + x_{(l+1):n})/2) = 0.5$ if $n = 2l$ and $\hat{F}_n(x_{(l+1):n}) = 0.5$ if $n = 2l + 1$.

In Section 2 we recall some known quantile estimators and propose new estimators, based on known continuous versions of the EDF and an estimator based on a new distribution function estimator which satisfies PF1–PF6. The distribution function estimator proposed also satisfies

PF7: $\hat{F}_n(ct + a; cx_1 + a, \ldots, cx_n + a) = \hat{F}_n(t; x_1, \ldots, x_n)$ for all $a, t \in \mathbb{R}$ and $c > 0$.

The estimator $\hat{F}_n$ satisfying PF7 leads to the quantile function estimator $\hat{Q}_n$ which is equivariant with respect to affine transformations, i.e.,

P7: $\hat{Q}_n(p; cX_1 + a, \ldots, cX_n + a) = c\hat{Q}_n(p; X_1, \ldots, X_n) + a.$

Asymptotic properties of the estimators proposed are given in Section 3. In Section 4 some results from simulation studies are provided. Real data analysis is carried out in Section 5. The paper ends with some concluding remarks in Section 6.

2. Quantile estimation based on a distribution function estimator. In this section we recall some known distribution function estimators, propose a new estimator, and give formulae for the quantile estimators based on them.

2.1. Quantile estimation based on the empirical distribution function. A traditional nonparametric estimator of the distribution function is the EDF given by (1.1). Accordingly, a nonparametric estimator of x_p is the empirical quantile

$$\hat{x}^E_{p,n} = X_{([np]+1):n}, \tag{2.1}$$

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

Remark 2.1. The function $\hat{Q}^E_n(p) = \hat{x}^E_{p,n}$ only has properties P2, P7, and for n odd, property P6.

The estimator $\hat{x}^E_{p,n}$ has the desired asymptotic properties, described in the following two theorems.

Theorem 2.2 ([19, Theorem 2.3.2]). *Let $X_1, \dots, X_n$ be i.i.d. random variables with a c.d.f. F satisfying $F(x_p - \epsilon) < p < F(x_p + \epsilon)$ for any $\epsilon > 0$. Then, for every $\epsilon > 0$ and $n = 1, 2, \dots,$*

$$P(|\hat{x}^E_{p,n} - x_p| > \epsilon) \le 2\exp(-2n\delta^2_\epsilon),$$

where $\delta_\epsilon = \min\{F(x_p + \epsilon) - p, p - F(x_p - \epsilon)\}$.

Theorem 2.3 ([20, Theorem 5.10]). *Let $X_1, \dots, X_n$ be i.i.d. random variables with a c.d.f. F.*

(i) $\lim_{n\to\infty} P(\sqrt{n}(\hat{x}^E_{p,n} - x_p) \le 0) = \Phi(0) = 1/2$, *where Φ is the c.d.f. of the standard normal distribution.*

(ii) *If F is continuous at x_p and $F'(x_p-) > 0$ exists, then*

$$\lim_{n\to\infty} P(\sqrt{n}(\hat{x}^E_{p,n} - x_p) \le t) = \Phi(t/\sigma^-_F), \quad t < 0,$$

where $\sigma^-_F = \sqrt{p(1-p)}/F'(x_p-)$.

(iii) *If F is continuous at x_p and $F'(x_p+) > 0$ exists, then*

$$\lim_{n\to\infty} P(\sqrt{n}(\hat{x}^E_{p,n} - x_p) \le t) = \Phi(t/\sigma^+_F), \quad t < 0,$$

where $\sigma^+_F = \sqrt{p(1-p)}/F'(x_p+)$.

(iv) *If $F'(x_p)$ exists and is positive, then*

$$\sqrt{n}(\hat{x}^E_{p,n} - x_p) \to_d \mathcal{N}(0, \sigma^2_F), \quad \text{where } \sigma_F = \sqrt{p(1-p)}/F'(x_p).$$

The estimator $\hat{x}^{E}_{p,n}$ is not symmetric. We call an estimator $X_{k(p):n}$ of the pth quantile *symmetric* if $k(1-p) = n - k(p) + 1$ (see property P4). A rationale for this condition is that if a quantile of order p is estimated, say, by the smallest order statistic $X_{1:n}$, then the quantile of order $1-p$ should be estimated by the largest order statistic $X_{n:n}$. For $p = l/n$ we have $\hat{x}^{E}_{p,n} = X_{l:n}$, and for $p = 1 - l/n$ we have $\hat{x}^{E}_{p,n} = X_{(n-l):n}$ instead of $X_{(n-l+1):n}$. Another disadvantage of the estimator $\hat{x}^{E}_{p,n}$ is that if $p = 1/2$ and $n = 2m$ for an integer m, then $\hat{x}^{E}_{1/2,n}$ equals $X_{m:n}$ instead of being a combination of two central-order statistics $X_{m:n}$ and $X_{(m+1):n}$.

To overcome the disadvantages of $\hat{x}^{E}_{p,n}$, Zieliński [26] defined the estimator

$$\hat{x}^{EM}_{p,n} = X_{k(p):n}, \tag{2.2}$$

where

$$k(p) = \begin{cases} np & \text{if } np \text{ is an integer and } p < 0.5,\\ np+1 & \text{if } np \text{ is an integer and } p > 0.5,\\ n/2 + I_{(0,1/2]}(U) & \text{if } np \text{ is an integer and } p = 0.5,\\ [np]+1 & \text{if } np \text{ is not an integer,} \end{cases}$$

where U is a uniformly $\mathcal{U}(0,1)$ distributed random variable independent of the observations $X_1, \ldots, X_n$.

Remark 2.4. Note that $\hat{x}^{EM}_{p,n}$ may differ from $\hat{x}^{E}_{p,n}$ only when estimating quantiles of order $p = j/n$, $j = 1, \ldots, n$, i.e., if np is an integer. If $p = 0.5$, then $\hat{x}^{EM}_{p,n}$ is a median unbiased estimator of the median $x_{0.5}$; the estimator $\hat{x}^{E}_{p,n}$ does not have this property.

2.2. Quantile estimation based on a level crossing empirical distribution function. Relying on the concept of level crossing, Huang and Brill [9] constructed a level crossing empirical distribution function (LCEDF) of the form

$$\hat{F}^{HB}_{n}(t) = \sum_{i=1}^{n} w_{n,i} \mathbf{1}_{(-\infty,t]}(X_i),$$

where

$$w_{n,i} = \begin{cases} \dfrac{1}{2}\left(1 - \dfrac{n-2}{\sqrt{n(n-1)}}\right) & \text{for } i = 1, n,\\ \dfrac{1}{\sqrt{n(n-1)}} & \text{for } i = 2, \ldots, n-1. \end{cases}$$

Remark 2.5. With probability one,

$$\sup_{t\in\mathbb{R}} |\hat{F}^{HB}_{n}(t) - \hat{F}^{E}_{n}(t)| \le 1/n.$$

In [10] an efficiency function for the LCEDF relative to the EDF has been derived, and it has been shown that the LCEDF gives more efficient estimates than the EDF in the tails of any underlying continuous distribution, for both small and large sample sizes.

Making use of the LCEDF, we obtain the quantile estimator

$$\hat{x}_n^{HB}(p) = X_{([b]+2):n}, \tag{2.3}$$

where

$$b = \sqrt{n(n-1)}\left[p - \frac{1}{2}\left(1 - \frac{n-2}{\sqrt{n(n-1)}}\right)\right].$$

Remark 2.6. The function $\hat{Q}_n^{HB}(p) = \hat{x}_{p,n}^{HB}$ only has properties P2, P7, and for n odd, property P6.

2.3. Quantile estimation based on a kernel estimator of the distribution function. It seems unnatural to estimate a continuous distribution function by a step function. In the abundant literature of the subject, one can find different approaches to smoothing empirical distribution functions. The kernel distribution function estimator, first introduced by Nadaraya [16], provides an alternative to the EDF. It is defined by

$$\hat{F}_n^K(t) = \frac{1}{n}\sum_{j=1}^{n} \mathcal{K}\left(\frac{t - X_j}{h_n}\right),$$

where $\mathcal{K}(v) = \int_{-\infty}^{v} K(z)\,dz$, K is a kernel function, and h_n is a bandwidth parameter.

The corresponding estimator of the quantile x_p is then defined by

$$\hat{x}_p^K = \inf\{t \in \mathbb{R} : \hat{F}_n^K(t) \geq p\}. \tag{2.4}$$

Under some assumptions on K, f and h, it is shown in [16] that $\hat{x}_p^K$ (appropriately normalized) has an asymptotic standard normal distribution. Another notable property of $\hat{x}_p^K$, almost sure consistency, was proved in [22]. Azzalini [2] considered second-order properties of $\hat{x}_p^K$. In [17] necessary and sufficient conditions for the asymptotic normality of $\hat{x}_p^K$ are given.

2.4. Quantile estimation based on a kernel estimator with random bandwidth. In kernel distribution function estimation, Zieliński [27] proposed to replace the standard smoothing parameter h_n by a random bandwidth

$$H_n = \min\{X_{j:n} - X_{(j-1):n} : j = 2, \ldots, n\}.$$

Assuming that

$$\mathcal{K}(t)=\begin{cases}0 & \text{for } t\le -1/2,\\ 1/2 & \text{for } t=0,\\ 1 & \text{for } t\ge 1/2,\end{cases}$$

and $\mathcal{K}$ is continuous and nondecreasing in $(0,1)$, he obtained a continuous estimator of the unknown distribution function with asymptotic properties similar to the empirical distribution function. We will denote the estimator proposed in [27] by $\hat{F}_n^Z$.

In the special case when

$$\mathcal{K}(t)=\begin{cases}0 & \text{for } t\le -1/2,\\ t+1/2 & \text{for } -1/2<t<1/2,\\ 1 & \text{for } t\ge 1/2,\end{cases}$$

the estimator $\hat{F}_n^Z(x)$ equals

$$\begin{cases}\frac{1}{n}\big(\frac{x-X_{k:n}}{H_n}+\frac{1}{2}\big)+\frac{k-1}{n} & \text{for } x\in[X_{k:n}-H_n/2, X_{k:n}+H_n/2],\\ \frac{k}{n} & \text{for } x\in(X_{k:n}+H_n/2, X_{(k+1):n}-H_n/2),\end{cases}$$

for $k=1,\dots,n$, and accordingly

$$\hat{x}_{p,n}^Z=\hat{Q}_n^Z(p)=X_{k:n}+H_n(np-k+1/2) \tag{2.5}$$

for $p\in((k-1)/n, k/n]$ and $k=1,\dots,n$.

Remark 2.7. With probability one,

$$\sup_{t\in\mathbb{R}}|\hat{F}_n^Z(t)-\hat{F}_n^E(t)|\le 1/(2n).$$

Remark 2.8. The function $\hat{Q}_n^Z(p)$ only fails property P1 and property P6 if n is even.

To the best of our knowledge, $\hat{x}_{p,n}^Z$, based on the distribution function estimator $\hat{F}_n^Z$, has not been considered in the literature as a quantile estimator.

2.5. Quantile estimation based on a continuous and strictly increasing estimator of the distribution function. The estimator $\hat{F}_n^Z$ proposed in [27] is constant on some intervals of the real line $\mathbb{R}$. Consequently, the quantile function estimator $\hat{Q}_n^Z$ based on $\hat{F}_n^Z$ is not continuous. In [12] a construction of a continuous and easily invertible estimator of the distribution function is proposed, based on Zieliński's idea [27]. Denote by $X_{0:n}$, $X_{(n+1):n}$ random variables such that $X_{0:n}\le X_{1:n}$ and $X_{(n+1):n}\ge X_{n:n}$

almost surely,

$$M_j(\mathbf{X}_n) = \frac{X_{(j-1):n} + X_{j:n}}{2}, \qquad j = 1, \dots, n+1,$$

$$R_j(\mathbf{X}_n) = M_{j+1}(\mathbf{X}_n) - M_j(\mathbf{X}_n) = \frac{X_{(j+1):n} - X_{(j-1):n}}{2}, \qquad j = 1, \dots, n.$$

With this notation, in [12] a distribution function estimator was defined by

$$\hat{F}_n^{JP}(t) = \frac{1}{n}\sum_{j=1}^{n} T\bigg(\frac{t - M_j(\mathbf{X}_n)}{R_j(\mathbf{X}_n)}\bigg),$$

where

$$T(x) = \begin{cases} 0 & \text{for } x < 0, \\ r(x) & \text{for } 0 \le x \le 1, \\ 1 & \text{for } x > 1, \end{cases}$$

where $r\colon [0,1] \to [0,1]$ is a continuous, strictly increasing function such that $r(0) = 0$, $r(1) = 1$, e.g., $r(x) = x$. In comparison with the kernel estimator, they replaced the bandwidth h_n by the differences $R_j(\cdot) = M_{j+1}(\cdot) - M_j(\cdot)$. The plain order statistics have been replaced by the statistics $M_j(\mathbf{X}_n)$, indicating the centers of the intervals between the consecutive order statistics.

When $r(x) = x$,

$$\hat{F}_n^{JP}(t) = \begin{cases} 0 & \text{for } t < M_1(\mathbf{X}_n), \\ \dfrac{t - M_k(\mathbf{X}_n)}{nR_k(\mathbf{X}_n)} + \dfrac{k-1}{n} & \text{for } M_k(\mathbf{X}_n) \le t \le M_{k+1}(\mathbf{X}_n), \\ 1 & \text{for } t > M_{n+1}(\mathbf{X}_n), \end{cases}$$

for $k = 1, \dots, n$. It is easy to see that

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

if

$$\frac{X_{(k-1):n} + X_{k:n}}{2} \le t \le \frac{X_{k:n} + X_{(k+1):n}}{2} \qquad \text{for } k = 1, \dots, n.$$

Remark 2.9. With probability one,

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

The corresponding quantile function estimator $\hat{Q}_n^{JP}$ 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 $(k-1)/n < p \leq k/n$ for $k = 1, \ldots, n$, and

$$\hat{x}_{p,n}^{JP} = \hat{Q}_n^{JP}(p). \tag{2.6}$$

REMARK 2.10. The function $\hat{Q}_n^{JP}$ satisfies P1, P2, P5, P7.

REMARK 2.11. For $k/n < p \leq (k+1)/n$, $k = 0, \ldots, n-2$, and for $(n-1)/n < p < 1$, with probability one,

$$\hat{x}_{p,n}^{E} - \frac{X_{(k+1):n} - X_{k:n}}{2} < \hat{x}_{p,n}^{JP} \leq \hat{x}_{p,n}^{E} + \frac{X_{(k+2):n} - X_{(k+1):n}}{2}.$$

The function $\hat{Q}_n^{JP}(p)$ fails properties P3, P4 and P6. This is a consequence of the estimator $\hat{F}_n^{JP}$ failing conditions PF3, PF4 and PF6. Therefore, we propose modifying $\hat{F}_n^{JP}$ in the following way. Denote

$$X_{0:n} = X_{1:n} - \frac{X_{2:n} - X_{1:n}}{2} = \frac{3}{2}X_{1:n} - \frac{1}{2}X_{2:n},$$
$$X_{(n+1):n} = X_{n:n} + \frac{X_{n:n} - X_{(n-1):n}}{2} = \frac{3}{2}X_{n:n} - \frac{1}{2}X_{(n-1):n}.$$

If $n = 2l+1$ set $\hat{F}_n^M(X_{(l+1):n}) = 1/2$, and for $k \in \{1, \ldots, l\}$, also when $n = 2l$,

$$\hat{F}_n^M(X_{k:n}) = [\hat{F}_n^{JP}(X_{k:n}) + (1 - \hat{F}_n^{JP}(X_{(n-k+1):n}))]/2 = 1 - \hat{F}_n^M(X_{(n-k+1):n}).$$

Moreover for $j \in \{1, \ldots, n-1\}$,

$$\hat{F}_n^M((X_{j:n} + X_{(j+1):n})/2) = j/n,$$

$\hat{F}_n^M(X_{0:n}) = 0$, $\hat{F}_n^M(X_{(n+1):n}) = 1$. The estimator $\hat{F}_n^M$ is constructed by linearly interpolating between the points

$$(X_{0:n}, 0), \quad (X_{1:n}, \hat{F}_n^M(X_{1:n})), \quad (X_{1:n} + X_{2:n})/2,$$
$$\hat{F}_n^M((X_{1:n} + X_{2:n})/2)), \ \ldots, \ (X_{n:n}, \hat{F}_n^M(X_{n:n})), \quad (X_{(n+1):n}, 1).$$

REMARK 2.12. With probability one,

$$\sup_{t \in \mathbb{R}} |\hat{F}_n^M(t) - \hat{F}_n^E(t)| \leq 1/n.$$

The corresponding quantile function estimator $\hat{Q}_n^M$ is the inverse function of $\hat{F}_n^M$, i.e.

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

for $k = 1, \dots, n-1$ and

$$\hat{x}_{p,n}^M = \hat{Q}_n^M(p). \tag{2.7}$$

REMARK 2.13. The function $\hat{Q}_n^M(p)$ has properties P1–P7.

REMARK 2.14. For $k/n < p \le (k+1)/n$, $k = 0, \dots, n-2$, and for $(n-1)/n < p < 1$, with probability one,

$$\hat{x}_{p,n}^E - \frac{X_{(k+1):n} - X_{k:n}}{2} < \hat{x}_{p,n}^M \le \hat{x}_{p,n}^E + \frac{X_{(k+2):n} - X_{(k+1):n}}{2}.$$

3. Asymptotic properties of the quantile estimators based on continuous versions of the empirical distribution function. Let $\tilde{F}_n$ be an estimator of F such that with probability one,

$$\sup_{t\in\mathbb{R}} |\tilde{F}_n(t) - \hat{F}_n^E(t)| \le 1/n. \tag{3.1}$$

Denote by

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

the quantile estimator based on $\tilde{F}_n$.

LEMMA 3.1 ([8]). *Let $Y_1, \dots, Y_n$ be independent random variables satisfying $P(a \le Y_i \le b)) = 1$, $i = 1, \dots, n$, where $a < b$. Then, for $t > 0$,*

$$P\Big(\sum_{i=1}^n Y_i - \sum_{i=1}^n E(Y_i) \ge nt\Big) \le \exp\{-2nt^2/(b-a)^2\}.$$

THEOREM 3.2. *Let $X_1, \dots, X_n$ be i.i.d. random variables with a c.d.f. F satisfying $F(x_p - \epsilon) < p < F(x_p + \epsilon)$ for any $\epsilon > 0$ and $p \in (0,1)$. Denote $\tilde{\delta}_\epsilon = \min\{F(x_p+\epsilon)-p-1/n,\ p-F(x_p-\epsilon)-1/n\}$, $n_0(\epsilon, F, p) := \min\{n \in \mathbb{N} : \tilde{\delta}_\epsilon > 0\}$. Then, for every $\epsilon > 0$ and $n \ge n_0(\epsilon, F, p)$,*

$$P(|\tilde{x}_{p,n} - x_p| > \epsilon) \le 2\exp(-2n\tilde{\delta}_\epsilon^2).$$

Proof. Let $\epsilon > 0$. Write

$$P(|\tilde{x}_{p,n} - x_p| > \epsilon) = P(\tilde{x}_{p,n} > x_p + \epsilon) + P(\tilde{x}_{p,n} < x_p - \epsilon).$$

The first summand above can be written as

$$\begin{aligned} P(\tilde{x}_{p,n} > x_p + \epsilon) &= P(p > \tilde{F}_n(x_p + \epsilon)) \\ &= P\big(p > \hat{F}_n^E(x_p + \epsilon) - \hat{F}_n^E(x_p + \epsilon) + \tilde{F}_n(x_p + \epsilon)\big) \\ &\le P(\hat{F}_n^E(x_p + \epsilon) - 1/n < p), \end{aligned}$$

where the last inequality follows from (3.1). But

$$\begin{aligned} &P\big(\hat{F}_n^E(x_p + \epsilon) - 1/n < p\big) \\ &\quad = P\Big(\sum_{i=1}^n I(X_i > x_p + \epsilon) - n[1 - F(x_p + \epsilon)] > n[F(x_p + \epsilon) - p - 1/n]\Big), \end{aligned}$$

and when $F(x_p + \epsilon) - p - 1/n > 0$, from Lemma 3.1 we have

$$P(\tilde{x}_{p,n} > x_p + \epsilon) \le \exp\{-2n[F(x_p + \epsilon) - p - 1/n]^2\}.$$

In an analogous way we can show that

$$P(\tilde{x}_{p,n} < x_p - \epsilon) \le \exp\{-2n[p - F(x_p - \epsilon) - 1/n]^2\},$$

and the proof is complete. ∎

In view of Remarks 2.7, 2.9 and 2.12, we have

Corollary 3.3. *The conclusion of Theorem* 3.2 *holds for* $\hat{x}_{p,n}^Z$, $\hat{x}_{p,n}^{JP}$ *and* $\hat{x}_{p,n}^M$, *i.e.*,

$$\begin{aligned} P(|\hat{x}_{p,n}^Z - x_p| > \epsilon) &\le 2\exp(-2n\tilde{\delta}_\epsilon^2), \\ P(|\hat{x}_{p,n}^{JP} - x_p| > \epsilon) &\le 2\exp(-2n\tilde{\delta}_\epsilon^2), \\ P(|\hat{x}_{p,n}^M - x_p| > \epsilon) &\le 2\exp(-2n\tilde{\delta}_\epsilon^2). \end{aligned}$$

Corollary 3.4. *Under the assumption of Theorem* 3.2, *the estimators* $\hat{x}_{p,n}^Z$, $\hat{x}_{p,n}^{JP}$, $\hat{x}_{p,n}^M$ *are strongly consistent for* x_p.

Denote by $F'(x_p-)$ and $F'(x_p+)$ the left and right derivatives of F at x_p, respectively.

Theorem 3.5. *Let* $p \in (0,1)$. *Suppose that* F *is continuous at* x_p.

(i) *If* $F'(x_p-) > 0$ *exists, then for* $t < 0$,

$$\lim_{n\to\infty} P\bigg(\frac{n^{1/2}(\tilde{x}_{p,n} - x_p)}{[p(1-p)]^{1/2}/F'(x_p-)} \le t\bigg) = \Phi(t).$$

(ii) *If* $F'(x_p+) > 0$ *exists, then for* $t > 0$,

$$\lim_{n\to\infty} P\bigg(\frac{n^{1/2}(\tilde{x}_{p,n} - x_p)}{[p(1-p)]^{1/2}/F'(x_p+)} \le t\bigg) = \Phi(t).$$

(iii) $\lim_{n\to\infty} P\big(n^{1/2}(\tilde{x}_{p,n} - x_p) \le 0\big) = \Phi(0) = \frac{1}{2}$.

Proof. Fix t. Let $A > 0$ be a normalizing constant to be specified later. Denote

$$G_n(t) = P\big(n^{1/2}(\tilde{x}_{p,n} - x_p)/A \le t\big).$$

From the fact that $\tilde{F}_n^{-1}(p) \le t$ if and only if $\tilde{F}_n(t) \ge p$, we have

$$G_n(t) = P(\tilde{x}_{p,n} \le x_p + tAn^{-1/2}) = P\big(p \le \tilde{F}_n(x_p + tAn^{-1/2})\big).$$

The estimator $\tilde{F}_n$ satisfies (3.1), therefore

$$G_n(t) \le P\big(p \le \hat{F}_n^E(x_p + tAn^{-1/2}) + 1/n\big), \tag{3.2}$$

$$G_n(t) \ge P\big(p \le \hat{F}_n^E(x_p + tAn^{-1/2}) - 1/n\big). \tag{3.3}$$

We will show that under some assumption, $\lim_{n\to\infty} G_n(t) = \Phi(t)$ for all $t \in \mathbb{R}$ and for $A > 0$ appropriately chosen. Denote by $S_n(\Delta)$ a binomial $\mathcal{B}(n, \Delta)$ random variable,

$$S_n^*(\Delta) = \frac{S_n(\Delta) - n\Delta}{[n\Delta(1-\Delta)]^{1/2}},$$

and set

$$\Delta_{nt} = F(x_p + tAn^{-1/2}).$$

We have

$$\begin{aligned} P\big(p \le \hat{F}_n^E(x_p + tAn^{-1/2}) + 1/n\big) &= P(S_n(\Delta_{nt}) \ge np - 1) \\ &= P(S_n^*(\Delta_{nt}) \ge -c_{nt}), \end{aligned} \tag{3.4}$$

where

$$c_{nt} = \frac{n^{1/2}(\Delta_{nt} - p - 1/n)}{[\Delta_{nt}(1-\Delta_{nt})]^{1/2}}. \tag{3.5}$$

Applying the Berry–Esséen theorem (see e.g. [19, Theorem 1.9.5]), we can write

$$\sup_{x\in\mathbb{R}} |P(S_n^*(\Delta) < x) - \Phi(x)| \le C\frac{\rho_\Delta}{\sigma_\Delta^3 n^{1/2}},$$

where C is a universal constant, $\sigma_\Delta^2 = \mathrm{Var} S_1(\Delta) = \Delta(1-\Delta)$, and $\rho_\Delta = E|S_1(\Delta) - \Delta|^3 = \Delta(1-\Delta)((1-\Delta)^2 + \Delta^2)$. Therefore

$$\sup_{x\in\mathbb{R}} |P(S_n^*(\Delta) < x) - \Phi(x)| \le C\frac{(1-\Delta)^2 + \Delta^2}{[\Delta(1-\Delta)]^{1/2} n^{1/2}}. \tag{3.6}$$

From the triangle inequality and (3.6) we have

$$\begin{aligned}(3.7)\qquad |P(S_n^*(\varDelta_{nt}) \ge -c_{nt}) - \varPhi(t)| &\le |\varPhi(-c_{nt}) - P(S_n^*(\varDelta_{nt}) < c_{nt})| + |\varPhi(c_{nt} - \varPhi(t)| \\ &\le C\frac{(1-\varDelta_{nt})^2 + \varDelta_{nt}^2}{[\varDelta_{nt}(1-\varDelta_{nt})]^{1/2} n^{1/2}} + |\varPhi(c_{nt}) - \varPhi(t)|.\end{aligned}$$

Since F is continuous at x_p, the first summand tends to zero as $n \to \infty$. Writing

(3.8)

$$c_{nt} = \frac{tA}{[\varDelta_{nt}(1-\varDelta_{nt})]^{1/2}} \, \frac{F(x_p + tAn^{-1/2}) - F(x_p)}{tAn^{-1/2}} - \frac{n^{-1/2}}{[\varDelta_{nt}(1-\varDelta_{nt})]^{1/2}}$$

by (3.5), we see that if either $t < 0$ and $A = [p(1-p)]^{1/2}/F(x_p-)]$, or $t > 0$ and $A = [p(1-p)]^{1/2}/F(x_p+)]$, or $t = 0$ and $A = 1$, then $c_{nt} \to t$ as $n \to \infty$. From (3.4), (3.7) and (3.8) we infer that if either $t < 0$ and $A = [p(1-p)]^{1/2}/F(x_p-)]$, or $t > 0$ and $A = [p(1-p)]^{1/2}/F(x_p+)]$, or $t = 0$ and $A = 1$, then

$$(3.9)\qquad P\big(p \le \hat{F}_n^E(x_p + tAn^{-1/2}) + 1/n\big) \to \varPhi(t) \quad \text{as } n \to \infty.$$

Analogously we can show that if either $t < 0$ and $A = [p(1-p)]^{1/2}/F(x_p-)]$, or $t > 0$ and $A = [p(1-p)]^{1/2}/F(x_p+)]$, or $t = 0$ and $A = 1$, then

$$(3.10)\qquad P\big(p \le \hat{F}_n^E(x_p + tAn^{-1/2}) - 1/n\big) \to \varPhi(t) \quad \text{as } n \to \infty.$$

From (3.2), (3.3), (3.9) and (3.10) we deduce that if either $t < 0$ and $A = [p(1-p)]^{1/2}/F(x_p-)]$, or $t > 0$ and $A = [p(1-p)]^{1/2}/F(x_p+)]$, or $t = 0$ and $A = 1$, then

$$G_n(t) \to \varPhi(t) \quad \text{as } n \to \infty,$$

and the theorem is proved. ∎

COROLLARY 3.6. *Let $0 < p < 1$. If F is differentiable at x_p and $F'(x_p) > 0$, then*

$$\tilde{x}_{p,n} \text{ is } \mathcal{AN}\left(x_p, \frac{p(1-p)}{[F'(x_p)]^2 n}\right).$$

COROLLARY 3.7. *Let $0 < p < 1$. If F has a density f in a neighborhood of x_p and f is positive and continuous at x_p, then*

$$\tilde{x}_{p,n} \text{ is } \mathcal{AN}\left(x_p, \frac{p(1-p)}{f^2(x_p) n}\right).$$

Taking into account Remarks 2.7, 2.9 and 2.12, we obtain

COROLLARY 3.8. *The conclusion of Theorem 3.5 holds for $\hat{x}_{p,n}^Z$, $\hat{x}_{p,n}^{JP}$ and $\hat{x}_{p,n}^M$.*

4. A simulation study. A simulation experiment was conducted in order to investigate the accuracy of the following ten quantile estimators:

- $\hat{x}^{E}_{p,n}$ given by (2.1), in tables denoted by E,
- $\hat{x}^{EM}_{p,n}$ given by (2.2), denoted by EM,
- $\hat{x}^{HB}_{p,n}$ given by (2.3), denoted by HB,
- $\hat{x}^{K}_{p,n}$ given by (2.4), denoted by K,
- $\hat{x}^{Z}_{p,n}$ given by (2.5), denoted by Z,
- $\hat{x}^{JP}_{p,n}$ given by (2.6), denoted by JP,
- $\hat{x}^{M}_{p,n}$ given by (2.7), denoted by M,
- $\hat{x}^{H}_{p,n}$ based on plotting positions given by (1.2), denoted by H,
- $\hat{x}^{HF}_{p,n}$ based on plotting positions given by (1.3), denoted by HF,
- $\hat{x}^{WG}_{p,n}$ based on plotting positions given by (1.4), denoted by WG,

in the case of small sample sizes.

We generated samples of sizes $n = 15$ and $n = 30$ from the following distributions:

- Student $\mathcal{T}(r)$ with $r = 3$ and $r = 15$ degrees of freedom,
- generalized Pareto distribution $\mathcal{GP}(\xi, \sigma)$ with distribution function

$$F_{\xi,\sigma}(t) = \begin{cases} [1-(1+\xi t/\sigma)^{-1/\xi}]I_{[0,\infty)}(t) & \text{for } \xi > 0,\ \sigma > 0, \\ [1-(1+\xi t/\sigma)^{-1/\xi}]I_{[0,-\sigma/\xi]}(t) & \text{for } \xi < 0,\ \sigma > 0, \\ [1-\exp(-t/\sigma)]I_{[0,\infty)}(t) & \text{for } \xi = 0,\ \sigma > 0, \end{cases} \tag{4.1}$$

with $\sigma = 1$ and $\xi \in \{-1/4, 1/4, 2/3\}$.

The mean and variance of $\mathcal{GP}(\xi, \sigma)$ are given by

$$E(X_{\xi,\sigma}) = \frac{\sigma}{1-\xi} \qquad \text{for } \xi < 1,$$

$$\mathrm{Var}(X_{\xi,\sigma}) = \frac{\sigma^2}{(1-\xi)^2(1-2\xi)} \qquad \text{for } \xi < 1/2$$

(see e.g. [14]). When $\xi \geq 1$ the mean is not defined, and when $\xi \geq 1/2$ the variance is not defined. For $\sigma = 1$ and $\xi = -1/4, 1/4, 2/3$ we have

$$\begin{aligned} E(X_{-1/4,1}) &= 4/5, & \mathrm{Var}(X_{-1/4,1}) &= 32/75, \\ E(X_{1/4,1}) &= 4/3, & \mathrm{Var}(X_{1/4,1}) &= 32/9, \\ E(X_{2/3,1}) &= 3, & \mathrm{Var}(X_{2/3,1}) &\text{ is not defined.} \end{aligned}$$

All the computations were done in the statistical computing language R. The *quantile* function from the *stats* package in R was used to calculate the quantile estimates based on plotting positions, i.e. $\hat{x}^{H}_{p,n}$, $\hat{x}^{HF}_{p,n}$, $\hat{x}^{WG}_{p,n}$. The $\hat{x}^{K}_{p,n}$ estimates were obtained using the *qkde* function from the *ks* package in R. The remaining quantile estimates were calculated with the use of our own implementations.

Since we do not assume that the observable random variable has a finite expected value, to compare the estimators considered, the differences between the median of the estimators and the true value of the quantiles, denoted by ME, and the differences between the upper and lower quartile of the estimators, denoted by IQR, were used to measure the accuracy of the estimators. The ME and IQR were estimated by generating 10000 random samples for each distribution and each n.

The simulation results for the Student and generalized Pareto distribution are presented in Tables 1, 2 and 3, 4, respectively, where the estimated ME and IQR of each estimator considered are given. The exact theoretical quantiles of the Student distribution have been read off from statistical tables, while in the case of the generalized Pareto distribution, they have been calculated as appropriate values of the inverse of the distribution functions (4.1), i.e.

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

Table 1. Simulation results for the Student distribution and $n = 15$

		p									
		0.05		0.25		0.5		0.75		0.95	
		ME	IQR	ME	IQR	ME	IQR	ME	IQR	ME	IQR
$r = 3$	E	−0.1062	1.7696	−0.0395	0.5898	−0.0061	0.4716	0.0363	0.5953	0.1326	1.8464
	EM	−0.1062	1.7696	−0.0395	0.5898	−0.0061	0.4716	0.0363	0.5953	0.1326	1.8464
	H	0.0849	1.5464	0.0142	0.5603	−0.0061	0.4716	−0.0178	0.5664	−0.0639	1.5875
	HF	−0.0349	1.6785	−0.0036	0.5712	−0.0061	0.4716	0.0009	0.5738	0.0585	1.7366
	WG	−0.1062	1.7696	−0.0395	0.5898	−0.0061	0.4716	0.0363	0.5953	0.1326	1.8464
	HB	−0.1062	1.7696	−0.0395	0.5898	−0.0061	0.4716	0.0363	0.5953	0.1326	1.8464
	K	−0.1392	1.5015	−0.1286	0.5617	−0.0308	0.4306	0.0697	0.5607	0.1027	1.5481
	Z	−0.1018	1.7698	−0.0349	0.5879	−0.0061	0.4716	0.0316	0.5942	0.1287	1.8438
	JP	0.1366	1.4799	0.0051	0.5532	−0.0061	0.4554	−0.0043	0.5588	−0.1194	1.5303
	M	0.1366	1.4799	0.0076	0.5591	−0.0061	0.4716	−0.0097	0.5654	−0.1194	1.5303
$r = 15$	E	−0.0525	0.8451	−0.037	0.4859	−0.0029	0.4331	0.0336	0.4834	0.0618	0.8354
	EM	−0.0525	0.8451	−0.037	0.4859	−0.0029	0.4331	0.0336	0.4834	0.0618	0.8354
	H	0.0651	0.7573	0.0133	0.47	−0.0029	0.4331	−0.0175	0.4644	−0.0569	0.7542
	HF	−0.0079	0.8086	−0.0051	0.4746	−0.0029	0.4331	0.0009	0.4684	0.0166	0.8017
	WG	−0.0525	0.8451	−0.037	0.4859	−0.0029	0.4331	0.0336	0.4834	0.0618	0.8354
	HB	−0.0525	0.8451	−0.037	0.4859	−0.0029	0.4331	0.0336	0.4834	0.0618	0.8354
	K	−0.169	0.7476	−0.0891	0.4576	−0.0185	0.3861	0.0524	0.444	0.1366	0.7318
	Z	−0.0477	0.8446	−0.0329	0.4856	−0.0029	0.4331	0.0295	0.484	0.0583	0.8347
	JP	0.0966	0.7368	0.0077	0.458	−0.0021	0.4173	−0.011	0.4573	−0.086	0.732
	M	0.0966	0.7368	0.0059	0.4676	−0.0029	0.4331	−0.0137	0.4614	−0.086	0.732

It follows from the simulation study that

- among the three estimators $\hat{x}_{p,n}^{HF}$, $\hat{x}_{p,n}^{WG}$ and $\hat{x}_{p,n}^{H}$, based on plotting positions, in most of the cases considered, the first two, which are recommended in the literature, have greater estimated IQR than $\hat{x}_{p,n}^{H}$;
- in almost all cases considered the estimator $\hat{x}_{p,n}^{HF}$ has the smallest absolute values of the estimated ME (it is known that $\hat{x}_{p,n}^{HF}$ is median unbiased of order $o(n^{-1/2})$, see [18]);

Table 2. Simulation results for the Student distribution and $n = 30$

		p									
		0.05		0.25		0.5		0.75		0.95	
		ME	IQR	ME	IQR	ME	IQR	ME	IQR	ME	IQR
$r = 3$	E	0.1051	1.088	0.017	0.407	0.0433	0.328	−0.0139	0.4062	−0.1143	1.1058
	EM	0.1051	1.088	0.017	0.407	−0.0012	0.3318	−0.0139	0.4062	−0.1143	1.1058
	H	0.1051	1.088	0.017	0.407	−0.002	0.3231	−0.0139	0.4062	−0.1143	1.1058
	HF	−0.0896	1.1985	0.0049	0.4041	−0.002	0.3231	−0.0018	0.4032	0.0792	1.2179
	WG	−0.4243	1.478	−0.0184	0.4069	−0.002	0.3231	0.021	0.4062	0.4106	1.4982
	HB	0.1051	1.088	0.017	0.407	0.0433	0.328	−0.0139	0.4062	−0.1143	1.1058
	K	−0.1188	0.9713	−0.1225	0.3935	−0.0313	0.3005	0.0662	0.3899	0.0641	0.9838
	Z	0.1051	1.088	0.017	0.407	0.041	0.3281	−0.0139	0.4062	−0.1143	1.1058
	JP	−0.09	1.1926	0.0101	0.3983	−0.002	0.3231	−0.0081	0.3927	0.0885	1.192
	M	−0.0574	1.1767	0.0105	0.3988	−0.002	0.3231	−0.0105	0.3996	0.0559	1.1861
$r = 15$	E	0.0512	0.5653	0.009	0.343	0.0425	0.3148	−0.0047	0.343	−0.0493	0.5716
	EM	0.0512	0.5653	0.009	0.343	0.0005	0.3211	−0.0047	0.343	−0.0493	0.5716
	H	0.0512	0.5653	0.009	0.343	−0.0004	0.309	−0.0047	0.343	−0.0493	0.5716
	HF	−0.0291	0.5635	−0.0004	0.3407	−0.0004	0.309	0.0042	0.3418	0.0288	0.5761
	WG	−0.1781	0.6264	−0.0191	0.3422	−0.0004	0.309	0.0224	0.3396	0.1841	0.618
	HB	0.0512	0.5653	0.009	0.343	0.0425	0.3148	−0.0047	0.343	−0.0493	0.5716
	K	−0.173	0.5219	−0.0905	0.3178	−0.0188	0.2723	0.0516	0.3158	0.1314	0.5088
	Z	0.0512	0.5653	0.009	0.343	0.0405	0.3142	−0.0047	0.343	−0.0493	0.5716
	JP	−0.018	0.5462	0.0066	0.3345	−0.0004	0.309	−0.0015	0.3346	0.0191	0.5446
	M	−0.0062	0.5565	0.0064	0.3388	−0.0004	0.309	−0.0021	0.3394	0.0079	0.5646

Table 3. Simulation results for the GP distribution and $n = 15$

		p									
		0.05		0.25		0.5		0.75		0.95	
		ME	IQR	ME	IQR	ME	IQR	ME	IQR	ME	IQR
$\xi = \frac{-1}{4}$, $\sigma = 1$	E	−0.0053	0.0728	−0.015	0.1799	−0.0041	0.287	0.0325	0.4219	0.048	0.6928
	EM	−0.0053	0.0728	−0.015	0.1799	−0.0041	0.287	0.0325	0.4219	0.048	0.6928
	H	0.0139	0.0767	0.0085	0.181	−0.0041	0.287	−0.0092	0.4009	−0.0526	0.6184
	HF	0.002	0.0731	0.0011	0.1804	−0.0041	0.287	0.0057	0.4079	0.0095	0.6601
	WG	−0.0053	0.0728	−0.015	0.1799	−0.0041	0.287	0.0325	0.4219	0.048	0.6928
	HB	−0.0053	0.0728	−0.015	0.1799	−0.0041	0.287	0.0325	0.4219	0.048	0.6928
	K	−0.196	0.146	−0.0071	0.15	0.0233	0.2495	0.0248	0.384	0.0169	0.628
	Z	−0.003	0.0727	−0.0125	0.1805	−0.0041	0.287	0.0308	0.422	0.0458	0.6942
	JP	0.0187	0.0791	0.0101	0.1773	0.0016	0.2731	−0.0033	0.3869	−0.0795	0.6029
	M	0.0187	0.0791	0.0088	0.179	−0.0041	0.287	−0.0062	0.395	−0.0795	0.6029
$\xi = \frac{1}{4}$, $\sigma = 1$	E	−0.0047	0.0739	−0.0156	0.2038	0.0003	0.4094	0.0601	0.8778	0.1852	3.3828
	EM	−0.0047	0.0739	−0.0156	0.2038	0.0003	0.4094	0.0601	0.8778	0.1852	3.3828
	H	0.0151	0.0789	0.0103	0.2093	0.0003	0.4094	−0.0141	0.8145	−0.1691	2.8817
	HF	0.0028	0.0746	0.001	0.2064	0.0003	0.4094	0.0095	0.8334	0.0475	3.18
	WG	−0.0047	0.0739	−0.0156	0.2038	0.0003	0.4094	0.0601	0.8778	0.1852	3.3828
	HB	−0.0047	0.0739	−0.0156	0.2038	0.0003	0.4094	0.0601	0.8778	0.1852	3.3828
	K	−0.3249	0.2395	−0.0127	0.184	0.0664	0.3889	0.0317	0.7993	0.0125	3.1803
	Z	−0.0021	0.0739	−0.0131	0.2049	0.0003	0.4094	0.0573	0.8778	0.1834	3.3796
	JP	0.0201	0.0816	0.0134	0.2083	0.0097	0.3987	0.0088	0.8008	−0.268	2.7549
	M	0.0201	0.0816	0.0104	0.2083	0.0003	0.4094	−0.0059	0.8067	−0.268	2.7549
$\xi = \frac{2}{3}$, $\sigma = 1$	E	−0.0044	0.0751	−0.0161	0.2337	−0.0018	0.5452	0.108	1.616	0.82	13.7029
	EM	−0.0044	0.0751	−0.0161	0.2337	−0.0018	0.5452	0.108	1.616	0.82	13.7029
	H	0.0156	0.0809	0.0144	0.2418	−0.0018	0.5452	−0.0253	1.4923	−0.2886	11.3304
	HF	0.0034	0.076	0.0056	0.239	−0.0018	0.5452	0.0197	1.5331	0.385	12.7587
	WG	−0.0044	0.0751	−0.0161	0.2337	−0.0018	0.5452	0.108	1.616	0.82	13.7029
	HB	−0.0044	0.0751	−0.0161	0.2337	−0.0018	0.5452	0.108	1.616	0.82	13.7029
	K	−0.5082	0.4168	−0.0429	0.2186	0.1062	0.5668	0.0196	1.4752	0.4154	13.3591
	Z	−0.001	0.0753	−0.0125	0.2351	−0.0018	0.5452	0.1036	1.6145	0.8158	13.703
	JP	0.0207	0.0842	0.018	0.2394	0.0215	0.539	0.0457	1.5125	−0.5865	10.7577
	M	0.0207	0.0842	0.0138	0.2405	−0.0018	0.5452	0.001	1.5077	−0.5865	10.7577

Table 4. Simulation results for the GP distribution and $n = 30$

		p 0.05 ME	0.05 IQR	0.25 ME	0.25 IQR	0.5 ME	0.5 IQR	0.75 ME	0.75 IQR	0.95 ME	0.95 IQR
	E	0.0055	0.057	0.0034	0.1328	0.0258	0.2084	−0.0077	0.2944	−0.0436	0.4864
	EM	0.0055	0.057	0.0034	0.1328	−0.0015	0.2065	−0.0077	0.2944	−0.0436	0.4864
	H	0.0055	0.057	0.0034	0.1328	−0.0017	0.2004	−0.0077	0.2944	−0.0436	0.4864
	HF	0.0012	0.053	0.0005	0.1309	−0.0017	0.2004	0.0007	0.2933	0.0145	0.4753
$\xi = \frac{-1}{4}$	WG	−0.0083	0.0448	−0.0059	0.1275	−0.0017	0.2004	0.0181	0.2951	0.1355	0.4852
$\sigma = 1$	HB	0.0055	0.057	0.0034	0.1328	0.0258	0.2084	−0.0077	0.2944	−0.0436	0.4864
	K	−0.1628	0.0873	−0.0034	0.105	0.0149	0.1797	0.0181	0.2783	0.0063	0.4414
	Z	0.0055	0.057	0.0034	0.1328	0.0244	0.2081	−0.0077	0.2944	−0.0436	0.4864
	JP	0.0075	0.0514	0.0049	0.1279	−0.0017	0.2004	−0.0031	0.2882	−0.0019	0.439
	M	0.0054	0.0531	0.0038	0.1293	−0.0017	0.2004	−0.0063	0.2926	−0.0199	0.4546
	E	0.0053	0.0597	0.001	0.1544	0.038	0.2952	−0.0165	0.5894	−0.1576	2.1609
	EM	0.0053	0.0597	0.001	0.1544	−0.0007	0.2912	−0.0165	0.5894	−0.1576	2.1609
	H	0.0053	0.0597	0.001	0.1544	0.0019	0.2814	−0.0165	0.5894	−0.1576	2.1609
	HF	0.0009	0.0553	−0.0025	0.1519	0.0019	0.2814	0.001	0.5916	0.2306	2.3616
$\xi = \frac{1}{4}$	WG	−0.0081	0.0469	−0.0095	0.1478	0.0019	0.2814	0.0364	0.5965	0.8679	2.8967
$\sigma = 1$	HB	0.0053	0.0597	0.001	0.1544	0.038	0.2952	−0.0165	0.5894	−0.1576	2.1609
	K	−0.2689	0.1372	−0.0123	0.1292	0.0392	0.2703	0.0188	0.561	−0.1384	2.1011
	Z	0.0053	0.0597	0.001	0.1544	0.0367	0.2947	−0.0165	0.5894	−0.1576	2.1609
	JP	0.0076	0.054	0.002	0.151	0.0019	0.2814	−0.0039	0.5775	0.2455	2.3157
	M	0.0049	0.0555	0.0008	0.1523	0.0019	0.2814	−0.0117	0.5822	0.0725	2.227
	E	0.0064	0.06	0.0031	0.1743	0.0562	0.4063	−0.0326	1.0656	−0.6529	7.1551
	EM	0.0064	0.06	0.0031	0.1743	−0.0013	0.3881	−0.0326	1.0656	−0.6529	7.1551
	H	0.0064	0.06	0.0031	0.1743	0.0034	0.3787	−0.0326	1.0656	−0.6529	7.1551
	HF	0.0017	0.0555	−0.0014	0.1715	0.0034	0.3787	−0.0011	1.0764	1.2442	9.2688
$\xi = \frac{2}{3}$	WG	−0.008	0.0466	−0.0097	0.1668	0.0034	0.3787	0.0637	1.1057	3.7078	13.4398
$\sigma = 1$	HB	0.0064	0.06	0.0031	0.1743	0.0562	0.4063	−0.0326	1.0656	−0.6529	7.1551
	K	−0.4297	0.2558	−0.0486	0.1615	0.0369	0.3831	−0.0327	1.0253	−0.7062	7.0855
	Z	0.0064	0.06	0.0031	0.1743	0.0548	0.4059	−0.0326	1.0656	−0.6529	7.1551
	JP	0.0087	0.0547	0.004	0.1683	0.0034	0.3787	−0.0009	1.0542	1.4456	9.802
	M	0.0049	0.0554	0.0023	0.1703	0.0034	0.3787	−0.017	1.0557	0.6897	8.7422

- no estimator is uniformly best for all the parameter values and all quantile levels, but the proposed estimators perform well compared to existing estimators.

5. Real data analysis. To illustrate all the nonparametric estimators under study, we apply them to the *Flood* data from the *extRemes* package in *R*. This data presents United States total economic damage (in billions of U.S. dollars) caused by floods by hydrologic year from 1932–1997.

Table 5. Quantile estimates based on real data

p	0.05	0.25	0.5	0.75	0.95
E	0.2816	0.6862	1.4177	3.3917	8.0099
EM	0.2816	0.6862	1.4177	3.3917	8.0099
H	0.2557	0.6862	1.3956	3.3917	8.0794
HF	0.2362	0.6843	1.3956	3.4009	8.1316
WG	0.1974	0.6806	1.3956	3.4192	8.2359
HB	0.2816	0.6862	1.4177	3.3917	8.0099
K	−0.106	0.6505	1.4159	3.4539	7.9734
Z	0.2813	0.6862	1.417	3.3917	8.0102
JP	0.243	0.6846	1.3956	3.4085	8.0158
M	0.2538	0.6836	1.3956	3.4045	8.0845

Table 5 presents the quantile estimates considered in this article.

The estimates of quantiles of level 0.25, 0.5, 0.75 do not differ very much. The value q_{WG} of $\hat{x}^{WG}_{0.95,66}$ is the greatest and the value q_K of $\hat{x}^{K}_{0.05,66}$ is the smallest and much smaller than others.

6. Conclusions and some prospects. In this article certain estimators $\hat{x}^{Z}_{p,n}$, $\hat{x}^{JP}_{p,n}$ and $\hat{x}^{M}_{p,n}$ of quantiles in the nonparametric setting have been proposed. The estimator $\hat{x}^{Z}_{p,n}$ is based on the distribution function estimator $\hat{F}^{Z}_{n}$ proposed by Zieliński [27], and $\hat{x}^{JP}_{p,n}$ is based on the distribution function estimator $\hat{F}^{JP}_{n}$ considered in [12] in the context of ROC curve estimation. The estimators $\hat{F}^{Z}_{n}$, $\hat{F}^{JP}_{n}$ do not satisfy all conditions PF1–PF7, and therefore the estimators $\hat{x}^{Z}_{p,n}$, $\hat{x}^{JP}_{p,n}$ do not have the desired properties P1–P7. We have proposed a nonparametric distribution function estimator which satisfies all conditions PF1–PF7, and provides a quantile estimator $\hat{x}^{M}_{p,n}$ enjoying all properties P1–P7. Under some assumptions, the estimators proposed are asymptotically equivalent to the empirical quantile estimator and also to the quantile estimators recommended in the literature. In simulations the accuracy of the proposed estimators and seven others have been investigated when the small sizes samples were generated from the Student and Pareto distributions. Although no estimator is uniformly best for all the parameter values and all quantile levels, the proposed estimators perform well compared to existing estimators.

References

[1] R. Alemany, C. Bolancé and M. Guillén, *A nonparametric approach to calculating value-at-risk*, Insurance Math. Economics 52 (2013), 255–262.

[2] A. Azzalini, *A note on the estimation of a distribution function and quantiles by a kernel method*, Biometrika 68 (1981), 326–328.

[3] C. Cheng, *The Bernstein polynomial estimator of a smooth quantile function*, Statist. Probab. Lett. 24 (1995), 321–330.

[4] C. Cheng and E. Parzen, *Unified estimators of smooth quantile and quantile density functions*, J. Statist. Planning Inference 59 (1997), 291–307.

[5] Z. Ciesielski, *Nonparametric polynomial density estimation*, Probab. Math. Statist. 9 (1988), 1–10.

[6] E. J. Gumbel, *La probabilité des hypothèses*, C. R. Acad. Sci. Paris 209 (1939), 645–647.

[7] A. Hazen, *Storage to be provided in impounding reservoirs for municipal water supply (with discussion)*, Trans. Amer. Soc. Civil Engineers 77 (1914), 1539–1669.

[8] W. Hoeffding, *Probability inequalities for sums of bounded random variables*, J. Amer. Statist. Assoc. 58 (1963), 13–30.

[9] M. L. Huang and P. H. Brill, *A level crossing quantile estimation method*, Statist. Probab. Lett. 45 (1999), 111–119.

[10] M. L. Huang and P. H. Brill, *A distribution estimation method based on level crossings*, J. Statist. Planning Inference 124 (2004), 45–62.

[11] R. J. Hyndman and Y. Fan, *Sample quantiles in statistical packages*, Amer. Statistician 50 (1996), 361–365.
[12] A. Jokiel-Rokita and M. Pulit, *Nonparametric estimation of the ROC curve based on smoothed empirical distribution function*, Statist. Computing 23 (2013), 703–712.
[13] E. Langford, *Quartiles in elementary statistics*, J. Statist. Education 14 (2006), no. 3, 20 pp.
[14] E. B. L. Mackay, P. G. Challenor and A. S. Bahaj, *A comparison of estimators for the generalised Pareto distribution*, Ocean Engineering 38 (2011), 1338–1346.
[15] L. Makkonen and M. Pajari, *Defining sample quantiles by the true rank probability*, J. Probab. Statist. 2014, art. 326579, 6 pp.
[16] E. A. Nadaraya, *Some new estimates for distribution function*, Theory Probab. Appl. 9 (1964), 497–500.
[17] S. S. Ralescu and S. Sun, *Necessary and sufficient conditions for the asymptotic normality of perturbed sample quantiles*, J. Statist. Planning Inference 35 (1993), 55–64.
[18] R. D. Reiss, *Approximate Distributions of Order Statistics with Applications to Nonparametric Statistics*, Springer, 1989.
[19] R. J. Serfling, *Approximation Theorems of Mathematical Statistics*, Wiley, New York, 1980.
[20] J. Shao, *Mathematical Statistics*, Springer, New York, 2003.
[21] W. Weibull, *The Phenomenon of Rupture in Solids*, Ingeniörs Vetenskaps Akademien Handlingar 153, Generalstabens Litografiska Anstalts Förlag, Stockholm, 1939.
[22] H. Yamato, *Uniform convergence of an estimator of a distribution function*, Bull. Math. Statist. 15 (1973), 69–78.
[23] D. Zelterman, *Smooth nonparametric estimation of the quantile function*, J. Statist. Planning Inference 26 (1990), 339–352.
[24] R. Zieliński, *Estimating median and other quantiles in nonparametric models*, Appl. Math. (Warsaw) 23 (1995), 363–370.
[25] R. Zieliński, *Best equivariant nonparametric estimator of quantile*, Statist. Probab. Lett. 45 (1999), 79–84.
[26] R. Zieliński, *Small-sample quantile estimators in a large nonparametric model*, Comm. Statist. Theory Methods 35 (2006), 1223–1241.
[27] R. Zieliński, *Kernel estimators and the Dvoretzky–Kiefer–Wolfowitz inequality*, Appl. Math. (Warsaw) 34 (2007), 401–404.
[28] R. Zieliński, *Optimal estimation of high quantiles in a large nonparametric model*, Appl. Math. (Warsaw) 39 (2012), 137–142.

Alicja Jokiel-Rokita
Faculty of Pure and Applied Mathematics
Wrocław University of Science and Technology
Wybrzeże Wyspiańskiego 27
50-370 Wrocław, Poland
E-mail: alicja.jokiel-rokita@pwr.edu.pl

Agnieszka Siedlaczek
Institute of Mathematics
and Computer Science
University of Opole
Pl. Kopernika 11a
45-040 Opole, Poland
E-mail: asiedlaczek@uni.opole.pl