DOI: 10.17516/1997-1397-2022-15-4-510-522
УДК 519.2

Estimating the Inverse Distribution Function at the Boundary

Nassima Almi*
Abdallah Sayah†
Laboratory of Applied Mathematics
Mohamed Khider University
Biskra, Algeria

Received 31.01.2022, received in revised form 21.03.2022, accepted 06.05.2022

Abstract. Most of existing quantile estimators have problems of inefficiency in extreme quantiles. To solve this problem, in this paper we suggested an alternative estimator and provided its asymptotic behaviour when quantile near the boundary value. A simulation studies and two real data applications were included to demonstrate the efficiency and reliability of our theoretical results.

Keywords: Kernel quantile estimation, Mean Square Error, Optimal Bandwidth, Boundary Quantiles.

Citation: N. Almi, A. Sayah, Estimating the Inverse Distribution Function at the Boundary, J. Sib. Fed. Univ. Math. Phys., 2022, 15(4), 510–522. DOI: 10.17516/1997-1397-2022-15-4-510-522.

1. Introduction and preliminaries

The estimation of population quantiles is of great interest when a parametric form for the underlying distribution is not available. It plays an important role in both statistical and probabilistic applications, namely: the goodness-of-fit, the computation of extreme quantiles and Value-at-Risk in insurance which are important measures of random performance, business and financial risk management, in reliability and medical studies, quantiles are adopted for characterize the survival distribution. Also, a large class of actuarial risk measures can be defined as functionals of quantiles (see, e.g. [1]).

Let $X_1, ..., X_n$ be independent and identically distributed with an unknown density $f(.)$ and absolutely continuous distribution function $F(.)$, while $X_{(1)} \leqslant ... \leqslant X_{(n)}$ denote the corresponding order statistics. The quantile function $Q(.)$ is defined to be the left-continuous inverse of $F(.)$ as follows:

$$Q(p) = \inf\{x : F(x) \geqslant p\} = F^{-1}(p), \ 0 < p < 1. \tag{1}$$

Indeed, to estimate a quantile function we need an estimator of the distribution function.

We recall two classical estimators. Traditionally, the estimator of the distribution function is the empirical function $F_n(.)$, which is defined as

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{]-\infty,x]}(X_i),$$

where the indicator function $\mathbb{1}_{]-\infty,x]}(X_i) = 1$ if $X_i \leqslant x$ and 0 otherwise. Theoretical properties of $F_n(.)$ have been investigated by several authors, (see, e.g. [2,3] and [4]). It is well known that

*almi.nassima@gmail.com
†sayahabdel@yahoo.fr

$F_n(.)$ is less smoothing, this fact leads to the effort to obtain a smooth estimate. Rosenblatt [5], Parsen [6] and Nadaraya [7] introduced the kernel estimators of $f(.)$ and $F(.)$ at x by:

$$\tilde{f}_n(x) = \frac{1}{nh}\sum_{i=1}^{n} k\left(\frac{x - X_i}{h}\right),$$

and

$$\tilde{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right),$$

respectively, where $h = h_n$ is the smoothing parameter (or the bandwidth) since it controls the amount of smoothness in the estimator, and satisfy $h := h_n \to 0$ as $n \to \infty$, $k(.)$ is a kernel function which is a predetermined density function symmetric about 0 and the function K is defined from a kernel k as

$$K(x) = \int_{-\infty}^{x} k(t)\,dt.$$

When the support of the variable is bounded, the asymptotic properties are not satisfactory when the data is near the endpoints of the support, due to so-called boundary problem. To remove this boundary problem several methods have been proposed, (see, e.g. [8–13]). As a result, the corresponding estimators of the quantile function have been proposed and studied extensively, in references can be found in the books of Galambos [14] and David [15].

A basic estimator of $Q(.)$ is the empirical quantile or the sample quantiles which is given by

$$Q_n(p) = \inf\{x : F_n(x) \geqslant p\} = X_{([np])},$$

where [.] denotes the integer part. The corresponding estimator of the quantile function $Q = F^{-1}$ is then defined by

$$\tilde{Q}_n(p) = \inf\{x : \tilde{F}_n(x) \geqslant p\}, \quad 0 < p < 1. \tag{2}$$

Nadaraya [7] showed under some assumptions for k, f and h, $\tilde{Q}_n(p)$ has an asymptotic standard normal distribution. The almost sure consistency, was obtained by Yamato [2]. Ralescu and Sun [16] obtained the necessary and conditions for the asymptotic normality. Azzalini [17] obtains the asymptotic mean squared error of $\tilde{Q}_n(p)$:

$$AMSE\left(\tilde{Q}_n(p)\right) = \frac{h^4}{4}\left(\frac{Q^{(2)}(p)}{\left(Q^{(1)}(p)\right)^2}\right)^2 \mu_2^2 + \frac{p(1-p)}{n}\left(Q^{(1)}(p)\right)^2 - \frac{h}{n}Q^{(1)}(p)\,\psi(k), \tag{3}$$

where $\psi(k) = 2\int\limits_{-\infty}^{\infty} tk(t)K(t)\,dt$, $\mu_2 = \int\limits_{-\infty}^{\infty} t^2 k(t)\,dt$ and $Q^{(1)}$, $Q^{(2)}$ are the first, the second derivative of Q respectively. It can be seen that the optimal bandwidth for minimizing (3) has the form

$$\tilde{h}_{opt} = \left(\frac{\left(Q^{(1)}(p)\right)^5 \psi(k)}{n\left(Q^{(2)}(p)\right)^2 \mu_2^2}\right)^{1/3}.$$

Parzen [6] proposed a version of the kernel quantile estimator as below:

$$\hat{Q}_n(p) = \sum_{i=1}^{n}\left[\int_{\frac{i-1}{n}}^{\frac{i}{n}} \frac{1}{h}k\left(\frac{x-p}{h}\right)dx\right]X_{(i)}. \tag{4}$$

In practice, the following approximation to $\hat{Q}_n(p)$ is often used:

$$\hat{Q}_n^a(p) = \frac{1}{nh}\sum_{i=1}^{n} X_{(i)}k\left(\frac{\frac{i}{n} - p}{h}\right). \tag{5}$$

Under suitable conditions on F, Falk [4] proposed the following kernel type quantile estimator

$$\check{Q}_n(p) = \frac{1}{h}\int_0^1 Q_n(p)\, k\left(\frac{x-p}{h}\right) dx. \tag{6}$$

This kernel-type quantile estimate can then be approximated by $\hat{Q}_n(p)$. Yang [18] provided the asymptotic normality property and the mean squared consistency of $\hat{Q}_n(p)$ and proved that $\hat{Q}_n(p)$ and $\hat{Q}_n^a(p)$ are asymptotically equivalent in terms of mean square errors. Falk [4] showed that the asymptotic performance of $\hat{Q}_n(p)$ is better than that of the empirical sample quantile $Q_n(p)$ in terms of relative deficiency for appropriately chosen kernels and sufficiently smooth distribution functions. Building on Falk [4], Sheater et al [19] gave the asymptotic mean squared error of $\hat{Q}_n(p)$.

If the second derivative of Q is continuous in a neighborhood of p and if f is not symmetric or f is symmetric but $p \neq \frac{1}{2}$ then asymptotic mean squared error of $\hat{Q}_n(p)$ is

$$AMSE\left(\hat{Q}_n(p)\right) = \frac{p(1-p)}{n}\left(Q^{(1)}(p)\right)^2 + \frac{h^4}{4}\left(Q^{(2)}(p)\right)^2\mu_2^2 - \frac{h}{n}\left(Q^{(1)}(p)\right)^2\psi(k). \tag{7}$$

The optimal bandwidth for $AMSE\left(\hat{Q}_n(p)\right)$ is

$$\hat{h}_{opt} = \left(\frac{\left(Q^{(1)}(p)\right)^2\psi(k)}{n\left(Q^{(2)}(p)\right)^2\mu_2^2}\right)^{1/3}. \tag{8}$$

When F is symmetric and $p = 1/2$ then the asymptotic mean squared error of $\hat{Q}_n(p)$

$$AMSE\left(\hat{Q}_n(p)\right) = n^{-1}\left(Q^{(1)}(1/2)\right)^2\left[0.25 - 0.5h\psi(k) + (nh)^{-1}\rho(k)\right],$$

where $\rho(k) = \int\limits_{-\infty}^{\infty} k^2(x)\,dx$.

But all these estimators have a large bias when p is close to boundary. In order to correct the bias problems in the case of extreme quantiles, Harrell et al [20] and Park [21] suggest to use asymmetric kernel, namely the Beta-type kernel. In particular in the case of heavy tailed distributions and for the same aim, Charpentier et al [22] suggested several nonparametric quantile estimators based on the beta-kernel and applied them to transformed data. Sayah et al [23] propose a new approach based on the modified Champernowne distribution. The main objective of this paper is to propose a new estimator to improve the asymptotic problems of extreme quantiles.

The paper orginised as follows: In Section 2, we propose our estimator and drive its asymptotic properties. In Section 3, a simulation study was conducted where we compare the performance of our proposed estimator with both the empirical and the classical quantile estimators at specific values of p. In Section 4, we compare graphically the mentioned estimators by using two real data applications. The paper is finalized with brief conclusion.

2. Main results

According to the work of Almi et al [8], our estimator based on a self elimination between the Bias $\hat{Q}_n(p)$ of the estimator from it self

$$\bar{Q}_n(p) = \hat{Q}_n(p) - B\hat{ias}\left(\hat{Q}_n(p)\right), \tag{9}$$

then the explicit form of our estimator is given by

$$\bar{Q}(p) = \frac{1}{h}\sum_{i=1}^{n}\left(\int_{\frac{i-1}{n}}^{\frac{i}{n}}\left(k\left(\frac{x-p}{h}\right)-\frac{1}{2}\mu_2 k^{(2)}\left(\frac{x-p}{h}\right)\right)dx\right)X_{(i)}.$$

The following theorem shows that the Bias of $\bar{Q}_n(p)$ is of order $O(h^4)$, while that of $\hat{Q}_n(p)$ is $O(h^2)$, and it gives the expressions for the bias and the variance of the proposed estimator

Theorem 2.1. *Assume that Q has four bounded, continuous derivatives in a neighborhood of p and the kernel function k is a continuous bounded density, symmetric about zero, then if $0 < h \to 0, nh^4 \to \infty$, for all fixed $p \in]0,1[$, we have*

$$Bias\left(\bar{Q}_n(p)\right) = \frac{h^4}{24}Q^{(4)}(p)\left(\mu_4 - 6\mu_2^2\right) + o\left(h^4\right),$$

and

$$Var\left(\bar{Q}_n(p)\right) = \frac{p(1-p)}{n}\left(Q^{(1)}(p)\right)^2 - \frac{h}{n}\left(Q^{(1)}(p)\right)^2\Psi(k) + o\left(\frac{h}{n}\right) + o(1),$$

where

$$\Psi(k) = \psi(k) - \frac{1}{4}\mu_2^2\int_{-\infty}^{+\infty}\left(k^{(1)}(t)\right)^2 dt - \mu_2\left(\int_{-\infty}^{+\infty} tk(t)\left(\int_{-\infty}^{t} k^{(2)}(t)dt\right)dt + \int_{-\infty}^{+\infty} tk^{(2)}(t)\left(\int_{-\infty}^{t} k(t)dt\right)dt\right),$$

$$\mu_4 = \int_{-\infty}^{\infty} t^4 k(t)\,dt.$$

Proof. As a result, our proposed estimator is

$$\bar{Q}_n(p) = \hat{Q}_n(p) - \frac{1}{2}h^2\hat{Q}_n^{(2)}(p)\mu_2.$$

We can easily see that

$$E\left(\bar{Q}_n(p)\right) = E\left(\hat{Q}_n(p)\right) - \frac{1}{2}h^2\mu_2 E\left(\hat{Q}_n^{(2)}(p)\right),$$

by a Taylor expansion, we have

$$\begin{aligned} E\left(\hat{Q}_n(p)\right) &= \frac{1}{h}\int_0^1 Q(x)\,k\left(\frac{p-x}{h}\right)dx = \\ &= \int_{-\infty}^{\infty} k(t)\left(Q(p-ht)\right)dt = \\ &= Q(p) + \frac{1}{2}h^2 Q^{(2)}(p)\mu_2 + \frac{h^4}{24}Q^{(4)}(p)\mu_4 + o\left(h^4\right). \end{aligned}$$

Moreover

$$\begin{aligned} E\left(\hat{Q}_n^{(2)}(p)\right) &= \frac{1}{h^3}\int_0^1 Q(x)\,k^{(2)}\left(\frac{p-x}{h}\right)dx = \\ &= \frac{1}{h}\int_0^1 Q^{(2)}(x)\,k\left(\frac{p-x}{h}\right)dx = \\ &= \int_{-\infty}^{\infty} k(t)\,Q^{(2)}(p-ht)\,dt = \\ &= Q^{(2)}(p) + \frac{1}{2}h^2 Q^{(4)}(p)\mu_2 + o\left(h^2\right). \end{aligned}$$

Thus, we have

$$Bias\left(\bar{Q}_n(p)\right) = \frac{h^4}{24} Q^{(4)}(p)\left(\mu_4 - 6\mu_2^2\right) + o\left(h^4\right).$$

On the other hand

$$\begin{aligned} Var\left(\bar{Q}_n(p)\right) &= Var\left(\hat{Q}_n(p) - \frac{1}{2}h^2 \hat{Q}_n^{(2)}(p)\,\mu_2\right) = \\ &= Var\left(\hat{Q}_n(p)\right) + \frac{1}{4}h^4\mu_2^2 Var\left(\hat{Q}_n^{(2)}(p)\right) - h^2\mu_2 Cov\left(\hat{Q}_n(p), \hat{Q}_n^{(2)}(p)\right), \end{aligned}$$

the variance of $\hat{Q}_n(p)$ can be computed as

$$\begin{aligned} Var\left(\hat{Q}_n(p)\right) &= \frac{1}{n}\left(Q^{(1)}(p)\right)^2\left(-p^2 + 2\int_{-\infty}^{\infty}(p-ht)\,k(t)\,K(t)\,dt\right) + o\left(\frac{h}{n}\right) = \\ &= \frac{p(1-p)}{n}\left(Q^{(1)}(p)\right)^2 - \frac{h}{n}\left(Q^{(1)}(p)\right)^2\psi(k) + o\left(\frac{h}{n}\right) \end{aligned}$$

and

$$Var\left(\hat{Q}_n^{(2)}(p)\right) = \frac{1}{nh^3}\left(Q^{(1)}(p)\right)^2\int_{-\infty}^{\infty}\left(k^{(1)}(t)\right)^2 dt + o\left(\frac{1}{nh^3}\right) + o(1).$$

Now we will calculate the third term on the right hand side of $Var\left(\bar{Q}_n(p)\right)$. We have

$$\begin{aligned} Cov\left(\hat{Q}_n(p), \hat{Q}_n^{(2)}(p)\right) &= E\left(\frac{1}{h}\left(\int_0^1 Q_n(x)\,k\left(\frac{p-x}{h}\right)dx - \int_0^1 Q(x)\,k\left(\frac{p-x}{h}\right)dx\right)\times\right.\\ &\left.\times\frac{1}{h^3}\left(\int_0^1 Q_n(x)\,k^{(2)}\left(\frac{p-x}{h}\right)dx - \int_0^1 Q(x)\,k^{(2)}\left(\frac{p-x}{h}\right)dx\right)\right) = \\ &= \frac{1}{h^2}E\left(\left(\int_{-\infty}^{\infty} k(t)\left((p-ht) - \bar{F}_n(p-ht)\right)Q^{(1)}(p-ht)\,dt\right)\times\right.\\ &\left.\times\left(\int_{-\infty}^{\infty} k^{(2)}(t)\left((p-ht) - \bar{F}_n(p-ht)\right)Q^{(1)}(p-ht)\,dt\right)\right), \end{aligned}$$

where $\bar{F}_n$ is the empirical distribution function according to n independent, uniformly on $[0,1]$ distributed random variables. Furthermore

$$\begin{aligned} Cov\left(\hat{Q}_n(p), \hat{Q}_n^{(2)}(p)\right) &= \frac{1}{nh^2}\int_0^1\left(\int_{-\infty}^{\infty} k(t)\left((p-ht) - 1_{]0,p-ht[}(y)\right)Q^{(1)}(p-ht)\,dt\right)\times\\ &\times\left(\int_{-\infty}^{\infty} k^{(2)}(t)\left((p-ht) - 1_{]0,p-ht[}(y)\right)Q^{(1)}(p-ht)\,dt\right)dy, \end{aligned}$$

$$\begin{aligned} Cov\left(\hat{Q}_n(p), \hat{Q}_n^{(2)}(p)\right) &= \frac{1}{nh^2}\left(\int_0^1\left(\int_{-\infty}^{\infty} k(t)\left((p-ht) - 1_{]0,p-ht[}(y)Q^{(1)}(p)dt\right)\times\right.\right.\\ &\left.\times\left(\int_{-\infty}^{\infty} k^{(2)}(t)\left((p-ht) - 1_{]0,p-ht[}(y)\right)Q^{(1)}(p)dt\right)dy + o(h)\right) = \\ &= \frac{\left(Q^{(1)}(p)\right)^2}{nh^2}\int_0^1\left(\int_{\frac{p-1}{h}}^{\frac{p-y}{h}} k(t)\,dt\int_{\frac{p-1}{h}}^{\frac{p-y}{h}} k^{(2)}(t)\,dt\right)dy + o\left(\frac{1}{nh}\right) + o(1). \end{aligned}$$

By integration by part we find

$$
\begin{aligned}
Cov\left(\hat{Q}_n(p), \hat{Q}_n^{(2)}(p)\right) &= \frac{\left(Q^{(1)}(p)\right)^2}{nh^3} \int_0^1 y \left(k\left(\frac{p-y}{h}\right) \int_{\frac{p-1}{h}}^{\frac{p-y}{h}} k^{(2)}(t)\,dt + \right.\\
&\left. + k^{(2)}\left(\frac{p-y}{h}\right) \int_{\frac{p-1}{h}}^{\frac{p-y}{h}} k(t)\,dt\right) dy + o\left(\frac{1}{nh}\right) + o(1) = \\
&= \frac{\left(Q^{(1)}(p)\right)^2}{nh^2} \left(\int_{\frac{p-1}{h}}^{\frac{p}{h}} (p-ht)\,k(t)\left(\int_{\frac{p-1}{h}}^{t} k^{(2)}(t)\,dt\right) dt + \right.\\
&\left. + \int -\frac{p-1}{h}^{\frac{p}{h}} (p-ht)\,k^{(2)}(t)\left(\int_{\frac{p-1}{h}}^{t} k(t)\,dt\right) dt\right) + o\left(\frac{1}{nh}\right) + o(1) = \\
&= \frac{-1}{nh}\left(Q^{(1)}(p)\right)^2 \left(\int_{-\infty}^{\infty} tk(t)\left(\int_{-\infty}^{t} k^{(2)}(t)\,dt\right) dt + \right.\\
&\left. + \int_{-\infty}^{\infty} tk^{(2)}(t)\left(\int_{-\infty}^{t} k(t)\,dt\right) dt\right) + o\left(\frac{1}{nh}\right) + o(1).
\end{aligned}
$$

By adding up all these terms we have the desired result for the variance of $\bar{Q}_n$. □

Corollary 2.1. *Suppose that the conditions of previous theorem2.1 hold. The asymptotic mean squared error of $\bar{Q}_n(p)$ is given by*

$$
AMSE\left(\bar{Q}_n(p)\right) = \left(\frac{h^4}{24} Q^{(4)}(p)\left(\mu_4 - 6\mu_2^2\right)\right)^2 + \frac{p(1-p)}{n}\left(Q'(p)\right)^2 - \frac{h}{n}\left(Q'(p)\right)^2 \Psi(k).
$$

It can be seen that the optimal bandwidth for minimizing $AMSE\left(\bar{Q}_n(p)\right)$ is both of order $O\left(n^{-1/7}\right)$ and has the form

$$
\bar{h}_{opt} = \left(\frac{72\left(Q'(p)\right)^2 \Psi(k)}{n\left(Q^{(4)}(p)\left(\mu_4 - 6\mu_2^2\right)\right)^2}\right)^{1/7},
$$

and the associated asymptotic mean squared error is given by

$$
AMSE_{opt}\left(\bar{Q}_n(p)\right) = \frac{p(1-p)}{n}\left(Q^{(1)}(p)\right)^2 - 7\left(\frac{\left(\frac{1}{8}\left(Q^{(1)}(p)\right)^2 \Psi(k)\right)^8}{\left(\left(\frac{1}{24} Q^{(4)}(p)\left(\mu_4 - 6\mu_2^2\right)\right)^2\right)}\right)^{1/7} n^{-8/7}.
$$

3. Simulation study

In this section, we report results of a Monte Carlo study which was conducted to compare the performance of our proposed estimator $\bar{Q}_n(p)$ with the classical $\hat{Q}_n(p)$ and the empirical quantile estimators $\tilde{Q}_n(p)$, by computing the Bias and Mse for specific values of p where $p \in \{0.025, 0.05, 0.10, 0.20, 0.40, 0.60, 0.80, 0.90, 0.95, 0.975\}$. It is well known that bandwidth plays a critical role in the kernel estimation, for this reason we use the optimal bandwidth for $AMSE\left(\hat{Q}_n(p)\right)$ on each p-values, by using triweight kernel $\frac{35}{32}\left(1-t^2\right)^3 \mathbb{1}_{|t|\leqslant 1}$. In order to account for different cases, we generate a thousand samples of two sizes $n = 50$ and $n = 200$ from different distributions listed in the Tab. 1, results of the comparison are shown in Tabs. 2 to 11.

Table 1. Distributions used in the simulation study

Distribution	Theorical quantille $Q(p)$
Weibull $\left(\frac{3}{2}, 1\right)$	$(-\log(1-p))^{\frac{2}{3}}$
Log-normal $(0, \frac{1}{2})$	$\exp\left(\frac{1}{2}\phi^{-1}(p)\right)$
Chi-square (1)	$\left(\phi^{-1}\left(\frac{p+1}{2}\right)\right)^2$
Log-logistic $(1, 3)$	$\frac{1}{3}\left(\frac{p}{1-p}\right)$
Pareto $(3, 1)$	$\left(\frac{1}{1-p}\right)^{\frac{1}{2}} - 1$

where ϕ^{-1} denote the Inverse of standard normal distribution.

Table 2. Bias(MSE) values for Weibull distribution, n=50, Results are re-scaled by the factor 0.0001.

p	$\tilde{Q}$	$\hat{Q}$	$\bar{Q}$
0.025	921.8(851.8)	67.49(8.130)	33.35(4.673)
0.050	872.0(760.4)	29.16(8.812)	32.21(10.97)
0.100	787.2(619.4)	17.64(15.67)	16.25(14.80)
0.200	642.2(412.4)	43.34(34.93)	26.63(24.63)
0.400	371.0(137.7)	22.15(20.78)	15.45(12.27)
0.600	297.0(267.9)	19.31(12.89)	9.298(3.918)
0.800	264.6(218.4)	19.11(6.300)	5.810(5.128)
0.900	307.7(125.4)	27.62(11.86)	12.41(11.31)
0.950	547.1(544.6)	96.32(47.93)	33.57(20.53)
0.975	261.9(159.2)	105.3(102.7)	98.05(65.27)

Table 3. Bias(MSE) values for Weibull distribution, n=200, Results are re-scaled by the factor 0.0001.

p	$\tilde{Q}$	$\hat{Q}$	$\bar{Q}$
0.025	91.84(37.23)	15.85(6.271)	8.210(3.764)
0.050	86.69(75.16)	18.75(6.188)	17.36(5.389)
0.100	78.19(61.14)	6.156(8.972)	4.025(7.439)
0.200	63.71(40.59)	5.627(8.628)	1.929(6.379)
0.400	36.60(33.39)	1.103(19.49)	0.137(17.56)
0.600	21.03(34.17)	0.955(31.91)	0.128(30.42)
0.800	63.66(405.3)	1.263(59.20)	0.134(57.26)
0.900	27.23(80.20)	2.177(10.92)	0.165(10.80)
0.950	62.74(60.91)	3.333(20.22)	0.318(20.04)
0.975	147.9(102.8)	87.32(78.69)	29.48(34.66)

After examining all tables, the classical estimator $\hat{Q}_n$ does not perform as well at near boundary points $p = 0.025$ to 0.10 and $p = 0.90$ to 0.975 as at interior points from $p = 0.20$ to $p = 0.80$. However, it can be observed that our proposed estimator $\bar{Q}_n$ produces lower Bias(MSE) for almost values of p specifically extreme values in all distributions considered, except for Weibull distribution in the case where $p = 0.05$ the performance of the classical estimator is better than our estimator for the small size. Both estimators are more efficient than the empirical quantile estimator $\tilde{Q}_n$ in most situations.

Table 4. Bias (MSE) values for Log-normal distribution, n=50, Results are re-scaled by the factor 0.0001.

p	$\tilde{Q}$	$\hat{Q}$	$\bar{Q}$
0.025	692.5(874.8)	251.9(222.9)	224.6(183.1)
0.050	656.6(849.7)	167.9(205.2)	120.5(56.68)
0.100	478.1(496.2)	152.4(171.9)	98.47(22.09)
0.200	368.6(257.6)	148.6(249.4)	159.5(253.2)
0.400	124.0(254.2)	88.39(83.43)	75.51(68.43)
0.600	98.36(96.20)	48.36(47.50)	38.51(32.16)
0.800	99.57(95.47)	79.02(73.13)	65.15(60.21)
0.900	354.8(135.2)	281.6(156.2)	102.9(98.25)
0.950	97.22(48.21)	47.00(45.21)	31.42(27.52)
0.975	106.8(231.2)	94.27(194.2)	56.22(152.7)

Table 5. Bias(MSE) values for Log-normal distribution, n=200, Results are re-scaled by the factor 0.0001.

p	$\tilde{Q}$	$\hat{Q}$	$\bar{Q}$
0.025	525.8(973.5)	209.2(214.4)	146.1(143.0)
0.050	365.6(511.6)	119.1(147.2)	101.3(137.7)
0.100	278.1(437.0)	109.9(204.2)	83.91(129.3)
0.200	368.6(188.5)	27.7(28.46)	5.168(19.84)
0.400	124.0(254.2)	88.39(83.48)	68.43(75.71)
0.600	87.50(255.1)	8.891(31.66)	4.475(14.99)
0.800	486.8(410.2)	12.91(9.126)	9.848(7.638)
0.900	264.8(301.4)	18.98(35.20)	7.667(18.07)
0.950	686.7(261.6)	305.2(184.1)	76.70(156.6)
0.975	437.0(545.1)	568.0(315.5)	387.9(231.7)

Table 6. Bias (MSE) values for Chi-square distribution, n=50, Results are re-scaled by the factor 0.0001.

p	$\tilde{Q}$	$\hat{Q}$	$\bar{Q}$
0.025	57.08(41.67)	28.72 (25.07)	24.32 (20.59)
0.050	20.10(18.72)	1 3.46 (12.74)	8. 029 (6.447)
0.100	44.23(38.56)	26.27(24.28)	14.082(10.35)
0.200	94.08(97.85)	56.06(53.25)	19.084 (17.83)
0.400	88.94(84.51)	37.12(32.35)	32.35(20.47)
0.600	80.89(72.16)	29.18 (27.32)	11.00(10.31)
0.800	85.51(75.92)	30.83(27.46)	14.68(12.10)
0.900	96.79(95.75)	41.97(37.65)	7.484(5.908)
0.950	195.6(193.8)	185.6(176.1)	158.6(155.7)
0.975	196.4(195.5)	185.4(134.4)	153.0(144.2)

Table 7. Bias (MSE) values for Chi-square distribution, n=200, Results are re-scaled by the factor 0.0001.

p	$\tilde{Q}$	$\hat{Q}$	$\bar{Q}$
0.025	257.0(246.7)	150.2(122.4)	144.7(132.4)
0.050	18.07(15.84)	5.725(3.278)	0.487(0.222)
0.100	28.92(19.78)	16.43(12.70)	6.843(4.683)
0.200	9.408(8.851)	4.544(2.065)	2.201 (2.035)
0.400	13.29(13.23)	8.518(7.256)	6.235(5.524)
0.600	29.66(28.80)	9.882(9.504)	1.182(1.087)
0.800	40.78(36.63)	16.47(12.71)	4.830(2.333)
0.900	45.02(44.27)	20.38(19.23)	4.845(2.201)
0.950	46.00(43.11)	21.94(14.61)	3.487(2.516)
0.975	136.30(131.80)	115.31(112.40)	101.3(98.45)

Table 8. Bias (MSE) values for Log-logistic distribution, n=50, Results are re-scaled by the factor 0.0001.

p	$\tilde{Q}$	$\hat{Q}$	$\bar{Q}$
0.025	849.7(7.222)	229.6(5.246)	83.33(0.694)
0.050	987.4(9.751)	205.2(4.214)	56.68(0.321)
0.100	967.9(9.370)	171.9(2.95)	22.09(0.048)
0.200	188.8(35.66)	162.1(2.624)	19.84(0.039)
0.400	490.2(43.03)	284.6(8.102)	75.90(0.576)
0.600	487.4(57.24)	384.1(46.80)	221.6(4.914)
0.800	479.1(105.0)	279.3(77.91)	80.53(64.85)
0.900	462.4(234.6)	444.3(171.6)	130.9(127.7)
0.950	429.1(489.2)	198.1(392.7)	152.7(233.2)
0.975	107.2(1214)	92.13(848.8)	99.90(458.7)

Table 9. Bias (MSE) values for Log-logistic distribution, n=200, Results are re-scaled by the factor 0.0001.

p	$\tilde{Q}$	$\hat{Q}$	$\bar{Q}$
0.025	85.547(0.7305)	110,20(1.214)	38.083(0.145
0.050	334.21(111.69)	84.898(0.720)	14.621(0.021)
0.100	967,98(937.00)	55.155(0.304)	4.789(0.002)
0.200	921.69(849.51)	49.941(0.494)	8.084(0.006)
0.400	782.80(612.78)	80.428(0.646)	11.409(0.013
0.600	505.02(255.02)	182,61(3.334)	49.413(0.244
0.800	640.53(410.28)	569.59(32.44)	122.44(1.499
0.900	545.72(297.87)	404.6(99.81)	276.37(7.638)
0.950	3549.0(301.4)	2411.4(304,7)	990.34(98.07)
0.975	511.55(626.85)	409.28(475,1)	109.844(120.6)

4. Application

In this section, we compare the performance of our proposed estimator with the empirical and the classical estimators by using the graphical representation of two real data sets. The first data set consist of 100 observations of breaking stress of carbon fibres (in Gba) given by Nichols and Padget [24] and the second data set consist of 63 observations relates to the strength

Table 10. Bias (MSE) values for Pareto distribution, n=50, Results are re-scaled by the factor 0.0001.

p	$\tilde{Q}$	$\hat{Q}$	$\bar{Q}$
0.025	946.5(975.1)	257.7(421.4)	150.7(52.13
0.050	950.9(937.1)	261.0(654.2)	79.53(258.7)
0.100	45.83(583.7)	44.56(331.5)	39.39(250. 02)
0.200	48.95(96.10)	26.61(85.10)	12.73(83.8)
0.400	48.66(93.20)	39.63(93.00)	29.14(85.04)
0.600	48.01(158.2)	9.677(51.20)	4.853 (23. 55)
0.800	78.00(278.4)	26.99(258.4)	2.474(20.01)
0.900	457.7(673.5)	32.46(140.6)	19.51(76.42)
0.950	2439(1954)	308.8(531.1)	278.4(445.9
0.975	3135(3298)	848.8(1104)	87.25(300.1)

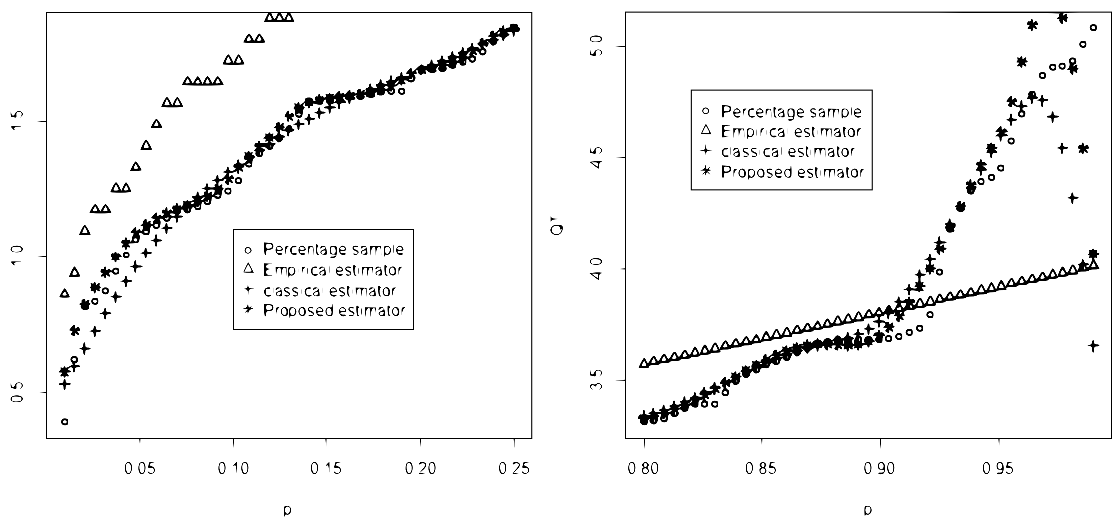
Table 11. Bias (MSE) values for Pareto distribution, n=200, Results are re-scaled by the factor 0.0001.

p	$\tilde{Q}$	$\hat{Q}$	$\bar{Q}$
0.025	503.7(763.5)	232.2(404.5)	105.0(43.85)
0.050	201.4(528.6)	159.3(401.7)	34.69(36.87)
0.100	9.509(280.6)	9.489(400.1)	9.109(36.76)
0.200	8.869(86.30)	1.693(39.96)	0.163(35.75)
0.400	10.71(58.61)	3.885(39.99)	2.917(36.72)
0.600	4.238(58.61)	1.421(40.11)	1.302(36.65)
0.800	72.47(256.0)	34.20(41.08)	5.521(36.73
0.900	472.3(301.7)	272.6(145.8)	96.73(101.2)
0.950	2037(1582)	1051(780.0)	803.2(386.6)
0.975	671.9(330.0	347.8(459.5)	187.5(221.4)

of carbon fibers tested under tension at gauge lengths of 10 mm, The data has been recently reported and analyzed by Bi and Gui [25] among others, the choice of the bandwidth bases to coss validation method. The results are shown in Figs. 1 and 2 respectively. It's remarkably clear that our newly proposed estimator is closer to the unknown quantile function as compared to both estimators the classical and the empirical kernel estimators, this yeild that our estimator improve the performance of the classical estimator in extreme quantiles.

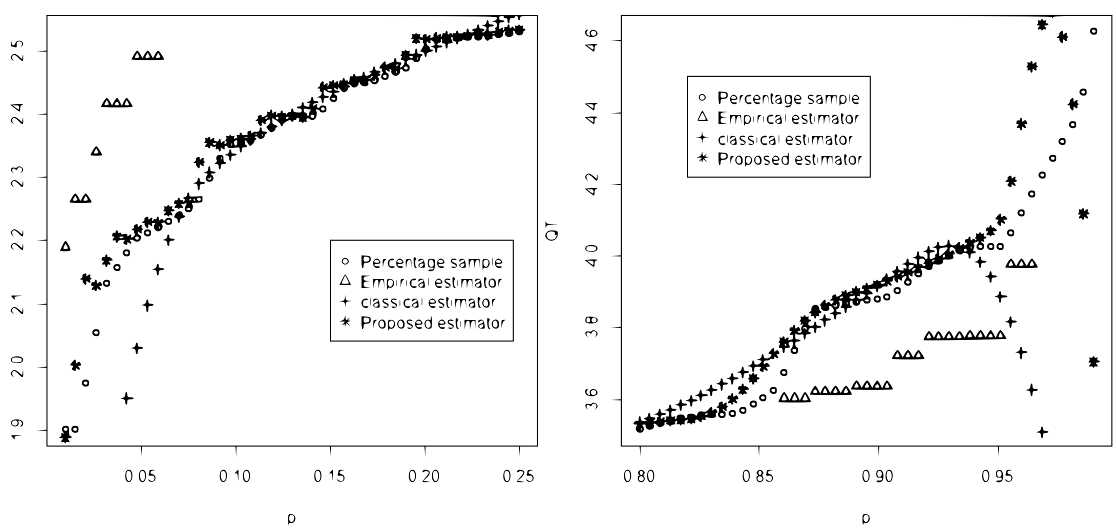
Conclusion

This paper proposes a smooth estimator of the quantile function to improve the efficiency of the classical kernel estimator at extreme quantiles. Depending on the theoretical it turned out that the Bias of our proposed estimator is of fourth-order power of the bandwidth, while that of the classical is second order. The numerical results are summarized in Tabs. 2 to 11 and the Figs. 1 and 2 conducted that our proposed estimator is better than both the classical and the empirical quantile estimators in the meaning of Bias and Mse for almost all p-values and specifically at extremes. These numerical results coincide with the theoretical results in Theorem 2.1.



(a) Left extreme quantile (b) Right extreme quantile

Fig. 1. Performance of considering estimators at extreme quantiles of breaking stress of carbon fibres data set



(a) Left extreme quantile (b) Right extreme quantile

Fig. 2. Performance of considering estimators at extreme quantiles relates to the strength of carbon fibers tested under tension at gauge lengths

References

[1] M.Denuit, J. Dhaene , Actuarial Theory for Dependent Risk: Measures, Orders and Models, Wiley, New York, 2005.

[2] H.Yamato, Uniform convergence of an estimator of a distribution function, *Bull.Math.Statist*, **15**(1973), 69–78.

[3] R.D.Reiss, Nonparametric estimation of smooth distribution functions, *Scand. J.Statist*, **8**(1981), 116–119.

[4] M.Falk, Relative efficiency and deficiency of kernel type estimators of smooth distribution functions, *Statist. Neerlandica*, **37**(1983), 73–83.

[5] M.Rosenblatt, Remarks on Some Nonparametric Estimates of a Density Function, *The Annals of Mathematical Statistics*, **27**(1956), 832–837.

[6] E.Parzen, Nonparametric Statistical Data Modelling, *Journal of the American Statistical Association*, **74**(1979), 105–131.

[7] E.A.Nadaraya, Some new estimates for distribution function,*Theory of Probab*, **9**(1964), 497–500.

[8] N.Almi, A.Sayah, nonparametric kernel distribution function estimation near endpoints, *Advances in Mathematics: Scientific Journal*, **10**(2021), 3679–3697.
DOI: 10.37418/amsj.10.12.10

[9] M.Tour, A.Sayah, Y.Djebrane, A Modified Champernowne Transformation to Improve Boundary Effect in Kernel Distribution Estimation, *Afrika Statistika*, **12**(2017), 1219–1233.
DOI: 10.16929/as/2017.1219.101

[10] C.Tenreiro, Boundary Kernels for Distribution Function Estimation, *REVSTAT Statistical Journal*, **11**(2013), 169–190.

[11] C.Tenreiro, A note on boundary kernels for distribution function estimation, arXiv preprint arXiv:1501.04206, 2015.

[12] J.Kolácek, R.J.Karunamuni, On boundary correction in kernel estimation of ROC curves, *Austrian Journal of Statistics*, **38**(2009), 17–32. DOI: 10.17713/ajs.v38i1.257

[13] S.Zhang, L.Zhong, Z.Zhang, Estimating a Distribution Function at the Boundary, *Austrian Journal of Statistics*, **49**(2020), 1–23. DOI: 10.17713/ajs.v49i1.801

[14] J.Galambos, The asymptotic theory of extreme order statistics, *Krieger, Malabar, Florida*, 1978.

[15] H.A.David, Order Statistics, 2nd Edition, New York, John Wiley, 1981.

[16] S.S.Ralescu, S.Sun, Necessary and sufficient conditions for the asymptotic normality of perturbed sample quantiles, *J. Statist. Plann. Inference*, **35**(1993), 55–64.

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

[18] S.S.Yang, A Smooth Nonparametric Estimator of a Quantile Function, *Journal of the American Statistical Association*, **80**(1985), 1004–1011.

[19] S.J.Sheater, J.S.Marron, Kernel quatile estimtors, *Journal of the American Statistical Associatio*, **85**(1990), 410–416.

[20] F.E.Harrell, C.E.Davis, A New Distribution-Free Quantile Estimator, *Biometrika*, **69**(1982), 635–640.

[21] C.Park, Smooth nonparametric estimation of a quantile function under right censoring using beta kernels, Technical Report (TR 2006-01-CP), Department of Mathematical Sciences, Clemson University, 2006.

[22] A.Charpentier, A.Oulidi, Beta kernel quantile estimators of heavy-tailed loss distributions, *Stat. Comput,* **20**(2010), 35–55. DOI: 10.1007/s11222-009-9114-2

[23] A.Sayah, Y.Djebrane, A Necir, Champernowne transformation in kernel quantile estimation for heavy-tailed distributions, *Afrika Statistika*, **5**(2010), 288–296.

[24] M.D.Nichols, WJ.Padgett, A bootstrap control chart for Weibull percentiles, *Qual Reliab Eng Int*, **22**(2006), 141–151. DOI: 10.1002/qre.691

[25] Q.Bi, W.Gui, Bayesian and classical estimation of stress-strength reliability for inverse Weibull lifetime models,*Algorithms*, **10**(2017), 71. DOI: 10.3390/a10020071

Оценка обратной функции распределения на границе

Нассима Алми
Абдалла Сайах
Лаборатория прикладной математики
Университет Мохамеда Хидера
Бискра, Алжир

Аннотация. Большинство существующих квантильных оценок имеют проблемы неэффективности в экстремальных квантилях. Чтобы решить эту проблему, в этой статье мы предложили альтернативную оценку и представили ее асимптотическое поведение, когда квантиль близок к граничному значению. Для демонстрации эффективности и надежности наших теоретических результатов были включены симуляционные исследования и два приложения с реальными данными.

Ключевые слова: оценка квантилей ядра, среднеквадратическая ошибка, оптимальная пропускная способность, граничные квантили.