ELSEVIER

Statistics & Probability Letters 52 (2001) 207–213

STATISTICS & PROBABILITY LETTERS

www.elsevier.nl/locate/stapro

Sharp distribution-free bounds on the bias in estimating quantiles via order statistics

Andrzej Okolewski[a], Tomasz Rychlik[b, *]

[a]Institute of Mathematics, Technical University of Łódź, Żwirki 36, 90924 Łódź, Poland
[b]Institute of Mathematics, Polish Academy of Sciences, Chopina 12, 87100 Toruń, Poland

Received August 2000; received in revised form November 2000

Abstract

Sharp distribution-free lower and upper bounds on the bias in estimating quantiles by the sample counterparts are obtained by the use of Moriguti's greatest convex minorant approach.

MSC: 62G30

Keywords: Quantile; Sample quantile; Order statistic; Bias; Bernstein polynomial; The greatest convex minorant; Variation diminishing property

1. Results

Suppose that $X, X_1, \ldots, X_n$ are i.i.d. random variables with F, F^{-1}, μ and $\sigma < \infty$ denoting common distribution function, (upper) quantile function, mean and standard deviation, respectively. Let $X_{j:n}$ stand for the jth-order statistic. The aim of this paper is to present the lower and upper bounds on the bias in estimating a quantile $F^{-1}(p)$ by a sample quantile $X_{j:n}$, such that j/n is close to p. Although some more sophisticated smooth versions of quantile estimates based on kernels and Bernstein polynomials have been proposed (see, e.g. Sheather and Marron, 1990; Cheng, 1995, respectively), the sample quantiles are the most natural estimates justified by the strong consistency property under the mild conditions that $j/n \to p$ and $F^{-1}(p)$ is unique. Optimality of order statistics in quantile estimation for finite samples under various criteria was proved in Zieliński (1988, 1999, 2001). Here we also exploit a nonasymptotic approach. To fix the ideas, we assume that $j/n = p$ precisely. Applying Moriguti's (1953) concept of the greatest convex minorant and the Schwarz inequality we prove that for the majority of j and n, it holds

$$-\frac{F_{j:n}(p)}{\sqrt{p(1-p)}} \leqslant \frac{E_F X_{j:n} - F^{-1}(p)}{\sigma} \leqslant \frac{1-F_{j:n}(p)}{\sqrt{p(1-p)}}, \tag{1}$$

* Corresponding author.
E-mail addresses: oko@ck-sg.p.lodz.pl (A. Okolewski), trychlik@impan.gov.pl (T. Rychlik).

0167-7152/01/$ - see front matter
PII: S0167-7152(00)00242-X

where $F_{j:n}(x)$ is the distribution function of the jth-order statistic from the standard uniform sample with the density function $f_{j:n}(x)=nB_{j-1,n-1}(x),\ x\in(0,1)$, and

$$B_{r,m}(x)=\binom{m}{r}x^r(1-x)^{m-r},\quad 0\leqslant r\leqslant m<\infty,$$

are the classic Bernstein polynomials.

Proposition 1. (i) *If* $1\leqslant j<n<\infty$, *then the upper bound in* (1) *holds.*

(ii) *If either* $2=j<n\leqslant 6$ *or* $3\leqslant j<n$, *then the lower inequality in* (1) *is satisfied.*

Both the bounds (1) *are attained for the two point distribution*

$$P(X=\mu-\sigma\sqrt{(1-p)/p})=p=1-P(X=\mu+\sigma\sqrt{p/(1-p)}).\tag{2}$$

The lower bounds for the cases of j and n not considered in Proposition 1 are provided in Propositions 2 and 3.

Proposition 2. *For* $j=1<n<\infty$, *we have*

$$\frac{E_FX_{1:n}-F^{-1}(1/n)}{\sigma}\geqslant -A(n),\tag{3}$$

where

$$A^2(n)=\frac{n^2}{2n-1}F_{1:2n-1}\left(\frac{1}{n}\right)+\frac{n}{n-1}F^2_{1:n}\left(\frac{1}{n}\right).$$

The equality is attained if

$$F(x)=\begin{cases}0 & \text{for } \dfrac{x-\mu}{\sigma}<-\dfrac{n}{A(n)},\\[2ex] 1-\left[A(n)\dfrac{\mu-x}{n\sigma}\right]^{1/(n-1)} & \text{for } -\dfrac{n}{A(n)}\leqslant\dfrac{x-\mu}{\sigma}<-\dfrac{(n-1)^{n-1}}{n^{n-2}A(n)},\\[2ex] \dfrac{1}{n} & \text{for } -\dfrac{(n-1)^{n-1}}{n^{n-2}A(n)}\leqslant\dfrac{x-\mu}{\sigma}<\dfrac{nF_{1:n}(1/n)}{(n-1)A(n)},\\[2ex] 1 & \text{for } \dfrac{x-\mu}{\sigma}\geqslant\dfrac{nF_{1:n}(1/n)}{(n-1)A(n)}.\end{cases}\tag{4}$$

Proposition 3. *For* $j=2$ *and* $n\geqslant 7$, *we have*

$$\frac{E_FX_{2:n}-F^{-1}(\frac{2}{n})}{\sigma}\geqslant -B(n),\tag{5}$$

where

$$B^2(n)=x_nf^2_{2:n}(x_n)+\frac{n}{n-2}F^2_{2:n}\left(\frac{2}{n}\right)+\frac{(n-1)n^2}{(2n-1)(2n-3)}\left[F_{3:2n-1}\left(\frac{2}{n}\right)-F_{3:2n-1}(x_n)\right]$$

with $x_n\in(1/(n-1),2/n)$ *being the unique solution to the following equation*:

$$[(n-1)^2x^2+(n-2)x+1](1-x)^{n-2}=1.\tag{6}$$

The equality is attained if

$$F(x)=\begin{cases} 0 & \textit{for } \dfrac{x-\mu}{\sigma}<\dfrac{-f_{2:n}(x_n)}{B(n)},\\ f_{2:n}^{-1}\left(B(n)\dfrac{\mu-x}{\sigma}\right) & \textit{for } -\dfrac{f_{2:n}(x_n)}{B(n)}\leqslant\dfrac{x-\mu}{\sigma}<-\dfrac{f_{2:n}(2/n)}{B(n)},\\ \dfrac{2}{n} & \textit{for } -\dfrac{f_{2:n}(2/n)}{B(n)}\leqslant\dfrac{x-\mu}{\sigma}<\dfrac{nF_{2:n}(2/n)}{(n-2)B(n)},\\ 1 & \textit{for } \dfrac{x-\mu}{\sigma}\geqslant\dfrac{nF_{2:n}(2/n)}{(n-2)B(n)}. \end{cases} \tag{7}$$

2. Remarks

1. Distribution functions (2), (4) and (7) have a jump at the point $F^{-1}(p)$ from p to 1 and they are constant on some left-hand neighborhood of $F^{-1}(p)$. Hence equalities in bounds (1), (3) and (5) are actually attained for the corresponding lower quantiles. However, we can slightly modify (2), (4) and (7) replacing the atom of weight $1-p$ by the same contribution of a uniform distribution on a small neighborhood of the atom. Shrinking the support of the uniform distribution, we attain the bounds in limit for unique quantiles and so the right ones as well.
2. The difference between the bounds in (1) is $1/\sqrt{p(1-p)}$. It follows that it is easier to estimate central quantiles than extreme ones. Surprisingly, the bias oscillation does not depend on the sample size. The only positive effect of taking large samples is that the bias becomes symmetric and so its absolute value decreases, but it does not vanish, though. We have

   $$\lim_{n\to\infty}\sup_F\left|\frac{E_F X_{j:n}-F^{-1}(j/n)}{\sigma}\right|=\frac{1}{2\sqrt{p(1-p)}}>0, \tag{8}$$

   where the supremum is taken over all distributions whose second moments are finite. By Remark 1, it also suffices to take the distributions with finite density at $F^{-1}(p)$. Formula (8) is a consequence of the de Moivre–Laplace theorem which implies $F_{j:n}(j/n)\to\frac{1}{2}$ as $n\to\infty$. We finally point out that the bias oscillation in the cases treated in Propositions 2 and 3 are greater than those of Proposition 1.
3. If F has a positive density $f(F^{-1}(p))$ at $F^{-1}(p)$, then

   $$\mathscr{L}(\sqrt{n}[X_{j:n}-F^{-1}(p)])\to\mathscr{N}\left(0,\frac{p(1-p)}{f^2(F^{-1}(p))}\right).$$

   By Remark 1 and (8), for such distributions one cannot expect that $|X_{j:n}-F^{-1}(j/n)|/\sigma\to 0$ uniformly. A simple condition for the uniform almost sure convergence is stated in Zieliński (1998). Problems in estimating nonunique quantiles are discussed in Feldman and Tucker (1966).
4. Since $F^{-1}(p)$ is usually estimated by $X_{[np]+1:n}$, the jth-order statistic is used for estimating $F^{-1}(p)$ for $p\in((j-1)/n,j/n]$. Applying the method from the proof of Proposition 1, one can verify that both the inequalities of (1) hold true and are optimal for $p=(j-1)/(n-1)\in((j-1)/n,j/n)$ and all $1<j<n<\infty$. In fact, the lower bound in (1) holds for $p\in[(j-1)/n,(j-1)/(n-1)]$ and so does the upper one for $p\in[(j-1)/(n-1),j/n]$.
5. Using the Hőlder inequality in (1), (3) and (5), instead of the Schwarz one, we could obtain bias evaluations expressed in scale units generated by absolute central moments of various orders.

3. Proofs

Let us first recall some auxiliary lemmas. We formulate Theorem 1 of Moriguti (1953) in the form which we shall use further.

Lemma 1. *Let $\Phi : [a,b] \to \mathscr{R}$ be a right-continuous function of finite variation. Then the relation*

$$\int_a^b x(t)\,\mathrm{d}\Phi(t) \leqslant \int_a^b x(t)\bar{\varphi}(t)\,\mathrm{d}t$$

holds for any nondecreasing function $x : (a,b) \to \mathscr{R}$ for which the integrals exist and are finite, where $\bar{\varphi}$ is the right-hand derivative of the greatest convex minorant $\bar{\Phi}$ of Φ. The equality holds if and only if x is constant on each connected interval from the set $\{t \in (a,b) : \bar{\Phi}(t) < \Phi(t)\}$.

The following variation diminishing property of the Bernstein polynomials is presented in Schoenberg (1959) (see also Gajek and Rychlik, 1998).

Lemma 2. *Let $a_1,\ldots,a_n$ be real numbers such that at least one of them is not equal to zero. The number of zeros in $(0,1)$ of a linear combination $\sum_{j=1}^n a_j B_{j,n}$ of Bernstein polynomials does not exceed the number of sign changes of the sequence $a_1,\ldots,a_n$. The first and the last sign of $\sum_{j=1}^n a_j B_{j,n}$ are identical with the signs of the first and last nonzero elements of $a_1,\ldots,a_n$, respectively.*

Lemma 3 is cited from Arnold and Balakrishnan (1989, Lemma 3.31), which was concluded from a more general result in Barlow and Proschan (1966, Lemma 3.5).

Lemma 3. *Suppose that $g : [0,1] \to \mathscr{R}$ changes sign from $+$ to $-$ exactly once in the interval $[0,1]$. If the relevant integrals converge, then*

$$a_{j,n} = \int_0^1 g(u) f_{j:n}(u)\,\mathrm{d}u$$

changes sign at most once (from $+$ to $-$) as n is fixed and j increases from 1 to n. Similarly, sequence $a_{j,n}$ changes sign at most once (from $-$ to $+$) as n increases from fixed j to ∞.

Lemma 4. *Set*

$$b(j,n) = F_{j:n}\left(\frac{j}{n}\right) - \frac{j}{n} f_{j:n}\left(\frac{j}{n}\right), \quad 2 \leqslant j < n < \infty.$$

Then $b(j,n)$ are positive if and only if $j = 2$ and $n \geqslant 7$.

Proof. For small $\varepsilon > 0$ define

$$b(j,n,\varepsilon) = F_{j:n}\left(\frac{j}{n}\right) - \frac{j}{n}\frac{F_{j:n}(j/n+\varepsilon) - F_{j:n}(j/n)}{\varepsilon}$$

$$= \int_0^1 \left[\mathbf{1}_{[0,j/n)}(x) - \frac{j}{n\varepsilon}\mathbf{1}_{[j/n,j/n+\varepsilon)}(x)\right] f_{j:n}(x)\,\mathrm{d}x.$$

Lemma 3 implies that for arbitrary j and ε, sequence $b(j,\cdot,\varepsilon)$ is either positive (is $+$, for short) or negative $(-)$, or changes sign from $-$ to $+$ $(-\,+)$. Since $b(j,n,\varepsilon) \nearrow b(j,n)$ as $\varepsilon \searrow 0$, each limiting sequence

$b(j,\cdot)$, $n>j$, is also either $-$ or $+$, or $-\,+$. Consequently, $b(j)=\lim_{n\to\infty} b(j,n)<0$ implies $b(j,n)<0$ for every $n>j$. Note that

$$b(j)=1-\mathrm{e}^{-j}\sum_{k=0}^{j-1}\frac{j^k}{k!}-\mathrm{e}^{-j}\frac{j^j}{(j-1)!}\equiv 1-c(j)-a(j),$$

with $1-c(j)\in(0,1)$, because $c(j)=P(X<j)$ for a random variable X with the Poisson distribution with parameter j. It follows that condition $a(j)>1$ implies $b(j)<0$. We can easily check that sequence $a(j)$, $j\geqslant 2$, is increasing and $a(6)<1<a(7)$, and therefore $b(j)<0$ for every $j\geqslant 7$. The same conclusion can be drawn by direct inspection for $j=3,4,5,6$. It follows that $b(j,n)<0$ for all $3\leqslant j<n$. In the case $j=2$ we verify that $b(2)>0$ and $b(2,6)<0<b(2,7)$. Referring to Lemma 3 again, we finally conclude that $b(2,n)>0$ for $n\geqslant 7$ and $b(2,n)<0$ for $n=3,4,5,6$. □

Proof of Proposition 1. (i) Observe that for all $1<j<n<\infty$

$$E_F X_{j:n}-F^{-1}(j/n)=\int_0^1[F^{-1}(x)-\mu]H_{j:n}(\mathrm{d}x)$$

with

$$H_{j:n}(x)=\begin{cases}F_{j:n}(x) & \text{for } x<p,\\ F_{j:n}(x)-1 & \text{for } x\geqslant p.\end{cases}\tag{9}$$

Function (9) increases from 0 at 0 to $F_{j:n}(p)>0$, jumps down to $F_{j:n}(p)-1<0$ at p, and is ultimately concave and increasing to 0 at 1. Therefore, its greatest convex minorant is a two-piece broken line with knots at $(0,0)$, $(p,F_{j:n}(p)-1)$ and $(1,0)$, and the derivative satisfying

$$\overline{H}'_{j:n}(x)=\begin{cases}\dfrac{F_{j:n}(p)-1}{p} & \text{for } x<p,\\[2mm] \dfrac{1-F_{j:n}(p)}{1-p} & \text{for } x\geqslant p\end{cases}$$

with

$$\|\overline{H}'_{j:n}\|^2=\frac{[1-F_{j:n}(p)]^2}{p(1-p)}.$$

By Lemma 1 and the Schwarz inequality we obtain

$$E_F X_{j:n}-F^{-1}(j/n)\leqslant\int_0^1[F^{-1}(x)-\mu]\overline{H}'_{j:n}(x)\,\mathrm{d}x\leqslant\|\overline{H}'_{j:n}\|\sigma,$$

which proves (i). The equality is attained in the upper inequality of (1) if

$$F^{-1}(x)-\mu=\frac{\overline{H}'_{j:n}(x)}{\|\overline{H}'_{j:n}\|}\sigma,$$

which implies (2).

(ii) Suppose that either $j\geqslant 3$, or $j=2$ with $n=2,3,4,5,6$, and note that

$$F^{-1}(j/n)-E_F X_{j:n}=\int_0^1[F^{-1}(x)-\mu](-H_{j:n})(\mathrm{d}x).$$

The negative of (9) is concave–convex decreasing from 0 at 0 to $-F_{j:n}(p)$ at $p=j/n$ with the inflection point $(j-1)/(n-1)$, and then positive decreasing to 0 at 1. If the line segment joining $(0,0)$ with $(p,-F_{j:n}(p))$ entirely passes below the graph of $-H_{j:n}$, this provides the part of the greatest convex minorant of $-H_{j:n}$, and the other is the linear piece joining $(p, -F_{j:n}(p))$ with $(1,0)$. This occurs if $-H_{j;n}(p-) = -F_{j:n}(p)/p \geqslant H'_{j:n}(p-) = -f_{j:n}(p)$, or, equivalently, if $b(j,n) \leqslant 0$. This is so in the cases we consider here. Therefore the derivative of the greatest convex minorant $\overline{-H}_{j:n}$ of $-H_{j:n}$ has the following form:

$$\overline{-H}'_{j:n}(x) = \begin{cases} -\dfrac{F_{j:n}(p)}{p} & \text{for } x < p, \\ \dfrac{F_{j:n}(p)}{1-p} & \text{for } x \geqslant p. \end{cases}$$

By Lemma 1 we get

$$F^{-1}(j/n) - E_F X_{j:n} \leqslant \int_0^1 [F^{-1}(x) - \mu](\overline{-H}_{j:n})'(x)\,\mathrm{d}x \leqslant \|\overline{-H}'_{j:n}\|\sigma,$$

where

$$\|\overline{-H}'_{j:n}\|^2 = \frac{F^2_{j:n}(p)}{p(1-p)}.$$

The lower inequality in (1) is attained if

$$\frac{F^{-1}(x) - \mu}{\sigma} = \frac{\overline{-H}'_{j:n}(x)}{\|\overline{-H}'_{j:n}\|},$$

which leads to (2) again. □

Proof of Proposition 2. Function $-H_{1:n}$ starts from 0 at 0, is convex decreasing on $[0, 1/n)$, and positive decreasing to 0 on $[1/n, 1]$. The conclusion is that the derivative of the greatest convex minorant $\overline{-H}_{1:n}$ of $-H_{1:n}$ has the form

$$\overline{-H}'_{1:n}(x) = \begin{cases} -f_{1:n}(x) & \text{for } x < 1/n, \\ \dfrac{nF_{1:n}(1/n)}{n-1} & \text{for } x \geqslant 1/n. \end{cases}$$

Arguments similar to those of Proposition 1 give us the statement of Proposition 2. □

Proof of Proposition 3. Suppose that $j = 2$ and $n \geqslant 7$. Our basic purpose is determining the greatest convex minorant $\overline{-H}_{2:n}$ of $-H_{2:n}$. We recall properties of $-H_{2:n}$ from the proof of Proposition 1. Now condition $b(2,n) > 0$ derived in Lemma 4 implies that the greatest convex minorant consists of a linear part on some $[0, x_n] \subset [0, p)$ with x_n determined by $H_{2:n}(x_n-)/x_n = H'_{2:n}(x_n-)$ (which can be rewritten as (6)), and $-H_{2:n}$ itself on $[x_n, p]$, and the straight line running through $(p, -F_{2:n}(p))$ and $(1,0)$ on $[p, 1]$. The uniqueness of x_n follows from an application of Lemma 2 to

$$G_{j,n}(x) = -(j-1)B_{j:n}(x) + \sum_{k=j+1}^{n} B_{k:n}(x).$$

Consequently,

$$\overline{-H}_{2:n}(x) = \begin{cases} -f_{2:n}(x_n)x & \text{for } x \in [0, x_n], \\ -F_{2:n}(x) & \text{for } x \in [x_n, 2/n], \\ \dfrac{nF_{2:n}(2/n)}{n-2}(x-1) & \text{for } x \in [2/n, 1]. \end{cases}$$

The rest of the proof runs analogously to the proofs of Propositions 1 and 2. □

Acknowledgements

This work was done under the support of the Polish State Committee for Scientific Research (KBN) Grant 2 P03A 014 13.

References

Arnold, B.C., Balakrishnan, N., 1989. Relations, Bounds and Approximations for Order Statistics. Springer, Berlin.

Barlow, R.E., Proschan, F., 1966. Inequalities for linear combinations of order statistics from restricted families. Ann. Math. Statist. 37, 1574–1591.

Cheng, C., 1995. The Bernstein polynomial estimator of a smooth quantile function. Statist. Probab. Lett. 24, 321–330.

Feldman, D., Tucker, H.G., 1966. Estimation of non-unique quantiles. Ann. Math. Statist. 37, 451–457.

Gajek, L., Rychlik, T., 1998. Projection method for moment bounds on order statistics from restricted families. II. Independent case. J. Multivariate Anal. 64, 156–182.

Moriguti, S., 1953. A modification of Schwarz's inequality with applications to distributions. Ann. Math. Statist. 24, 107–113.

Schoenberg, I.J., 1959. On variation diminishing approximation methods, In: Langer, R.E. (Ed.), On Numerical Approximation: Proceedings of a Symposium, Madison, WI, April 21–23, Publ. No. 1, Mathematics Research Center, US Army. University of Winsconsin Press, Madison, WI.

Sheather, S.J., Marron, J.S., 1990. Kernel quantile estimators. J. Amer. Statist. Assoc. 80, 410–416.

Zieliński, R., 1988. A distribution-free median-unbiased quantile estimator. Statistics 19, 223–227.

Zieliński, R., 1998. Uniform strong consistency of sample quantiles. Statist. Probab. Lett. 37, 115–119.

Zieliński, R., 1999. Best equivariant nonparametric estimator of a quantile. Statist. Probab. Lett. 45, 79–84.

Zieliński, R., 2001. PMC-optimal nonparametric quantile estimator. Statistics, to appear.