Ryszard Zieliński

OPTIMAL QUANTILE ESTIMATORS
SMALL SAMPLE APPROACH

IMPAN 2004

Contents

1

The Problem

The problem of quantile estimation has a very long history and abundant literature: in out booklet we shall quote only the sources which we directly refer to.

We are interested in **small sample** and **nonparametric** quantile estimators.

"Small sample" is here used as an opposite to "asymptotic" and it is meant that the statistical inference will be based on independently and identically distributed observations $X_1, \ldots, X_n$ for **a fixed** n. A short excursion to asymptotics is presented in Chapter 6.

"Nonparametric" is here used to say that observations $X_1, \ldots, X_n$ come from an unknown distribution $F \in \mathcal{F}$ with $\mathcal{F}$ being the class of all continuous and strictly increasing distribution functions and, for a given $q \in (0,1)$, we are interested in estimating the qth quantile $x_q = x_q(F)$ of the distribution F. If q is fixed (for example, if we are interested in estimating the median only), the conditions for $\mathcal{F}$ may be relaxed and $\mathcal{F}$ may be considered as the class of all locally at x_q continuous and strictly increasing distributions; we shall not exploit this trivial comment. The nonparametric class $\mathcal{F}$ of distributions is rather large and one can hardly expect to get many strict mathematical theorems which hold simultaneously for all distributions $F \in \mathcal{F}$. An example of such a theorem is the celebrated Glivenko–Cantelli theorem for the Kolmogorov distance $\sup|F_n - F|$. It appears that the class $\mathcal{F}$ is too large to say something useful concerning the behavior of L-estimators; classical estimators and their properties are discussed in Chapter 2. The natural class of estimators in $\mathcal{F}$ is the class $\mathcal{T}$ of estimators which are equivariant under monotonic transformations of data; under different criteria of optimality, the best estimators in $\mathcal{T}$ are constructed in Chapter 3.

Our primary interest is **optimal** nonparametric estimation. Constructions of optimal estimators are presented in Chapter 3 followed by their applications to parametric models (Chapter 4) and some results concerning their asymptotic properties (Chapter 6). An excursion to optimal interval estimation is presented in Chapter 5. In Chapter 3 to Chapter 6 we present almost without changes previous results of the author published since 1987 in different journals, mainly in *Statistics*, *Statistics and Probability Letters*, and *Applicationes Mathematicae (Warszawa)*.

Observe that in the class $\mathcal{F}$ of distributions, the order statistic $(X_{1:n}, \ldots, X_{n:n})$, where $X_{1:n} \leq \ldots, \leq X_{n:n}$, is a complete minimal sufficient statistic. As a consequence we confine ourselves to estimators which are functions of $(X_{1:n}, \ldots, X_{n:n})$. Some further restrictions for the estimators will be considered in Chapter 3. We shall use $T(q)$ or shortly T as general symbols for estimators to be considered; the sample size n is fixed and in consequence we omit n in most notations.

2

Classical Approach (inverse of cdf)

For a distribution function F, the qth quantile $x_q = x_q(F)$ of F is defined as $x_q = F^{-1}(q)$ with

$$F^{-1}(q) = inf\{x : F(x) \geq q\}. \tag{1}$$

For $F \in \mathcal{F}$ and $q \in (0,1)$ it always exists and is uniquely determined. The well recognized *generalized method of moments* or *method of statistical functionals* gives us formally

$$T(q) = F_n^{-1}(q) = inf\{x : F_n(x) \geq q\} \tag{2}$$

as an estimator $T(q)$ of x_q; here F_n is an empirical distribution function. Different definitions of F_n lead of course to different estimators T. One can say that the variety of definitions of F_n (left- or right-continuous step functions, smoothed versions, F_n as a kernel estimator of F, etc) is what produces the variety of estimators to be found in abundant literature in mathematical statistics. We shall use the following definition of the empirical distribution function:

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1_{(-\infty,x]}(X_i), \tag{3}$$

where the indicator function $1_{(-\infty,x]}(X_i) = 1$ if $X_i \leq x$ and $= 0$ otherwise. Note that under the definition adapted, the empirical distribution function is right-continuous.

There are two kinds of estimators widely used. Given a sample, if $F_n(x)$ is a step function then estimator (2) as a function of $q \in (0,1)$ takes on a finite number of different values, typically the values of order statistics from the sample; if $F_n(x)$ is continuous and strictly increasing empirical distribution function, so is its inverse $Q_n(t) = F_n^{-1}(t)$, $t \in (0,1)$, the quantile function, and $T(q)$ can be considered as a continuous and strictly increasing function of $q \in (0,1)$. An example give us estimators presented in Fig. 2.2.1 (Sec. 2.2). In what follows we discuss both types of estimators.

A natural problem arises how one can assess quality of an estimator, compare distributions, or at least some parameters of distributions of different estimators of a given quantile, or even the distributions of a fixed estimator under different parent distributions F from the large class $\mathcal{F}$? In other words: how one can assess the quality of an estimator in very large nonparametric class $\mathcal{F}$ of distributions? No characteristics like bias (in the sense of mean), mean square error, mean absolute error, etc, are acceptable because not for all distributions $F \in \mathcal{F}$ they exist or if exist they may be infinite.

What is more: it appears that the model $\mathcal{F}$ is so large that assessing the quality of an estimator T of the q-th quantile $x_q(F)$ in terms of the difference $T - x_q(F)$ makes no sense. To see that take as an example the well known estimator of the median $m_F = x_{0.5}(F)$ of an unknown distribution $F \in \mathcal{F}$ from a sample of size $2n$, defined as the arithmetic mean of two central observations $M_{2n} = (X_{n:2n} + X_{n+1:2n})/2$. Let $Med(F, M_{2n})$ denote a median of the distribution of M_{2n} if the sample comes from the distribution F.

Theorem 1 (Zieliński 1995). *For every $C > 0$ there exists $F \in \mathcal{F}$ such that*

$$Med(F, M_{2n}) - m_F > C.$$

Proof. The proof consists in constructing $F \in \mathcal{F}$ for a given $C > 0$. Let $\mathcal{F}_0$ be the class of all strictly increasing continuous functions G on $(0,1)$ satisfying $G(0) = 0$, $G(1) = 1$. Then $\mathcal{F}$ is the class of all functions F satisfying $F(x) = G((x-a)/(b-a))$ for some a and b $(-\infty < a < b < +\infty)$, and for some $G \in \mathcal{F}_0$.

For a fixed $t \in (\frac{1}{4}, \frac{1}{2})$ and a fixed $\varepsilon \in (0, \frac{1}{4})$, let $F_{t,\varepsilon} \in \mathcal{F}_0$ be a distribution function such that

$$F_{t,\varepsilon}\left(\frac{1}{2}\right) = \frac{1}{2}, \quad F_{t,\varepsilon}(t) = \frac{1}{2} - \varepsilon,$$
$$F_{t,\varepsilon}(t - \frac{1}{4}) = \frac{1}{2} - 2\varepsilon, \quad F_{t,\varepsilon}(t + \frac{1}{4}) = 1 - 2\varepsilon.$$

Let $Y_1, Y_2, \ldots, Y_{2n}$ be a sample from $F_{t,\varepsilon}$. We shall prove that for every $t \in (\frac{1}{4}, \frac{1}{2})$ there exists $\varepsilon > 0$ such that

$$Med\left(F_{t,\varepsilon}, \frac{1}{2}(Y_{n:2n} + Y_{n+1:2n})\right) \leq t. \tag{4}$$

Consider two random events:

$$A_1 = \{0 \leq Y_{n:2n} \leq t, 0 \leq Y_{n+1:2n} \leq t\},$$
$$A_2 = \{0 \leq Y_{n:2n} \leq t - \frac{1}{4}, \frac{1}{2} \leq Y_{n+1:2n} \leq t + \frac{1}{4}\},$$

and observe that $A_1 \cap A_2 = \emptyset$ and

$$A_1 \cup A_2 \subseteq \{\frac{1}{2}(Y_{n:2n} + Y_{n+1:2n}) \leq t\}. \tag{5}$$

If the sample comes from a distribution G with a probability density function g, then the joint probability density function $h(x,y)$ of $Y_{n:2n}, Y_{n+1,2n}$ is given by the formula

$$h(x,y) = \frac{\Gamma(2n+1)}{\Gamma^2(n)} G^{n-1}(x) \left[1 - G(y)\right]^{n-1} g(x)g(y), \quad 0 \leq x \leq y \leq 1,$$

and the probability of A_1 equals

$$P_G(A_1) = \int_0^t dx \int_x^t dy\, h(x,y).$$

Using the formula

$$\frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \int_0^x t^{p-1}(1-t)^{q-1} dt = \sum_{j=p}^{p+q-1} \binom{p+q-1}{j} x^j (1-x)^{p+q-1-j},$$

we obtain

$$P_G(A_1) = \sum_{j=n+1}^{2n} \binom{2n}{j} G^j(t) \left(1 - G(t)\right)^{2n-j}.$$

For $P_G(A_2)$ we obtain

$$\begin{aligned} P_G(A_2) &= \int_0^{t-\frac{1}{4}} dx \int_{\frac{1}{2}}^{t+\frac{1}{4}} dy\, h(x,y) \\ &= \binom{2n}{n} G^n(t - \frac{1}{4}) \left[\left(1 - G\left(\frac{1}{2}\right)\right)^n - \left(1 - G(t + \frac{1}{4})\right)^n \right]. \end{aligned}$$

Define $C_1(\varepsilon) = P_{F_{t,\varepsilon}}(A_1)$ and $C_2(\varepsilon) = P_{F_{t,\varepsilon}}(A_2)$. Then

$$C_1(\varepsilon) = \sum_{j=n+1}^{2n} \binom{2n}{j} (\frac{1}{2} - \varepsilon)^j (\frac{1}{2} + \varepsilon)^{2n-j},$$

$$C_2(\varepsilon) = \binom{2n}{n} (\frac{1}{2} - 2\varepsilon)^n \left[\left(\frac{1}{2}\right)^n - (2\varepsilon)^n \right].$$

Observe that

$$C_1(\varepsilon) \nearrow \frac{1}{2} - \frac{1}{2}\binom{2n}{n}\left(\frac{1}{2}\right)^{2n} \quad \text{as } \varepsilon \searrow 0$$

and

$$C_2(\varepsilon) \nearrow \binom{2n}{n}\left(\frac{1}{2}\right)^{2n} \quad \text{as} \quad \varepsilon \searrow 0.$$

Let $\varepsilon_1 > 0$ be such that

$$(\forall \varepsilon < \varepsilon_1) \qquad C_1(\varepsilon) > \frac{1}{2} - \frac{3}{4}\binom{2n}{n}\left(\frac{1}{2}\right)^{2n}$$

and let ε_2 be such that

$$(\forall \varepsilon < \varepsilon_2) \qquad C_2(\varepsilon) > \frac{3}{4}\binom{2n}{n}\left(\frac{1}{2}\right)^{2n}.$$

Then for every $\varepsilon < \bar{\varepsilon} = min\{\varepsilon_1, \varepsilon_2\}$ we have $C_1(\varepsilon) + C_2(\varepsilon) > \frac{1}{2}$ and by (5) for every $\varepsilon < \bar{\varepsilon}$,

$$P_{F_{t,\varepsilon}}\{\frac{1}{2}\left(Y_{n:2n} + Y_{n+1:2n}\right) \leq t\} > C_1(\varepsilon) + C_2(\varepsilon) > \frac{1}{2},$$

which proves (4).

For a fixed $t \in (\frac{1}{4}, \frac{1}{2})$ and $\varepsilon < \bar{\varepsilon}$, let $Y, Y_1, \ldots, Y_{2n}$ be independent random variables identically distributed according to $F_{t,\varepsilon}$, and for a given $C > 0$, define

$$X = C \cdot \frac{\frac{1}{2} - Y}{\frac{1}{2} - t},$$

$$X_{i:2n} = C \cdot \frac{\frac{1}{2} - Y_{2n+1-i:2n}}{\frac{1}{2} - t}, \quad i = 1, \ldots, 2n.$$

Let F denote the distribution function of X. Then

$$P\{X \leq 0\} = P\{Y \geq \frac{1}{2}\} = \frac{1}{2}.$$

Hence $F^{-1}(\frac{1}{2}) = 0$ and

$$P\{\frac{1}{2}\left(X_{n:2n} + X_{n+1:2n}\right) \leq C\} = P\{\frac{1}{2}\left(Y_{n:2n} + Y_{n+1:2n}\right) \geq t\} \leq \frac{1}{2}.$$

Thus $Med\left(F, \frac{1}{2}\left(X_{n:2n} + X_{n+1:2n}\right)\right) \geq C$, which proves the Theorem. □

It is obvious from the proof of Theorem 1 that similar result holds for all non-trivial L-estimators; "non-trivial" means that two or more coefficients α in $\sum \alpha_j X_{j:n}$ do not equal zero.

We may overcome the difficulty as follows. If $T = T(q)$ is an estimator of the qth quantile of an unknown distribution $F \in \mathcal{F}$, then $F(T)$ may be considered as an estimator of the (known!) value q. The distribution of $F(T)$ is concentrated in the interval $(0,1)$ and we exactly know what it is that $F(T)$ estimates. Of course all moments of the distribution of $F(T)$ exist and we are able to assess quality of such estimators F in terms if their bias in mean (or bias), bias in median, mean square error ($MSE = \sqrt{E_F(F(T)-q)^2}$), mean absolute deviation ($MAD = E_F|F(T)-q|$), etc, as well as to compare quality of different estimators of that kind. Some estimators T have the property that $F(T)$ does not depend of the parent distribution $F \in \mathcal{F}$; they are "truly" nonparametrical (distribution-free) estimators. Estimators which do not share the property may perform very bad at least for some distribution $F \in \mathcal{F}$ and if the statistician does not know anything more about the parent distribution except that it belongs to $\mathcal{F}$, he is not able to predict consequences of his inference.

In this Chapter we discuss in details some well known and widely used estimators T and assess their quality in terms of $F(T)$.

2.1. Single order statistics

By (3) and (2), as an estimator of the qth quantile we obtain (cf David et al. 1986)

$$x_q^{(1)} = \begin{cases} X_{nq:n}, & \text{if } nq \text{ is an integer,} \\ X_{[nq]+1:n}, & \text{if } nq \text{ is not an integer.} \end{cases}$$

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

The estimator is defined for all $q \in (0,1)$ but due to a property of F_n as defined in (3) (continuous from the right and discontinuous from the left) it is not symmetric. We call an estimator of the q-th quantile $X_{k(q):n}$ symmetric if $k(1-q) = n - k(q) + 1$. A rationale for condition of symmetry for an estimator is that if a quantile of order q is estimated, say, by the smallest order statistic $X_{1:n}$ then the quantile of order $1-q$ should be estimated by the largest order statistic $X_{n:n}$. For estimator $x_q^{(1)}$, if nq is not an integer, and $(k-1)/n < q < k/n$ for some k, then $[nq] = k-1$, $x_q^{(1)} = X_{k:n}$, $[n(1-q)] = n-k$ and $x_{1-q}^{(1)} = X_{n-k+1:n}$. If, however, nq is an integer and $q = k/n$ then $x_q^{(1)} = X_{k:n}$ but $1-q = 1-k/n$, $[n(1-q)] = n-k$ and $x_{1-q}^{(1)} = X_{n-k:n}$. To remove the flaw we shall define $x_q^{(1)} = X_{nq}$ if nq is an integer and $q < 0.5$, and $x_q^{(1)} = X_{nq+1}$ if nq is an integer and $q > 0.5$. Another disadvantage (an asymmetry) of the estimator $x_q^{(1)}$ is that if $q = 1/2$ (estimating a median) and $n = 2m$ for an integer m, then the estimator equals $X_{m:n}$ instead of being a combination of two central order statistics $X_{m:n}$ and $X_{m+1:n}$. We may define, in full agreement with statistical tradition, $x_{0.5}^{(1)} = (X_{m:n} + X_{m+1:n})/2$ but that is not a single order statistic (see next Section) and we prefer to choose $X_{m:n}$ or $X_{m+1:n}$ at random, each with probability 1/2.

Eventually we define the estimator (we call it *standard*)

$$\hat{x}_q = X_{k(q):n} \tag{6}$$

where

$$k(q) = \begin{cases} nq, & \text{if } nq \text{ is an integer and } q < 0.5, \\ nq + 1, & \text{if } nq \text{ is an integer and } q > 0.5, \\ \dfrac{n}{2} + \mathbb{1}_{(0,1/2]}(U), & \text{if } nq \text{ is an integer and } q = 0.5, \\ [nq] + 1, & \text{if } nq \text{ is not an integer.} \end{cases}$$

Here U is a uniformly $U(0,1)$ distributed random variable independent of the observations $X_1, \ldots, X_n$, and $\mathbb{1}_{(a,b)}(x)$ is the indicator function which equals 1 if $x \in (a,b)$ and 0 otherwise. In other words: to estimate the median (i.e. for $q = 0.5$) take the central order statistic if the sample size n is odd or choose at random one of two central order statistics if n is even. Note that $\hat{x}_q$ may differ from the typical $x_q^{(1)}$ only when estimating the quantiles of order $q = j/n, j = 1, 2, \ldots, n$ i.e. if nq is an integer.

The distribution function of $\hat{x}_q$, if the sample comes from a distribution F, is given by the formula

$$P_F\{\hat{x}_q \le x\} =$$

$$= \begin{cases} \sum_{j=\frac{n}{2}+1}^{n} \binom{n}{j} F^j(x)[1-F(x)]^{n-j} + \frac{1}{2}\binom{n}{n/2} \left(F(x)[1-F(x)]\right)^{\frac{n}{2}}, & \text{if } nq \text{ is an integer and } q = 0.5 \\ \sum_{j=k(q)}^{n} \binom{n}{j} F^j(x)[1-F(x)]^{n-j}, & \text{otherwise.} \end{cases}$$

If $q = 0.5$ then $\hat{x}_q$ is a median unbiased estimator of the median $F^{-1}(1/2)$ and also $E\hat{x}_q$ equals the median, if the expectation exists. Estimator $x_q^{(1)}$ does not have that property.

Sometimes estimators $x_q^{(2)} = X_{[nq]:n}$, $x_q^{(3)} = X_{[(n+1)q]:n}$, or $x_q^{(4)} = X_{[(n+1)q]+1:n}$ are used. The statistics are however defective in a sense:

$x_q^{(2)} = X_{[nq]:n} = X_{0:n}$ for $q < 1/n$ so that the statistic is not defined for q close to zero, but it is well defined for all q in every vicinity of 1; an asymmetry arises. The order statistic $X_{n:n}$ is never used;

$x_q^{(3)} = X_{[(n+1)q]:n}$ is not symmetric and not defined for $q < 1/(n+1)$;

$x_q^{(4)} = X_{[(n+1)q]+1:n}$ is not symmetric and not defined for $q > n/(n+1)$ though well defined for all $q \in (0, n/(n+1))$.

One can argue that there is no sense to estimate quantiles of the order close to 0 or close to 1 if a sample is not large enough. Then, for example, the following estimators give us a remedy

$$\hat{x}_q = \begin{cases} X_{[nq]:n}, & \text{if } \{nq\} \le 0.5, \\ X_{[nq]+1:n}, & \text{if } \{nq\} > 0.5, \end{cases} \quad \text{or} \quad \hat{x}_q = \begin{cases} X_{[nq]:n}, & \text{if } \{nq\} < 0.5, \\ X_{[nq]+1:n}, & \text{if } \{nq\} \ge 0.5. \end{cases}$$

Here $\{x\} = x - [x]$ is the fractional part of x ("the nearest integer principle"). Another construction gives us

$$\hat{x}_q = \begin{cases} X_{[(n+1)q]:n}, & \text{if } \ q \le 0.5, \\ X_{[(n+1)q]+1:n}, & \text{if } \ q > 0.5. \end{cases} \quad \text{or} \quad \hat{x}_q = \begin{cases} X_{[(n+1)q]:n}, & \text{if } \ q < 0.5, \\ X_{[(n+1)q]+1:n}, & \text{if } \ q \ge 0.5. \end{cases}$$

The former is not defined outside of the interval $[1/n, 1/n)$, the latter outside the interval $[1/(n+1), 1 - 1/(n+1))$; observe that the intervals are not symmetric. However, a more serious problem is to choose between $\Big(\{nq\} \le 0.5, \{nq\} > 0.5\Big)$ or $\Big(\{nq\} < 0.5, \{nq\} \ge 0.5\Big)$ in the former case or between $\Big(q \le 0.5, q > 0.5\Big)$ and $\Big(q < 0.5, q \ge 0.5\Big)$ in the latter case; or perhaps introduce a new definition of the estimator for $q = 0.5$. A possible corrections of the definitions when estimating the median from a sample of size n, if n even, is to take the arithmetic mean of central observations, which is a common practice, but then the estimator is not a single order statistic which we discuss in this Section.

Another approach consists in defining an estimator as in (2) with a modified empirical distribution function, e.g.

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

("weighted empirical distribution function") instead of (3). For example, Huang and Brill (1999) considered

$$w_{i,n} = \begin{cases} \dfrac{1}{2}\left[1 - \dfrac{n-2}{\sqrt{n(n-1)}}\right], & i = 1, n, \\ \dfrac{1}{\sqrt{n(n-1)}}, & i = 2, 3, \ldots, n-1 \end{cases}$$

which gives us

$$\hat{x}_{HB}(q) = X_{[b]+2:n}, \ \ q \in (0,1),$$

with

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

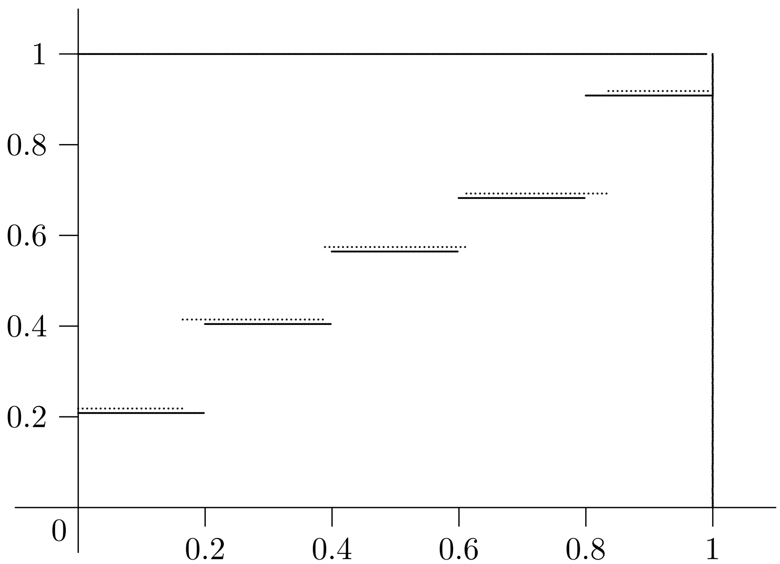


Fig.2.1.1. Two estimators
from the sample (0.2081, 0.4043, 0.5642, 0.6822, 0.9082)
generated from the uniform $U(0,1)$ distribution
Estimator $X_{[nq]+1:n}$ - solid line; Huang-Brill estimator -dots
Solid lines and dots are at the same levels $X_{1:n}, X_{2:n}$, etc

Note that both estimators take on the values of single order statistics (Fig. 2.1.1):

$$\hat{x}_q = X_{k:n} \quad \text{iff} \quad \frac{k-1}{n} < q < \frac{k}{n}$$

and

$$\hat{x}_{HB}(q) = X_{k:n} \quad \text{iff} \quad \frac{1}{2} + \frac{k - n/2 - 1}{\sqrt{n(n-1)}} < q < \frac{1}{2} + \frac{k - n/2}{\sqrt{n(n-1)}}.$$

with suitable modifications if nq is an integer. The Huang-Brill estimator $\hat{x}_{HB}(q)$ is defined on the interval $\left(0.5 - 0.5\sqrt{n/(n-1)}, \ 0.5 + 0.5\sqrt{n/(n-1)}\right) \supset (0,1)$.

How can we assess the quality of the estimators and to decide which estimator to choose?

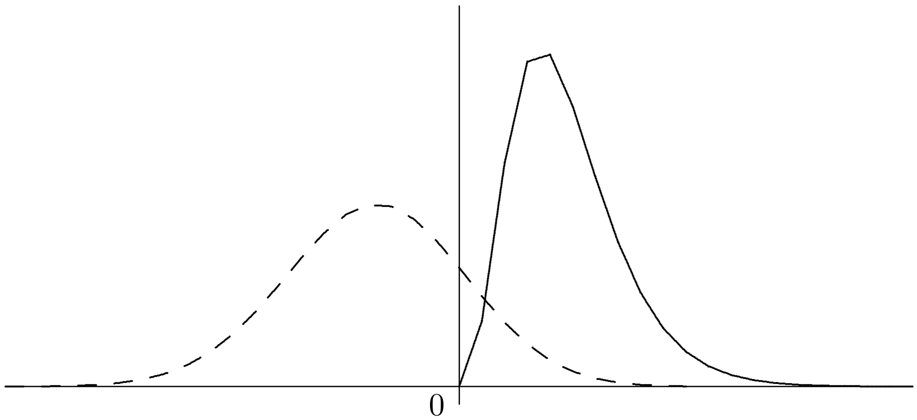


Fig.2.1.2.
Distribution of $\hat{x}_q$ for $N(0,1)$ [dashes] and $E(1)$ [solid] parent distributions

The variety of distributions leads of course to a variety of distributions of a given estimator. As an example consider distributions of $\hat{x}_q$ for $q = 0.3$ if the sample of size $n = 10$ comes from the normal $N(0,1)$ and from the Exponential $E(1)$ distributions (Fig. 2.1.2).

An advantage of single order statistics as quantile estimators T=T(q)= $X_{k:n}$ for some k is that if a sample $X_1, \ldots, X_n$ comes from a distribution $F \in \mathcal{F}$ then the distribution of $F(\mathrm{T}) = U_{k:n}$ does not depend on the parent distribution; here $U_{k:n}$ is the kth order statistic from the sample from the uniform $U(0,1)$ distribution. It follows that the distribution of $F(\mathrm{T})$ is the same for all $F \in \mathcal{F}$; that for $n = 10$ and $q = 0.3$ as above is presented in Fig. 2.1.3; the quality of the estimator in the whole class $\mathcal{F}$ is completely characterized by that distribution.

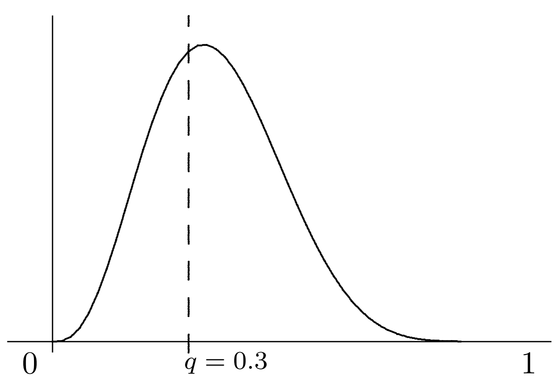
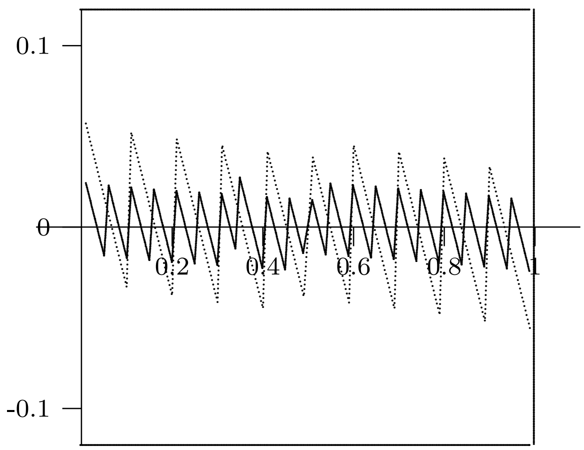


Fig.2.1.3.

Distribution of $F(\hat{x}_q)$ for $n = 10$ and $q = 0.3$
if the sample comes from any distribution $F \in \mathcal{F}$

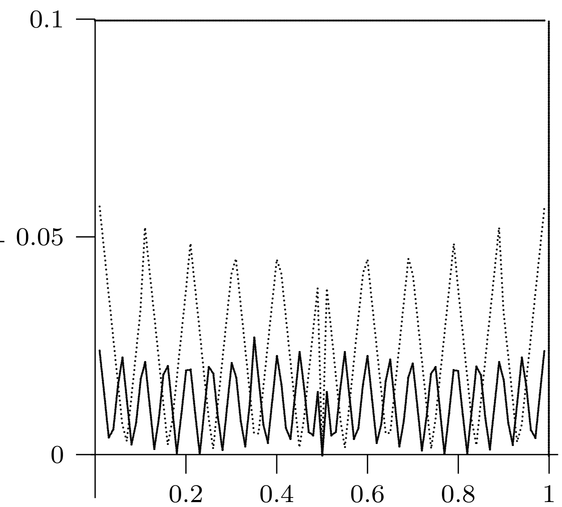
Bias, median-bias and their absolute values, MSE and MAD of the estimators are exhibited in Fig. 2.1.4 - 2.1.7.

In Fig. 2.1.8 and Fig. 2.1.9, MSE and MAD of both estimators are compared for samples of size $n = 10$ and $n = 20$ respectively. The figures demonstrate that manipulations with empirical distribution function may introduce some asymmetry in estimators as well as in their quality.



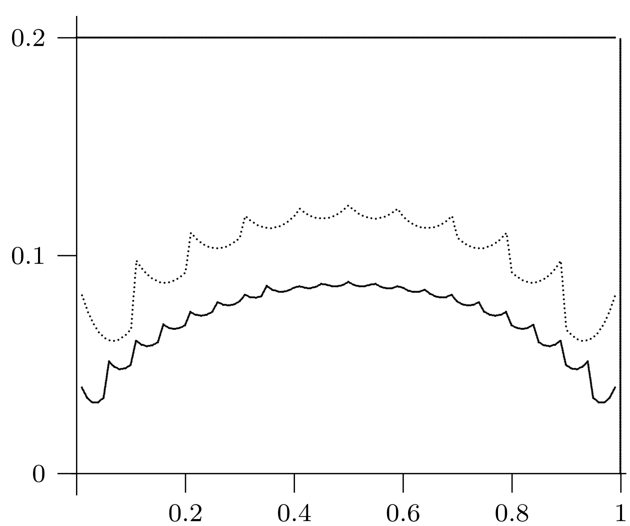
a) Median bias of $\hat{x}_q$

$n\!=\!10$ dots, $n\!=\!20$ solid



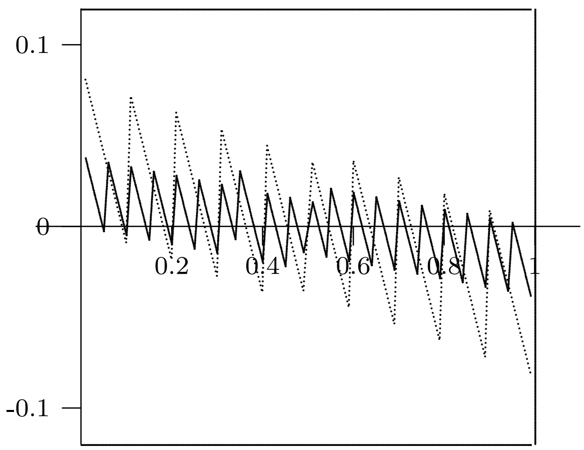
b) Absolute median bias of $\hat{x}_q$

$n\!=\!10$ dots, $n\!=\!20$ solid

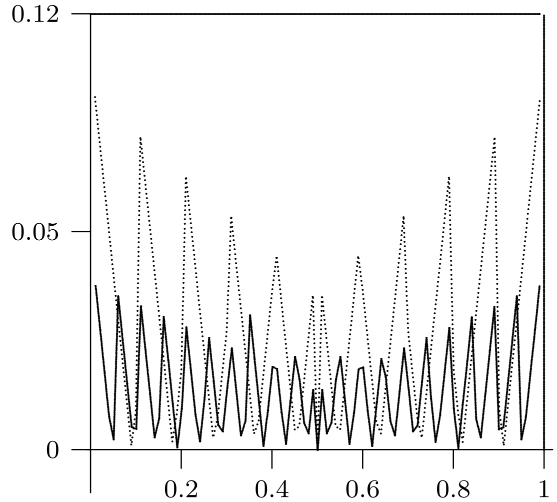


c) MAD of $\hat{x}_q$: $n\!=\!10$ dots, $n\!=\!20$ solid

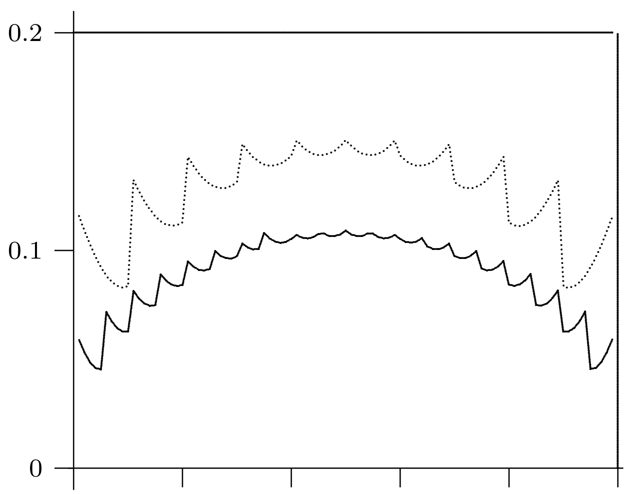
Fig. 2.1.4.



a) Bias of $\hat{x}_q$: $n = 10$ dots, $n = 20$ solid

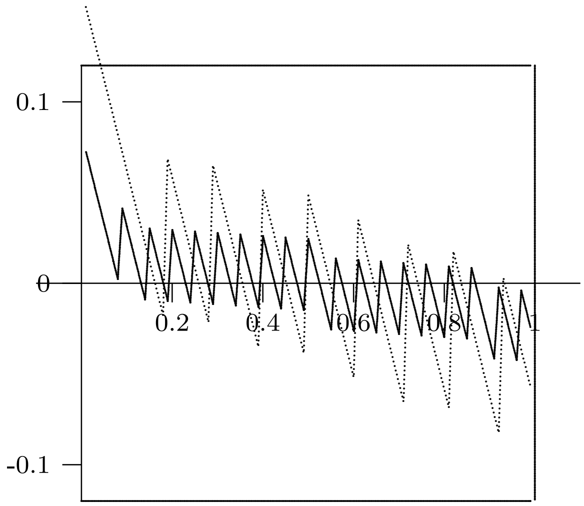


b) Absolute bias of $\hat{x}_q$: $n = 10$ dots, $n = 20$ solid



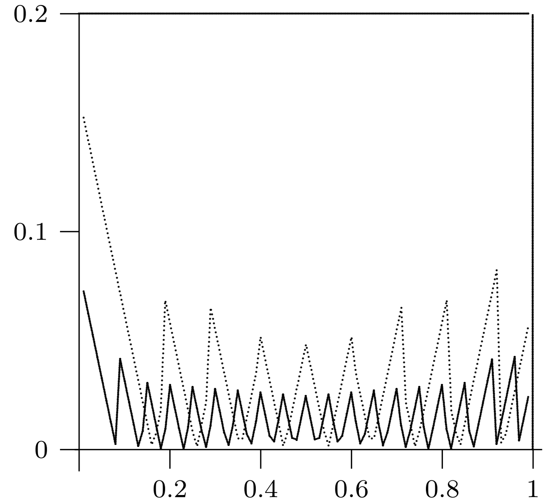
c) MSE of $\hat{x}_q$: $n = 10$ dots, $n = 20$ solid

Fig. 2.1.5.



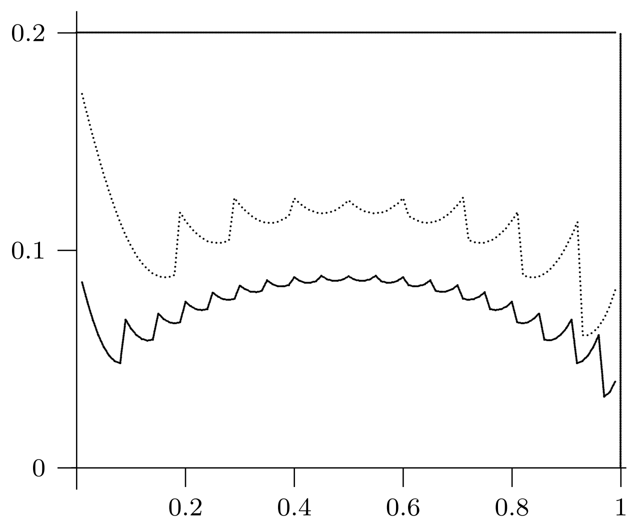
a) Median bias of $\hat{x}_{HB}$

$n=10$ dots, $n=20$ solid



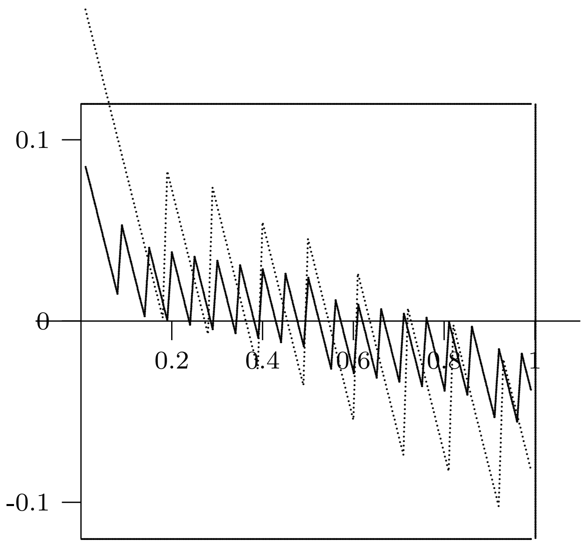
b) Absolute median bias of $\hat{x}_{HB}$

$n=10$ dots, $n=20$ solid

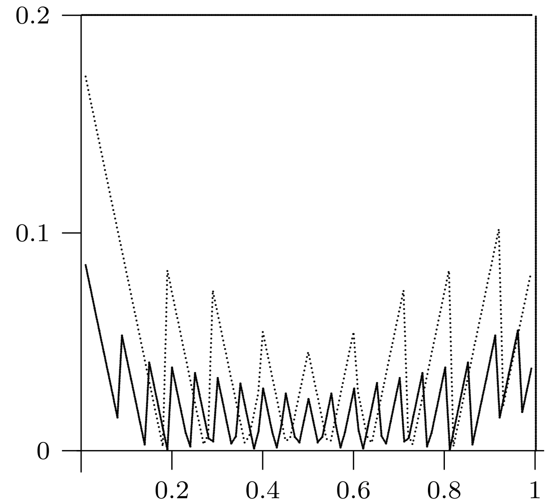


c) MAD of $\hat{x}_{HB}$: $n=10$ dots, $n=20$ solid

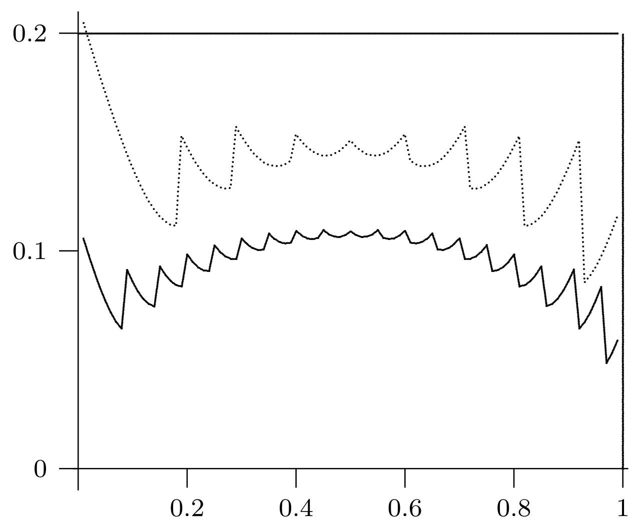
Fig. 2.1.6.



a) Bias of $\hat{x}_{HB}$: $n = 10$ dots, $n = 20$ solid



b) Absolute bias of $\hat{x}_{HB}$: $n = 10$ dots, $n = 20$ solid



c) MSE of $\hat{x}_{HB}$: $n = 10$ dots, $n = 20$ solid

Fig. 2.1.7.

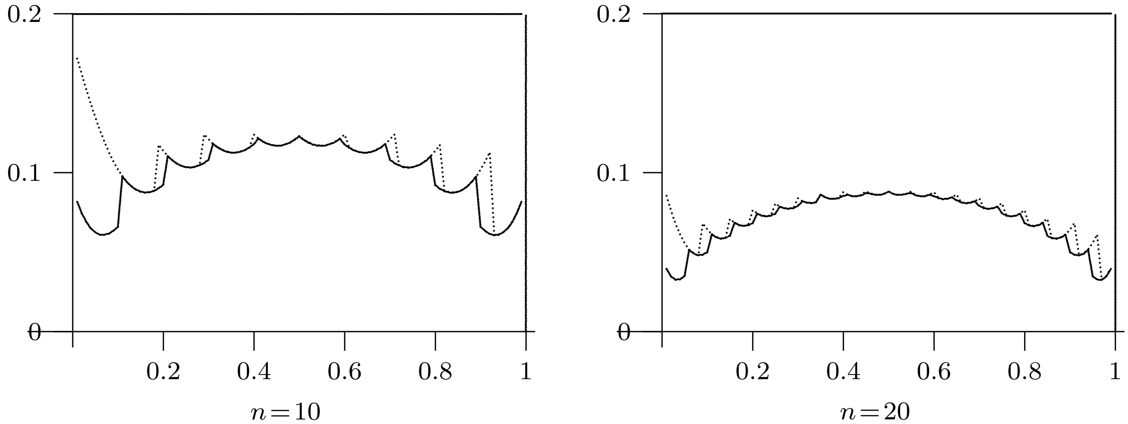


Fig. 2.1.8. MAD of $\hat{x}_{HB}$ - dots, $\hat{x}_q$ - solid

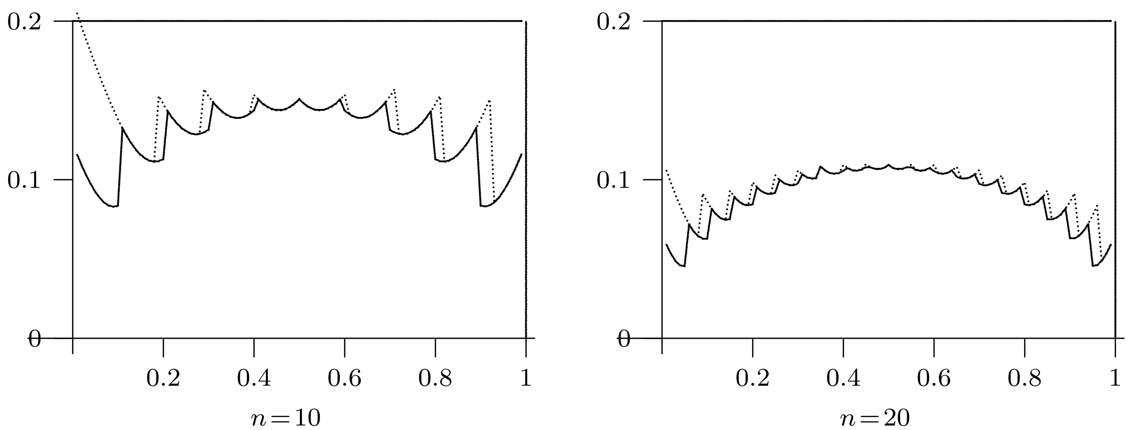


Fig. 2.1.9. MSE of $\hat{x}_{HB}$ - dots, $\hat{x}_q$ - solid

Though not fully satisfactory, we choose estimator $\hat{x}_q$ as a benchmark for assessing other estimators below.

To the end we return to estimators which we rejected as "defective" at the very beginning of this Section. Fig. 2.1.10 exhibits absolute median-bias and MAD of estimators $x_q^{(2)} = X_{[nq]:n}$, $x_q^{(3)} = X_{[(n+1)q]:n}$, $x_q^{(4)} = X_{[(n+1)q]+1:n}$, and the standard estimator $\hat{x}_q$.

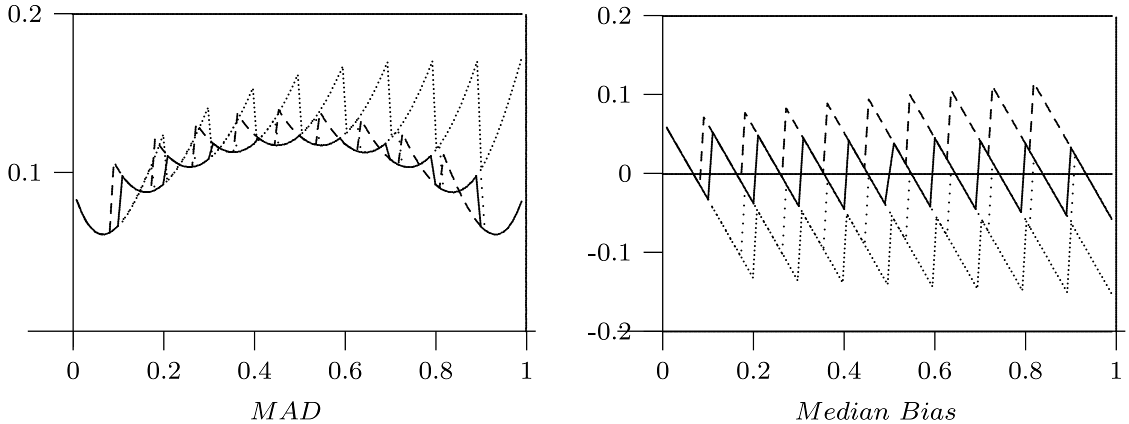


Fig. 2.1.10. Mean Absolute Deviation and Median Bias of four estimators

$\hat{x}_q$ .... $X_{[nq]:n}$ .... $X_{[(n+1)q]:n}$ - - - - - - $X_{[(n+1)q]+1:n}$

We clearly see that from all those estimators only the estimator $\hat{x}_q$ deserves some attention.

When estimating the median of an unknown distribution $F \in \mathcal{F}$ from a sample of an even size $2n$, estimator $\hat{x}_q$ is randomized: it chooses $X_{n:2n}$ or $X_{n+1:2n}$ with equal probability 1/2; otherwise the estimator is not randomized. Let $L_F(T)$ denote a loss function of an estimator T when estimating the median of F. Then the risk of the estimator T is $E_F L_F(T)$. For $\hat{x}_{0.5}$ from a sample of size $2n$ we have

$$
\begin{aligned}
E_F(\hat{x}_{0.5}) &= \frac{1}{2}\Big(E_F L(X_{n:2n}) + E_F L(X_{n+1:2n})\Big) \\
&= \frac{1}{2}\Big(EL(U_{n:2n}) + EL(U_{n+1:2n})\Big) \\
&= \frac{1}{2}\Big[\int_0^1 L(x)\frac{(2n)!}{(n-1)!n!}x^{n-1}(1-x)^n dx + \int_0^1 L(x)\frac{(2n)!}{n!(n-1)!}x^n(1-x)^{n-1}dx\Big] \\
&= \int_0^1 \frac{(2n-1)!}{(n-1)!(n-1)!}x^{n-1}(1-x)^{n-1}dx \\
&= EL(U_{n:2n-1}) = E_F L(X_{n:2n-1})
\end{aligned}
$$

which means that the risk of the randomized estimator $\hat{x}_{0.5}$ from a sample of size $2n$ is is equal to the risk of the non-randomized estimator $\hat{x}_{0.5}$ from the sample of size $2n-1$. It follows that instead of randomization we may reject one observation from the original sample: randomization for the median amounts to removing one observation.

2.2. Local Smoothing

Given q, the local smoothing idea consists in constructing an estimator of the qth quantile x_q on the basis of two consecutive order statistics from a neighborhood of $X_{[nq]+1:n}$. Perhaps the best known example is the sample median which for n being an even integer is defined as the arithmetic mean of two "central" observations: $(X_{\frac{n}{2}:n} + X_{\frac{n}{2}+1:n})/2$. A possible rationale for the choice is as follows. According to Definition (3)

$$F_n(X_{\frac{n}{2}:n}) = \lim_{0<t\to 0} F_n(X_{\frac{n}{2}+1:n} - t) = \frac{1}{2}.$$

The left-continuous version of the empirical distribution function

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

satisfies

$$\lim_{0<t\to 0} F_n'(X_{\frac{n}{2}:n} + t) = F_n'(X_{\frac{n}{2}+1:n}) = \frac{1}{2}$$

so that there is no reason to choose $X_{\frac{n}{2}:n}$ instead of $X_{\frac{n}{2}+1:n}$ or vice versa as an estimator for the median $x_{0.5}$ and to define the sample median depending on a choice of the right- or a left-continuous version of the empirical distribution function. Statistical tradition suggests to take the mean of both. Another point of view on the choice $(X_{\frac{n}{2}:n} + X_{\frac{n}{2}+1:n})/2$ as an estimator of the median was presented in the previous section when discussing the cases of $\{nq\} = 0.5$ or $q = 0.5$. It appears that the resulting estimator performs not very well in the very large statistical model $\mathcal{F}$ (see Theorem 1 above).

More generally, a simple linear smoothing based on two consecutive order statistics leads to the estimator

$$(7)\quad \hat{x}_{LS} = \Big(1-(n+1)q+[(n+1)q]\Big)X_{[(n+1)q]:n} + \Big((n+1)q-[(n+1)q]\Big)X_{[(n+1)q]+1:n}$$

which however is naturally defined for $q \in [1/(n+1), n/(n+1))$ only. A reason for choice of $(n+1)q$ in (7) instead of nq as in (6) is that as a special case of (7) we obtain the central value $X_{\frac{n}{2}+1:n}$ of the sample $(X_{1:n}, \ldots, X_{n:n})$ if n is odd, or the arithmetic mean $(X_{\frac{n}{2}:n}+X_{\frac{n}{2}+1:n})/2$ of two central observations if n is even. Examples of estimators (6) and (7) as functions of the order q of the quantile to be estimated, for a fixed sample from the uniform $U(0,1)$ distribution, are exhibited in Fig. 2.2.1.

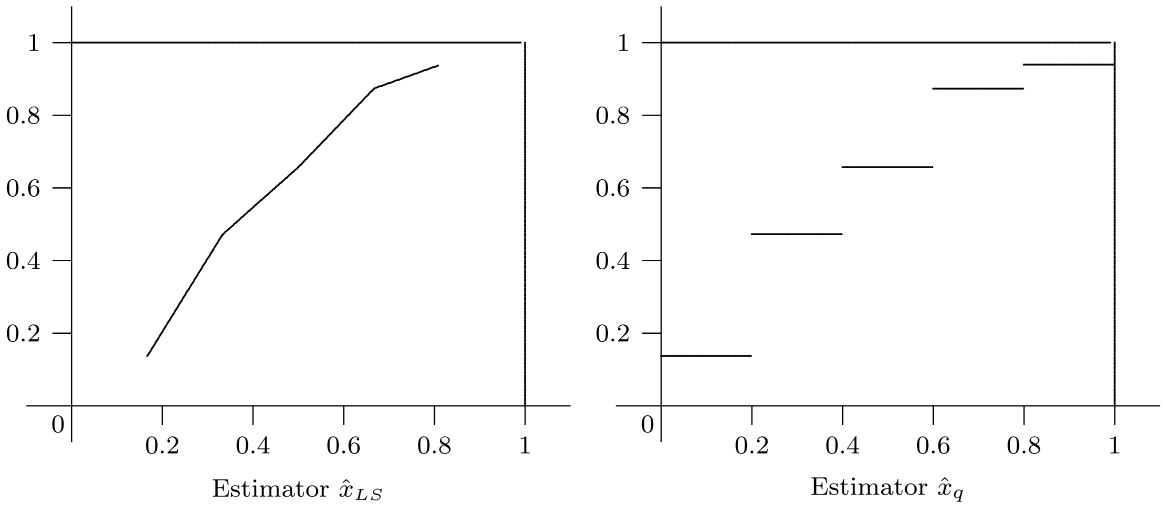


Fig.2.2.1
Estimators $\hat{x}_{LS}$ and $\hat{x}_q$ as functions of $q \in (0,1)$
from the sample (0.1375, 0.4716, 0.6575, 0.8741, 0.9398)
generated from the uniform $U(0,1)$ distribution

In opposite to the standard estimator $\hat{x}_q$ from the previous Section, $\hat{x}_{LS}$ is not "truly" nonparametric estimator: it is obvious that the distribution of $\hat{x}_{LS}$ from the sample depends on the parent distribution of that sample (which is also the case for $\hat{x}_q$), but it appears that the distribution of $F(\hat{x}_{LS})$ also depends on the parent distribution F. It is a result of the fact that the estimator is not equivariant with respect to monotonic transformations.

Theorem 2. *If the sample comes from the uniform $U(0,1)$ distribution, then for the estimator $\hat{x}_{LS}$ we have*

$$P\{\hat{x}_{LS}(q) \le s\} = B(s; k+1, n-k) +$$

$$+ \begin{cases} (-1)^{k+1} \left(\frac{\lambda}{1-\lambda}\right)^k \sum_{i=k+1}^{n} (-1)^i \binom{n}{i} (\frac{s}{\lambda})^i \bar{B}(\lambda; i-k, k+1), & \text{if } s \le \lambda \\ (-1)^k \left(\frac{\lambda}{1-\lambda}\right)^k \sum_{i=0}^{k} (-1)^i \binom{n}{i} (\frac{s}{\lambda})^i \bar{B}(s; k-i+1, n-k), & \text{if } s \ge \lambda \end{cases}$$

where $k = [(n+1)q]$, $\lambda = \{(n+1)q\}$, $B(x; p, q)$ is the distribution function of the Beta random variable $B(p,q)$ at x:

$$B(x; p, q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \int_0^x t^{p-1}(1-t)^{q-1} dt,$$

and $\bar{B}(x;p,q) = 1 - B(x;p,q)$ is the tail of that random variable. Another, perhaps more suitable for numerical calculations, formula is

$$P\{\hat{x}_{LS} \le s\} = B\big(s; k+1, n-k\big) + \frac{n!}{k!(n-k-1)!} \int_s^{\min\{1,s/\lambda\}} \left(\frac{s-\lambda u}{1-\lambda}\right)^k (1-u)^{n-k-1} du.$$

Proof. The formula may be obtained as follows. The join probability distribution function of $X_{k:n}$ and $X_{k+1:n}$ is

$$f(x,y) = \frac{n!}{(k-1)!(n-k-1)!} x^{k-1}(1-y)^{n-k-1}, \quad 0 \le x \le y \le 1$$

and

(8) $$P\{\hat{x}_{LS} \le s\} = \int_S \int f(x,y)dxdy,$$

where

$$S = \{(x,y) : (1-\lambda)x + \lambda y \le s,\ 0 \le x \le y \le 1\} = A \cup B$$

with

$$A = \Big\{0 \le y \le s,\ 0 \le x \le y\Big\}$$

and

$$B = \left\{ s \le y \le min\Big\{1, \frac{s}{\lambda}\Big\},\ 0 \le x \le \frac{s-\lambda y}{1-\lambda} \right\}$$

(see Fig.2.2.2).

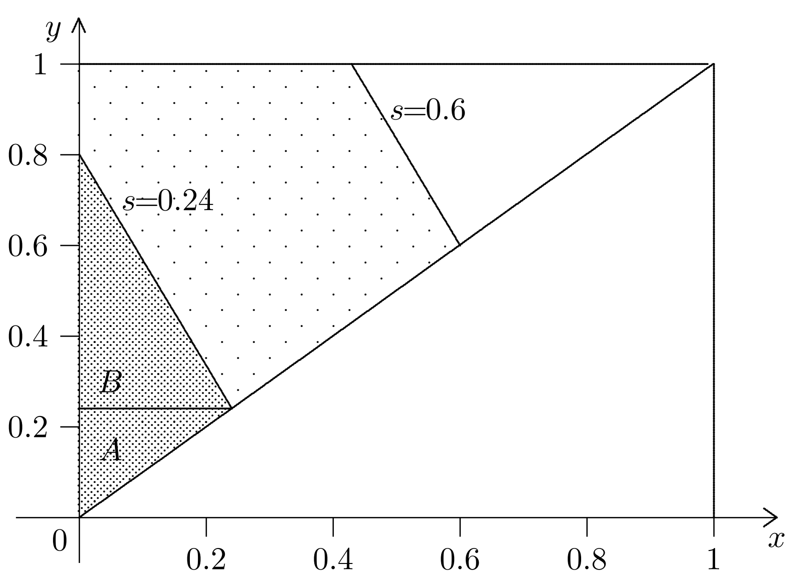


Fig.2.2.2.Integration areas in (8) for $\lambda = 0.3$

Integrating over A gives us

$$\int_A\int f(x,y)dxdy = \int_0^s dy \int_0^y dx f(x,y)$$
$$= \frac{n!}{k!(n-k-1)!}\int_0^s y^k(1-y)^{n-k-1}dy = B(s;k+1,n-k).$$

By the equality

$$\int\limits_0^{\frac{s-\lambda y}{1-\lambda}} x^{k-1}dx = \frac{1}{k}\left(\frac{s-\lambda y}{1-\lambda}\right)^k,$$

the integration over B gives us

$$\int_B\int f(x,y)dxdy = \frac{n!}{k!(n-k-1)!}\int\limits_s^{\min\{1,s/\lambda\}}\left(\frac{s-\lambda u}{1-\lambda}\right)^k(1-u)^{n-k-1}du$$

which gives us the second formula. Performing simple calculations we obtain

$$\int_B\int f(x,y)dxdy =$$

$$= \begin{cases} \dfrac{n!}{k!(n-k-1)!}\left(\dfrac{s}{1-\lambda}\right)^k\displaystyle\int_s^{s/\lambda}(1-y)^{n-k-1}\left(1-\frac{\lambda}{s}y\right)^k dy, & \text{if } s\le\lambda \\ \dfrac{n!}{k!(n-k-1)!}\left(\dfrac{\lambda}{1-\lambda}\right)^k\displaystyle\int_s^{1}(1-y)^{n-k-1}\left(\frac{\lambda}{s}-y\right)^k dy, & \text{if } s>\lambda \end{cases}$$

$$= \begin{cases} \dfrac{n!}{k!(n-k-1)!}\left(\dfrac{s}{1-\lambda}\right)^k\displaystyle\sum_{j=0}^{n-k-1}(-1)^j\binom{n-k-1}{j}\left(\frac{s}{\lambda}\right)^{j+1}\int\limits_\lambda^1 u^j(1-u)^k du, & \text{if } s\le\lambda \\ \dfrac{n!}{k!(n-k-1)!}\left(\dfrac{\lambda}{1-\lambda}\right)^k\displaystyle\sum_{j=0}^{k}(-1)^{k-j}\binom{k}{j}\left(\frac{s}{\lambda}\right)^{j}\int\limits_s^1(1-y)^{n-k-1}y^{k-j}du, & \text{if } s>\lambda \end{cases}$$

and the result follows. □

Probability distribution function is given by the formula

$$\frac{1}{1-\lambda}\cdot\frac{n!}{(k-1)!(n-k-1)!}\int\limits_s^{\min\{1,s/\lambda\}}\left(\frac{s-\lambda u}{1-\lambda}\right)^{k-1}(1-u)^{n-k-1}du$$

which may be useful in theoretical analysis of properties of the estimator $\hat{x}_{LS}$.

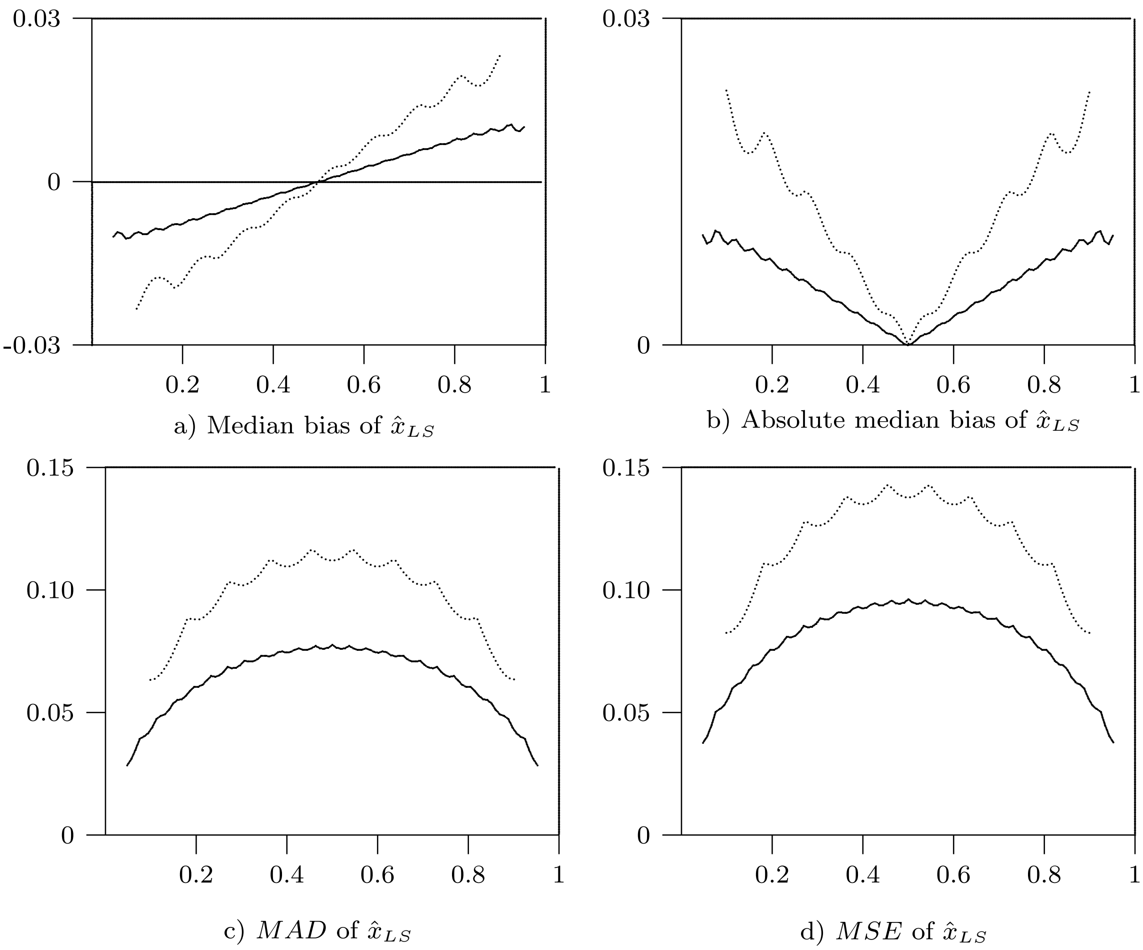


Fig. 2.2.3. ($n=10$ dots, $n=25$ solid)

If a sample comes from a distribution $F \in \mathcal{F}$ then the distribution function of $\hat{x}_{LS}$ is given by the formula:

$$\begin{aligned} P_F\{\hat{x}_{LS} \le s\} =& B\big(F(s); k+1, n-k\big) \\ &+ \frac{n!}{k!(n-k-1)!} \int_s^\infty F^k\left(\frac{s-\lambda u}{1-\lambda}\right)(1-F(u))^{n-k-1} f(u)du \\ =& B\big(F(s); k+1, n-k\big) \\ &+ \frac{n!}{k!(n-k-1)!} \int_{F(s)}^{F(s/\lambda)} F^k\left(\frac{s-\lambda F^{-1}(t)}{1-\lambda}\right)(1-t)^{n-k-1} dt \end{aligned}$$

The formula may be easily obtained by the method used in the proof of Theorem 2.

If a sample comes from the uniform $U(0,1)$ distribution then the estimator $\hat{x}_{LS}$ of the q-th quantile is unbiased:

$$
\begin{aligned}
&E(\hat{x}_{LS}) = \\
&= \Big(1-(n+1)q+[(n+1)q]\Big)EX_{[(n+1)q]:n} + \Big((n+1)q-[(n+1)q]\Big)EX_{[(n+1)q]+1:n} \\
&= \Big(1-(n+1)q+[(n+1)q]\Big)\frac{[(n+1)q]}{n+1} + \Big((n+1)q-[(n+1)q]\Big)\frac{[(n+1)q]+1}{n+1} = q.
\end{aligned}
$$

Median bias, absolute median bias, Mean Absolute Deviation, and standard deviation of the estimator are exhibited in Fig. 2.2.3.

2.3. Global smoothing

Global smoothing consists in constructing a smooth (continuous and strictly increasing) version Q_n of the empirical quantile function and estimating the q-quantile x_q by $Q_n(q)$. Typically,

$$
Q_n(u) = \sum_{j=1}^{n} \psi(u,j,n)X_{j:n}, \quad 0<u<1,
$$

with suitably chosen functions ψ which may be considered as "kernels" (then an additional parameter - the windows width - is usually added) or simply as functional coefficients associated with order statistics. We shall not consider general kernels estimators here; all what is known in general case are asymptotic properties which are out of our concern. In the fixed sample size case the estimators may be considered as L-estimators; below and in Sec. 2.4, we present a choice.

2.3.1. Harrell-Davis estimator.

Harrell-Davis estimator (Harrell and Davis (1982), David and Steinberg (1986)) is based on the following reasoning:

> *Since $E(X_{(n+1)q})$ converges to $F^{-1}(q)$ for $q\in(0,1)$, we take as our estimator of $x_q = F^{-1}(q)$ something which estimates $E(X_{(n+1)q})$ whether or not $(n+1)q$ is an integer.*

Expectation of $X_{j:n}$ is given by the formula

$$E(X_{j:n}) = \frac{1}{B(i, n-i+1)} \int_{-\infty}^{+\infty} xF^{j-1}(x)(1-F(x))^{n-j} dF(x)$$
$$= \frac{1}{B(i, n-i+1)} \int_0^1 F^{-1}(t)t^{j-1}(1-t)^{n-j} dt,$$

and *"something which estimates* $E(X_{j:n})$*"* is the "sample mean"

$$\frac{1}{B(i, n-i+1)} \int_0^1 F_n^{-1}(t)t^{j-1}(1-t)^{n-j} dt.$$

Assuming $j = (n+1)q$ and

$$F_n^{-1}(t) = X_{j:n} \quad \text{if} \quad \frac{(j-1)}{n} < t \leq \frac{j}{n},$$

the Harrell-Davis estimator of x_q takes on the form

$$\hat{x}_{HD}(q) = \sum_{j=1}^{n} W_{q,j,n} X_{j:n}, \quad 0 < q < 1, \tag{9}$$

where

$$W_{q,j,n} = \frac{1}{B((n+1)q, (n+1)(1-q))} \int_{(j-1)/n}^{j/n} t^{(n+1)q-1}(1-t)^{(n+1)(1-q)-1} dt.$$

Harrell-Davis estimator is defined for all $q \in (0,1)$. By the well known equality relating beta and binomial distributions

$$\frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \int_0^x t^{p-1}(1-t)^{q-1} dt = \sum_{j=p}^{p+q-1} \binom{p+q-1}{j} x^j (1-x)^{p+q-1-j}$$

we have

$$W_{q,1,n} = \sum_{j=(n+1)q}^{n} \binom{n}{j} \left(\frac{1}{n}\right)^j \left(1 - \frac{1}{n}\right)^{n-j} \to 1, \quad \text{as } q \to 0$$

and similarly

$$W_{q,n,n} \to 1, \quad \text{as } q \to 1.$$

Hence

$$\hat{x}_{HD}(q) \to X_{1:n} \text{ as } q \to 0 \quad \text{and} \quad \hat{x}_{HD}(q) \to X_{n:n} \text{ as } q \to 1.$$

Estimators $\hat{x}_{HD}(q)$ as functions of $q \in (0,1)$ for three samples generated from the uniform $U(0,1)$ distribution are exhibited in Fig. 2.3.1.

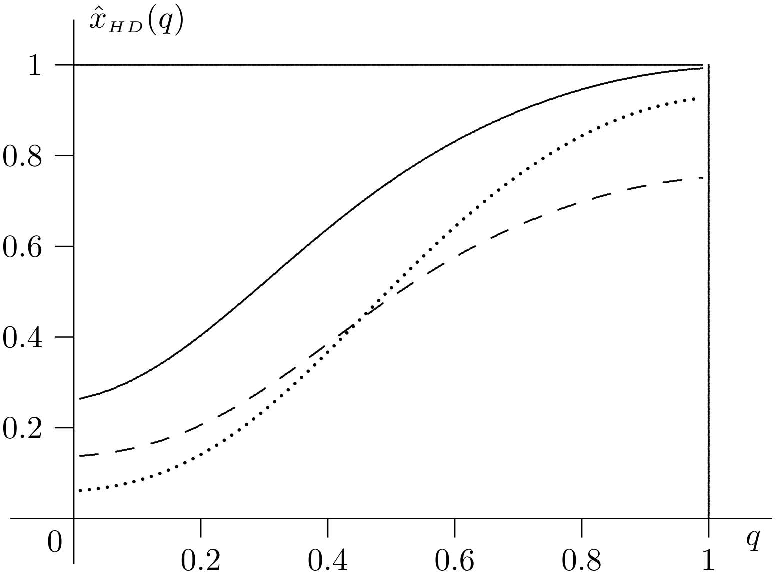


Fig.2.3.1. Harrell-Davis estimators $\hat{x}_{HD}(q)$ as functions of $q \in (0,1)$
Random samples from the uniform $U(0,1)$ distribution
Sample (0.261, 0.513, 0.822, 0.919, 0.993) - solid
Sample (0.137, 0.189, 0.592, 0.657, 0.752) - dashes
Sample (0.060, 0.117, 0.580, 0.807, 0.929) - dots

The estimator, like other non-trivial L-statistics, is not equivariant with respect to monotonic transformations of data so that the distribution of $F(\hat{x}_{HD})$ depends on the parent distribution $F \in \mathcal{F}$; that is demonstrated in Fig. 2.3.2, for sample of size $n = 10$, for parent distributions uniform $U(0,1)$, normal $N(0,1)$, and Pareto $F_\alpha(x) = 1 - x^{-\alpha}, x > 0$, for $\alpha = 0.5$ and $\alpha = 1$.

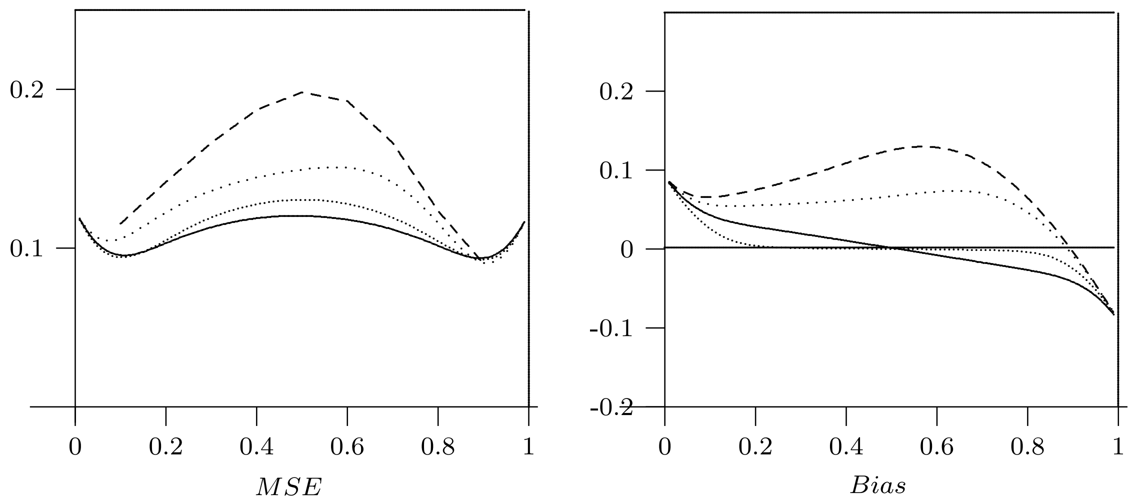


Fig. 2.3.2. Standard Deviation and Bias of $\hat{x}_{HD}$

$U(0,1)$ $N(0,1)$ $Pareto(1)$ $Pareto(0.5)$

To compare the estimator with some other traditional ones (Section 2.5) we shall consider its behavior for the uniform $U(0,1)$ distribution of data.

2.3.2. Bernstein polynomial estimator.

The literature of the estimator is rather abundant, most results however concern asymptotic properties; some recent studies of the estimator appeared in Cheng (1995), where the estimator of the qth quantile is defined as

$$\hat{x}_{Be}(q) = \sum_{j=1}^{n} \left[\binom{n-1}{j-1} q^{j-1}(1-q)^{n-j} \right] X_{j:n}, \quad 0 < q < 1. \tag{10}$$

Estimator $\hat{x}_{Be}(q)$ as a function of $q \in (0,1)$ for three samples generated from the uniform $U(0,1)$ distribution are exhibited in Fig. 2.3.3.

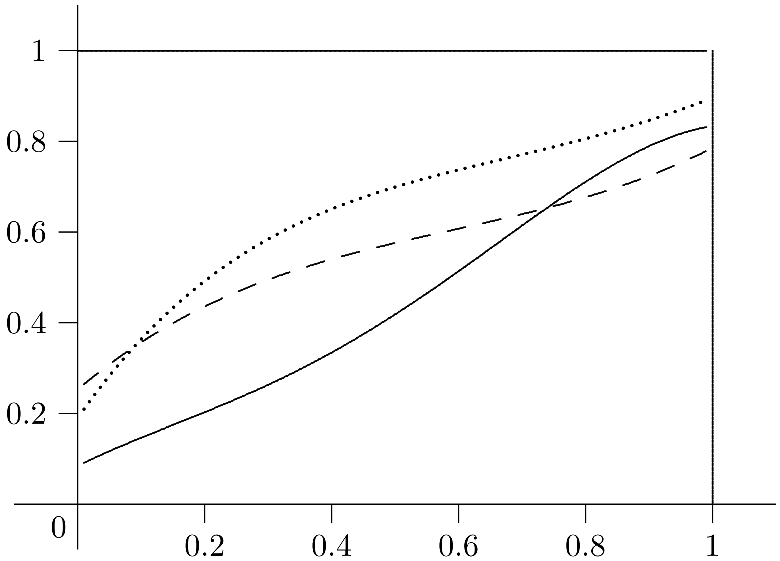


Fig.2.3.3. Bernstein polynomial estimators $\hat{x}_{Be}(q)$ as functions of $q \in (0, 1)$
Random samples from the uniform $U(0, 1)$ distribution
Sample (0.084 0.257 0.267 0.785 0.833) - solid
Sample (0.252 0.547 0.59 0.61 0.785) - dashes
Sample (0.189 0.691 0.717 0.758 0.896) - dots

The estimator is well defined for all $q \in (0, 1)$ and formally even for $q = 0$ and $q = 1$ under the convention that $0^0 = 1$; then $\hat{x}_{Be}(0) = X_{1:n}$ and $\hat{x}_{Be}(1) = X_{n:n}$.

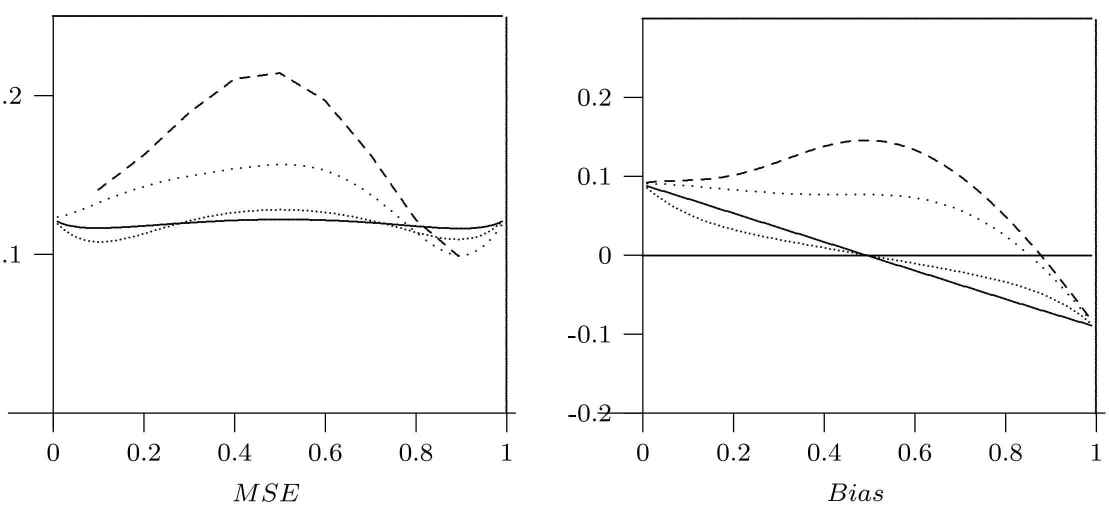


Fig. 2.3.4. Standard Deviation and Bias of $\hat{x}_{Be}$
———— $U(0, 1)$ ········ $N(0, 1)$ ·········· $Pareto(1)$ – – – – – – $Pareto(0.5)$

The estimator, like other non-trivial L-statistics, is not equivariant with respect to monotonic transformations of data so that the distribution of $F(\hat{x}_{Be})$ depends on the parent distribution $F \in \mathcal{F}$; that is demonstrated in Fig. 2.3.4, for sample of size $n = 10$, for parent distributions uniform $U(0,1)$, normal $N(0,1)$, and Pareto $F_\alpha(x) = 1 - x^{-\alpha}, x > 0$, for $\alpha = 0.5$ and $\alpha = 1$. To compare the estimator with some other traditional ones (Section 2.5) we shall consider its behavior for the uniform $U(0,1)$ distribution of data.

2.4. Kaigh-Lachenbruch estimator

The celebrated Kaigh-Lachenbruch estimator (twice in *Encyclopedia of Statistical Sciences, Vol. 7, 1986 Wiley*: David, C.E. and Steinberg, S.M. (1986), Keating, J.P. and Tripathi, R.C. (1986)) was introduced in the original paper (Kaigh and Lachenbruch 1982) as follows (with minor changes in notation):

> *For a fixed integer k satisfying $1 \le k \le n$, consider the selection of a simple random sample (without replacement) from the complete sample $X_1, \ldots, X_n$ and denote the ordered observations in the subsample by $Y_{1:k;n}, \ldots, Y_{k:k;n}$. An elementary combinatorial argument shows that for each integer r satisfying $1 \le r \le k$*
>
> $$P\{Y_{r:k;n} = X_{j:n}\} = \frac{\binom{j-1}{r-1}\binom{n-j}{k-r}}{\binom{n}{k}}, \quad r \le j \le r + n - k.$$
>
> *For $0 < q < 1$ a sample quantile estimator of x_q based on the observations in a single subsample would be $Y_{[(k+1)q]:k;n}$. We define the alternative quantile estimator $KL(q)$ to be the subsample quantile averaged over all $\binom{n}{k}$ subsamples of size k so that the estimator of the qth quantile takes on the form*
>
> $$\sum_{j=r}^{r+n-k} \frac{\binom{j-1}{r-1}\binom{n-j}{k-r}}{\binom{n}{k}} X_{j:n}, \quad r = [(k+1)q].$$

A peculiarity of the estimator consists in that, given a sample and estimating any quantile x_q, $q \in (0,1)$, it can take on only k different values and it is defined for q from the interval $[1/(k+1), 1)$ only.

For example (Fig. 2.4.1), for the sample $(0.2081, 0.4043, 0.5642, 0.6822, 0.9082)$ of size $n = 5$ and $k = 3$, the estimator takes on three different values only, e.g. the value 0.3025 when estimating the qth quantile of any order $q \in (0, 0.333)$. If $k = 1$ then no quantile x_q of order $q < 0.5$ is estimable and each quantile x_q of order $q \ge 0.5$ is

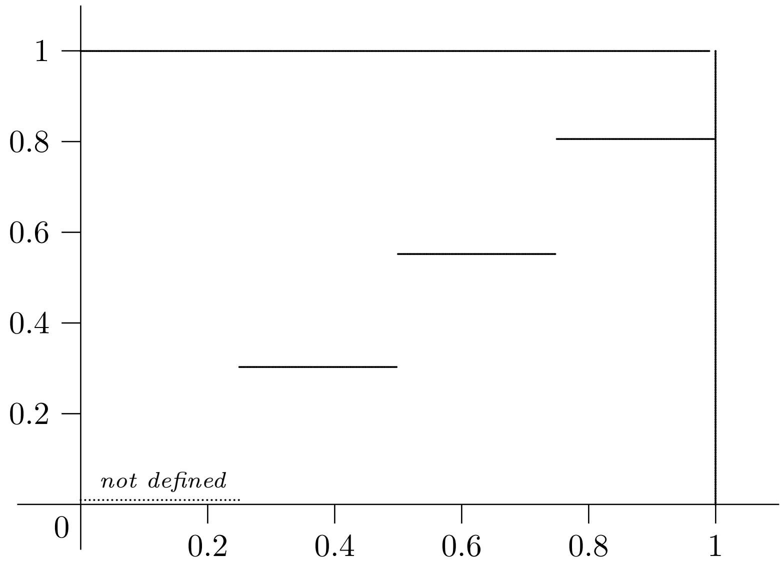


Fig.2.4.1. Original Kaigh-Lachenbruch estimator as a function of $q \in (0,1)$ for $n = 5, k = 3$from the sample (0.2081, 0.4043, 0.5642, 0.6822, 0.9082) generated from the uniform $U(0,1)$ distribution

estimated by the same value $\sum_{j=1}^{n} X_{j:n}/n$. For $k = n$ the estimator equals $X_{(n+1)q:n}$ (see the comment following Formula (6)). The drawbacks do not play much role in the asymptotic theory but for fixed size samples a serious problem in applications arises. As a remedy we may define, like as in (6), $r = r(q)$ rather than $r = [(k+1)q]$ and consider the modification

$$\hat{x}_{KL} = \sum_{j=r}^{r+n-k} \frac{\binom{j-1}{r-1}\binom{n-j}{k-r}}{\binom{n}{k}} X_{j:n} \tag{11}$$

with

$$r = r(q) = \begin{cases} kq, & \text{if } kq \text{ is an integer and } q < 0.5, \\ kq+1, & \text{if } kq \text{ is an integer and } q > 0.5, \\ \dfrac{k}{2} + 1_{(0,1/2]}(U), & \text{if } kq \text{ is an integer and } q = 0.5, \\ [kq]+1, & \text{if } kq \text{ is not an integer.} \end{cases}$$

The modified estimator $\hat{x}_{KL}$, like the estimator (6), is defined for all $q \in (0,1)$ (Fig. 2.4.2). For $k = n$ the estimator is identical with the standard estimator $\hat{x}_q$ defined by (6).

Both in original and modified estimator there is no general rule concerning a choice of parameter k; bias and standard deviation of estimators for $n = 10$ and $k = 2, 4, 6, 8, 10$ are exhibited in Fig. 2.4.3 (original estimator) and Fig. 2.4.4 (modified estimator), all calculations for the uniform $U(0,1)$ parent distribution. No choice of k is uniformly better than other.

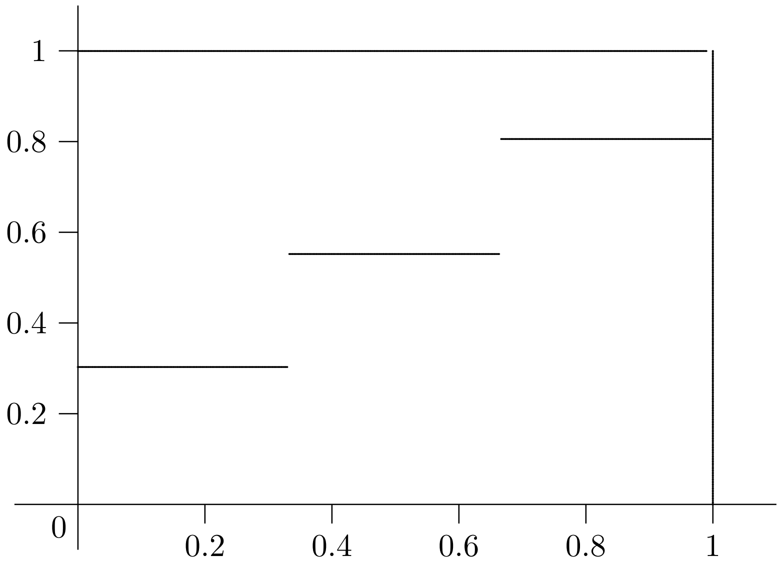


Fig.2.4.2. Modified Kaigh-Lachenbruch $KL(q)$ estimator (11) as a function of $q \in (0,1)$ for $(n = 5, k = 3)$ from the sample (0.2081, 0.4043, 0.5642, 0.6822, 0.9082) generated from the uniform $U(0,1)$ distribution

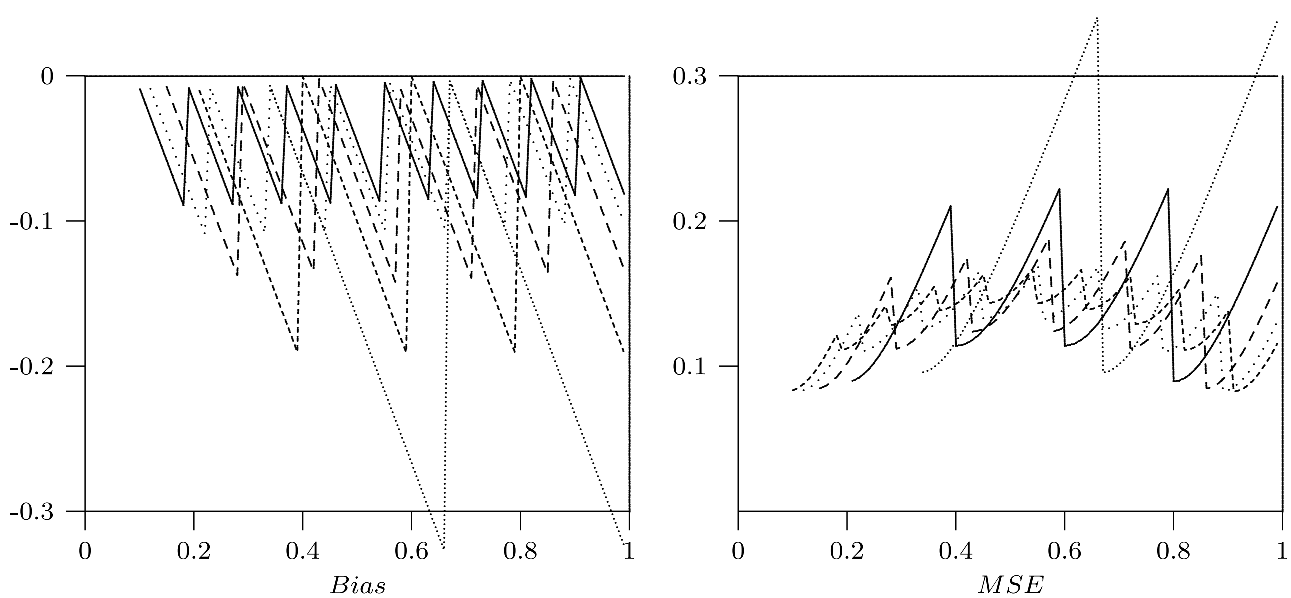
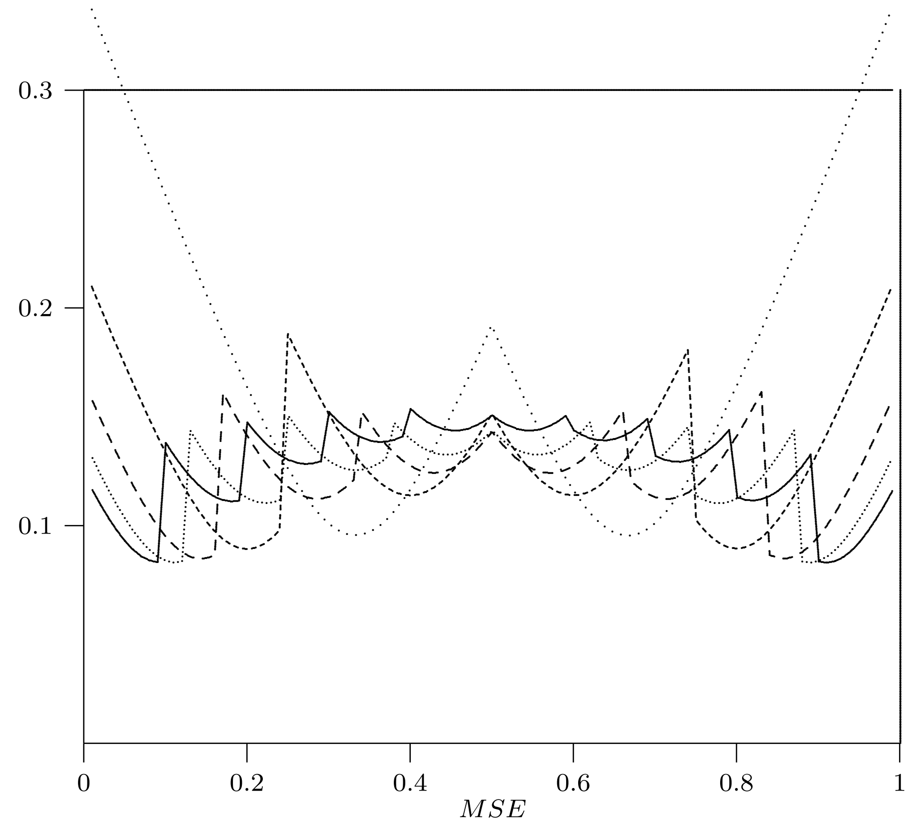


Fig.2.4.3. Original Kaigh-Lachenbruch estimator. Uniform parent distribution. Sample size $n = 10$

$k = 2$ $k = 4$ $k = 6$ $k = 8$ $k = 10$



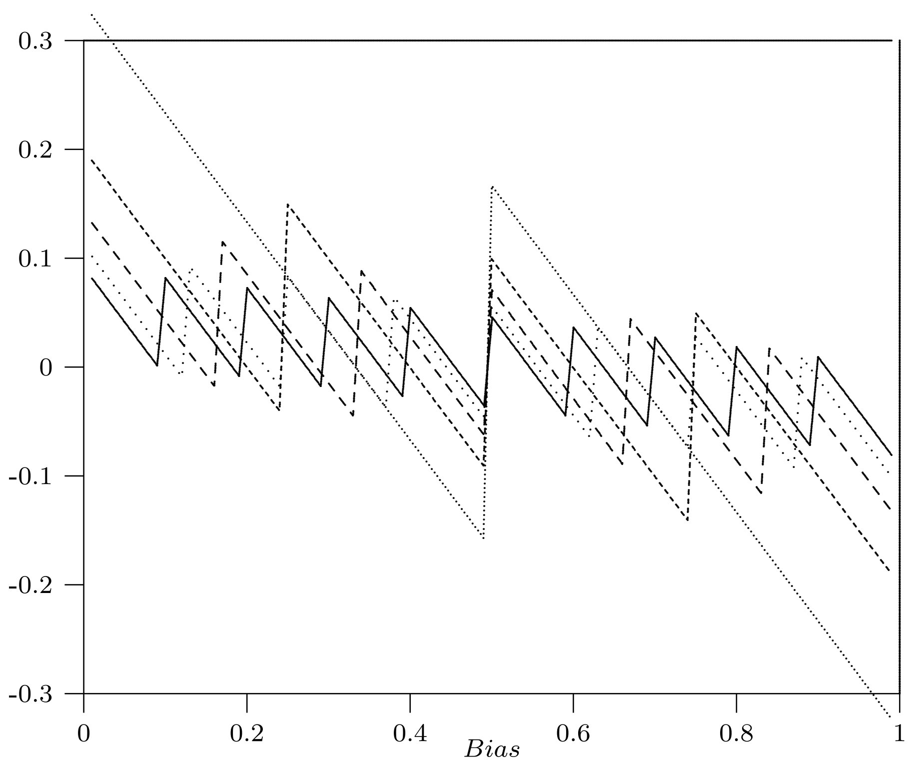


Fig.2.4.4. Modified Kaigh-Lachenbruch estimator. Uniform parent distribution. Sample size $n = 10$

$k = 2$ $k = 4$ $k = 6$ $k = 8$ $k = 10$

A more careful insight into above results suggests to consider the Kaigh-Lachenbruch estimator $\hat{x}_{KL}$ with k depending on q in an "optimal" way, for example in such a way, that, given q, the Mean Square Error of the estimator is minimized. Numerical calculations for uniform $U(0,1)$ parent distribution, for $n = 10$ and $q = 0.01j$, $j = 1, 2, \ldots, 99$, give us the optimal k as in the following table:

q		k
0.01 - 0.09	0.91 - 0.99	10
0.10 - 0.10	0.90 - 0.90	9
0.11 - 0.12	0.88 - 0.89	8
0.13 - 0.13	0.87 - 0.87	7
0.14 - 0.16	0.84 - 0.86	6
0.17 - 0.19	0.81 - 0.83	5
0.20 - 0.22	0.78 - 0.80	4
0.23 - 0.29	0.71 - 0.77	3
0.30 - 0.39	0.61 - 0.70	2
0.40 - 0.42	0.58 - 0.60	4
0.43 -	- 0.57	1

The Mean Square Error of the Kaigh-Lachenbruch estimator with k adapted to q as in the Table above is presented in Fig. 2.4.5; for comparison, the Mean Square Error of the standard estimator $\hat{x}_q$ is also presented.

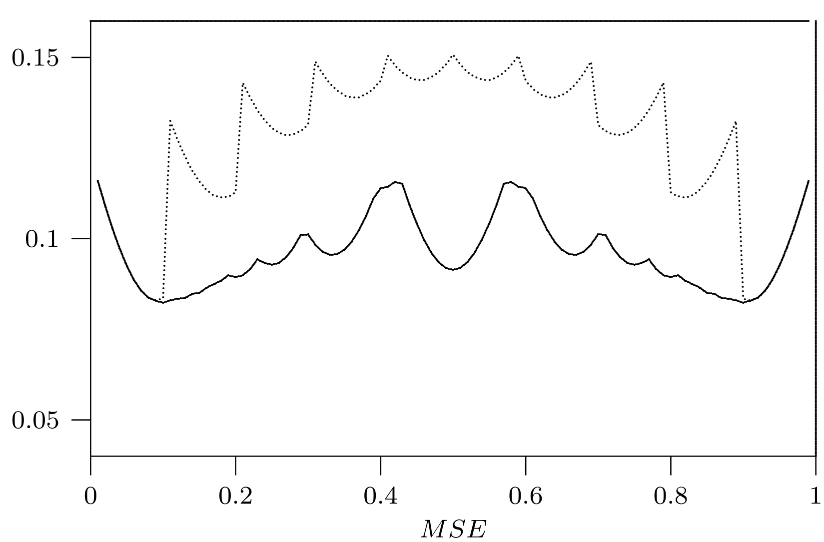


Fig.2.4.5. Modified Kaigh-Lachenbruch estimator with optimal k - solid and the standard estimator $\hat{x}_q$ - dotted. Uniform $U(0,1)$ parent distribution

The solid curve in Fig. 2.4.5 presents the lower envelope of curves of $\sqrt{MSE}$ from Fig. 2.4.4.

Like other nontrivial L-statistics, the estimator is not distribution free in the sense that $F(\hat{x}_{KL})$ depends on the distribution $F \in \mathcal{F}$ of data; if a statistician knows only that the unknown parent distribution F is a member of the family $\mathcal{F}$, then, like in the case of $\hat{x}_{LS}$, $\hat{x}_{HD}$, and $\hat{x}_{Be}$, he is not able to predict the error of his estimation. What is more: a statistician is not able neither to choose an adequate $k = 1, 2, \ldots, n$ nor to optimize the choice by adopting the best $k = k(q)$ depending on the order of the quantile to be estimated. Let us consider the problem in details for the case of the exponential parent distribution.

Fig. 2.4.6 exhibits the MSE of the Kaigh-Lachenbruch estimator for $k = 2, 4, 6, 8, 10$ for the Exponential $F(x) = 1 - e^{-1}$ parent distribution.

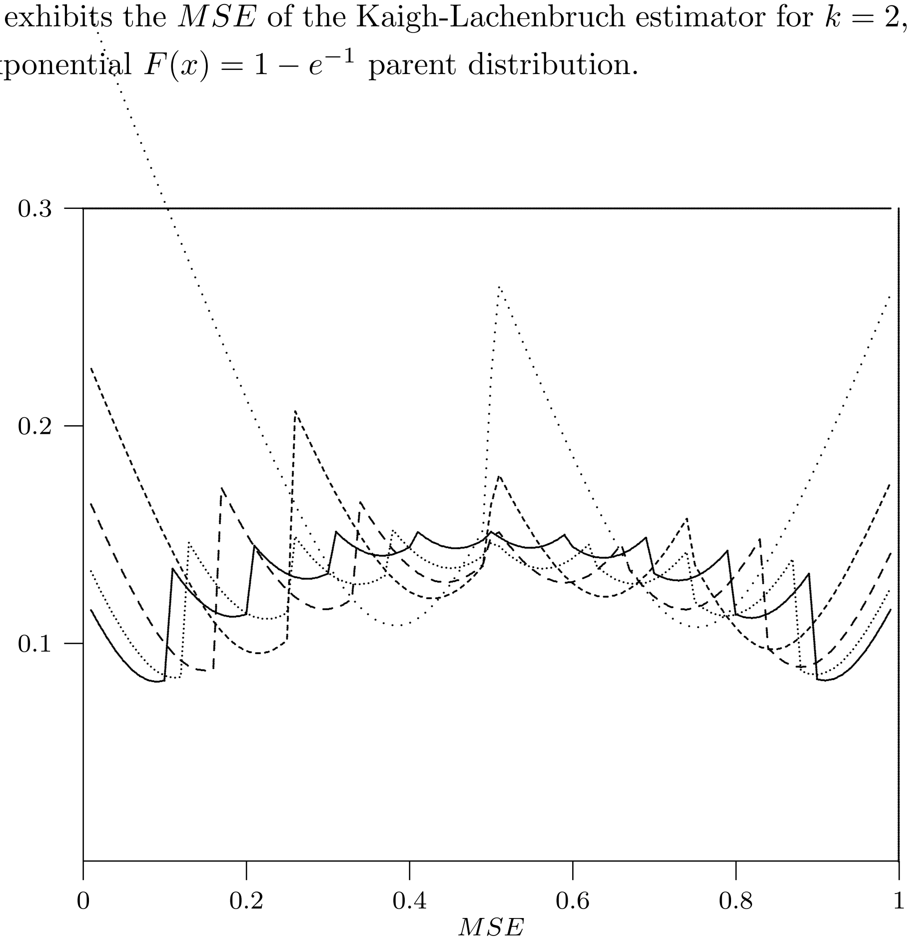


Fig.2.4.6. Modified Kaigh-Lachenbruch estimator. Exponential parent distribution. Sample size $n = 10$

$k = 2$ $k = 4$ $k = 6$ $k = 8$ $k = 10$

Now optimal $k = k(q)$ for $q = 0.01, 0.02, \ldots, 0.99$ are as in the following table

q		k
0.01 - 0.09	0.91 - 0.99	10
0.10 - 0.11	0.89 - 0.90	9
0.12 - 0.12	0.88 - 0.88	8
0.13 - 0.14	0.87 - 0.87	7
0.15 - 0.16	0.85 - 0.86	6
0.17 - 0.20	0.84 - 0.84	5
0.21 - 0.25	0.82 - 0.83	4
0.26 - 0.33	0.78 - 0.81	3
0.34 - 0.43	0.70 - 0.77	2
0.44 - 0.48		4
0.49 - 0.49	0.67 - 0.69	5
0.50 - 0.56		3
0.57 - 0.66		1

Fig. 2.4.7 exhibits the Mean Square Error of the Kaigh-Lachenbruch estimator with $k = k(q)$ optimal for the exponential parent distribution.

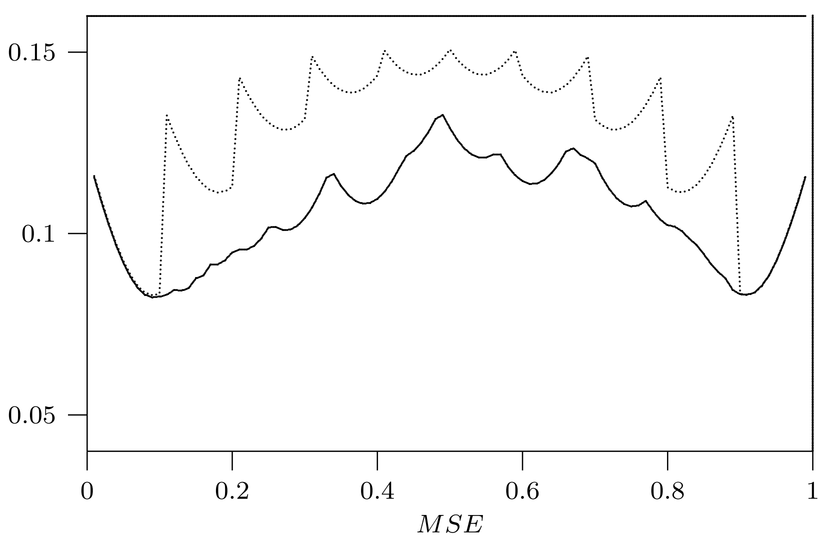


Fig.2.4.7. Modified Kaigh-Lachenbruch estimator with optimal k - solid and the standard estimator $\hat{x}_q$ - dotted. Exponential parent distribution

Like as in Fig. 2.4.5, the optimal Kaigh-Lachenbruch estimator performs better than the standard estimator $\hat{x}_q$, however to construct the Kaigh-Lachenbruch estimator an illegal information has been used, that about the exponential parent distribution. In our nonparametric model $\mathcal{F}$ all that we know about the parent distribution is that the distribution function is continuous and strictly increasing. We may try to base the Kaigh-Lachenbruch estimator on $k = k(q)$ calculated, for example, for the uniform $U(0,1)$ parent distribution and to use such "suboptimal" estimator in the general nonparametric model $\mathcal{F}$. The Mean Square Error of the "suboptimal" estimator if the parent distribution is exponential $F(x) = 1 - e^{-x}$ is presented in Fig. 2.4.8

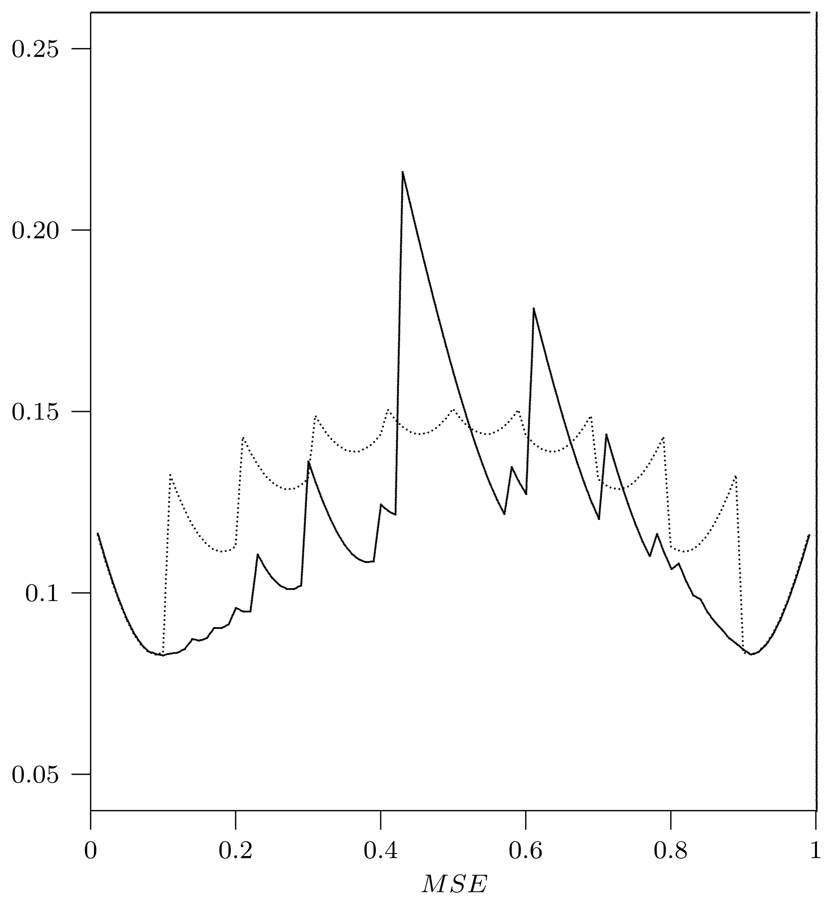


Fig.2.4.8. Suboptimal Kaigh-Lachenbruch estimator for exponential parent distribution - solid and the standard estimator $\hat{x}_q$ - dotted

In Sec.2.5 we shall compare the suboptimal Kaigh-Lachenbruch estimator with other estimators considered above.

2.5. Comparisons of estimators

Facing the problem of estimating the q-th quantile of an unknown distribution of which it is known only that it belongs to the nonparametric family $\mathcal{F}$, we have to decide which estimator to choose.

First of all observe that the class $\mathcal{F}$ of distributions which we consider in our nonparametric model is very large: it consists of all distribution with continuous and strictly increasing distribution functions. The variety of distributions leads of course to a variety of distributions of a given estimator.

If T is a quantile estimator, then it is obvious that to the fact, that the model is very large, there are no possibilities to compare the bias (in the sense of mean value), variance, quadratic risk etc, because such parameters typically do not exist, and if exist then they heavily depend on the unknown distribution $F \in \mathcal{F}$. Let us recall our considerations which followed Theorem 1. To overcome the difficulties we consider $F(T)$ instead of T; all moments of $F(T)$ exist and, for example, the theory of minimum variance unbiased estimators may be fully applied. In Chapter 3 we consider the problem of optimal estimators in a class of estimators T for which properties of $F(T)$ do not depend on the distribution F, so that optimal estimators are uniformly optimal in the class considered. Estimators $\hat{x}_{LS}$, $\hat{x}_{HD}$, $\hat{x}_{Be}$, and $\hat{x}_{KL}$ do not share that property and we confine ourselves to comparing the estimators for some parent distributions. As a criterion of comparison we have chosen the bias and the Mean Square Error of $F(T)$.

All comparisons below are based on simulation results; the exact theoretical distribution is explicitly known only for $\hat{x}_q$ (Section 2.1) and $\hat{x}_{LS}$ (Section 2.2). We decided to present results for small sample of size $n = 10$ and to simulate the behavior of estimators for 10^5 runs. To assess the accuracy of simulated results, Table 2.5.1 presents the exact and simulated values of some quantiles of distributions of the estimator under consideration

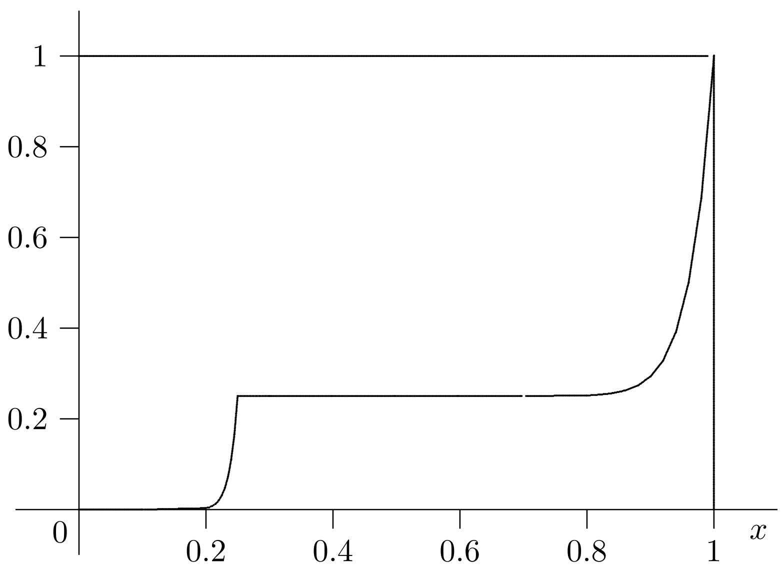
Table 2.5.1. Quantiles

Estimator		0.05	0.25	0.5	0.75	0.95
$\hat{x}_q$	theor	0.15003	0.26085	0.35510	0.45770	0.60622
	simul	0.14956	0.26092	0.35543	0.45785	0.60661
$\hat{x}_{LS}$	theor	0.11492	0.20575	0.28761	0.38171	0.52764
	simul	0.11538	0.20572	0.28819	0.38148	0.52565

The results allow us to consider the simulation results as sufficiently exact.

Fig. 2.5.1-2.5.12 exhibit simulation results for the parent distributions: uniform $U(0,1)$, normal $N(0,1)$, Pareto with distribution function $1-1/x$ and Pareto with distribution function $1-1/\sqrt{x}$, exponential, and a "special" distribution H with distribution function

$$H(x) = \begin{cases} q\left(\dfrac{x}{q}\right)^{\alpha}, & \text{if } 0 < x \le q \\ q + (1-q)\left(\dfrac{x-q}{1-q}\right)^{\alpha}, & \text{if } q < x < 1, \end{cases}$$



Distribution function H

which appeared to be very illustrative for the behavior of estimators (Zieliski 1995). Mean Square Error and Bias of the estimators considered here for the distribution H with parameters $q = 0.25$ and $\alpha = 20$ are presented in Fig. 2.5.11 and 2.5.12.

As a conclusion we obtain that the only estimator stable (robust) with respect to its bias and Mean Square Error under changing the parent distribution is $\hat{x}_q$. Perhaps however a preferable estimator is $\hat{x}_{LS}$ which has smaller bias and Mean Square Error; these parameters do not change very much (Fig. 2.5.13 and Fig. 2.5.14) and even under as exotic parent distribution as the distribution H remain quite acceptable. What is more, both estimators have explicitly given distribution function which may be treated numerically.

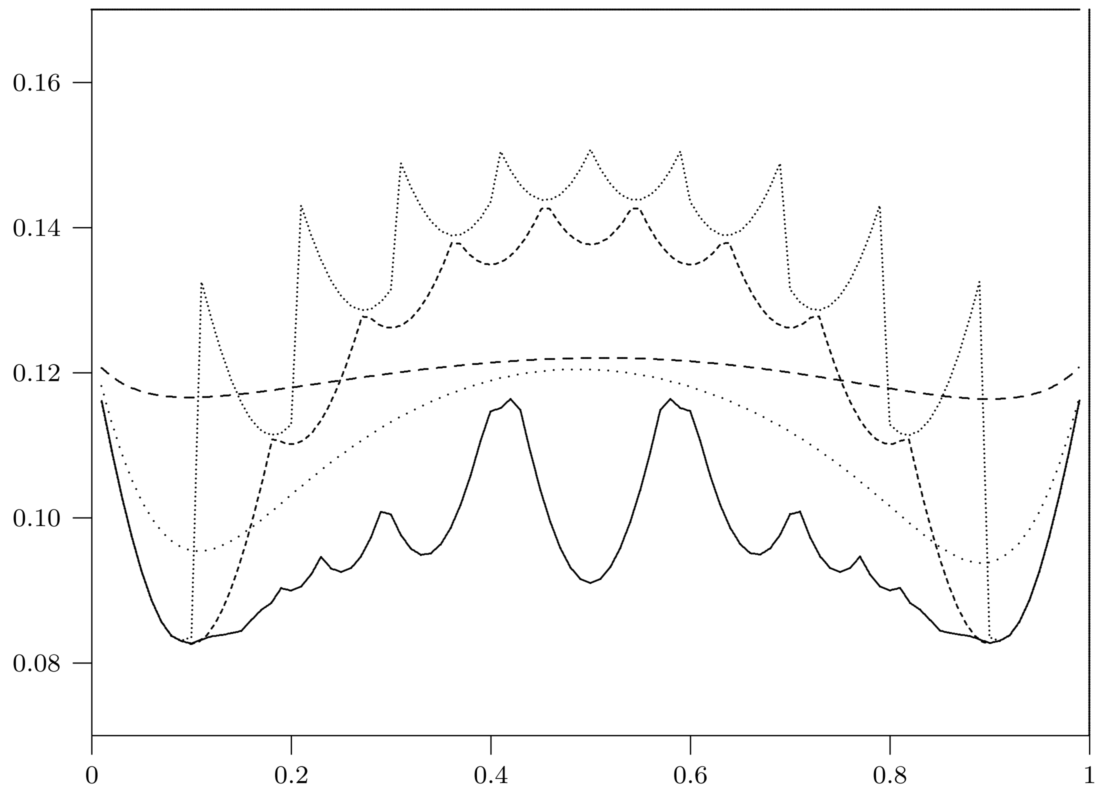


Fig.2.5.1. MSE for the parent distribution $F = U(0,1)$

$\hat{x}_q$ $\hat{x}_{LS}$ $\hat{x}_{Be}$ $\hat{x}_{HD}$ $\hat{x}_{KL}$

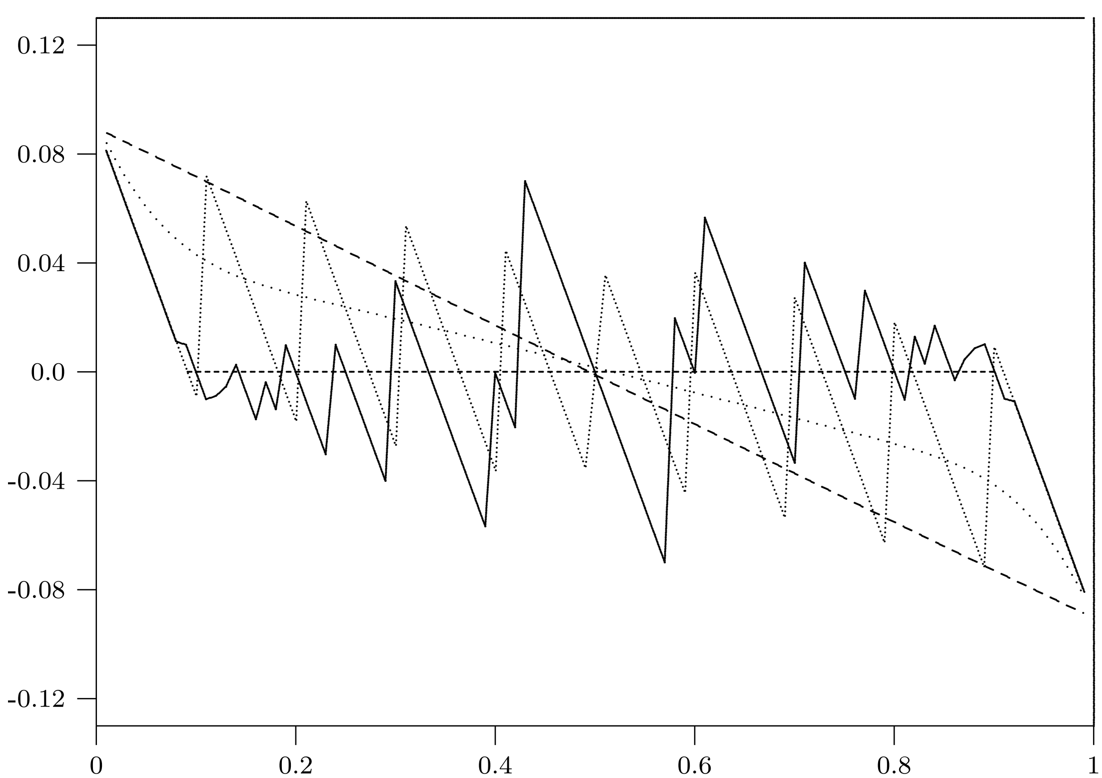


Fig.2.5.2. $Bias$ for the uniform parent distribution $F = U(0,1)$

$\hat{x}_q$ $\hat{x}_{LS}$ $\hat{x}_{Be}$ $\hat{x}_{HD}$ $\hat{x}_{KL}$

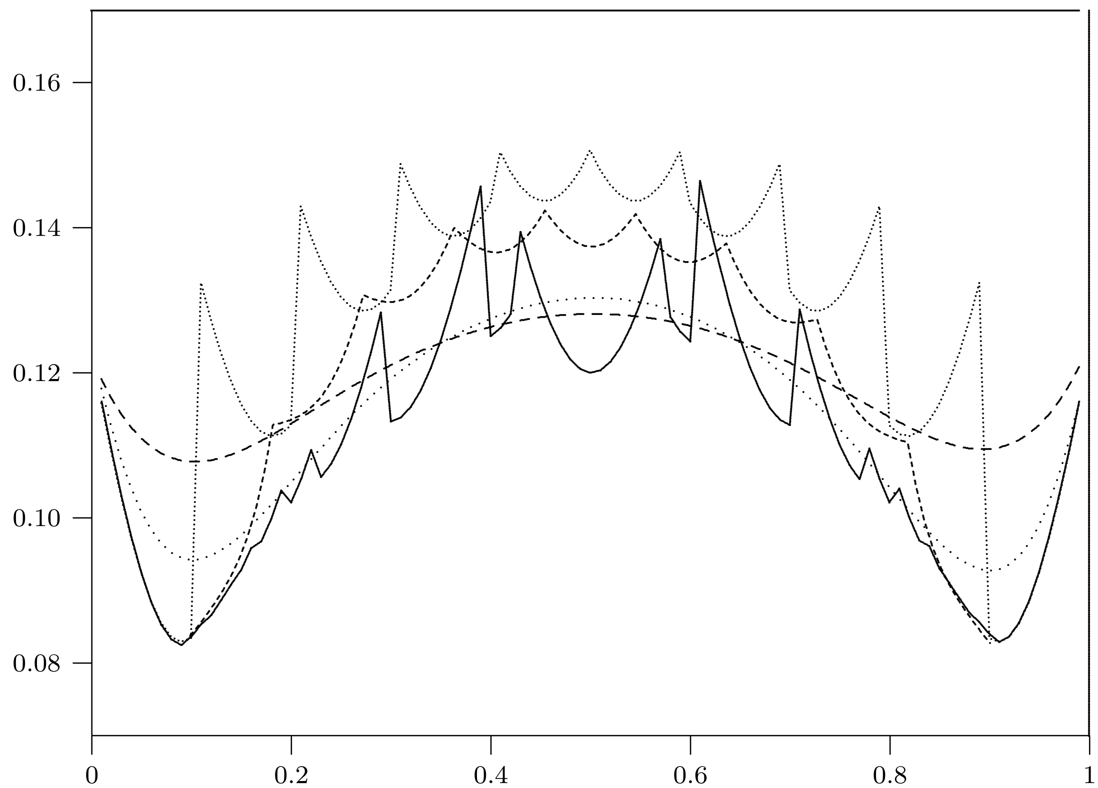


Fig.2.5.3. *MSE* for the normal parent distribution $F = N(0,1)$

$\hat{x}_q$ $\hat{x}_{LS}$ $\hat{x}_{Be}$ $\hat{x}_{HD}$ $\hat{x}_{KL}$

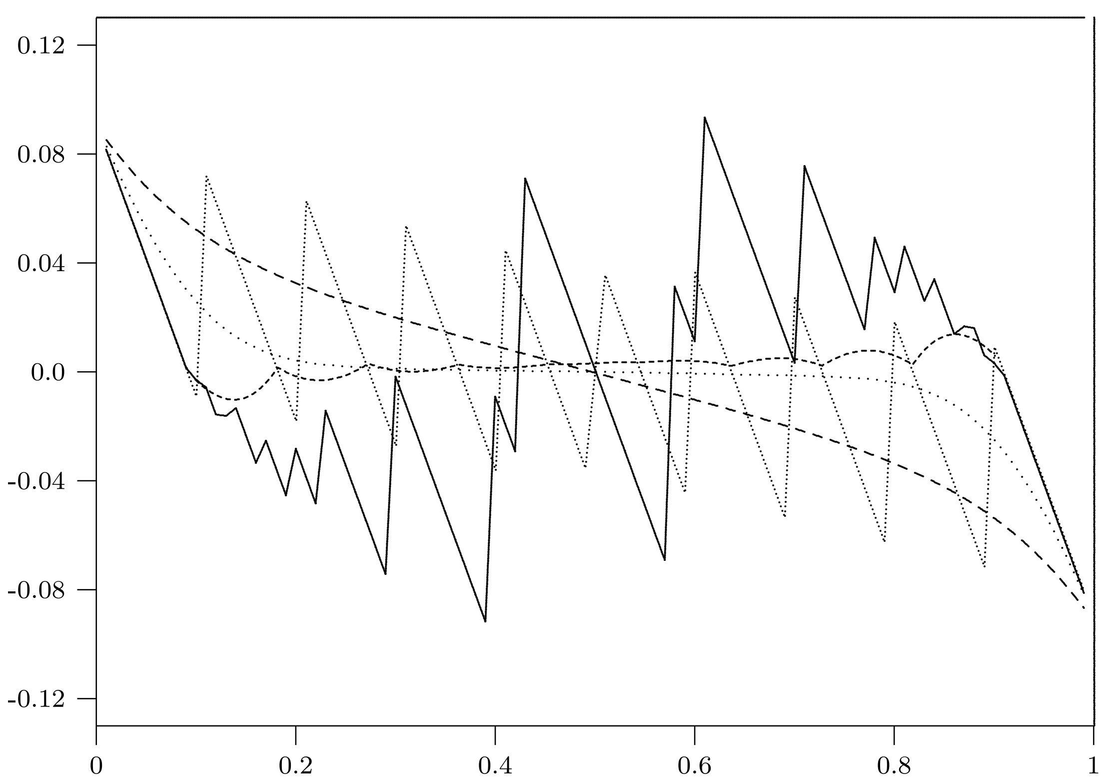


Fig.2.5.4. *Bias* for the normal parent distribution $F = N(0,1)$

$\hat{x}_q$ $\hat{x}_{LS}$ $\hat{x}_{Be}$ $\hat{x}_{HD}$ $\hat{x}_{KL}$

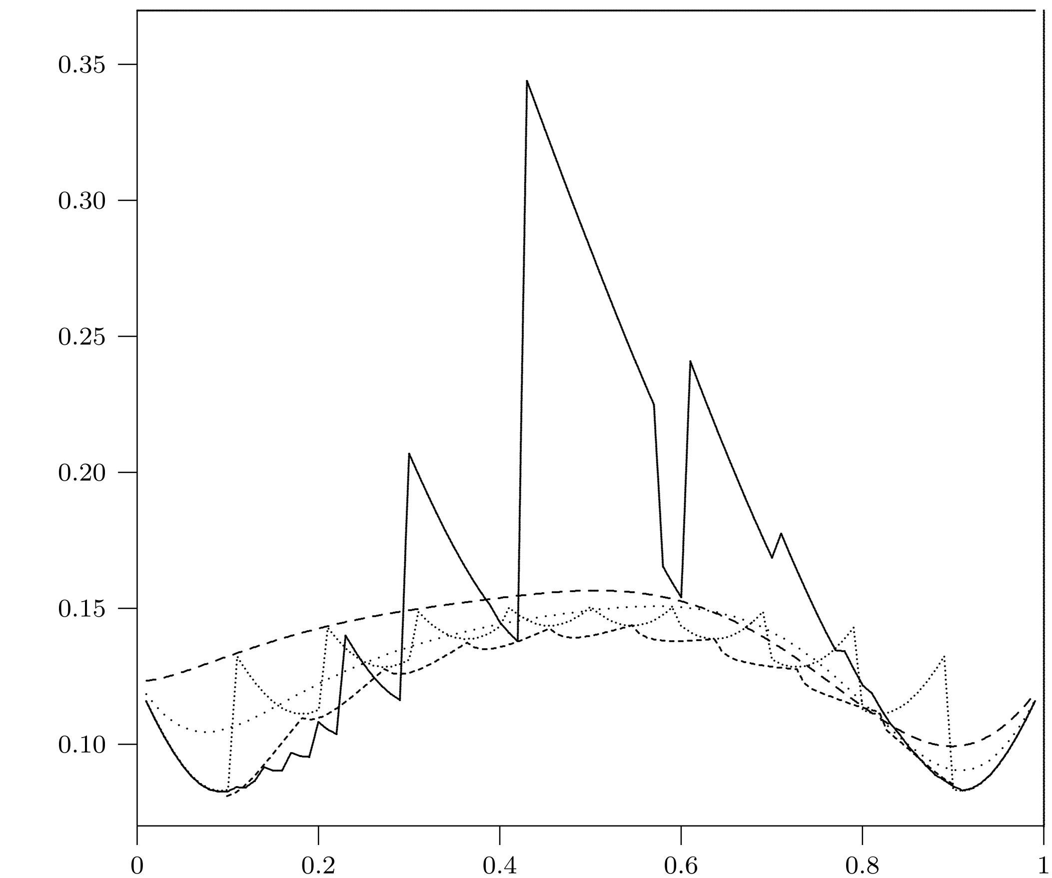


Fig. 2.5.5. MSE for the Pareto(1) parent distribution $F(x) = 1 - 1/x, x > 0$

$\hat{x}_q$ $\hat{x}_{LS}$ $\hat{x}_{Be}$ $\hat{x}_{HD}$ $\hat{x}_{KL}$

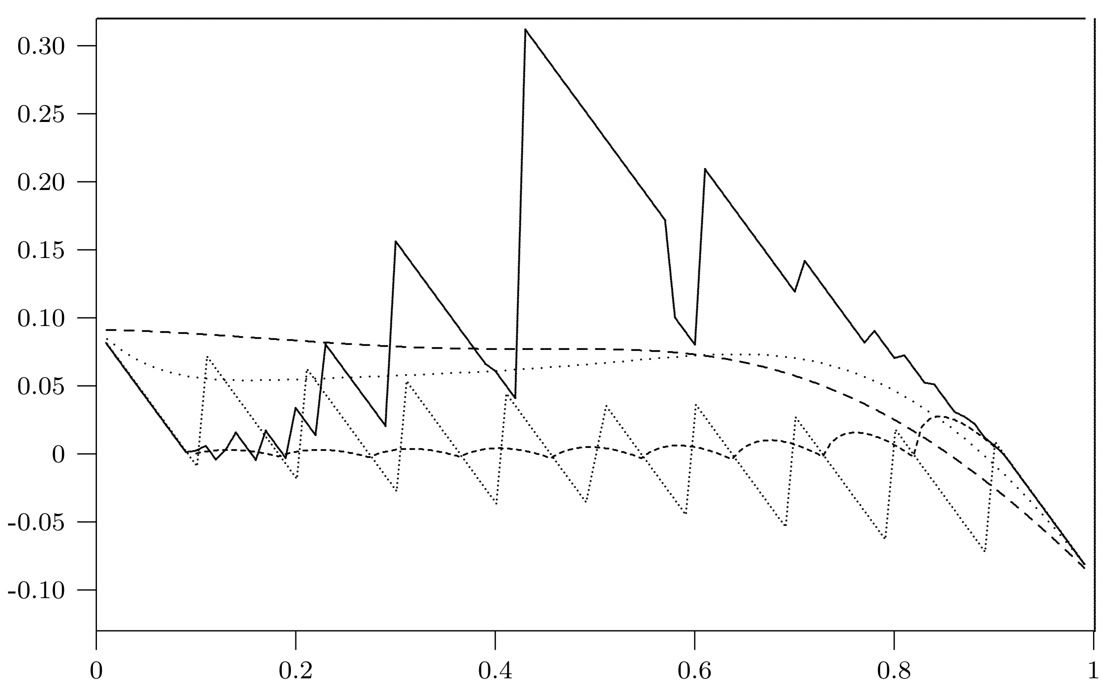


Fig.2.5.6. $Bias$ for the Pareto(1) parent distribution $F(x) = 1 - 1/x, x > 0$

$\hat{x}_q$ $\hat{x}_{LS}$ $\hat{x}_{Be}$ $\hat{x}_{HD}$ $\hat{x}_{KL}$

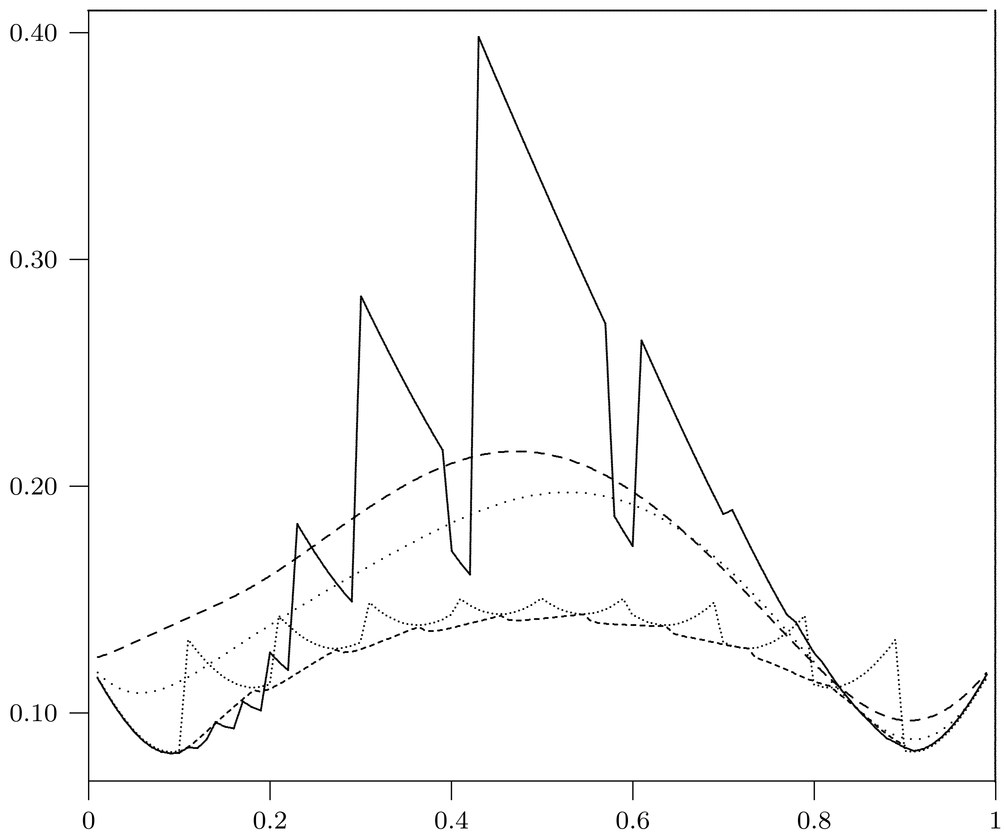


Fig.2.5.7.MSE for the Pareto(0.5) parent distribution $F(x) = 1 - 1/\sqrt{x}, x > 0$

$\hat{x}_q$ $\hat{x}_{LS}$ $\hat{x}_{Be}$ $\hat{x}_{HD}$ $\hat{x}_{KL}$

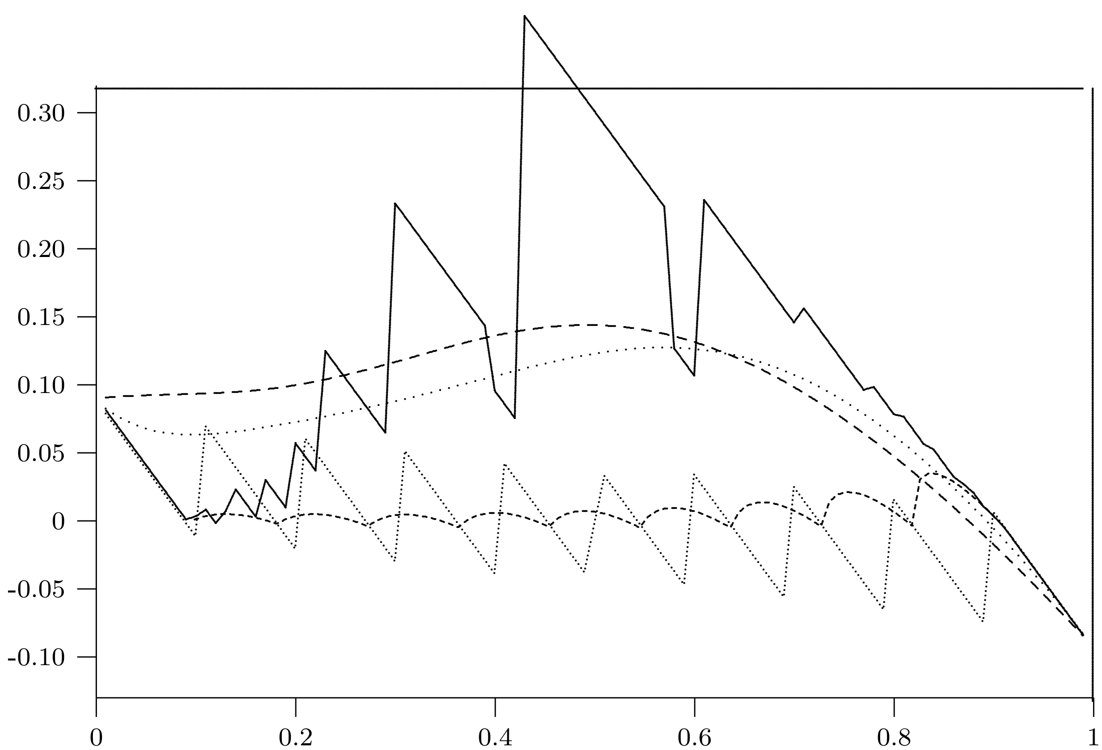


Fig.2.5.8. $Bias$ for the Pareto(0.5) parent distribution $F(x) = 1 - 1/\sqrt{x}, x > 0$

$\hat{x}_q$ $\hat{x}_{LS}$ $\hat{x}_{Be}$ $\hat{x}_{HD}$ $\hat{x}_{KL}$

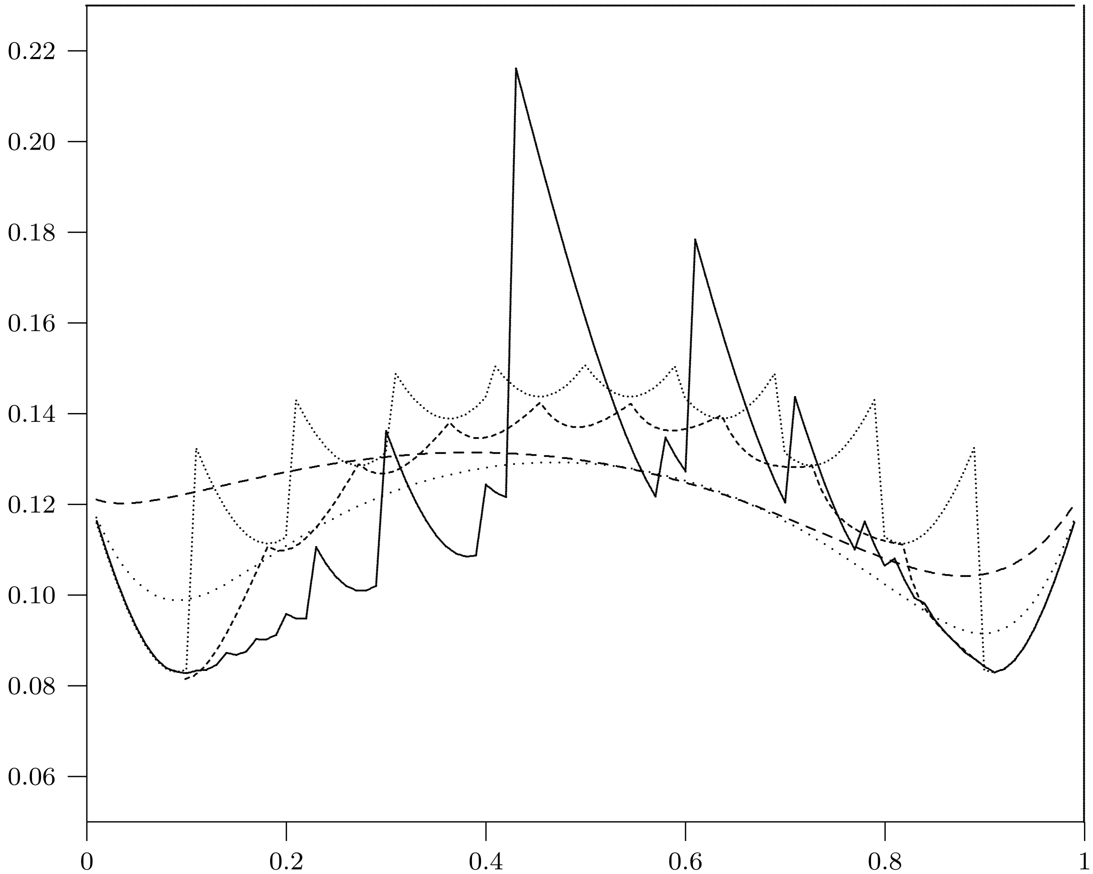


Fig.2.5.9. MSE for the exponential distribution $F(x) = 1 - e^{-x}$

$\hat{x}_q$ $\hat{x}_{LS}$ $\hat{x}_{Be}$ $\hat{x}_{HD}$ $\hat{x}_{KL}$

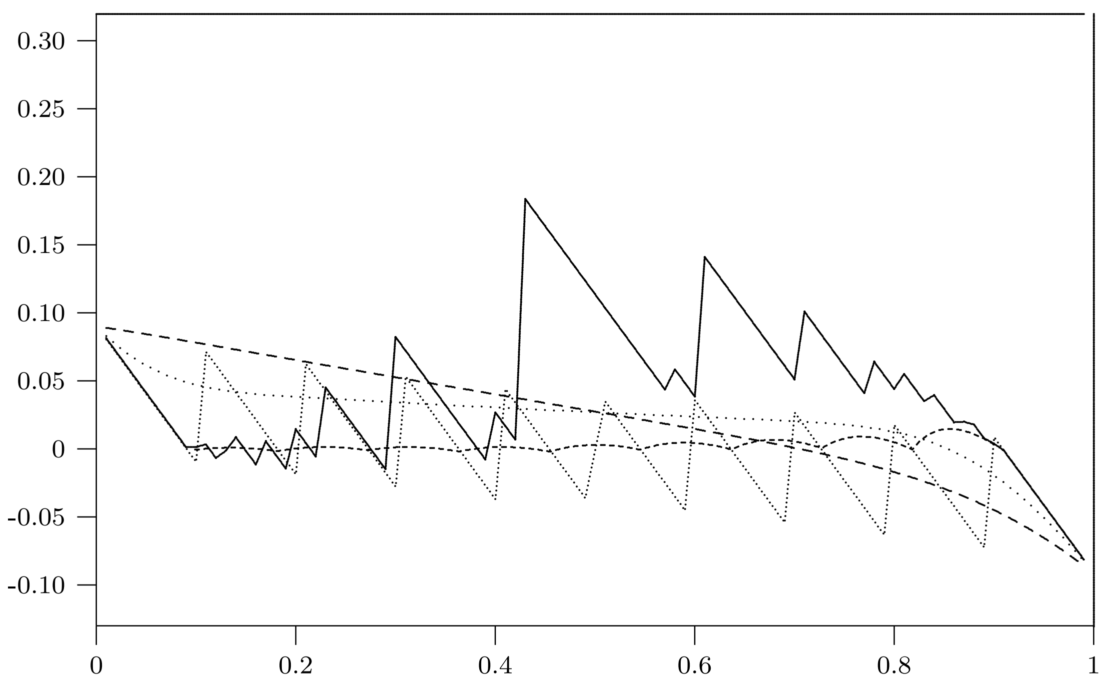


Fig.2.5.10. $Bias$ for the exponential distribution $F(x) = 1 - e^{-x}$

$\hat{x}_q$ $\hat{x}_{LS}$ $\hat{x}_{Be}$ $\hat{x}_{HD}$ $\hat{x}_{KL}$

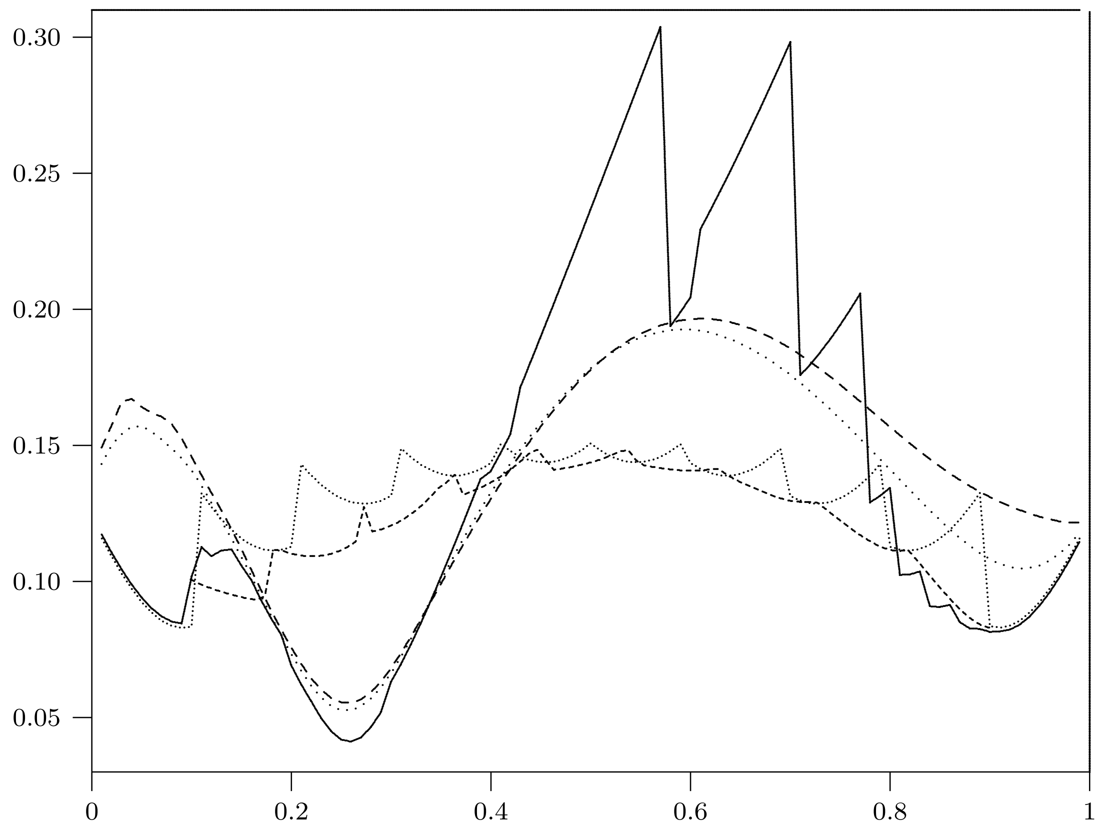


Fig.2.5.11. *MSE* for the distribution *H*

$\hat{x}_q$ $\hat{x}_{LS}$ $\hat{x}_{Be}$ $\hat{x}_{HD}$ $\hat{x}_{KL}$

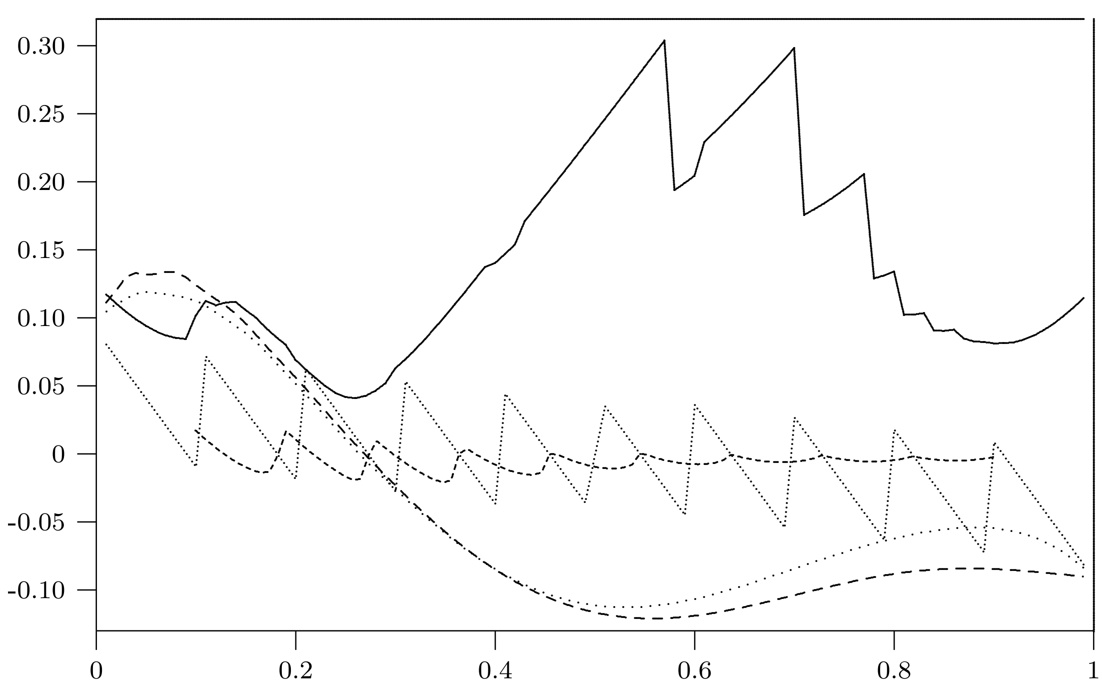


Fig.2.5.12. *Bias* for the distribution *H*

$\hat{x}_q$ $\hat{x}_{LS}$ $\hat{x}_{Be}$ $\hat{x}_{HD}$ $\hat{x}_{KL}$

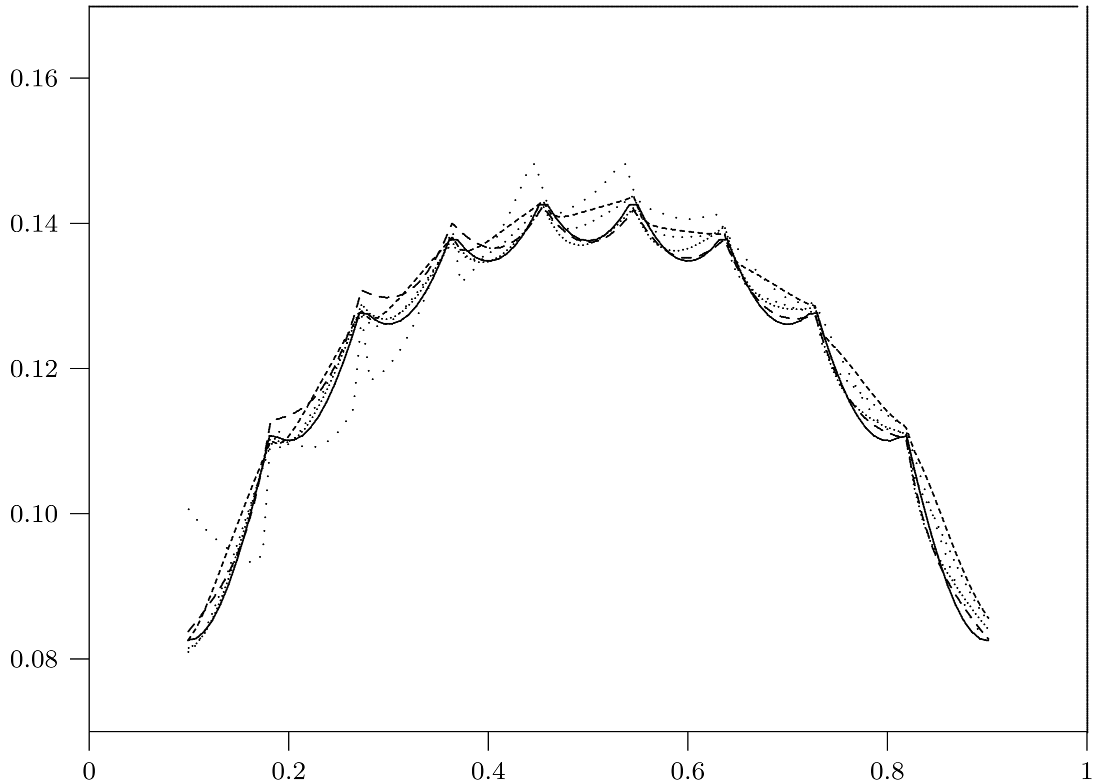


Fig.2.5.13. MSE of $\hat{x}_{LS}$ for different parent distributions

E $Pa(0.5)$ $N(0,1)$ $Pa(1)$ $U(0,1)$ H

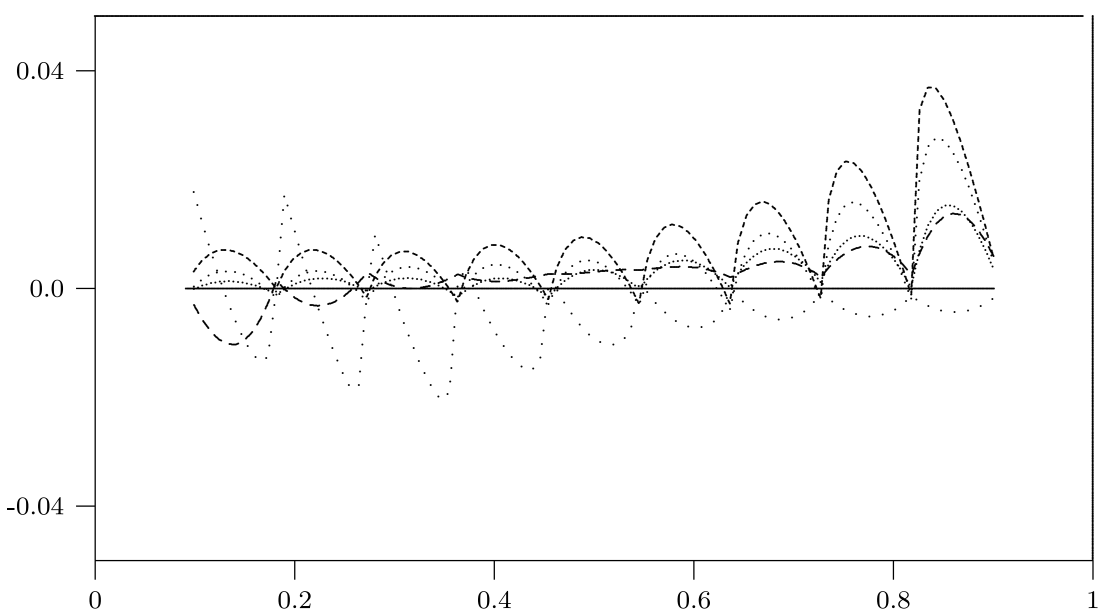


Fig.2.5.14. $Bias$ of $\hat{x}_{LS}$ for different parent distributions

E $Pa(0.5)$ $N(0,1)$ $Pa(1)$ $U(0,1)$ H

3

Optimal estimation

In the classical approach (Chapter 2), a heuristically constructed estimator is a starting point for investigations. Sometimes rationale for a choice of estimator are seemingly very natural: for example sample quantile is taken as an estimator of the population quantile. Not all statisticians agree concerning what is "qth sample quantile": $X_{[nq]:n}$, or perhaps $X_{[(n+1)q]:n}$, or something else. Sometimes rationale are more sophisticated, like celebrated Harell-Davies or Kaigh-Lachenbruch estimators. Generally, an estimator is a result of a heuristic process.

In optimal estimation approach, which is basic for this chapter, an estimator is formally constructed as a solution of a problem of finding a minimum or a maximum of a criterion defined on a given set. "A given set" which we have in mind is the class of estimators under consideration. Typically the criterion generates an ordering in the class of estimators and the problem is to find the best element with respect to this order. Perhaps the most difficult problem arises from the fact that we interested in an ordering such that if an estimator T is better than estimator S for a given distribution $F \in \mathcal{F}$ then T is better than S for all model distributions $F \in \mathcal{F}$ (uniformly in $\mathcal{F}$). The class of distributions $\mathcal{F}$ under consideration is very large (recall: it is the class of all continuous and strictly increasing distribution functions) and such "uniform orderings" can be constructed only in some very specific and rather narrow classes of estimators.

3.1. The class of estimators

First of all, according to what was said in Chapter 1, we consider estimators which are functions of order statistic $X_{1:n}, \ldots, X_{n:n}$. We are going to dramatically further restrict the class of estimators to be considered; "dramatically" means e.g. that nontrivial L-statistics will be excluded. To justify such restriction observe that our nonparametric model $\mathcal{F}$ is is so large that if a random variable X has a distribution $F \in \mathcal{F}$ and $g : R^1 \to R^1$ is a continuous and strictly monotonic transformation of the real line, then the random variable $Y = g(X)$ has a distribution which belongs to $\mathcal{F}$. Denote by $\mathcal{G}$ the class of all such transformations g; every $g \in \mathcal{G}$ leaves the model $\mathcal{F}$ invariant. We are not going to present here the full theory of invariance of statistical models (an elementary and excellent exposition one can find in Lehmann 1983); a general conclusion from the theory is that a natural class of estimators in such models is the class of equivariant estimators: an estimator $T = T(x_1, x_2, \ldots, x_n)$ is said to be equivariant under a transformation g of the sample space if

$$(1) \qquad T(g(x_1), g(x_2), \ldots, g(x_n)) = g\left(T(x_1, x_2, \ldots, x_n)\right), \quad x_1 \leq x_2 \leq \ldots \leq x_n.$$

In what follows we confine ourselves to the class $\mathcal{T}$ of estimators which are equivariant under every $g \in \mathcal{G}$. Property (1) holds for all $g \in \mathcal{G}$ if and only if $T(x_1, x_2, \ldots, x_n) = x_j$ for an arbitrarily fixed j. "Arbitrarily fixed" means that we may consider j as a random variable with values in $\{1, 2, \ldots, n\}$, independent of $X_{1:n}, \ldots, X_{n:n}$. As a consequence, the class $\mathcal{T}$ of estimators which we consider for a given n we identify with the class of all randomized order statistics $X_{J:n}$:

$$(2) \qquad T \in \mathcal{T} \quad \text{iff} \quad T = X_{J:n} \text{ for a random variable } J \text{ on } \{1, 2, \ldots, n\}.$$

A consequence of the equivariance is that if the sample comes from a distribution $F \in \mathcal{F}$, then for every q

$$(3) \qquad P_F\{T \leq x_q(F)\} = P_F\{X_{J:n} \leq x_q(F)\} = P_F\{F(X_{J:n}) \leq q\} = P\{U_{J:n} \leq q\},$$

which does not depend on F; here P without index F denotes the probability under the uniform distribution $U(0,1)$.

By (2) and (3), constructing an optimal estimator in the class $\mathcal{T}$ amounts to constructing the optimal distribution on $\{1, 2, \ldots, n\}$ for the random variable J. By (3) if an estimator is better than any other for a specific parent distribution then it is better for all distributions $F \in \mathcal{F}$, i.e. uniformly in $\mathcal{F}$.

The class $\mathcal{T}$ of estimators is rather narrow which is a consequence of the fact that the class $\mathcal{F}$ of distributions is very large. If we wish to include L-estimators into the class of estimators that we consider, we must restrict the nonparametric class of distribution which we take into account (see Fig. 2.3.2 for Harrell-Davis estimator or Fig. 2.4.4 for Kaigh-Lachenbruch estimator); otherwise we will be not able to say anything concerning errors of our estimation. Some global restrictions, like unimodality (which excludes Beta distributions) or distributions with densities f which satisfy the condition $f(x) \geq \eta > 0$ with a known η (which excludes normal distribution and most distributions of practical interest) seems to be hardly tractable. Local restrictions, for example $f(x_q) \geq \eta > 0$ at the quantile of interest x_q, need some additional knowledge which is rather unattainable for the statistician. To obtain practical conclusions for the very large nonparametric family $\mathcal{F}$ of distribution in our model we confine ourselves to the above class $\mathcal{T}$ of estimators. The class seems to be the largest class of estimators that can be reasonably treated in the nonparametric model $\mathcal{F}$.

3.2. Criteria

If $T \in \mathcal{T}$ is an estimator of q-th quantile $x_q(F)$ of a distribution $F \in \mathcal{F}$ then obviously the distribution of T should be located near $x_q(F)$ and should be concentrated around $x_q(F)$. A full discussion on measuring location and spread in nonparametric models is presented in series of papers by Bickel and Lehmann 1975-1979.

As a measure of location we choose a median and as the optimal estimator we choose the median–unbiased estimator which is most concentrated around the median in the class of all median–unbiased estimators. A case of particular interest is the interquartile range of a distribution F defined as $F^{-1}(3/4) - F^{-1}(1/4)$. If a median–unbiased estimator T_1 is more concentrated than a median–unbiased estimator T_2 then its interquartile range is of course smaller than that of T_2. The most concentrated median–unbiased estimator is constructed in Sec. 3.3.1.

Other optimality criteria (Uniformly Minimum Variance estimators in Sec. 3.3.2, Uniformly Minimum Absolute Deviation estimators in Sec. 3.3.3, and Optimal estimators in the sense of Pitman's Measure of Closeness in Sec. 3.3.4) are considered in terms of $F(T)$; for T they do not have any sense (cf introduction to Chapter 2 above).

3.3. Optimal estimators

3.3.1. The most concentrated median–unbiased estimator

Choose a level $\alpha \in (0,1)$. An estimator T^* is said to be the most concentrated estimator of the q-th quantile $x_q(F)$ at a level α in a given class $\mathcal{T}$ of estimators if

$$1^0. \quad P_F\{T^* \le x_q(F)\} = \alpha$$

$$(4) \qquad 2^0. \quad P_F\{T^* \le t\} \le P_F\{T \le t\} \quad \text{for } t \le x_q(F) \quad \text{and}$$
$$P_F\{T^* \le t\} \ge P_F\{T \le t\} \quad \text{for } t \ge x_q(F) \quad \text{for all } T \in \mathcal{T}$$

(see Fig. 1). Estimators concentrated at the level $\alpha = 0.5$ are median unbiased. Typically we are interested in constructing the most concentrated median unbiased estimator. Observe that without the condition 1^0 above every constant is a most concentrated estimator; the condition 1^0 plays here similar role as (mean-)unbiasedness of the estimator with minimum variance.

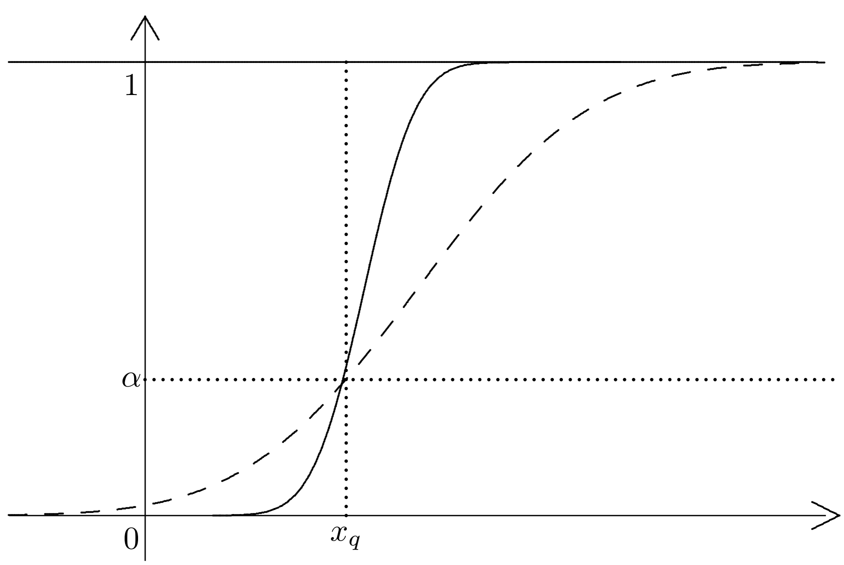


Fig.3.3.1. Estimator of x_q with solid cdf
is more concentrated at the level α than that with dashed cdf

If estimators T and S are distributed symmetrically around an estimand θ, which is rather exceptional when estimating quantiles, there is an alternative way of arriving at the criterion. Following Bickel and Lehmann (1976) one can consider T as more dispersed about θ than S if $|T-\theta|$ is stochastically larger $|S-\theta|$. Then the most concentrated median unbiased estimator of the quantile x_q is that which is the less dispersed about x_q in the class of estimators under consideration. Full analogy to minimum variance unbiased estimators is obvious.

In what follows, following Zieliński (1988), we construct effectively the most concentrated median-unbiased estimator of the qth quantile $x_q(F)$ of an unknown distribution $F \in \mathcal{F}$.

Denote

$$\pi_k(q) = P_F\{X_{k:n} \leq x_q(F)\} = \sum_{j=k}^{n} \binom{n}{j} q^j (1-q)^{n-j}$$

and observe that a median unbiased estimator exists iff $\pi_1(q) \geq 1/2 \geq \pi_n(q)$. Given q, the smallest $n = n(q)$ for which a median-unbiased estimator exists is given by the formula $n(q) = \min\{n : n \geq -\log 2/\log(\max\{q, 1-q\})\}$. On the other hand, given n, the order q of a quantile to be estimated should satisfy $1-(1/2)^{1/n} < q < (1/2)^{1/n}$.

Given $q \in (0,1)$, we define the class $\mathcal{T}(q) \subset \mathcal{T}$ of all median-unbiased estimators of the qth quantile:

$$\mathcal{T}(q) = \left\{ X_{J:n} : \ P\{J=j\} = \lambda_j, \ \sum_{j=1}^{n} \lambda_j \pi_j(q) = \frac{1}{2} \right\}$$

Given q and $n \geq n(q)$, let k be an integer such that

$$\pi_k(q) > \frac{1}{2} > \pi_{k+1}(q).$$

The case that $\pi_k(q) = \frac{1}{2}$ for some k will be discussed later on.

Let

$$\lambda_k^* = \frac{\frac{1}{2} - \pi_{k+1}(q)}{\pi_k(q) - \pi_{k+1}(q)}, \quad \lambda_{k+1}^* = 1 - \lambda_k^*, \quad \lambda_i^* = 0 \text{ for } i \notin \{k, k+1\}$$

and let $T^* = X_{J^*:n}$, where J^* has the distribution $(\lambda_1^*, \lambda_2^*, \ldots, \lambda_n^*)$.

Lemma 1. For all $T \in \mathcal{T}(q)$, all $F \in \mathcal{F}$, and $\eta > 0$ such that $q + \eta < 1$,

$$P_F\{T^* \leq x_{q+\eta}(F)\} \geq P_F\{T \leq x_{q+\eta}(F)\}.$$

Proof. Consider the following linear programming problem: given $\eta > 0$, find $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_n$ which maximizes

$$P_F\{T \leq x_{q+\eta}(F)\} = \sum_{j=1}^{n} \lambda_j \pi_j(q+\eta) \tag{5}$$

under the restrictions

$$\begin{cases} \sum_{j=1}^{n} \lambda_j \pi_j(q) = \frac{1}{2}, \\ \sum_{j}^{n} \lambda_j = 1, \\ \lambda_1, \lambda_2, \ldots, \lambda_n \geq 0. \end{cases} \tag{6}$$

All vertices of the polyhedron (6) are of the form: all but two coordinates $\lambda_i, i = 1, 2, \ldots, n$, equal zero, and those two, say λ_r and λ_s, $r < s$, satisfy

$$\begin{cases} \lambda_r \pi_r(q) + \lambda_s \pi_x(q) = \frac{1}{2}, \\ \lambda_r + \lambda_s = 1. \end{cases}$$

Due to the sequence of inequalities

$$\pi_1(q) > \ldots > \pi_k(q) > \frac{1}{2} > \pi_{k+1}(q) > \ldots > \pi_n(q) > 0$$

and due to the first of the conditions (6), we have $r \leq k$ and $s \geq k+1$. Given r and s, the criterion (5) takes on the form

$$W_{r,s} = \frac{\frac{1}{2} - \pi_s(q)}{\pi_r(q) - \pi_s(q)} \pi_r(q+\eta) + \left(1 - \frac{\frac{1}{2} - \pi_s(q)}{\pi_r(q) - \pi_s(q)}\right) \pi_s(q+\eta).$$

Given s, $W_{r,s}$ can be written as

$$\left(\frac{1}{2} - \pi_s(q)\right) \cdot \frac{\pi_r(q+\eta) - \pi_s(q+\eta)}{\pi_r(q) - \pi_s(q)} + \pi_s(q+\eta)$$

so that maximizing $W_{r,s}$ with respect to $r \le k$ with a fixed $s \ge k+1$ amounts to maximizing

$$V_s(r) = \frac{\pi_r(q+\eta) - \pi_s(q+\eta)}{\pi_r(q) - \pi_s(q)}.$$

By the definition of $\pi_r(q)$ we obtain

$$V_s(r) = \frac{\alpha_r + \alpha_{r+1} + \ldots + \alpha_{s-1}}{\beta_r + \beta_{r+1} + \ldots + \beta_{s-1}},$$

where $\alpha_j = \binom{n}{j}(q+\eta)^j(1-q-\eta)^{n-j}$ and $\beta_j = \binom{n}{j}q^j(1-q)^{n-j}$. Observe that

$$\frac{\alpha_j}{\beta_j} < \frac{\alpha_{j+1}}{\beta_{j+1}}, \quad j = 1, 2, \ldots, n-1$$

and hence by Lemma 3 below

$$V_s(r) < V_s(k), \quad r < k,$$

uniformly in $\eta > 0$ such that $q + \eta < 1$.

Now, given $r = k$, $W_{r,s}$ may be written as

$$\pi_k(q+\eta) - \left(\pi_k(q) - \frac{1}{2}\right) \frac{\pi_k(q+\eta) - \pi_s(q+\eta)}{\pi_k(q) - \pi_s(q)},$$

so that maximizing $W_{k,s}$ with respect to $s \ge k+1$ with a fixed r amounts to minimizing

$$U_k(s) = \frac{\pi_k(q+\eta) - \pi_s(q+\eta)}{\pi_k(q) - \pi_s(q)}.$$

Similarly as in the first part of the proof,

$$U_k(s) = \frac{\alpha_k + \alpha_{k+1} + \ldots + \alpha_{s-1}}{\beta_k + \beta_{k+1} + \ldots + \beta_{s-1}},$$

and by Lemma 3 below and arguments as above

$$U_k(k+1) < U_k(s), \quad s > k+1,$$

uniformly in $\eta > 0$ such that $q+\eta < 1$. It follows that $W_{r,s}$ is maximized by $r = k$ and $s = k+1$, which ends the proof of Lemma 1.

Lemma 2. For all $T \in \mathcal{T}(q)$, all $F \in \mathcal{F}$, and $\xi > 0$ such that $q - \xi > 0$

$$P_F\{T^* \le x_{q-\xi}(F)\} \le P_F\{T \le x_{q-\xi}(F)\}.$$

Proof is similar to that of Lemma 1 except that now, for $\alpha_j = \binom{n}{j}(q-\xi)^j(1-q-\xi)^{n-j}$ and $\beta_j = \binom{n}{j}q^j(1-q)^{n-j}$, we have $\alpha_j/\beta_j > \alpha_{j+1}/\beta_{j+1}$. □

Suppose now that $\pi_k(q) = \dfrac{1}{2}$ for some $k = 1, 2, \ldots, n-1$. Take a positive h small enough to have

$$\pi_k(q+h) > \frac{1}{2} > \pi_{k+1}(q+h).$$

Given h, by Lemmas 1 and 2 the optimal estimator of the $(q+h)$-quantile is that with

$$\lambda_k^*(h) = \frac{\frac{1}{2} - \pi_{k+1}(q+h)}{\pi_k(q+h) - \pi_{k+1}(q+h)}, \quad \lambda_{k+1}^*(h) = 1 - \lambda_k^*(h)$$

All quantities under consideration are continuous in q and hence, by passing to limits as $h \to 0$, the solution is given by $\lambda_k^* = 1$. To establish the result for $k = n$, it is enough to take a negative h and consider the inequality $\pi_{n-1}(q+h) > \dfrac{1}{2} > \pi_n(q+h)$.

As a simple consequence we obtain

Theorem 1. *In the class $\mathcal{T}(q)$ of median-unbiased estimators of the qth quantile, T^* is the most concentrated one.* □

For the sake of uniformity of notation, the median-unbiased most concentrated estimator T^* will be denoted as $\hat{x}_{RZ}$ or, if q is specified, as $\hat{x}_{RZ}(q)$.

In the above proofs, the following Lemma 3 was used.

Lemma 3. Let $\alpha_1, \alpha_2, \ldots, \alpha_m, \beta_1, \beta_2, \ldots, \beta_m$ be positive numbers. If

$$\frac{\alpha_1}{\beta_1} < \frac{\alpha_2}{\beta_2} < \ldots < \frac{\alpha_m}{\beta_m}$$

then for $i < m$ and $j > 1$

$$\frac{\alpha_1 + \alpha_2 \ldots \alpha_i}{\beta_1 + \beta_2 + \ldots + \beta_i} < \frac{\alpha_1 + \alpha_2 \ldots \alpha_m}{\beta_1 + \beta_2 + \ldots + \beta_m} < \frac{\alpha_j + \alpha_{j+1} \ldots \alpha_m}{\beta_j + \beta_{j+1} + \ldots + \beta_m}$$

Proof of the Lemma is elementary. □

If G is the distribution function of an estimator T, the interquartile range $G^{-1}(3/4) - G^{-1}(1/4)$ of the distribution G will be called the interquartile range of the estimator T. As a simple consequence of Theorem 1 we have the following Corollary.

Corollary 1 . *The interquartile range of the estimator $\hat{x}_{RZ}$ is smaller than that of any other median-unbiased estimator $T \in \mathcal{T}(q)$.* □

The property stated in Corollary 1 gives us a full nonparametric analogue of the idea if uniformly minimum variance unbiased estimators in classical (parametric) statistics. It appears that the estimator $\hat{x}_{RZ}$ uniformly minimizes Mean Absolute Deviation - see Sec. 3.3.3. A comment on optimality of $\hat{x}_{RZ}$ in a context of nonparametric testing $H : F^{-1}(q) \leq u$ against $K : F^{-1}(q) > u$ for a fixed u, one can find in Reiss (1989).

3.3.2. Uniformly Minimum Variance Unbiased Estimator.

If $\hat{\theta}$ is an estimator of a real parameter θ, the estimator $\hat{\theta}$ is said to be unbiased if $E_\theta \hat{\theta} = \theta$ for all θ and it is said to be uniformly minimum variance unbiased (UMVU) if $Var_\theta \hat{\theta} \leq Var_\theta \tilde{\theta}$ for all unbiased estimators $\tilde{\theta}$.

If T is an estimator of the qth quantile, in our model with the family $\mathcal{F}$ of distributions F the expectation $E_F T$ of the estimator, and in consequence the variance, may not exist and the UMVU criterion as above makes no sense. According to what we said above, we modify the criterion as follows: an estimator T of the qth quantile is said to be unbiased if $E_F F(T) = q$ for all $F \in \mathcal{F}$, and it is said to be uniformly minimum variance unbiased if $Var_F F(T) \leq Var_F F(\tilde{x}_q)$ for all $F \in \mathcal{F}$ and all unbiased estimators $\tilde{x}_q$. Rationale behind the modified criterion is that if T is to be a good estimator of the qth quantile $x_q = x_q(F)$ and F is a continuous and strictly increasing function, then $F(T)$ should be a good estimator of $F(x_q) = q$.

Estimators $T \in \mathcal{T}$ from a sample of size n are of the form $T = X_{J:n}$, where J is a random integer with a distribution $P\{J = j\} = \lambda_j, j = 1, 2, \ldots, n$. If the sample $X_1, \ldots, X_n$ comes from a distribution $F \in \mathcal{F}$ then $F(X_{J:n}) = U_{J:n}$, where $U_1, \ldots, U_n$ is a sample from the uniform distribution $U(0, 1)$. Hence

$$E_F F(X_{J:n}) = \sum_{j=1}^{n} \lambda_j E U_{j:n} = \frac{1}{n+1} \sum_{j=1}^{n} j\lambda_j$$

and

$$Var_F F(X_{J:n}) = \frac{1}{(n+1)(n+2)} \sum_{j=1}^{n} j(j+1)\lambda_j - E_F^2 F(X_{J:n})$$

Construction of the minimum variance unbiased estimator of the qth quantile, reduces to minimizing $Var_F F(X_{J:n})$ under the condition $E_F F(X_{J:n}) = q$, which amounts to solving the following linear programming problem: find $\lambda_1, \lambda_2 \ldots, \lambda_n$ which minimize

$$\sum_{j=1}^{n} j(j+1)\lambda_j$$

under the restrictions

$$\begin{cases} \frac{1}{n+1} \sum_{j=1}^{n} j\lambda_j = q \\ \sum_{j=1}^{n} \lambda_j = 1 \\ \lambda_1, \lambda_2 \ldots, \lambda_n \geq 0 \end{cases}$$

Due to the fact that

$$\frac{1}{n+1} \leq \frac{1}{n+1} \sum_{j=1}^{n} j\lambda_j \leq \frac{n}{n+1},$$

unbiased estimators of the q-th quantile exist only if $q \in \left[\frac{1}{n+1}, \frac{n}{n+1}\right]$. If that is the case, routine technique for linear programming problems (or the method applied in Sec. 3.3.1) gives us

$$\begin{aligned} \lambda_k &= k + 1 - (n+1)q, \\ \lambda_{k+1} &= (n+1)q - k, \\ \lambda_j &= 0 \ \text{ if } \ j \neq k \text{ and } j \neq k+1 \end{aligned}$$

where $k = [(n+1)q]$.

The estimator has been introduced by Uhlmann (1963); we shall denote it by $\hat{x}_{WU}$ or $\hat{x}_{WU}(q)$ if needed. As a consequence we obtain the following theorem

Theorem 2. *In the class of all unbiased estimators T of the q-th quantile, i.e. estimators such that $E_F F(T) = q$ for all $F \in \mathcal{F}$, the randomized estimator*

$$\hat{x}_{WU}(q) = \begin{cases} k, & \text{with probability } \lambda_k = k + 1 - (n+1)q, \\ k+1, & \text{with probability } 1 - \lambda_k, \end{cases}$$

has the uniformly smallest variance. □

The variance $V_n(q) = Var_F F(\hat{x}_{WU})$ of the optimal estimator for the sample of size n is given by the formula

$$V_n(q) = \frac{([(n+1)q] + 1)\big(2(n+1)q - [(n+1)q]\big)}{(n+1)(n+2)} - q^2.$$

Another formula for the variance of the minimum variance unbiased estimator is

$$V_n(q) = \frac{((n+1)q - [(n+1)q])\,([(n+1)q] + 1 - (n+1)q)}{(n+1)(n+2)} + \frac{q(1-q)}{n+2};$$

the formula has been given by Uhlmann (1963). As one could expect, the variance is a symmetric function of q in the sense that $V_n(q) = V_n(1-q)$ and if $k = [(n+1)q]$ is optimal for q then $n - k + 1$ is optimal for $1 - q$. Fig. 3.3.2 exhibits the variance $V_n(q)$ as a function of $q \in [1/(n+1), n/(n+1)]$ for some values of the sample size n.

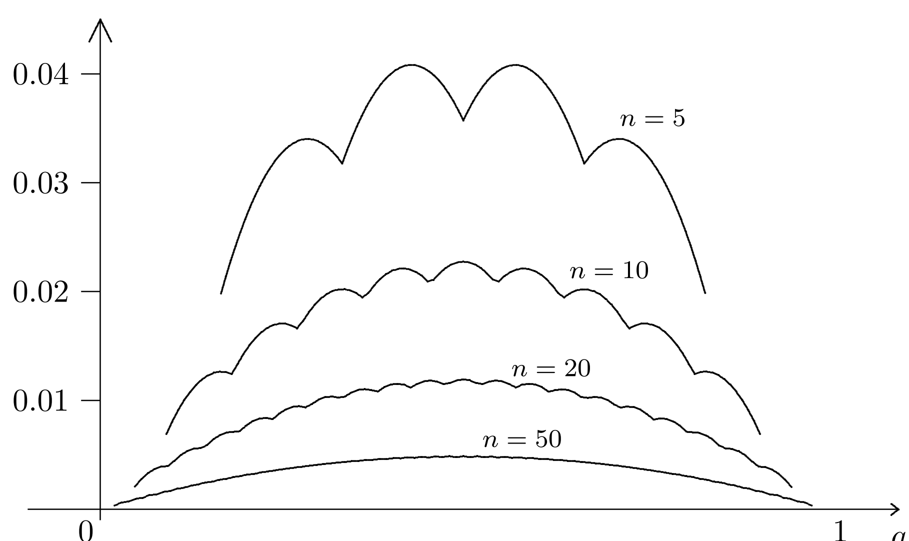


Fig.3.3.2.

Variances $V_n(q)$ of the estimator $\hat{x}_{WU}$ for some sample sizes n

Observe, that for every n, in the class $\mathcal{T}$ there exist an estimator that minimizes Mean Square Error $MSE_n(q)$ without the additional condition of unbiasedness. For an estimator $T = X_{J:n}$ we have

$$\begin{aligned}
MSE_n(q) &= E_F\Big(F(X_{J:n}) - q\Big)^2 = E(U_{J:n}) - q)^2 \\
&= \sum_{j=1}^{n} \lambda_j E(U_{j:n} - q)^2 \\
&= \sum_{j=1}^{n} \frac{\Gamma(n+1)}{\Gamma(j)\Gamma(n-j+1)} \int_0^1 (x-q)^2 x^{j-1}(1-x)^{n-j}\,dx \\
&= \frac{1}{(n+1)(n+2)} \sum_{j=1}^{n} j\big(j+1-2(n+2)q\big)\lambda_j + q^2
\end{aligned}$$

Estimator that, for fixed n and q, minimizes Mean Square Error is $X_{j^*:n}$, where j^* minimizes $j\big(j+1-2(n+2)q\big)$. Observe that $x\big(x+1-2(n+2)q\big)$ for real x is a quadratic function which has its minimum at $x_{\min} = (n+2)q - 0.5$. Hence

$$j^* = \begin{cases} 1, & \text{if } q \le \dfrac{2}{n+2} \\ NI[(n+2)q - 0.5], & \text{if } \dfrac{2}{n+2} < q < \dfrac{n}{n+2} \\ n, & \text{if } q \ge \dfrac{n}{n+2} \end{cases} \tag{7}$$

where $NI[x]$ denotes the integer nearest to x:

$$NI[x] = \begin{cases} [x], & \text{if } \{x\} \le \dfrac{1}{2} \\ [x]+1, & \text{if } \{x\} > \dfrac{1}{2} \end{cases}$$

One may change $(\le, >)$ to $(<, \ge)$ in the definition of $NI[x]$ above; it does not influence the value of Mean Square Error. As a consequence we obtain the following result

Theorem 3. In the class of all estimators $\mathcal{T}$, the estimator $X_{j^*:n}$ of the q-th quantile with j^* defined by (7) has uniformly minimal Mean Square Error. □

The estimator which minimizes Mean Square Error will be denoted by $\hat{x}_{MSE}$ or by $\hat{x}_{MSE}(q)$ if needed.

In opposite to unbiased estimators, estimator which minimizes Mean Square Error $MSE_n(q)$ is well defined for all $q \in (0, 1)$; Fig. 3.3.3 exhibits $MSE_n(q)$ for selected values of the sample size n.

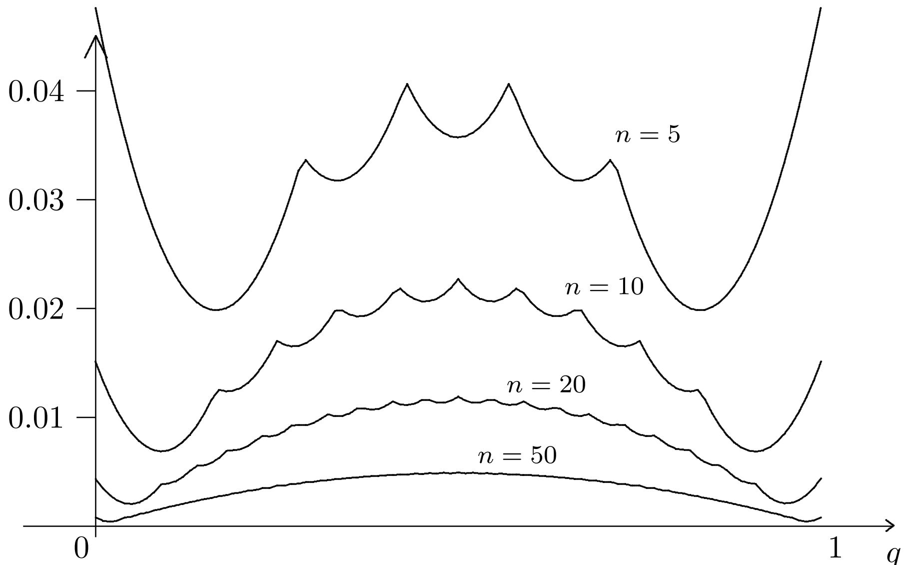


Fig.3.3.3.
MSE of the optimal estimator $\hat{x}_{MSE}$ for some n

Mean Square Error of optimal estimators are obviously smaller than that of unbiased estimators; that is presented in Fig. 3.3.4.

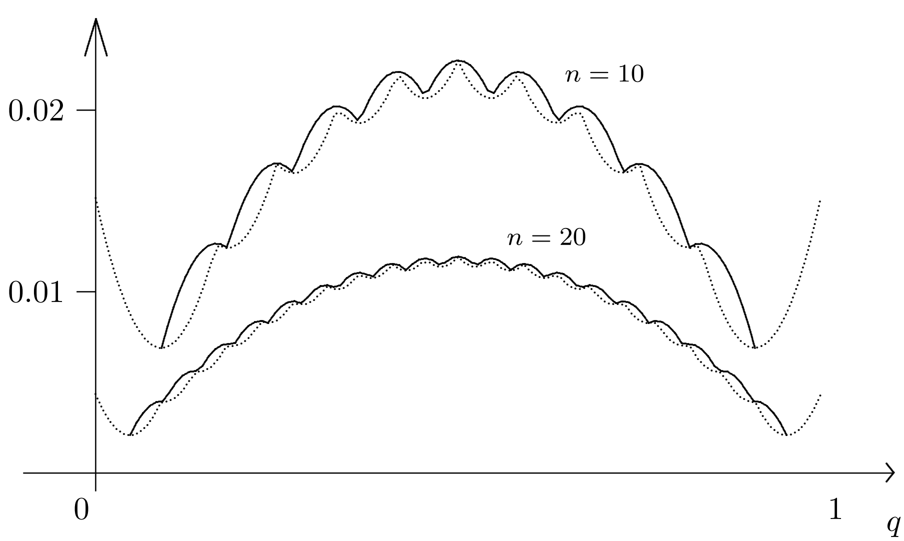


Fig.3.3.4.

MSE of the minimum variance unbiased estimator $\hat{x}_{WU}$ (solid) and the estimator $\hat{x}_{MSE}$ minimizing MSE (dots)

For $q = j/(n+1)$, $j = 1, 2, \ldots, n$, the estimator that minimizes MSE is unbiased and then $MSE_n(q) = V_n(q)$. At those values of q both Mean Square Error and the variance of the Minimum Variance Unbiased Estimator have their local minima.

3.3.3. Uniformly Minimum Absolute Deviation Estimator.

For a given estimator $T = T(X_{1:n}, \ldots, X_{n:n})$ of the q-th quantile from a sample of size n, define its Mean Absolute Deviation as

$$MAD_n(q) = E_F\Big|F\big(T(X_{1:n}, \ldots, X_{n:n})\big) - q\Big|.$$

For estimators $T \in \mathcal{T}$ of the form $T = X_{J:n}$ we have

$$MAD_n(q) = E_F\Big|F\big(X_{J:n}\big) - q\Big| = E\Big|U_{J:n} - q\Big| = \sum_{j=1}^{n} \lambda_j u_{j:n}(q),$$

where $u_{j:n}(q) = E|U_{j:n} - q|$. The obvious estimator which minimizes $MAD_n(q)$ is $X_{j^*:n}$ where j^* minimizes $u_{j:n}(q)$. To construct j^* explicitly, observe that

$$\begin{aligned}
u_{j:n}(q) &= \frac{\Gamma(n+1)}{\Gamma(j)\Gamma(n-j+1)} \int_0^1 |x-q| x^{j-1}(1-x)^{n-j} dx \\
&= \frac{\Gamma(n+1)}{\Gamma(j)\Gamma(n-j+1)} \int_q^1 (x-q) x^{j-1}(1-x)^{n-j} dx \\
&\qquad\qquad - \frac{\Gamma(n+1)}{\Gamma(j)\Gamma(n-j+1)} \int_0^q (x-q) x^{j-1}(1-x)^{n-j} dx \\
&= 2\left[B(q; j, n-j+1) - \frac{j}{n+1} B(q; j+1, n-j+1)\right] + \left(\frac{j}{n+1} - q\right)
\end{aligned}$$

where

$$B(x; \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^x t^{\alpha-1}(1-t)^{\beta-1} dt$$

is the incomplete Gamma function.

Let $Q(k;n,x)=\sum_{j=k}^{n}\binom{n}{j}x^j(1-x)^{n-j}$ be the tail of the Binomial distribution; making use of the equality

$$B(x;\alpha,\beta)=Q(\alpha;\alpha+\beta-1,x)$$

we obtain

$$u_{j:n}(q)=2\left[q\cdot Q(j;n,q)-\frac{j}{n+1}\cdot Q(j+1;,n+1,q)\right]+\left(\frac{j}{n+1}-q\right).$$

To find the smallest $u_{j:n}(q)$ observe that

$$u_{j+1:n}(q)-u_{j:n}(q)=\frac{1}{n+1}\left[1-2Q(j+1;n+1,q)\right],\quad j=1,2,\ldots,n-1$$

and hence

$$u_{j+1:n}(q)\geq u_{j:n}(q)\quad\text{iff}\quad Q(j+1;n+1,q)\leq\frac{1}{2}.$$

By the definition of $Q(j;n,x)$ we have

$$Q(1;n+1,q)>Q(2;n+1,q)>\ldots>Q(n+1;n+1,q)$$

and it follows that if $u_{j+1:n}(q)\geq u_{j:n}(q)$ then $u_{j+2:n}(q)\geq u_{j+1:n}(q)$ and consequently $u_{k:n}(q)\geq u_{j:n}(q)$ for all $k=j+1,j+2,\ldots,n$. Similarly, if $u_{j-1:n}(q)\geq u_{j:n}(q)$ then $u_{k:n}(q)\geq u_{j:n}(q)$ for all $k=1,2,\ldots,j-1$. As a result, as the optimal j^* we obtain

$$\begin{aligned}&j^*=1,\quad\text{if}\quad Q(2;n+1,q)\leq\frac{1}{2},\\ &j^*=n,\quad\text{if}\quad Q(n;n+1,q)\geq\frac{1}{2};\end{aligned}\tag{8}$$

otherwise j^* is any integer satisfying

$$\begin{aligned}&Q(j^*;n+1,q)\geq\frac{1}{2}>Q(j^*+1;n+1,q)\\ &\qquad\text{or } Q(j^*;n+1,q)>\frac{1}{2}\geq Q(j^*+1;n+1,q).\end{aligned}\tag{9}$$

As a result we obtain the following theorem.

Theorem 4. *In the class $\mathcal{T}$ of estimators, the estimator $X_{j^*:n}$ of the q-th quantile with j^* defined by (8) and (9) has uniformly minimal Mean Absolute Deviation.* □

The optimal j^* can be easily calculated with standard statistical tables or computer programs for the binomial distribution. In a sense, j^* defined by (8) is a median of the Binomial distribution with parameters $(n+1, q)$ (Neumann 1963, 1965). The estimator which minimizes Mean Absolute Deviation will be denoted by $\hat{x}_{MAD}$ or by $\hat{x}_{MAD}(q)$ if needed.

The risk of the optimal estimator $\hat{x}_{MAD}$ is given by the simple formula $MAD_n(q) = u_{j^*:n}(q)$ and is exhibited in Fig. 3.3.5.

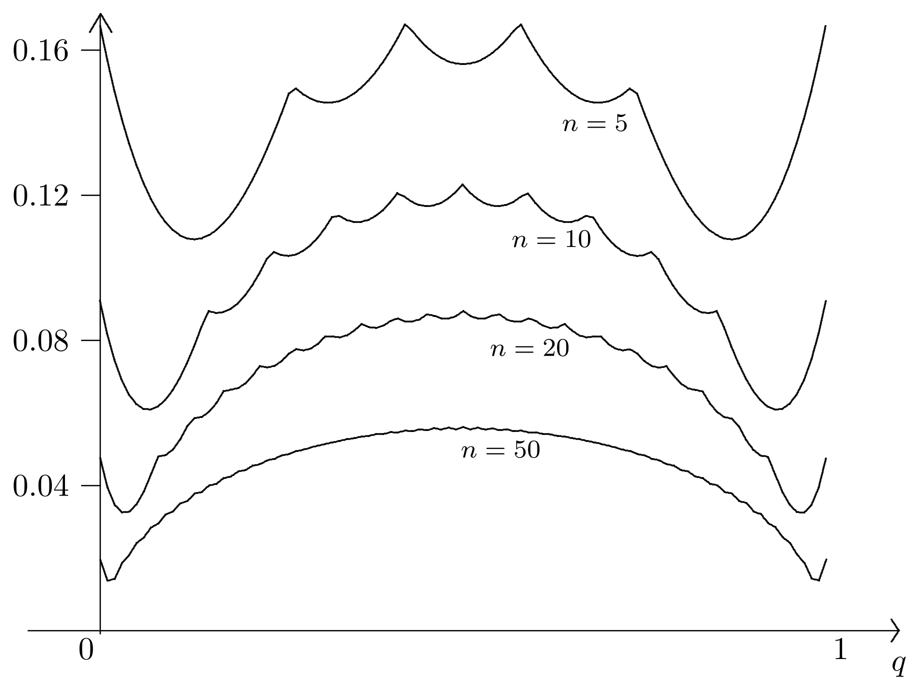


Fig.3.3.5.

MAD of the optimal estimator $\hat{x}_{MAD}$

If one is interested in unbiased estimation, the natural problem is to consider median-unbiased estimators with minimum MAD. Recall (see Sec. 3.3.1) that, given n and q, a median unbiased estimator exists iff $Q(1; n, q) \geq 1/2 \geq Q(n; n, q)$. Formally, the minimum MAD median-unbiased estimator is a solution of the following linear programming problem

$$\textit{minimize} \quad \sum_{j=1}^{n} \lambda_j u_{j:n}(q),$$

under the following conditions

$$\begin{cases} \sum_{j=1}^{n} \lambda_j Q(j;n,q) = \dfrac{1}{2}, \\ \sum_{j}^{n} \lambda_j = 1, \\ \lambda_1, \lambda_2, \ldots, \lambda_n \geq 0. \end{cases}$$

It is obvious that the median-unbiased minimum mean absolute deviation estimator is of the form $X_{J:n}$ where the random index has the distribution

$$P\{J = i^*\} = \lambda_{i^*}, \quad P\{J = i^* + 1\} = \lambda_{i^*+1}, \quad \lambda_{i^*} + \lambda_{i^*=1} = 1,$$

where i^* is the index such that $u_{i^*:n}$ is the smallest and $u_{i^*+1:n}$ is the second smallest element of the set $\{u_{1:n}, u_{2:n}, \ldots, u_{n:n}\}$. To find i^* we shall use the solution j^* of the problem of minimizing MAD without restriction on unbiasedness. We consider three following cases: $j^* = 1$, $j^* = n$, and $1 < j^* < n$. We will make use of the following inequalities:

(10) $\quad Q(i-1;n,q) > Q(i;n+1,q) > Q(i;n,q) > Q(i+1;n+1,q) > Q(i+1;n,q)$

Case $j^* = 1$. If $Q(1;n,q) = \dfrac{1}{2}$, then obviously $i^* = 1$. If $Q(1;n,q) > \dfrac{1}{2}$ then by the very construction of j^* and (10) we have $Q(2;n,q) \leq \frac{1}{2}$. Hence the optimal median unbiased estimator is $X_{J:n}$, where

$$P\{J = 1\} = \lambda_1 \quad \text{and} \quad P\{J = 2\} = 1 - \lambda_1$$

with

$$\lambda_1 = \frac{\dfrac{1}{2} - Q(2;n,q)}{Q(1;n,q) - Q(2:n,q)}.$$

If $Q(1;n,q) < 1/2$ then median-unbiased estimators of the qth quantile do not exist ("the sample size n is too small").

Case $j^* = n$. If $Q(n;n,q) = \dfrac{1}{2}$ then $i^* = n$. If $Q(n;n,q) < \dfrac{1}{2}$ then the optimal median unbiased estimator is $X_{J:n}$ with

$$P\{J = n-1\} = \lambda_{n-1} \quad \text{and} \quad P\{J = n\} = 1 - \lambda_{n-1}$$

where

$$\lambda_{n-1} = \frac{\frac{1}{2} - Q(n; n, q)}{Q(n-1; n, q) - Q(n : n, q)}.$$

If $Q(n; n, q) > \frac{1}{2}$, then median-unbiased estimators do not exist.

Case $1 < j^* < n$. Recall that the j^* satisfies (10). If $Q(j^*; n, q) = \frac{1}{2}$ then obviously $i^* = j^*$. If $Q(j^*; n, q) > \frac{1}{2}$ then $Q(j^*+1; n, q) < \frac{1}{2}$ and the median-unbiased minimum Mean Square Error estimator is $X_{J:n}$ for J such that

$$P\{J = i^*\} = \lambda_{i^*} \quad \text{and} \quad P\{J = i^* + 1\} = 1 - \lambda_{i^*},$$

where $i^* = j^*$ and

$$\lambda_{i^*} = \frac{\frac{1}{2} - Q(i^* + 1; n, q)}{Q(i^*; n, q) - Q(i^* + 1 : n, q)}.$$

Similarly, if $Q(j^*; n, q) < \frac{1}{2}$ then $Q(j^* - 1; n, q) > \frac{1}{2}$ and the optimal estimator is $X_{J:n}$ with

$$P\{J = i^* - 1\} = \lambda_{i^*-1} \quad \text{and} \quad P\{J = i^*\} = 1 - \lambda_{i^*-1}$$

where

$$\lambda_{i^*-1} = \frac{\frac{1}{2} - Q(i^*; n, q)}{Q(i^* - 1; n, q) - Q(i^* : n, q)}.$$

As a result we obtain the following theorem.

Theorem 5. *The minimum Mean Absolute Deviation median-unbiased estimator is $X_{J:n}$ with*

$$P\{J = i^*\} = \lambda_{i^*} \quad \text{and} \quad P\{J = i^* + 1\} = 1 - \lambda_{i^*}$$

where i^* is the unique integer such that

$$Q(i^*; n, q) \geq \frac{1}{2} \geq Q(i^* + 1; n, q)$$

and

$$\lambda_{i^*} = \frac{\frac{1}{2} - Q(i^* + 1; n, q)}{Q(i^*; n, q) - Q(i^* + 1 : n, q)}.$$

□

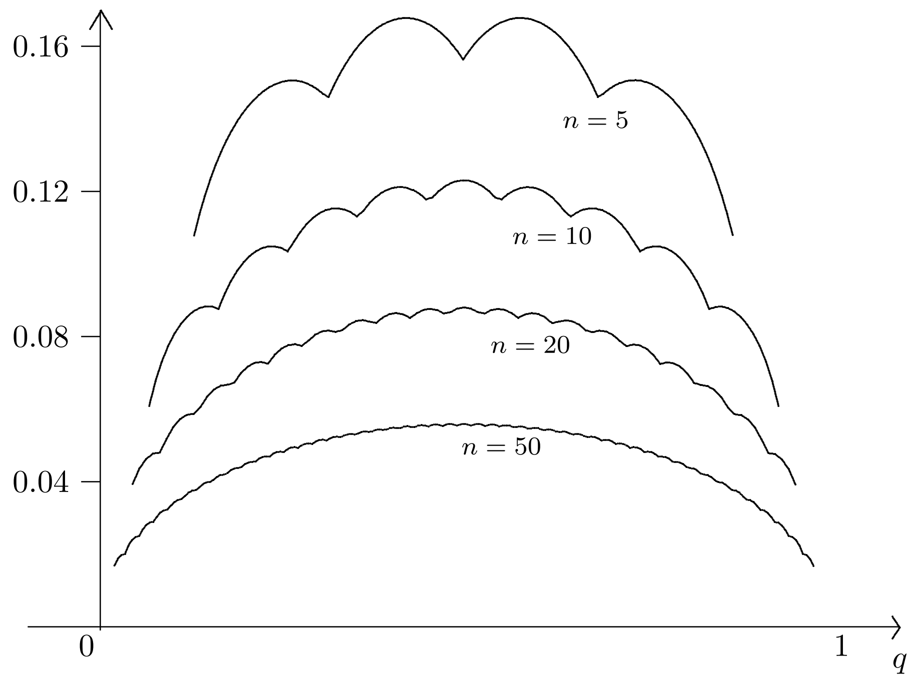


Fig.3.3.6.
MAD of the optimal median-unbiased estimator $\hat{x}_{UMAD}$

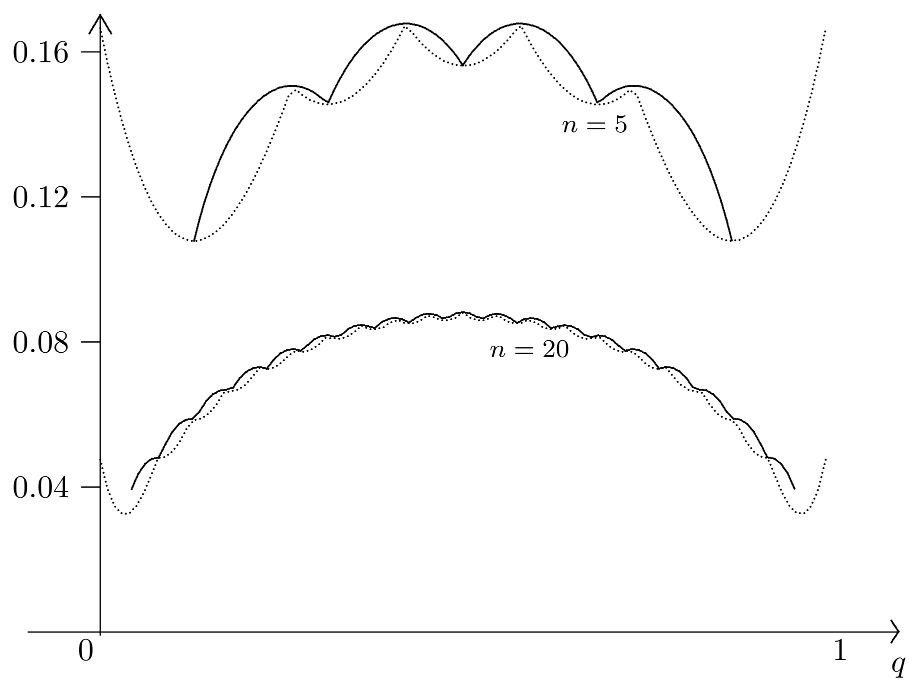


Fig.3.3.7.

MAD of the optimal estimator $\hat{x}_{MAD}$ (dots)
and that of the median-unbiased estimator $\hat{x}_{UMAD}$ (solid)

The minimum Mean Absolute Deviation median-unbiased estimator will be denoted by $\hat{x}_{UMAD}$ or $\hat{x}_{UMAD}(q)$ if needed. Mean Absolute Deviations of $\hat{x}_{UMAD}$ for some n are exhibited in Fig. 3.3.6

Mean Absolute Deviation of the unbiased estimator $\hat{x}_{UMAD}$ is obviously greater than that of the estimator $\hat{x}_{MAD}$ without restrictions to unbiasedness (Fig. 3.3.7.)

It is interesting to observe that the minimum MAD median-unbiased estimator $\hat{x}_{UMAD}$ is identical with the most concentrated median-unbiased estimator $\hat{x}_{RZ}$ constructed in Sec. 3.3.1.

3.3.4. Optimal estimator in the sense of Pitman's Measure of Closeness

According to Pitman's Measure of Closeness, if T_1 and T_2 are two estimators of a real parameter θ, then T_1 is better than T_2 if $P_\theta\{|T_1-\theta| \leq |T_2-\theta|\} \geq 1/2$ for all θ. It may however happen that while T_1 is better than T_2 and T_2 is better than T_3, T_3 is better than T_1 (Keating et al. 1991, 1993). A rationale for the criterion is that T_1 is better than T_2 if with probability at least one half the (absolute) error $|T_1-\theta|$ is smaller than the error $|T_2-\theta|$.

It may however happen that in a given statistical model and in a given class of estimators there exists one which is better than any other. We define such estimator as *PMC-optimal* or, for use in this Section, shortly optimal.

In full analogy to our considerations in previous sections, we shall measure the error of the estimator $T = T(X_1, X_2, \ldots, X_n)$ of the q-th quantile $x_q(F)$ in terms of differences $|F(T)-q|$ rather than in terms of differences $|T-x_q(F)|$. Then according to the Pitman's Measure of Closeness, an estimator T is better than S if

$$P_F\{|F(T)-q| \leq |F(S)-q|\} \geq 1/2 \text{ for all } F \in \mathcal{F} \tag{11}$$

(for more fine definitions see Keating *et al.* 1993). Consequently, an estimator T^* which satisfies

$$P_F\{|F(T^*)-q| \leq |F(S)-q|\} \geq 1/2 \text{ for all } F \in \mathcal{F} \text{ and for all } S \in \mathcal{T} \tag{12}$$

is *PMC-optimal*.

We use $\leq$ in the first inequality in the above definition because for $T=S$ we prefer to have the left hand side of (11) to be equal to one rather than to zero; otherwise the part "for all $T \in \mathcal{T}$" in (12) would not be true. For example two different estimators $X_{[nq]:n}$ and $X_{[(n+1)q]:n}$ are identical for $n = 7$ when estimating qth quantile for $q = 0.2$.

One can easily conclude from the results below that if there are exactly two different optimal estimators T_1^* and T_2^* (we will see that it may happen), then $P_F\{|F(T_1^*)-q| \leq |F(T_2^*)-q|\} = \frac{1}{2}$ and $P_F\{|F(T_1^*) - q| \leq |F(T) - q|\} > \frac{1}{2}$ for all other estimators $T \in \mathcal{T}$.

Denote $p(T, S) = P_F\{|F(T) - q| \leq |F(S) - q|\}$ and observe that to construct optimal estimator T^* it is enough to find T' such that

$$\min_{S\in\mathcal{T}} p(T', S) = \max_{T\in\mathcal{T}} \min_{S\in\mathcal{T}} p(T, S) \text{ for all } F \in \mathcal{F}$$

and take $T^* = T'$ if $\min_{S\in\mathcal{T}} p(T', S) \geq \frac{1}{2}$ for all $F \in \mathcal{F}$; if the inequality does not hold then the optimal estimator T^* does not exist. In what follows we construct the optimal estimator T^*. It would be convenient to formulate the results as a Theorem proceeded by two Lemmas. The results come from Zieliński 2001.

For estimators $T, S \in \mathcal{T}$ we have $T = X_{J(\lambda):n}$ and $S = X_{J(\mu):n}$ where

$$P\{J(\lambda) = j\} = \lambda_j, \quad \lambda_j \geq 0, \quad j = 1, 2, \ldots, n, \quad \sum_{j=1}^{n} \lambda_j = 1$$

and

$$P\{J(\mu) = j\} = \mu_j, \quad \mu_j \geq 0, \quad j = 1, 2, \ldots, n, \quad \sum_{j=1}^{n} \mu_j = 1.$$

If the sample $X_1, X_2, \ldots, X_n$ comes from a distribution function F then $F(T) = U_{J(\lambda):n}$ and $F(S) = U_{J(\mu):n}$, respectively, where $U_{1:n}, U_{2:n}, \ldots, U_{n:n}$ are the order statistics from a sample $U_1, U_2, \ldots, U_n$ drawn from the uniform distribution $U(0, 1)$. Denote

$$w_q(i, j) = P\{|U_{i:n} - q| \leq |U_{j:n} - q|\}, \quad 1 \leq i, j \leq n.$$

Then

$$p(T, S) = p(\lambda, \mu) = \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \mu_j w_q(i, j)$$

and $T^* = X_{J(\lambda^*):n}$ is optimal if

$$\min_{\mu} p(\lambda^*, \mu) \geq \frac{1}{2}.$$

For a fixed i, the sum $\sum_{j=1}^{n} \mu_j w_q(i, j)$ is minimal for $\mu_{j^*} = 1$, $\mu_j = 0, j \neq j^*$, where $j^* = j^*(i)$ is such that $w_q(i, j^*) \leq w_q(i, j)$, $j = 1, 2, \ldots, n$. Then the optimal λ^*

satisfies $\lambda_{i^*} = 1,\ \lambda_i = 0, i \neq i^*$, where i^* maximizes $w_q(i, j^*(i))$. It follows that the optimal estimator T^* is of the form $X_{i^*:n}$ with a suitable i^* and the problem reduces to finding i^*.

Denote $v_q^-(i) = w_q(i, i-1)$, $v_q^+(i) = w_q(i, i+1)$, and define $v_q^-(1) = v_q^+(n) = 1$. Proofs of all Lemmas and the Theorem below are postponed to the end of the Section.

Lemma 4. *For a fixed $i = 1, 2, \ldots, n$, we have* $\min_j w_q(i, j) = \min\{v_q^-(i), v_q^+(i)\}$.

By Lemma (4), the problem reduces to finding i^* which maximizes $\min\{v_q^-(i), v_q^+(i)\}$.

Lemma 5. *The sequence $v_q^+(i)$, $i = 1, 2, \ldots, n$, is increasing and the sequence $v_q^-(i)$, $i = 1, 2, \ldots, n$, is decreasing.*

By Lemma (5), to get i^* one should find $i' \in \{1, 2, \ldots, n-1\}$ such that

$$v_q^-(i') \geq v_q^+(i') \quad \text{and} \quad v_q^-(i'+1) < v_q^+(i'+1) \tag{13}$$

and then calculate

$$i^* = \begin{cases} i', & \text{if } v_q^+(i') \geq v_q^-(i'+1) \\ i'+1, & \text{otherwise.} \end{cases} \tag{14}$$

Eventually we obtain the following theorem.

Theorem 6. *Let i^* be defined by the formula*

$$i^* = \begin{cases} i', & \textit{if } v_q^+(i') \geq \dfrac{1}{2} \\ i'+1, & \textit{otherwise} \end{cases} \tag{15}$$

where

$$i' = \begin{cases} \textit{the smallest integer } i \in \{1, 2, \ldots, n-2\} \textit{ such that } Q(i+1; n, q) < \dfrac{1}{2} \\ n-1, \qquad\qquad\qquad\qquad\qquad \textit{if } Q(n-1, n, q) \geq \dfrac{1}{2}. \end{cases} \tag{16}$$

For i^* defined by (15) we have

$$P_F\{|F(X_{i^*:n}) - q| \leq |F(T) - q|\} \geq \frac{1}{2} \tag{17}$$

for all $F \in \mathcal{F}$ and for all estimators $T \in \mathcal{T}$ of the qth quantile, which means that i^* is optimal. □

Table 1

n	i								
	1	2	3	4	5	6	7	8	9
3	.3612								
4	.2800								
5	.2283	.4086							
6	.1926	.3450							
7	.1666	.2984	.4326						
8	.1467	.2628	.3811						
9	.1311	.2348	.3406	.4468					
10	.1184	.2122	.3077	.4038					
11	.1080	.1936	.2807	.3683	.4561				
12	.0993	.1779	.2580	.3385	.4192				
13	.0919	.1646	.2387	.3132	.3879	.4626			
14	.0855	.1532	.2221	.2914	.3609	.4304			
15	.0799	.1432	.2076	.2724	.3374	.4024	.4675		
16	.0750	.1344	.1949	.2558	.3168	.3778	.4389		
17	.0707	.1267	.1837	.2411	.2985	.3560	.4136	.4712	
18	.0669	.1198	.1737	.2279	.2823	.3367	.3911	.4455	
19	.0634	.1136	.1647	.2162	.2677	.3193	.3709	.4225	.4742
20	.0603	.1080	.1567	.2055	.2545	.3036	.3527	.4018	.4509

As a conclusion we obtain that $X_{i^*:n}$ is *PMC-optimal* in the class of all equivariant estimators $T \in \mathcal{T}$ of the qth quantile. *PMC-optimal* estimator will be denoted by $\hat{x}_{PMC}$ or $\hat{x}_{PMC}(q)$, respectively.

For practical applications, the index i' can be easily found by tables or suitable computer programs for Bernoulli or Beta distributions, but checking the condition in (15) needs a comment. First of all observe that $v_0^+(i) = 1$, $v_1^+(i) = 0$, and the first derivative of $v_q^+(i)$ with respect to q is negative. It follows that $v_q^+(i) \geq \frac{1}{2}$ iff $q \leq q_n(i)$ where $q_n(i)$ is the unique solution (with respect to q) of the equation $v_q^+(i) = \frac{1}{2}$. For $q \in (0,1)$, $v_q^+(i)$ is a strictly decreasing function with known values at both ends of the interval so that $q_n(i)$ can be easily found by a standard numerical routine. Table 1 gives us the values of $q_n(i)$ for $n = 3, 4, \ldots, 20$. Due to the equality

$$v_q^+(i) + v_{1-q}^+(n-i) = 1$$

we have $q \leq q_n(i)$ iff $1 - q \geq q_n(n-i)$ so that in Table 1 only the values $q_n(i)$ for $i < [n/2]$ are presented. Sometimes the following fact may be useful: if i^* is optimal for estimating the qth quantile from sample of size n, then $n - i^* + 1$ is optimal for estimating the $(1-q)$th quantile from the same sample.

As an example suppose we want to estimate the qth quantile with $q = 0.3$ from a sample of size $n = 10$. For the Bernoulli tail $Q(k, n; p) = \sum_{j=k}^{n} \binom{n}{j} p^j (1-p)^{n-j}$ we have

$$Q(4, 10; 0.3) = 0.3504 < \frac{1}{2} < Q(3, 10; 0.3)$$

hence $i' = 3$. Now, by Table (1), $q_{10}(3) = 0.3077$ so that $q < q_n(i')$, hence $i^* = 3$.

As another example, for $n = 8$ and $q = 0.75$ we have $Q(7, 8; 0.75) = 0.3671 < \frac{1}{2} < Q(6, 8; 0.75) = 0.6785$ and $i' = 6$. By Table (1) we have $q_8(6) = 1 - q_8(2) = 0.7372$; now $q > q_8(6)$ so that $i^* = i' + 1 = 7$.

It is interesting to observe that *PMC-optimal* estimator differs from that which minimizes Mean Absolute Deviation $E_F|F(T) - q|$; the latter has been constructed in the previous Section 3.3.3. For example, to estimate the quantile of order $q = 0.225$ from a sample of size $n = 10$, $X_{3:10}$ is *PMC-optimal*, while $X_{2:10}$ minimizes Mean Absolute Deviation.

To the end of the Section we present proofs of Lemmas and the Theorem.

Proof of Lemma (4). Suppose first that $i < j$ and consider the following events

$$A_1 = \{U_{i:n} > q\}, \quad A_2 = \{U_{i:n} \le q < U_{j:n}\}, \quad A_3 = \{U_{j:n} < q\}. \tag{18}$$

The events are pairwise disjoint and $P(A_1 \cup A_2 \cup A_3) = 1$. Hence

$$w_q(i, j) = \sum_{j=1}^{3} P\{|U_{i:n} - q| \le |U_{j:n} - q|, A_j\}.$$

For the first summand we have

$$P\{|U_{i:n} - q| \le |U_{j:n} - q|, A_1\} = P\{U_{i:n} > q\}.$$

The second summand can be written in the form

$$\begin{aligned} P\{|U_{i:n} - q| \le |U_{j:n} - q|, A_2\} &= P\{U_{i:n} + U_{j:n} \ge 2q, U_{i:n} \le q < U_{j:n}\} \\ &= P\{U_{i:n} \le q < U_{j:n}, U_{j:n} \ge 2q - U_{i:n}\}, \end{aligned}$$

and the third one equals zero.

If $j' > j$ then $U_{j':n} \geq U_{j:n}$, the event $\{U_{i:n} \leq q < U_{j:n}, U_{j:n} \geq 2q - U_{i:n}\}$ implies the event $\{U_{i:n} \leq q < U_{j':n}, U_{j':n} \geq 2q - U_{i:n}\}$, and hence

$$w_q(i, j') \geq w_q(i, j).$$

In consequence

$$\min_{j>i} w_q(i, j) = w_q(i, i+1) = v_q^+(i).$$

Similarly $\min_{j<i} w_q(i, j) = v_q^-(i)$, which ends the proof of Lemma (4). □

Proof of Lemma (5). Similarly as in the proof of Lemma (4), considering events (18) with $j = i + 1$, we obtain

$$v_q^+(i) = P\{U_{i:n} > q\} + P\{U_{i:n} + U_{i+1:n} \geq 2q, U_{i:n} \leq q < U_{i+1:n}\}$$

and by standard calculations

$$v_q^+(i) = \frac{n!}{(i-1)!(n-i)!} \left(\int_q^1 x^{i-1}(1-x)^{n-i} dx + \int_{(2q-1)^+}^{q} x^{i-1}(1-2q+x)^{n-i} dx \right)$$

where $x^+ = \max\{x, 0\}$. For $i = n - 1$ we obviously have $v_q^+(n-1) < v_q^+(n) = 1$. For $i \in \{1, 2, \ldots, n-1\}$ the inequality $v_q^+(i) < v_q^+(i+1)$ can be written in the form

$$i\Big(\int_q^1 x^{i-1}(1-x)^{n-i} dx + \int_{(2q-1)^+}^{q} x^{i-1}(1-2q+x)^{n-i} dx \Big) <$$

$$< (n-i)\Big(\int_q^1 x^{i}(1-x)^{n-i-1} dx + \int_{(2q-1)^+}^{q} x^{i}(1-2q+x)^{n-i-1} dx \Big).$$

Integrating left hand side by parts we obtain an equivalent inequality

$$2(n-i) \int_{(2q-1)^+}^{q} x^{i}(1-2q+x)^{n-i-1} dx > 0$$

which is obviously always true.

In full analogy to the calculation of $v_q^+(i)$, for $i \in \{2,3,\dots,n\}$ we obtain

$$v_q^-(i) = \frac{n!}{(i-1)!(n-i)!}\left(\int_0^q x^{i-1}(1-x)^{n-i}dx + \int_q^{min\{1,2q\}} (2q-x)^{i-1}(1-x)^{n-i}dx\right)$$

and the inequality $v_q^-(i-1) > v_q^-(i)$ can be proved as above, which ends the proof of Lemma (5). $\square$

Proof of Theorem (6). We shall use following facts

$$v_q^+(i) + v_q^-(i+1) = 1, \tag{19}$$

which follows from the obvious equality $w_q(i,j) + w_q(j,i) = 1$, and

$$v_q^+(i) + v_q^+(i+1) = 2\left(1 - Q(i+1;n,q)\right), \quad i = 1,2,\dots,n-1. \tag{20}$$

Equality (20) follows from integrating by parts both integrals in $v_q^+(i)$ and then calculating the sum $v_q^+(i) + v_q^+(i+1)$.

Let us consider condition (13) for $i = 1$, $i = n-1$, and $i \in \{2,3,\dots,n-2\}$, separately.

For $i = 1$ we have $v_1^-(1) = 1 > v_q^+(1)$ hence $i' = 1$ iff $v_q^-(2) < v_q^+(2)$ which by (9) amounts to $1 - v_q^+(1) < v_q^+(2)$ and by (20) to $2\left(1 - Q(2,n,q)\right) > 1$ or $Q(2,n,q) < \frac{1}{2}$. Now $i^* = 1$ if $v_q^+(1) \geq v_q^-(2)$ or $v_q^+(1) \geq 1 - v_q^+(1)$ or $v_q^+(1) \geq \frac{1}{2}$, and $i^* = 2$ if $v_q^+(1) < \frac{1}{2}$.

Due to the equality $v_q^-(n) < v_q^+(n) = 1$, by (13) we have $i' = n-1$ iff $v_q^-(n-1) \geq v_q^+(n-1)$ which by (19) amounts to $v_q^+(n-2) + v_q^+(n-1) \leq 1$, and by (20) to $Q(n-1;n,q) \geq \frac{1}{2}$. Now $i^* = n-1$ if $v_q^+(n-1) \geq v_q^-(n)$ or $v_q^+(n-1) \geq \frac{1}{2}$; otherwise $i^* = n$.

For $i \in \{2,3,\dots,n-2\}$, by (19), condition (13) can be written in the form

$$v_q^+(i-1) + v_q^+(i) \leq 1 \quad \text{and} \quad v_q^+(i) + v_q^+(i+1) > 1$$

and by (20) in the form

$$Q(i;n,q) \geq \frac{1}{2} \quad \text{and} \quad Q(i+1;n,q) < \frac{1}{2}.$$

Now by (14) and (19)

$$i^* = \begin{cases} i', & \text{if } v_q^+(i') \geq \frac{1}{2} \\ i'+1, & \text{otherwise.} \end{cases}$$

Summing up all above and taking into account that $Q(i;n,q)$ decreases in $i = 1,2,\ldots,n-1$, we obtain

$$i' = \begin{cases} \text{first } i \in \{1,2,\ldots,n-2\} \text{ such that } Q(i+1;n,q) < \frac{1}{2} \\ n-1, \text{if such } i \text{ does not exist.} \end{cases}$$

Then $i^* = i'$ if $v_q^+(i') \geq \frac{1}{2}$ and $i^* = i'+1$ otherwise, which gives us statements (15)-(16) of the Theorem.

To prove statement (17) of the Theorem observe that if $i^* = 1$ then $v_q^+(1) \geq \frac{1}{2}$ and if $i^* = n$ then $v_q^-(n) = 1 - v_q^+(n-1) \geq \frac{1}{2}$. For $i^* \in \{2,3,\ldots,n-1\}$ we have: 1) if $i^* = i'$ then by (15) $v_q^+(i^*) \geq \frac{1}{2}$ and by the first inequality in (13) $v_q^-(i^*) \geq v_q^+(i^*) \geq \frac{1}{2}$, hence $\min\{v_q^-(i^*), v_q^+(i^*)\} \geq \frac{1}{2}$ and 2) if $i^* = i'+1$ then by (15) $v_q^+(i^*-1) < \frac{1}{2}$ which amounts to $1 - v_q^-(i^*) < \frac{1}{2}$ or $v_q^-(i^*) > \frac{1}{2}$. Then by the second inequality in (13) we have $v_q^+(i^*) > v_q^-(i^*) > \frac{1}{2}$, so that again $\min\{v_q^-(i^*), v_q^+(i^*)\} \geq \frac{1}{2}$, which ends the proof of the Theorem. □

3.3.5. Comparison of estimators?

Each of estimators $\hat{x}_{RZ}$, $\hat{x}_{WU}$, $\hat{x}_{MSE}$, $\hat{x}_{MAD}$, $\hat{x}_{UMAD}$, and $\hat{x}_{PMC}$ considered in this Chapter is optimal with respect to a suitable criterion; if one chooses a criterion of optimality, then the estimator is uniquely determined. However, in large nonparametric model $\mathcal{F}$ that we consider, criterions which led to estimators $\hat{x}_{RZ} = \hat{x}_{UMAD}$ and $\hat{x}_{MAD}$, seem to be more natural than those based on mean-unbiasedness and Mean Square Error which gave us estimators $\hat{x}_{WU}$ and $\hat{x}_{MSE}$. An obvious reason is that $F(T)$ is a median-unbiased estimator of q if and only if T is a median-unbiased estimator of the q-th quantile $x_q(F)$ when the sample comes from a parent distribution $F \in \mathcal{F}$, while the analogous statement for mean-unbiasedness is not true. What is more, if $F(T)$ is the most concentrated median-unbiased estimator of q then T is the most concentrated median-unbiased of $x_q(F)$ when the sample comes from the distribution $F \in \mathcal{F}$.

Estimator $\hat{x}_{PMC}$ also needs a comment. Recall that $\hat{x}_{PMC}$ is optimal in the sense that it satisfies

$$(21) \qquad P_F\{|F(\hat{x}_{PMC}) - q| \leq |F(T) - q|\} \geq 1/2 \text{ for all } F \in \mathcal{F} \text{ and for all } T \in \mathcal{T}.$$

However the following statement, which is more interesting for practical applications,

$$P_F\{|\hat{x}_{PMC} - x_q(F)| \leq |T - x_q(F)|\} \geq 1/2 \text{ for all } F \in \mathcal{F} \text{ and for all } T \in \mathcal{T},$$

is not true. To see that consider the following numerical example. For $n = 10$ and $q = 0.225$ the optimal, in the sense of (21), estimator is $X_{3:10}$:

$$P_F\{|F(X_{3:10}) - q| \leq |F(T) - q|\} \geq 1/2 \text{ for all } F \in \mathcal{F} \text{ and for all } T \in \mathcal{T}.$$

However, for estimating the quantile of order $q = 0.225$ of the parent distribution $F(x) = x^{1/4}$ we have

$$P_F\{|X_{2:10}) - x_q(F)| \leq |X_{3:10}) - x_q(F)|\} = 0.5234$$

which means that $X_{2:10}$ is PMC-better than $X_{3:10}$.

Similar property share MAD-optimal estimators: they are optimal when the error is measured in terms of $|F(T) - q|$ and not in terms of $|T - x_q(F)|$. For example, to estimate the quantile of order $q = 0.3$ from a sample of size $n = 10$, the estimator which minimizes $E_F|F(T) - q|$ is $X_{3:10}$ with Mean Absolute Deviation equal to 0.1083 while for $F(x) = x^2,\ 0 < x < 1$ we have $E_F|X_{3:10} - F^{-1}(0.3)| = 0.1069$ and $E_F|X_{4:10} - F^{-1}(0.3)| = 0.1033$.

As a conclusion we see that the only property preserved under monotonic transformations of an estimator T, and especially under the transformation $F(T)$, is the property of being the most concentrated (in the sense of (4)) median-unbiased estimator.

4

Applications to parametric models

Typically specificity of a given parametric model allow us to construct effective statistical estimators. There are however some difficult problems (for example, constructing median-unbiased estimators) which may be solved through embedding the parametric model into a larger nonparametric model. Embedding parametric model in a larger (parametric or nonparametric) model is also typical in situation if our knowledge on the parametric model on hand is not exact and we wish to safeguard ourselves against "possible violations" of the specific parametric model under consideration: taking into account possible violations amounts embedding the model into a larger one and constructing suitable robust estimators.

4.1. Median-unbiased estimators

In this Section we consider the following problem: given a parametric model with a real (unknown) parameter $\theta \in \Theta$, we are interested in constructing an exactly median-unbiased estimator of the parameter, i.e. an estimator $\hat{\theta}$ such that a median of the estimator equals θ, uniformly over all $\theta \in \Theta$. Though the metod that we present below is general, effective constructions are available for some specific families of distributions only. We consider families of distributions $\mathcal{K} = \{K_\theta : \theta \in \Theta\}$, where Θ is an (finite or not) interval on the real line. The family $\mathcal{K}$ is assumed to be a family of distributions with continuous and strictly increasing distribution functions and stochastically ordered by θ so that for every $x \in supp\,\mathcal{K} = \bigcup_{\theta \in \Theta} supp\,K_\theta$ and for every $q \in (0,1)$, the equation $K_\theta(x) = q$ has exactly one solution in $\theta \in \Theta$; here $supp\,K$ denotes the support of the distribution K, which is a finite or infinite interval on the real line. It is obvious that the solution depends monotonically both on x and q. As we will see further on, the model represents a wide range of one-parameter families of distributions. Generally: every family of distributions $\{K_\theta : K_\theta(x) = K(x-\theta)\}$ with continuous and strictly increasing distribution function K and the location parameter θ satisfies the model assumptions. Similarly, every family $\{K_\theta : K_\theta(x) = K(x/\theta)\}$ with a continuous and strictly increasing distributions K on $(0, +\infty)$ with the scale parameter θ fits the model.

The method consists in

1) for a given $q \in (0,1)$, estimating the q-th quantile of the underlying distribution by a median-unbiased nonparametric estimator $\tilde{x}_q$. A restriction is that for a fixed n a median-unbiased estimator of the q-th quantile exists iff $1-(1/2)^{1/n} < q < (1/2)^{1/n}$ (cf Section 3.3.1); in our approach the restriction does not play any role because q may be chosen optionally;

2) solving the equation $K_\theta(\tilde{x}_q) = q$ with respect to θ; the solution $\hat{\theta}_q = \hat{\theta}_q(\hat{x}_q)$ is considered as an estimator of θ.

If $\tilde{x}_q$ is a median-unbiased estimator of x_q then, due to monotonicity of $\hat{\theta}_q(x)$ with respect to x, $\hat{\theta}_q$ is a median-unbiased estimator of θ. What is more, if $\tilde{x}_q$ is the median-unbiased estimator of x_q the most concentrated around x_q in the class of all median-unbiased estimators which are equivariant with respect to monotone transformations of data (briefly: the best estimator) then, due to monotonicity again, $\hat{\theta}_q$ is the most concentrated around θ median-unbiased estimator of θ in the class of all median-unbiased estimators which are equivariant with respect to monotone transformations of data (briefly: the best estimator).

As the best median-unbiased nonparametric estimator of the q-th quantile we take the estimator $\hat{x}_{RZ}$ from Sec. 3.3.1

$$\hat{x}_{RZ}(q) = X_{J:n}$$

where the random index J has distribution

$$P\{J = k\} = \lambda, \quad \{J = k+\} = 1 - \lambda$$

with

$$\begin{aligned} k &= k(q) \\ &= \textit{the unique integer such that } Q(k;n,q) \geq \frac{1}{2} \geq Q(k+1;n,q), \end{aligned}$$

$$\lambda = \lambda(q) = \frac{\frac{1}{2} - Q(k+1;n,q)}{Q(k;n,q) - Q(k+1;n,q)}.$$

Here $Q(k;n,q) = \sum_{j=k}^{n} \binom{n}{j} q^j (1-q)^{n-j}$.

The construction works for every $1-(1/2)^{1/n} < q < (1/2)^{1/n}$ and the problem arises how to choose an "optimal" q. To define a criterion of optimality let us recall (e.g. Lehmann 1983, Sec. 3.1) that a median-unbiased estimator $\hat{\theta}$ of a parameter θ satisfies

$$(1) \qquad E_\theta|\hat{\theta}-\theta| \le E_\theta|\hat{\theta}-\theta'| \quad for\ all\ \theta,\theta' \in \Theta$$

(the estimator is closer to the "true" value $\theta \in \Theta$ than to any other value $\theta' \in \Theta$ of the parameter). According to the property, we shall choose q_{opt} as that with minimal risk under the loss function $|\hat{\theta}-\theta|$, i.e. such that

$$E_\theta|\hat{\theta}_{q_{opt}}-\theta| \le E_\theta|\hat{\theta}_q-\theta|, \quad 1-(1/2)^{1/n} < q < (1/2)^{1/n}$$

for all $\theta \in \Theta$, if possible.

Using the fact that $\theta \in \Theta$ generates the stochastic ordering of the $\mathcal{K}$, we shall restrict our attention to finding q_{opt} which satisfies criterion (1) for a fixed θ, for example $\theta = 1$ (if θ is a scale or a shape parameter) or $\theta = 0$ if θ is a location parameter; then the problem reduces to minimizing

$$R(q) = E|\hat{\theta}_q - 1| \quad or \quad R(q) = E|\hat{\theta}_q|$$

with respect to $q \in (0,1)$, where $E = E_1$ or $E = E_0$, respectively. Formulas below are given for the case $\theta = 1$; the case of $\theta = 0$ can be treated in full analogy.

We obtain

$$R(q) = \lambda(q)\, E|\hat{\theta}_q(X_{k(q):n}) - 1| + (1-\lambda(q))\, E|\hat{\theta}_q(X_{k(q)+1:n}) - 1|.$$

By the fact that $R(q)$ is a convex combination of two quantities, it is obvious that q_{opt} satisfies

$$\lambda(q_{opt}) = 1$$

and

$$E|\hat{\theta}_{q_{opt}}\left(X_{k(q_{opt}):n}\right) - 1| \le E|\hat{\theta}_q\left(X_{k(q):n}\right) - 1|, \quad 1-(1/2)^{1/n} < q < (1/2)^{1/n}.$$

By the very definition of λ, $\lambda(q) = 1$ iff $q \in \{q_1, q_2, \ldots, q_n\}$ where q_i satisfies $Q(i; n, q_i) = 1/2$, and the problem reduces to finding the smallest element of the finite set

$$\{E|\hat{\theta}_{q_i}(X_{i:n}) - 1|, \quad i = 1, 2, \ldots, n\}.$$

If $X_{k:n}$ is the k-th order statistic from the sample $X_1, X_2, \ldots, X_n$ from a distribution function F, then $U_{k:n} = F(X_{k:n})$ is the k-th order statistic from the sample $U_1, U_2, \ldots, U_n$ from the uniform distribution on $(0, 1)$ which gives us

$$\begin{aligned} E|\hat{\theta}_{q_i}(X_{i:n}) - 1| &= E|\hat{\theta}_{q_i}(F^{-1}(U_{i:n})) - 1| \\ &= \frac{\Gamma(n)}{\Gamma(i)\Gamma(n-i+1)} \int_0^1 \left|\hat{\theta}_{q_i}\left(F^{-1}(t)\right) - 1\right| t^{i-1}(1-t)^{n-i} dt. \end{aligned} \tag{2}$$

The latter can be easily calculated numerically.

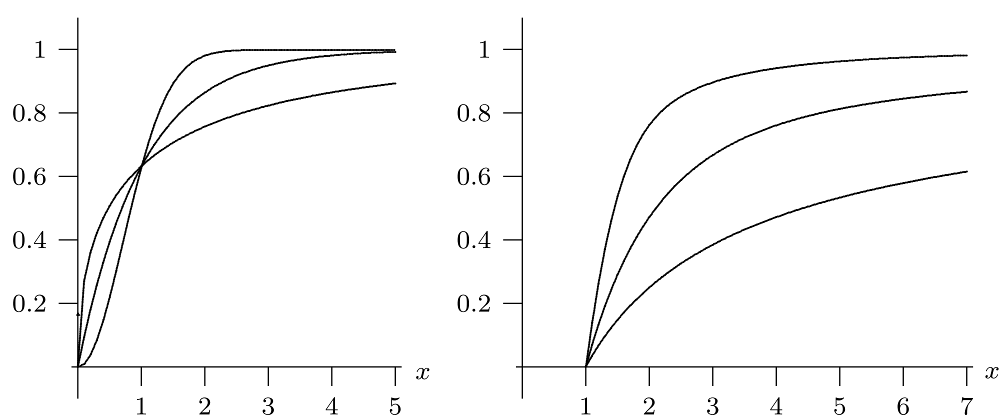
As an example consider constructing a median-unbiased estimator for the shape parameter α of the Weibull random variable Y with distribution function of the form

$$W_\alpha(y) = 1 - e^{-y^\alpha}, \quad y > 0, \ \alpha > 0.$$

The family of Weibull distributions $\{W_\alpha(x), \alpha > 0\}$ does not fit our model assumptions concerning stochastic ordering but we may consider transformed observations $X = \max\{Y, 1/Y\}$; those have a distribution of the form

$$F_\alpha(x) = e^{-x^{-\alpha}} - e^{-x^\alpha}, \quad x > 1, \ \alpha > 0.$$

The family of distributions $\{F_\alpha(x), \alpha > 0\}$ is stochastically ordered by the parameter α (Fig.4.1).



a) Distributions $G_\alpha, \alpha = 0.5, \ 1, \ 2$ b) Distributions $F_\alpha, \alpha = 0.5, \ 1, \ 2$

Fig. 4.1.

For the sample size $n = 10$ and for $F = F_1$ in (2), by numerical calculations we obtain $\min_i E|\hat{\theta}_{q_i}(X_{i:n}) - 1| = 0.2976$ for $i = 8$ and $q_{opt} = q_8 = 0.7414$. Hence the median-unbiased estimator of the parameter α of the distribution W_α is given as the (unique) solution with respect to α of the equation $F_\alpha(X_{8:10}) = 0.7417$ which can be explicitly written as $0.302/\log(X_{8:10})$.

Similarly we can obtain the following results (Zieliński 2003) (all numerical results for $n = 10$):

Uniform $U(\theta, \theta+1)$ distribution. Two equivalent median-unbiased estimators are $X_{1:10} - 0.067$ and $X_{10:10} - 0.933$.

Power distribution $F_\theta(x) = x^\theta, 0 < x < 1, \theta > 0$. A median-unbiased estimator is $-1.81854/\log(X_{2:10})$.

Cauchy distribution. If Y has a Cauchy distribution with the scale parameter λ and probability distribution function $(1/\lambda)\left(1 + (y/\lambda)^2\right)^{-1}$ then the family of distributions of the random variable $X = |Y|$ is stochastically ordered by the parameter λ. A median-unbiased estimator is $1.16456 \cdot X_{5:10}$.

Symmetric stable distribution. Symmetric stable distributions are defined by their characteristic functions of the form $\exp\{-|t|^\alpha\}, \alpha \in (0, 2]$. Typically explicite formulas neither for distribution functions nor for probability density functions for stable distributions are known, but distribution functions for the one-parameter family of symmetric stable distributions may by written in the form

$$F_\alpha(x) = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \frac{\sin y}{y} \exp\left\{-\left(\frac{y}{x}\right)^\alpha\right\} dy$$

A high accuracy algorithm for calculation of F_α is available in Zieliński 2000b and Zieliński 2001a. The construction of the estimator was presented in (Zieliński 2000a). In another context a method of estimating parameters of stable distributions through estimating appropriate quantiles was presented in Fama et al. (1971).

4.2.1. Estimating location parameter under ε-contamination

Let $F_\theta(x) = F(x-\theta), \theta \in \mathbf{R}^1$, be a location statistical model with a known continuous and strictly increasing unimodal distribution function F; the problem is to estimate θ, but what we observe is a random variable X with distribution $G(x) = (1-\varepsilon)F(x)+\varepsilon H(x)$, where H is any unknown distribution function. Let $T = T(X_1, \ldots, X_n)$ be a location equivariant estimator. i.e. $T(X_1+c, X_2+c, \ldots, X_n+c) = T(X_1, \ldots, X_n)+c$ for all $c \in \mathbf{R}^1$. We confine ourselves to the location invariant estimators. Let $M(G,T)$ denote a median of the distribution of T if $X_1, \ldots, X_n$ have the distribution G. A statistic T is a median-unbiased estimator of θ, if $M(F,T) = 0$ (or equivalently. if $M(F_\theta, T) = \theta$ for all $\theta \in \mathbf{R}^1$. Given $\varepsilon \in (0, 1/2)$, let $\pi_\varepsilon(F) = \{G = (1-\varepsilon)F(x) + \varepsilon H(x) : H \text{ any distribution}\}$ be the ε-contamination neighborhood of F, and let $B_\varepsilon(T) = \sup\{M(G_1,T) - M(G_2,T) : G_1, G_2 \in \pi_\varepsilon(F)\}$ be the oscillation of the bias of T when the "true" distribution G of observations runs over $\pi_\varepsilon(F)$. Given two median-unbiased estimators T_1 and T_2, we call T_1 more stable (or more robust) than T_2 if $B_\varepsilon(T_1) < B_\varepsilon(T_2)$. We call T the most stable (or most robust) median-unbiased estimator of θ under ε-contamination if $B_\varepsilon(T) \le B_\varepsilon(T')$ for all median-unbiased estimators T'. It appears that the most concentrated (See Sec. 3.3.1) most stable median-unbiased estimator of θ under ε-contamination is of the form $T^* = \hat{x}_{RZ}(q^*) - F^{-1}(q^*)$ for a suitable q^*, where $\hat{x}_{RZ}(q)$ is the most concentrated median-unbiased estimator constructed in Sec. 3.3.1.

Let us formulate the result formally.

(1) By unimodality of F, for every $\delta > 0$ there exists $x = x(\delta)$ such that $F(x) - F(x - \Delta^*) \le \varepsilon/(1-\varepsilon)$ is nondecreasing for $x < x(\delta)$ and nonincreasing for $x \ge x(\delta)$.

(2) The function

$$\Delta(z) = F^{-1}\left(\frac{z}{1-\varepsilon}\right) - F^{-1}\left(\frac{z-\varepsilon}{1-\varepsilon}\right), \quad \varepsilon < z < 1-\varepsilon$$

is continuous on $(\varepsilon, 1-\varepsilon)$. Let z^* be a point that minimizes $\Delta(z)$ and let $\Delta^* = \Delta(z^*)$. We assume that $z^* \in (\varepsilon, 1-\varepsilon)$; the condition ensures the uniqueness of the quantile of order z^* of each distribution $G \in \pi_\varepsilon(F)$.

(3) For $z \in (\varepsilon, 1-\varepsilon)$, define $x = F^{-1}\left(\dfrac{z}{1-\varepsilon}\right)$. Then $F^{-1}\left(\dfrac{z-\varepsilon}{1-\varepsilon}\right) = x - \Delta(z) \le x - \Delta^*$, and $F(x) - F(x-\Delta^*) \le \varepsilon/(1-\varepsilon)$. Denote $x^* = F^{-1}\left(\dfrac{z^*}{1-\varepsilon}\right)$. Then $F(x^*) - F(x^* - \Delta^*) = \varepsilon/(1-\varepsilon)$. By (1) it follows that $F(x) - F(x-\Delta^*)$ is nondecreasing for $x < x^*$ and nonincreasing for $x \ge x^*$.

(4) Let $q_G(z), z \in (\varepsilon, 1-\varepsilon)$, be the unique quantile of $G \in \pi_\varepsilon(F)$. Let $\hat{q}(z)$ be a distribution-free median-unbiased estimator of $q_G(z)$, i.e. an estimator such that $M(G, \hat{q}(z)) = q_G(z)$ for every $G \in \pi_\varepsilon(F)$ (such as constructed in Sect. 3.3.1).

Theorem 1. $T^* = \hat{q}(z^*) - q_F(z^*)$ is the most stable median-unbiased estimator of the location parameter θ.

Proof. The first part of the proof consists in demonstrating that $B_\varepsilon(T^*) \le \Delta^*$. In the second part we show that $B_\varepsilon(T) \ge \Delta^*$ for all median-unbiased estimators T.

For $G \in \pi_\varepsilon(F)$ we have $(1-\varepsilon)F \le G \le (1-\varepsilon)F + \varepsilon$ and in consequence

$$((1-\varepsilon)F + \varepsilon)^{-1}(z^*) \le q_G(z^*) \le ((1-\varepsilon)F)^{-1}(z^*)$$

or $x^* - \Delta^* \le q_G(z^*) \le x^*$ and consequently $x^* - q_F(z^*) - \Delta^* \le q_G(z^*) - q_F(z^*) \le x^* - q_F(z^*)$. Now, $q_G(z^*) - q_F(z^*)$ is the median of T^*, so that $B_\varepsilon(T^*) \le \Delta^*$.

To show that $B_\varepsilon(T) \ge \Delta^*$ for all median-unbiased estimators T it is enough to find two distributions G_1^* and G_2^* such that $M(G_2^*, T) = M(G_1^*) - \Delta^*$ whenever T is a median-unbiased estimator. To this end we define

$$H_1^*(x) = \begin{cases} 0, & \text{if } x < x^* \\ 1 - \dfrac{1-\varepsilon}{\varepsilon}\left(F(x) - F(x-\Delta^*)\right), & \text{if } x \ge x^* \end{cases}$$

and

$$H_2^*(x) = \begin{cases} \dfrac{1-\varepsilon}{\varepsilon}\left(F(x+\Delta^*) - F(x)\right), & \text{if } x < x^* - \Delta^* \\ 1, & \text{if } x \ge x^* - \Delta^* \end{cases}$$

and take $G_i^* = (1-\varepsilon)F + \varepsilon H_i^*, i = 1, 2$. Then $G_1^*, G_2^* \in \pi_\varepsilon(F)$, $G_2^*(x) = G_1^*(x + \Delta^*)$ and the result follows. □

As an obvious result we obtain that for F symmetric around zero and unimodal the most stable median-unbiased estimator of location is sample median. The asymptotic result was given by Huber (1981). The asymptotic result without restriction to symmetric distributions was given in Rychlik et al. (1985). The fixed sample size results presented above come from Zieliński (1988).

4.2.2. Estimating location parameter under ε-contamination with restrictions on contaminants

Consider as above the well know problem of estimating the location θ of the distribution $F_\theta(x) = F(x-\theta)$, where F is assumed to be symmetric around zero (i.e. $F(x) = 1 - F(-x),\ x \in \mathbf{R}^1$) and unimodal (mode $= 0$). As in the previous Section assume that the observations are $\varepsilon-$contaminated and their true distribution is $G_\theta(x) = G(x-\theta)$ such that $G = (1-\varepsilon)F + \varepsilon H$, where H is any distribution. We consider as estimators the statistics $T_n = T(G_n)$ derived from a translation equivariant functional T; here G_n is the empirical distribution function. Suppose that we are interested in finding such a T which minimizes the maximum asymptotic bias $\sup|T(G)-T(F)|$, where the supremum is taken over all H (Huber 1981, p. 11). The well known solution is the sample median (Huber 1981, Section 4.2). Generally, the commonly accepted opinion is that the sample median is the most robust estimator of location if contaminants may spoil the sample (e.g. Borovkov 1998, Browm 1985, Shervish 1995).

It appears that if the class $\mathcal{H}$ of contaminating distributions is smaller than the class of all distributions, the result may be quite different. Suppose that $\mathcal{H}$ is the class of distributions H such that

(i) support of each distribution $H \in \mathcal{H}$ is in the interval $\left(F^{-1}\left(\dfrac{1}{2(1-\varepsilon)}\right), +\infty\right)$,

(ii) contaminating distribution H has the finite expectation μ which satisfies the condition $\varepsilon\mu \le F^{-1}\left(\dfrac{1}{2(1-\varepsilon)}\right)$.

Theorem 2. (Zieliński 1987). For every $\varepsilon \in (0, 1/2)$ and for every contaminating distribution $H \in \mathcal{H}$, the bias of the sample mean, denoted by $T_1(G)$, is smaller than that of the sample median, denoted by $T_2(G)$.

Proof is trivial: $T_1(G) = \varepsilon\mu$ and $T_2(G) = F^{-1}\left(\dfrac{1}{2(1-\varepsilon)}\right)$. □

Conditions (i) and (II) are realistic. Condition (i) says that a "very small outlier is no outlier". For example, if F is normal distribution $N(0,\sigma^2)$ and $\varepsilon = 0.05$ than X may be considered as outlier (or contaminating observations) only if $X > 0.0660\sigma$; appropriate lower limits for outlier for $\varepsilon = 0.1$ and $\varepsilon = 0.01$ are 0.1397σ and 0.0127σ. If F is exponential distribution $1-\exp\{-x/\theta\}$ with mean θ than for $\varepsilon = 0.1,\ 0.05,\ 0.01$ an observation may be considered as outlier if it is greater than $0.811\theta,\ 0.747\theta,\ 0.703\theta$, respectively. Condition (ii) ensures that large outliers can occur with small probability.

Let us now relax the above conditions (i) and (ii) and suppose that $\mathcal{H}$ is the class of all contaminants which have a finite first moment (finite expectation). It appears that if we know that expectations of contaminants is absolutely not greater than a given $C > 0$, we are able to improve the functional T for our estimator $T_n = T(G_n)$. We assume that $C > C_0 = F^{-1}\left(\frac{1}{2(1-\varepsilon)}\right)$; observe that C_0 is the median of $G = (1-\varepsilon)F + \varepsilon H$ if H is so far to the right that the half of the mass of the distribution G comes from the distribution F.

For a given H, if $0<H(0)<1$, define

$$H^+(x) = \begin{cases} \dfrac{H(x) - H(0)}{1 - H(0)}, & \text{if } x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

and

$$H^-(x) = \begin{cases} \dfrac{H(x)}{H(0)}, & \text{if } x \leq 0 \\ 1, & \text{otherwise.} \end{cases}$$

If $H(0) = 0$ then define $H^-(x) = 0$ for $x \leq 0$ and if $H(0) = 1$ then define $H^+(0) = 1$ for $x \geq 0$. If a contaminant X has a distribution H, then H^+ is the distribution of $X^+ = X$, if $X \geq 0$ and $= 0$, otherwise. Similarly, H^- is the distribution of $X^- = -X$ if $X \leq 0$ and $= 0$ otherwise. We assume that $EX^+ < C$ and $EX^- < C$.

By the well known inequality for a positive random variable ξ with finite expectation $E\xi$

$$P\{\xi \geq t\} \leq \frac{E\xi}{t}, \quad t > 0,$$

we obtain

(3)
$$H^+(x) \geq 1 - \frac{C}{x}, \quad x > 0,$$
$$H^-(x) \leq -\frac{C}{x}, \quad x < 0.$$

Let

$$L(x) = (1-\varepsilon)F(x) + \varepsilon \begin{cases} 0, & \text{if } x \leq C \\ 1 - \dfrac{C}{x}, & \text{if } x > C \end{cases}$$

(4)

$$U(x) = (1-\varepsilon)F(x) + \varepsilon \begin{cases} -\dfrac{C}{x}, & \text{if } x \leq -C \\ 1, & \text{if } x > -C \end{cases}$$

and define

$$\mathcal{N}(\varepsilon, C) = \{G = (1-\varepsilon)F + \varepsilon H, L < G < U\}.$$

According to the definition, $\mathcal{N}(\varepsilon, C)$ is a "neighborhood" of the model distribution F that contains all distributions which can be obtained in the result of ε-contaminating F by any distribution with finite moment bounded by C. The appropriate ε-neighborhood of F without restrictions on contaminants is $\{G : (1-\varepsilon)F < G < (1-\varepsilon)F + \varepsilon\}$ (Fig. 4.2 with $F = N(0,1)$).

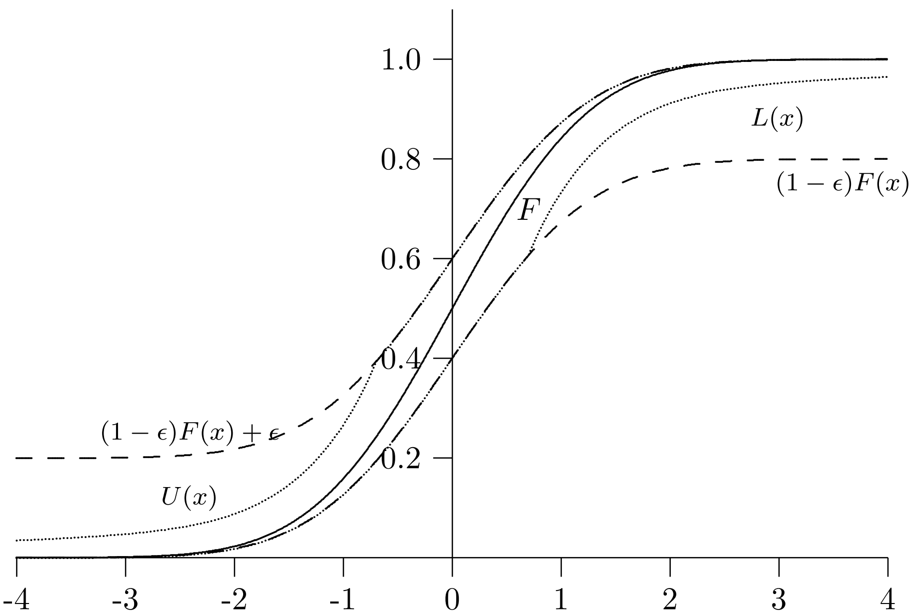


Fig.4.2

For $T \in \mathcal{T}$, let

$$B_{\varepsilon,C}(T) = \sup_{G_1, G_2 \in \mathcal{N}(\varepsilon, C)} |T(G_1) - T(G_2)|.$$

For $q \in (0,1)$ define

$$\delta(q) = L^{-1}(q) - U^{-1}(q)$$

and suppose that there exists $q^* \in (0,1)$ such that

$$\delta(q^*) \leq \delta(q), \quad q \in (0,1).$$

Let $\Delta(x) = \delta(L(x))$, $-\infty < x < \infty$, and denote $\Delta^* = \frac{1}{2}\Delta(L^{-1}(q^*))$. For $q = 0.5$ we have $\delta(q) = 2C_0$ so that $\Delta^* \leq C_0$. As an estimator of location θ we consider $\hat{\theta}_{q^*} = T_{q^*} - F^{-1}(q^*)$. Due to the fact that $|\hat{\theta}_{q^*}(G_1) - \hat{\theta}_{q^*}(G_2)| = |T_{q^*}(G1) - T_{q^*}(G_2)|$, to demonstrate the optimality of $\hat{\theta}_{q^*}$ it is enough to prove the following Theorem.

Theorem 3. $B_{\varepsilon,C}(T_{q^*}) \leq B_{\varepsilon,C}(T)$ for all $T \in \mathcal{T}$.

Proof. Define

$$G^U(x) = \begin{cases} L(x+2\Delta^*), & \text{if } x \leq -\Delta^*, \\ U(x), & \text{if } x > -\Delta^*. \end{cases}$$

By (4)

$$G^U(x) = \begin{cases} (1-\varepsilon)F(x) + \varepsilon \cdot \dfrac{1-\varepsilon}{\varepsilon}\left[F(x+2\Delta^*) - F(x)\right], & \text{if } x \leq -\Delta^*, \\ (1-\varepsilon)F(x) + \varepsilon, & \text{if } x > -\Delta^*. \end{cases}$$

The function

$$H_U^0(x) = \frac{1-\varepsilon}{\varepsilon}\left[F(x+2\Delta^*) - F(x)\right], x \leq -\Delta^*,$$

has the following properties:

1) $H_U^0(x) \geq 0$;

2) by symmetry and unimodality, $f(x+2\Delta^*) - f(x) > 0$ for $x \leq -\Delta^*$, so that $H_U^0(x)$ is increasing;

3) $H_U^0(-\Delta^*) = \dfrac{1-\varepsilon}{\varepsilon}\left[2F(\Delta^*) - 1\right] \leq \dfrac{1-\varepsilon}{\varepsilon}\left[2F(C_0) - 1\right] = 1.$

It follows that

$$H_U(x) = \begin{cases} H_U^0(x), & \text{if } x \leq -\Delta^*, \\ 1, & \text{if } x > -\Delta^*, \end{cases}$$

is a distribution function and in consequence $G^U(x)$ is a distribution function of the form $(1-\varepsilon)F(x) + \varepsilon H_U(x)$ and belongs to $\mathcal{N}(\varepsilon, C)$.

Define the function

$$G^L(x) = \begin{cases} L(x), & \text{if } x \leq \Delta^*, \\ U(x-2\Delta^*), & \text{if } x > \Delta^*. \end{cases}$$

By similar arguments to those concerning $G^U(x)$ we conclude that $G^L(x) \in \mathcal{N}(\varepsilon, C)$. It is easy to check that $G^U(x) = G^L(x+2\Delta^*)$ so that for $T \in \mathcal{T}$ we have $T(G^U) = T(G^L) + 2\Delta^*$ and in consequence $B_{\varepsilon,C}(T) \geq 2\Delta^*$ for all $T \in \mathcal{T}$.

For $G \in \mathcal{N}(\varepsilon, C)$ we have

$$T_{q^*}(U) \leq T_{q^*}(G) \leq T_{q^*}(L).$$

By the definition of q^* we have $T_{q^*}(L) - T_{q^*}(U) = 2\Delta^*$ so that $B_{\varepsilon,C}(T_{q^*}) \leq 2\Delta^*$. □

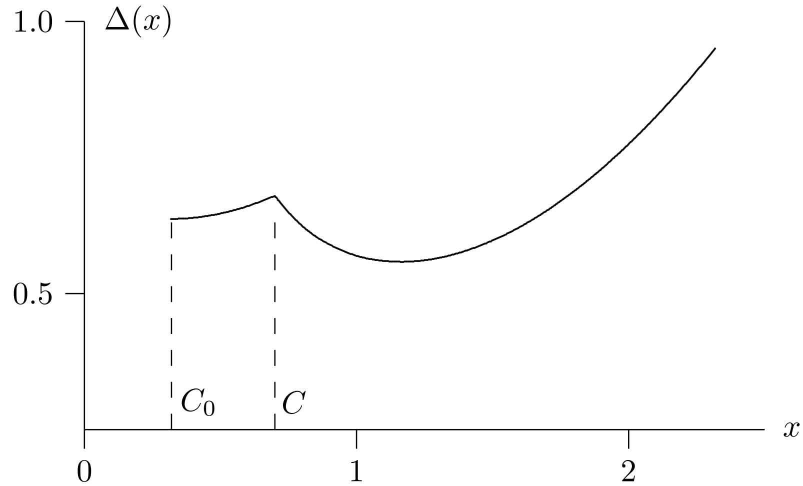


Fig. 4.3

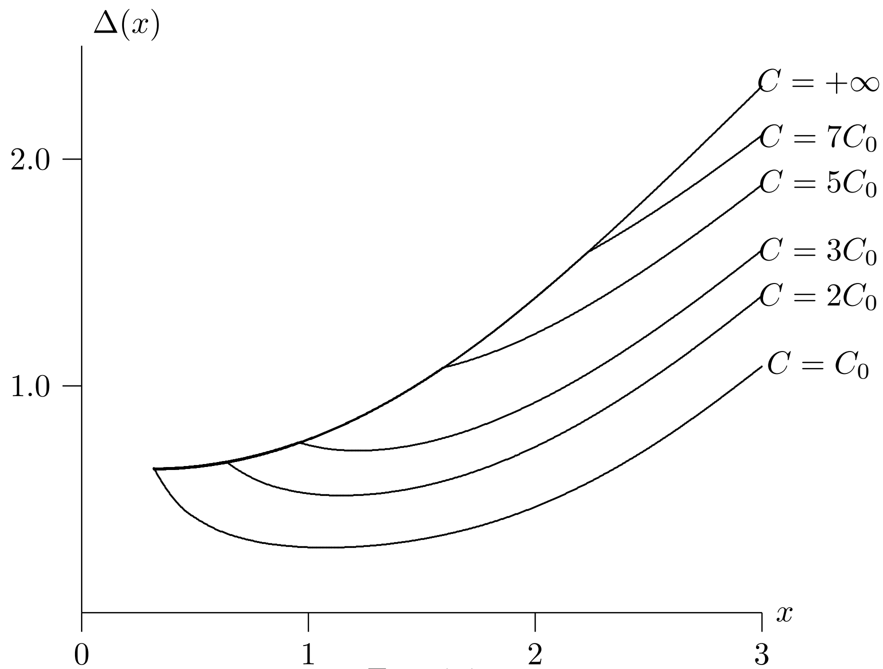


Fig. 4.4

If $x \in (-C, C)$ then $L(x) = (1-\varepsilon)F(x)$ and $U(x) = (1-\varepsilon)F(x) + \varepsilon$, so that

$$\min_{U(-C)\le q\le L(C)} \delta(q) = \min_{U(-C)\le q\le L(C)} \left[F^{-1}\left(\frac{q}{1-\varepsilon}\right) - F^{-1}\left(\frac{q-\varepsilon}{1-\varepsilon}\right)\right]$$
$$= 2F^{-1}\left(\frac{1}{2(1-\varepsilon)}\right) = 2C_0$$

for $q = \dfrac{1}{2}$. It follows that without a moment condition, i.e. for $C = +\infty$, we have $q^* = \dfrac{1}{2}$: then the best estimator is the median $T_{0.5}$.

If $C < +\infty$ then, given F and ε, it may happen that $\Delta(x)$ has some other minima in $\{x : x - \Delta(x) < -C\}$ or $\{x : x > C\}$, and the minima are smaller than $\Delta(C_0)$ for $q^* = \frac{1}{2}$. These minima give us more stable estimators. No general results for any class of F are known.

As an example we have chosen the case of normal $N(0,1)$ parent distribution F with $\varepsilon = 0.2$ and restriction on contaminants with $C = 0.7$. Now $C_0 = 0.3186$, so that for the optimal estimator, i.e. for the median, say M, in the model without restrictions we have $B_{0.2}(M) = 2C_0 = 0.6372$. In the model with restrictions, due to symmetry we may confine ourselves to considering the function $\Delta(x)$ on the interval $(C_0, +\infty)$ and to study its minimum on the interval $(C, +\infty)$. For $\varepsilon = 0.2$ and $C = 0.7$, the function is presented in Fig. 4.3. Numerical calculations give us $q^* = 0.7824$ with $B_{0.2,0.7}(T_{0.7824}) = 0.5589$ which significantly improves the estimator.

Functions $\Delta(x)$ for some other values of C are exhibited in Fig. 4.4. Numerical calculations give us the conclusion: if $C_0 \le C < 0.8245$ then the optimal estimator is T_{q^*} with some $q^* \neq \frac{1}{2}$ and the median is not the best choice. If the expected value of the contaminant is large enough ($C > 0.8245$), then the median is the most stable estimator.

4.3. Distribution-free quantile estimator in parametric models; how much do we lose?

Suppose we wish to estimate a quantile of an unknown distribution from a parametric family. Typically we construct a (best in a sense) estimator of the quantile. But if the parent distribution is heavily contaminated the estimator may be highly unsatisfactory. We may always estimate the quantile by a (best in a sense) nonparametric estimator. We gain in stability (robustness) of the estimator, but we lose in the effect of resigning from the information about the specific parametric family of distributions.

In what follows we present the problem for estimating quantiles of the one parameter exponential distribution with distribution function $W_\theta(x) = 1 - e^{-\theta x}$, $x > 0, \theta > 0$ similar problem, with an excursion to asymptotics, was considered in Wieczorkowski et al (1991). Let $X_1, \ldots, X_n$ be a sample from the distribution $W_\theta(x)$ and let $S_n = \sum_{j=1}^{n} X_j$. It is well known that S_n is a random variable distributed according to Gamma distribution $\Gamma(n, \theta)$ with the shape parameter n and the scale parameter θ. All statistics we consider below are equivariant with respect to the scale parameter so that in consequence we consider the case of $\theta = 1$ only. Let $Q_\Gamma(n, P)$ denote the P-th quantile of the distribution $\Gamma(n, 1)$.

The most concentrated median-unbiased estimator of the q-th quantile of the distribution $W_\theta(x)$ is given by $c_q S_n$ where $c_q = -\log(1-q)/Q_\Gamma(n, 1/2)$ (cf. Lehmann (1986), Ch.3.5). As a measure of accuracy of the estimator we take its interquartile range which is given by the formula

$$R_n = c_q \Big(Q_\Gamma(n, 3/4) - Q_\Gamma(n, 1/4) \Big).$$

In practical applications, the exponential distribution is usually an approximation only. But if the random variable under consideration is not exactly exponentially distributed, then the estimator $c_q S_n$ is not more median-unbiased. If we still insist to have a median-unbiased estimator, we may use the most concentrated median unbiased estimator $\hat{x}_{RZ}$ from Sec. 3.1.1. One may expect that the interquartile range of $\hat{x}_{RZ}$, denoted by $r_n^{(E)}$, in the exponential model with the parent distribution $E = W_1$, would be greater than R_n. A natural question arises: how large should be sample size N for $\hat{x}_{RZ}$ to get the accuracy R_n of the estimator $c_q S_n$ from a sample of size n. To say that formally: given the order q of the quantile to be estimated and a sample size n for the estimator $c_q S_n$, find the smallest sample size $N = N(q, n)$ for the estimator $\hat{x}_{RZ}$ such that $r_N^{(E)} \le R_n$.

As in Sec. 3.1.1, for the most concentrated median-unbiased nonparametric estimator $\hat{x}_{RZ}(q)$ we have

$$P_F\{\hat{x}_{RZ}(q) \le x_P(F)\} = \lambda_k P\{U_{k:n} \le P\} + (1-\lambda_k) P\{U_{k+1:n} \le P\}, \quad P \in (0,1),$$

where k is an integer satisfying

$$\pi_{k+1}(q) \ge \frac{1}{2} \ge \pi_k(q)$$

and

$$\lambda_k = \frac{\frac{1}{2} - \pi_{k+1}(q)}{\pi_k(q) - \pi_{k+1}(q)}$$

with

$$\pi_k(q) = \sum_{j=k}^{n} \binom{n}{j} q^j (1-q)^{n-j}.$$

It follows that

$$\begin{aligned} r_n^{(E)} &= x_{P(3/4,n)}(E) - x_{P(1/4,n)}(E) \\ &= \log\left(\frac{1-P(1/4,n)}{1-P(3/4,n)}\right) \end{aligned}$$

where, given n and q, $P(\tau, n)$ is the unique solution, with respect to P, of the equation

$$\lambda_k P\{U_{k:n} \le P\} + (1-\lambda_k) P\{U_{k+1:n} \le P\} = \tau.$$

Values of R_n and $r_n^{(E)}$ for some q and n are presented in Tab. 4.1 and $N(n)$ together with $N(n)/n$ in Tab. 4.2.

Tab. 4.1. R_n and $r_n^{(E)}$

n	$q = 0.1$		$q = 0.25$		$q = 0.5$		$q = 0.75$		$q = 0.9$	
	R_n	$r_n^{(E)}$	R_n	$r_n^{(E)}$	R_n	$r_n^{(E)}$	R_n	$r_n^{(E)}$	R_n	$r_n^{(E)}$
5	0.065	0.219	0.179	0.374	0.431	0.577	0.862	1.139	1.432	2.200
10	0.046	0.156	0.125	0.246	0.300	0.439	0.600	0.738	0.997	1.411
15	0.037	0.119	0.101	0.203	0.244	0.343	0.488	0.608	0.810	1.078
20	0.032	0.105	0.087	0.178	0.211	0.306	0.421	0.534	0.670	0.950
25	0.029	0.091	0.078	0.158	0.188	0.268	0.376	0.474	0.625	0.823
30	0.026	0.085	0.071	0.142	0.172	0.249	0.343	0.426	0.570	0.763
35	0.024	0.077	0.066	0.132	0.159	0.227	0.317	0.396	0.527	0.691
40	0.023	0.073	0.062	0.124	0.148	0.215	0.297	0.373	0.493	0.656
45	0.021	0.068	0.058	0.117	0.140	0.200	0.280	0.351	0.465	0.608
50	0.020	0.065	0.055	0.110	0.133	0.192	0.265	0.330	0.441	0.584

Tab. 4.2. $N(n)$ and $N(n)/n$

q	$n = 5$		$n = 10$		$n = 20$		$n = 50$	
0.1	-	-	98	9.8	199	9.95	499	9.98
0.2	25	5.0	50	5.0	100	5.0	251	5.02
0.3	17	3.4	34	3.4	68	3.4	168	3.36
0.4	13	2.6	26	2.6	51	2.55	128	2.56
0.5	11	2.2	21	2.1	41	2.05	105	2.10
0.6	9	1.8	18	1.8	36	1.8	89	1.78
0.7	8	1.6	16	1.6	32	1.6	81	1.62
0.8	8	1.6	17	1.7	32	1.6	78	1.56
0.9	-	-	17	1.7	35	1.75	86	1.72

Though the ratios $N(n)/n$ seem not to depend of n, no general result to this end is known. What is more, the results highly depend on the parametric family of distributions at hand: analogous results for the family $\{N(\mu, 1), \mu \in \mathbf{R}\}$ of normal distribution are presented in Tab. 4.3 and Tab. 4.4.

Tab. 4.3. R_n and $r_n^{(N)}$

n	$q = 0.1$		$q = 0.25$		$q = 0.5$		$q = 0.75$		$q = 0.9$	
	R_n	$r_n^{(N)}$	R_n	$r_n^{(N)}$	R_n	$r_n^{(N)}$	R_n	$r_n^{(N)}$	R_n	$r_n^{(N)}$
5	0.603	1.194	0.603	0.883	0.603	0.720	0.603	0.883	0.603	1.194
10	0.427	0.795	0.427	0.578	0.427	0.548	0.427	0.578	0.427	0.795
15	0.348	0.609	0.348	0.477	0.348	0.429	0.348	0.477	0.348	0.609
20	0.302	0.539	0.301	0.419	0.302	0.383	0.302	0.419	0.302	0.539
25	0.270	0.467	0.270	0.372	0.270	0.335	0.270	0.372	0.270	0.467
30	0.246	0.434	0.246	0.335	0.246	0.311	0.246	0.335	0.246	0.434
35	0.228	0.393	0.228	0.311	0.228	0.284	0.228	0.311	0.228	0.393
40	0.214	0.373	0.213	0.293	0.213	0.269	0.213	0.293	0.213	0.373
45	0.201	0.346	0.201	0.276	0.201	0.251	0.201	0.276	0.201	0.346
50	0.191	0.332	0.191	0.260	0.191	0.240	0.191	0.260	0.191	0.332

Tab. 4.4. $N(n)$ and $N(n)/n$ for the normal distribution

q	$n = 5$		$n = 10$		$n = 20$		$n = 50$	
0.1	-	-	32	3.2	61	3.05	146	2.92
0.2	12	2.4	22	2.2	42	2.1	103	2.06
0.3	9	1.8	18	1.8	35	1.75	88	1.76
0.4	9	1.8	17	1.7	33	1.65	81	1.62
0.5	9	1.8	17	1.7	31	1.55	79	1.58
0.6	9	1.8	17	1.7	33	1.65	81	1.62
0.7	9	1.8	18	1.8	35	1.75	88	1.76
0.8	12	2.4	22	2.2	42	2.1	103	2.06
0.9	-	-	32	3.2	61	3.05	146	2.92

From the examples above it follows that one can hardly expect to state a general rule concerning the loss in efficiency when using the simple universal nonparametric estimator of a quantile instead of a specific quantile estimator for a given family of distributions. In computer simulations we usually do not need be very restrictive with respect to the size of samples generated and then the simple nonparametric estimator may be preferred.

5

Optimal interval estimation

In this Chapter we present results from Zieliński at al (2004).

Nonparametric confidence intervals for quantiles are of the form $(X_{i:n}, X_{j:n})$ with suitably chosen order statistics $X_{i:n}$ and $X_{j:n}$. The problem we consider in this Section is to construct optimal confidence interval for a q-th quantile with exact predetermined confidence level $\gamma \in (0,1)$.

Given n and $q \in (0,1)$, it is well known that

$$P_F\{X_{i:n} \leq x_q(F) \leq X_{j:n}\} = p(i,j;n,q), \quad 1 \leq i < j \leq n,$$

where

$$p(i,j;n,q) = \sum_{s=i}^{j-1} \binom{n}{s} q^s (1-q)^{n-s}$$

does not depend on F. Then $(X_{i:n}, X_{j:n})$ with i and j such that

$$p(i,j;n,q) = \gamma$$

provides a nonparametric (distribution-free) confidence interval for the qth quantile $x_q(F)$ at the confidence level γ (see, for example, David 1981).

If F is the uniform distribution $U(0,1)$ we shall write P instead of P_F. By the "distribution-free property"

$$P_F\{X_{i:n} \leq x_q(F) \leq X_{j:n}\} = P\{U_{i:n} \leq q \leq U_{j:n}\},$$

all our considerations concerning confidence intervals $(X_{i:n}, X_{j:n})$, $(X_{k:n}, +\infty)$, or $(-\infty, X_{k:n})$ may be performed in terms of confidence intervals $(U_{i:n}, U_{j:n})$, $(U_{k:n}, 1)$, or $(0, U_{j:n})$, respectively, where $U_{i:n}, U_{j:n}$ are order statistics from the uniform distribution $U(0,1)$.

First of all observe that for a given γ two-sided confidence intervals exist if and only if

$$P\{U_{1:n} \le q \le U_{n:n}\} \ge \gamma.$$

The condition is equivalent to the condition

$$q^n + (1-q)^n \le 1 - \gamma. \tag{1}$$

For example, to get a two-sided confidence interval at the confidence level $\gamma = 0.95$ for the qth quantile with $q = 0.01$ the sample size n has to be not smaller than 299.

If condition (1) is satisfied than the possible exact levels of confidence intervals of the form $(X_{i:n}, X_{j:n})$ are determined by a discrete binomial distribution and typically the coverage probability cannot be rendered equal to a preselected value γ. To overcome the difficulty, or at least to construct confidence intervals with coverage probability as close as possible to the prescribed value, many different constructions have been proposed. The most recent ones can be find in Beran and Hall (1993) or in Hutson (1999), some others in references cited in those papers. The problem however is that under constructions proposed coverage probability is not exactly equal to the prescribed confidence level or/and does depend on the (unknown) distribution F. Our idea is to take two integer-valued random variables I and J, independent of the observed random variable X, such that $P\{I = i, J = j\} = \lambda_{ij}$, and consider (randomized) confidence intervals of the form $(X_{I:n}, X_{J:n})$. Now again $P_F\{X_{I:n} \le x_q(F) \le X_{J:n}\} = P\{U_{I:n} \le q \le U_{J:n}\} = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \lambda_{ij} p(i, j; n, q)$ does not depend on F. It is obvious that under condition (1) there exist infinitely many $\lambda_{ij}, 1 \le i < j \le n$, such that $P\{U_{I:n} \le q \le U_{J:n}\} = \gamma$, which gives us infinitely many exact nonparametric (randomized) confidence intervals for a given quantile at a given confidence level. The problem is to choose the best one. Similarly one-sided confidence intervals can be considered.

5.1. Optimal two-sided exact confidence intervals

In this Section we assume that condition (1) holds. For a confidence interval $(U_{I:n}, U_{J:n})$ with random indices I and J such that

$$P\{I = i, J = j\} = \lambda_{ij}, \quad 1 \le i < j \le n,$$
$$\lambda_{ij} \ge 0, \quad \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \lambda_{ij} = 1$$

we define (according to a suggestion by David (1981), p. 16) the length of the interval as

$$E(J-I)=\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(j-i)\lambda_{ij}.$$

To find the best random indices I and J we may proceed as follows.

Making use of unimodality of the binomial distribution, take k which maximizes $b(k;n,q)=\binom{n}{k}q^k(1-q)^{n-k}$; then $(U_{k:n},U_{k+1:n})$ is our first approximation to the confidence interval. Now take $(U_{k-1:n},U_{k+1:n})$ if $b(k-1;n,q)>b(k+1;n,q)$ and $(U_{k:n},U_{k+2:n})$ otherwise as the next approximation to the confidence interval. Proceed until

$$P\{U_{i:n}\le q\le U_{j:n}\}<\gamma \quad \text{and} \quad P\{U_{i':n}\le q\le U_{j':n}\}>\gamma$$

for some ($i'=i-1$ and $j'=j$) or ($i'=i$ and $j'=j+1$). Now calculate λ such that

$$\lambda P\{U_{i:n}\le q\le U_{j:n}\}+(1-\lambda)P\{U_{i':n}\le q\le U_{j':n}\}=\gamma$$

and as the confidence interval to be constructed take $(U_{i:n},U_{j:n})$ with probability λ or $(U_{i':n},U_{j':n})$ with probability $1-\lambda$.

If $P\{U_{i:n}\le q\le U_{j:n}\}=\gamma$ for some $1\le i<j\le n$ then the confidence interval to be constructed is $(U_{i:n},U_{j:n})$.

It is obvious that by the very construction the resulting confidence interval is the shortest one (in the nonparametric sense: minimum $E(J-I)$) at exactly the predetermined level γ.

For example, if $n=10$ and the proposed confidence level is $q=0.9$, then for the quantile of order $q=0.3$ we have

$$P_F\{X_{1:10}\le x_q(F)\le X_{5:10}\}=0.8216,\quad P_F\{X_{1:10}\le x_q(F)\le X_{6:10}\}=0.9245$$

and hence $\lambda=0.2381$. To get the optimal confidence interval for the q-th quantile $x_q(F)$ with $q=0.3$ take $(X_{1:10},X_{5:10})$ with probability $\lambda=0.2381$ or $(X_{1:10},X_{6:10})$ with probability $1-\lambda$.

3. Optimal one-sided confidence interval

The discussion of lower and upper bounds for confidence intervals is completely parallel; we confine ourselves to confidence intervals of the form $(U_{k:n}, 1)$.

First of all observe that if $P\{U_{1:n} \leq q\} < \gamma$ then the maximal nontrivial confidence interval $(U_{1:n}, 1)$ has confidence level smaller than the prescribed value γ and no nontrivial confidence interval at that level is available (nontrivial means other than (0,1)). In what follows we assume that $P\{U_{1:n} \leq q\} = 1 - (1-q)^n \geq \gamma$.

Let I be a random index with distribution $P\{I = i\} = \lambda_i, i = 1, 2, \ldots, n$. Following Lehmann (1986, Sec. 3.5) we define a confidence interval $(U_{I:n}, 1)$ as the uniformly most accurate confidence interval for the qth quantile at the confidence level γ if

$$P\{U_{I:n} \leq q\} = \gamma$$

and

$$P\{U_{I:n} \leq q'\} \leq P\{U_{J:n} \leq q'\}$$

for all $q' < q$ and for all random indices J. For rationale for the choice of the criterion, also in terms of an appropriate loss function, see Lehmann (1986): the idea is that the best lower confidence bound should underestimate the qth quantile by as little as possible.

Write $p_k(n,q) = p(k, n+1; n, q) = \sum_{s=k}^{n} \binom{n}{s} q^s (1-q)^{n-s} = P\{U_{k:n} \leq q\}$ and consider the following construction. If $p_k(n,q) = \gamma$ for some k, then take $(U_{k:n}, 1)$ as the confidence interval to be constructed. Otherwise, let k be an integer such that $p_k(n,q) > \gamma > p_{k+1}(n,q)$ and define

$$\lambda = \frac{\gamma - p_{k+1}(n,q)}{p_k(n,q) - p_{k+1}(n,q)}$$

Then $(U_{I:n}, 1)$ with the random index I such that

$$P\{I = k\} = \lambda, \quad P\{I = k+1\} = 1 - \lambda$$

is the uniformly most accurate confidence interval for the qth quantile at the confidence level γ.

A detailed proof with solving an appropriate linear programming problem is presented in Zieliński at al (2004). The proof may also be deduced from a construction of the most powerful test for testing $H : x_q(F) \le u$ against $K : x_q(F) > u$ (cf Reiss 1989).

It is interesting to note that the above confidence interval has an additional advantage: the lower confidence bound $U_{I:n}$ overestimates the qth quantile by as little as possible. To see that, it is enough to consider the linear programming problem maximizing $P\{U_{J:n} \le q'\}$ for $q' > q$ under the restrictions (2).

6

Asymptotics

Asymptotic theorems in statistical models differ from those in probability theory in that they should hold uniformly in the family of distributions specified by the model: all what statistician knows about the distribution that generates his observations is that it belongs to a specified family of distributions. An asymptotic theorem to be useful for statistician must be expressed in terms of the family of model distributions and not to use any specific information on the parent distribution at hand. An example of statistical asymptotic theorems is the limit law for Kolmogorov statistic

$$D_n = \sup_{-\infty<x<+\infty} |F_n(x) - F(x)|$$

where F_n is empirical distribution function from a sample of size n which comes from the parent distribution F. Theorems like that enable statisticians to do practical inference (tests, confidence intervals, etc) without going into details concerning the parent distribution at hand.

On the other hand, an example of a theorem of a rather moderate statistical applicability is the Berry-Esséen inequality (Serfling 1980)

$$\sup_{-\infty<x<+\infty} |F_n(x) - \Phi(x)| \leq \frac{33}{4} \frac{E|X - EX|^3}{\sigma^3 n^{1/2}}$$

To make use of the inequality a statistician has to know first three moments of the distribution of observations, or he has to confine himself to a rather artificial class of distribution with a know upper bound on the ratios $E|X - EX|^3/\sigma^3$.

Another example of an incorrect application of an asymptotic result is the well known Central Limit Law for Bernoulli trials: if S_n is the number of successes in n Bernoulli trials with probability of success in each trial θ then $(S_n - n\theta)/\sqrt{n\theta(1-\theta)}$ is asymptotically normally $N(0,1)$ distributed. However, for every n and for every $\eta > 0$ one can find θ and x such that $\left| P_\theta\{\frac{S_n - n\theta}{\sqrt{n\theta(1-\theta)}} \leq x\} - \Phi(x) \right| > \frac{1}{2} - \eta$ (non-uniform

Central Limit Theorem - Zieliński 2004). One may conclude that typical difficulties in constructing confidence intervals for θ (e.g. Brown et al. 2001), based on normal approximation, arise from the fact that CLT does not hold uniformly.

6.1. Uniform strong consistency of $X_{k:n}$

Concerning the problem of quantile estimation, it is well known that if $x_q(F)$ is the unique q-th quantile of a distribution function F, then $X_{k(n):n}$ with $k(n)/n \to q$ is a strongly consistent estimator of $x_q(F)$ (Serfling 1980). However, for every $\varepsilon > 0$ and for every, even very large n, $\sup_{F\in\mathcal{F}} P_F\{|X_{k(n):n} - x_q(F)| > \varepsilon\} = 1$. This is a consequence of the fact that in the family of all distribution functions with uniquely defined q-th quantile the almost sure convergence $X_{k(n):n} \to x_q(F)$ is not uniform in $\mathcal{F}$ (see Corollary 1 below).

A simple necessary and sufficient condition for the uniform strong consistency of $X_{k(n):n}$ is given in the following Theorem.

Theorem 1. (Zieliński 1998) *The sample quantile $X_{k(n):n}$ such that $k(n)/n \to q$ as $n \to \infty$ is an uniformly strongly consistent in $\mathcal{F}$ estimator of the q-th quantile $x_q = x_q(F)$ if and only if*

$$(1) \qquad (\forall \varepsilon > 0) \qquad \inf_{F\in\mathcal{F}} \min\{q - F(x_q - \varepsilon), F(x_q + \varepsilon) - q\} > 0.$$

Proof. *(Sufficiency)* Fix $\varepsilon > 0$ and let $\delta = \inf_{F\in\mathcal{F}} \min\{q - F(x_q - \varepsilon), F(x_q + \varepsilon) - q\}$.

In the proof we shall use the following result of Hoeffding (1963): *if $\xi_1, \xi_2, \ldots, \xi_n$ are independent random variables such that, for some finite a and b, $P\{a < \xi_j < b\} = 1$, $j = 1, 2, \ldots, n$, then for $t > 0$,*

$$P\left\{\frac{1}{n}\sum_{j=1}^{n}\xi_j - E\left(\frac{1}{n}\sum_{j=1}^{n}\xi_j\right) \geq t\right\} \leq \exp\{-2nt^2/(b-a)^2\}.$$

Take N such that $q - \delta/2 < k(n)/n < q + \delta/2$ if $n \geq N$. Denote by F_n the empirical distribution function generated by the sample $X_1, X_2, \ldots, X_n$ and by V_j the random variable equal to 1 if $X_j > x_q + \varepsilon$ and equal to 0 otherwise.

Then for $n \geq N$,

$$P_F\{X_{k(n):n} > x_q + \varepsilon\} \leq \exp\{-\frac{n\delta^2}{2}\}$$

and

$$P_F\{X_{k(n):n} < x_q - \varepsilon\} \leq \exp\{-\frac{n\delta^2}{2}\}$$

Hence for each $n \geq N$ and for each $F \in \mathcal{F}$,

$$P_F\{|X_{k(n):n} - x_q| > \varepsilon\} \leq 2\exp\{-\frac{n\delta^2}{2}\}.$$

It follows that for every $n \geq N$ and for each $F \in \mathcal{F}$,

$$P_F\{\sup_{m\geq n} |X_{k(m):m} - x_q| > \varepsilon\} \leq \frac{2\tau^n}{1-\tau},$$

where $\tau = \exp\{-\delta^2/2\}$. Now, the right-hand side tends to 0 as $n\to\infty$, independently of $F \in \mathcal{F}$, and hence $X_{k(n):n} \to x_q(F)$ a.s. uniformly in $\mathcal{F}$.

(Necessity) Suppose that the condition (1) does not hold so that the following condition is true:

$$(2) \qquad \begin{aligned} &(\exists\varepsilon > 0)(\forall\delta > 0)(\exists F \in \mathcal{F}) \\ &\qquad \min\{q - F(x_q - \varepsilon), F(x_q + \varepsilon) - q\} < \delta. \end{aligned}$$

We shall demonstrate that if the condition (2) holds then $(\exists\varepsilon > 0)$ $(\exists\eta > 0)$ $(\forall N)$ $(\exists n \geq N)$ $(\exists F)$ such that

$$P_F\{\sup_{m\geq n} |X_{k(m):m} - x_q(F)| > \varepsilon\} > \eta.$$

It is enough to prove that $(\exists\varepsilon > 0)(\exists\eta > 0)(\forall N)(\exists n \geq N)(\exists F)$ such that

$$P_F\{|X_{k(n):n} - x_q(F)| > \varepsilon\} > \eta.$$

We shall prove a somewhat stronger statement:

$$(3) \qquad \begin{aligned} &(\exists\varepsilon > 0)(\forall\eta > 0)(\forall N)(\exists n \geq N)(\exists F) \\ &\qquad P_F\{|X_{k(n):n} - x_q(F)| > \varepsilon\} > \frac{1}{2} - \eta. \end{aligned}$$

To this end, take ε as determined by (2). Fix any $\eta > 0$ and any N_0. We shall verify that $(\exists n \geq N_0)$ and $(\exists F)$ such that (3) holds.

Consider the incomplete beta-function

$$B(q; k, n-k+1) = \frac{\Gamma(n+1)}{\Gamma(k)\Gamma(n-k+1)} \int_0^q t^{k-1}(1-t)^{n-k} dt.$$

It is well known that $B(q; k, n-k+1) \to \frac{1}{2}$ as $n, k \to \infty$ in such a way that $k/n \to q$. It follows that for every fixed η there exists $N(\eta)$ such that for all $n \geq N(\eta)$,

$$\frac{1}{2} - \frac{1}{2}\eta < B(q; k, n-k+1) < \frac{1}{2} + \frac{1}{2}\eta.$$

Fix any $n \geq \max\{N_0, N(\eta)\}$. From the continuity of $B(q; k, n-k+1)$ for $q \in (0,1)$ it follows that for the given η there exists $\delta > 0$ such that $0 < q - \delta < q < q + \delta < 1$ and

$$B(q-\delta; k, n-k+1) > B(q; k, n-k+1) - \frac{1}{2}\eta > \frac{1}{2} - \eta,$$
$$B(q+\delta; k, n-k+1) < B(q; k, n-k+1) + \frac{1}{2}\eta < \frac{1}{2} + \eta.$$

Given $\delta > 0$, by (2) there exists F such that

$$F(x_q - \varepsilon) > q - \delta \quad \text{or} \quad F(x_q + \varepsilon) < q + \delta.$$

Hence

$$B(F(x_q - \varepsilon); k, n-k+1) > \frac{1}{2} - \eta \quad \text{or} \quad B(F(x_q + \varepsilon); k, n-k+1) < \frac{1}{2} + \eta.$$

Taking into account that

$$B(F(t); k, n-k+1) = P_F\{X_{k:n} \leq t\}$$

we obtain

$$P_F\{X_{k(n):n} \leq x_q(F) - \varepsilon\} \geq \frac{1}{2} - \eta$$

or

$$P_F\{X_{k(n):n} \leq x_q(F) + \varepsilon\} \leq \frac{1}{2} + \eta$$

and finally (3). □

Corollary 1. *In the nonparametric family $\mathcal{F}$ of all distributions with strictly increasing distribution functions, for every $q \in (0,1)$, for each $\varepsilon > 0$, for each $\eta > 0$, and for every sufficiently large n there exists a distribution F such that*

$$P_F\{|X_{k(n):n} - x_q(F)| > \varepsilon\} \geq 1 - \eta.$$

□

In other words: for every positive ε and for every sufficiently large n,

$$\sup_{F \in \mathcal{F}} P_F\{|X_{k(n):n} - x_q(F)| > \varepsilon\} = 1. \tag{4}$$

Observe that "sufficiently large n" is needed for $k(n)/n$ to be close to q. As a distribution F one can take any strictly increasing continuous distribution such that, under fixed $\varepsilon > 0$, both $F(x_q(F) - \varepsilon)$ and $F(x_q(F) + \varepsilon)$ are sufficiently close to $F(x_q(F)) = q$.

The next Corollary demonstrates that (4) may hold even in "small parametric families of distributions.

Corollary 2. (Boratyńska et al. 1997). *In the family of distributions*

$$F_\alpha(x) = \frac{1}{2}\left(1 + x^\alpha - (1-x)^\alpha\right), \quad 0 < x < 1, \quad \alpha \geq 1$$

the result (4) holds for $q = 0.5$, for each $\varepsilon \in (0, 0.5)$ and for every sufficiently large n.

□

Corollary 3. (ε-contamination). *Let F be a continuous and strictly increasing distribution function and, for a fixed $c \in (0, \frac{1}{2})$, let $\mathcal{G} = \{G = (1-c)F + H, H \in \mathcal{H}\}$, where $\mathcal{H}$ is the family of all distribution functions. Then for every $G \in \mathcal{G}$ and for every $q \in (0,1)$, if $k(n)/n \to q$, then $X_{k(n):n}$ is a strongly consistent estimator of the unique q-th quantile $x_q(G)$. The convergence is uniform in $\mathcal{G}$ iff $q \in (c, 1-c)$. In the Tukey model with the distribution function Φ of the standard normal distribution $N(0,1)$ as F and the family $\{N(0,\sigma), \sigma \geq 1\}$ as $\mathcal{H}$, the convergence is uniform iff $q \in (\frac{c}{2}, 1 - \frac{c}{2})$.* □

Corollary 4. *If F is a strictly increasing distribution function and $\mathcal{G} = \{G_\theta(x) = F(x-\theta), -\infty < \theta < \infty\}$ is the shift-family generated by F, then $X_{k(n):n}$, where $k(n)/n \to q$, is uniformly strong consistent estimator of the q-th quantile. If F is a strictly increasing distribution function and $\mathcal{G} = \{G_\sigma(x) = F(x/\sigma), \sigma > 0\}$ is the scale-family generated by F, then $X_{k(n):n}$ is not uniformly strong consistent estimator of the q-th quantile.* □

The result enables us to effective application of the strong convergence of sample quantiles in a smaller class of distribution functions. For a fixed $q \in (0,1)$, consider the class $\mathcal{F}(q, \vartheta)$ of all locally (at the qth quantile x_q) continuous and strictly increasing distributions F such that the densities f at the qth quantile x_q satisfy $f(x_q) \geq \vartheta > 0$.

Theorem 2. (Zieliński 2004). *For each $\varepsilon > 0$ and for each $\eta > 0$, there exists $N = N(\varepsilon, \eta)$ such that*

$$P_F\left\{\sup_{n \geq N} \left|X_{k(n):n} - x_q\right| > \varepsilon\right\} < \eta \quad \text{for all } F \in F(q, \vartheta)$$

and

$$N(\vartheta,\varepsilon,\eta) \geq -\frac{8\log\left(\frac{1}{2}\big(1-\exp\big\{-\frac{1}{8}\vartheta^2\varepsilon^2\big\}\big)\eta\right)}{\vartheta^2\varepsilon^2}.$$

Proof. If

$$\delta = \inf_{F\in\mathcal{H}} \min\{q - F(x_q-\varepsilon), F(x_q+\varepsilon)-q\}$$

for a class $\mathcal{H}$ of distributions, then for every $F \in \mathcal{H}$

$$P_F\left\{\sup_{n\geq N}\big|X_{k(n):n} - x_q\big| > \varepsilon\right\} < \frac{2\tau^N}{1-\tau}$$

with $\tau = \exp\{-\delta^2/2\}$ (Serfling 1980). In the class $\mathcal{F}(q,\vartheta)$ we have

$$\lim_{0<t\to 0}\frac{F(x_q+t)-q}{t} = \lim_{0<t\to 0}\frac{q-F(x_q-t)}{t} \geq \vartheta$$

so that there exists $t_0 > 0$ such that for all $t < t_0$

$$F(x_q+t)-q \geq \frac{1}{2}\vartheta t \quad \text{and} \quad q - F(x_q-t) \geq \frac{1}{2}\vartheta t$$

and in consequence, for all sufficiently small ε (for $\varepsilon < t_0$)

$$\delta = \min\{q-F(x_q-\varepsilon), F(x_q+\varepsilon)-q\} \geq \frac{1}{2}\vartheta\varepsilon.$$

Now

$$\tau = \exp\{-\delta^2/2\} \leq \exp\{-\frac{1}{8}\vartheta^2\varepsilon^2\}.$$

Solving, with respect to N, the equation

$$\frac{2\tau^N}{1-\tau} = \eta$$

we obtain the result. □

Table 1 below gives us an insight in how large samples are needed to get the prescribed accuracy of the asymptotic.

Table 1. $N(\vartheta,\varepsilon,\eta)$

		ε	
η	ϑ	0.05	0.10
0.1	0.5	159,398	35,414
	1.0	35,414	7,745
	2.0	7,745	1,660
0.01	0.5	188,871	42,782
	1.0	42,782	9,587
	2.0	9,587	2,120

6.2. Asymptotics of the most concentrated median-unbiased estimator

In Sec. 3.3.1 we constructed the most concentrated median-unbiased estimator of the form

$$\hat{x}_{RZ} = \begin{cases} X_{k:n}, & \text{with probability } \lambda, \\ X_{k+1:n}, & \text{with probability } 1-\lambda, \end{cases}$$

where k satisfies $\pi_k(q) \geq 1/2 > \pi_{k+1}(q)$, $\lambda_k^* = (1/2 - \pi_{k+1}(q))/(\pi_k(q) - \pi_{k+1}(q))$, and $\pi_k(q) = \sum_{j=k}^{n} \binom{n}{j} q^j (1-q)^{n-j}$. In this Section we consider the estimator for a fixed $q \in (0,1)$ and to avoid too tedious notation we shall denote shortly $T = \hat{x}_{RZ}$. The concentration function

$$C_n(\beta) = P_F\{T \leq x_\beta(F)\}$$

of the estimator does not depend of F but obviously depends on the size of the sample n (Fig. 6.1); we are interested in the asymptotic behavior of C_n as $n \to +\infty$.

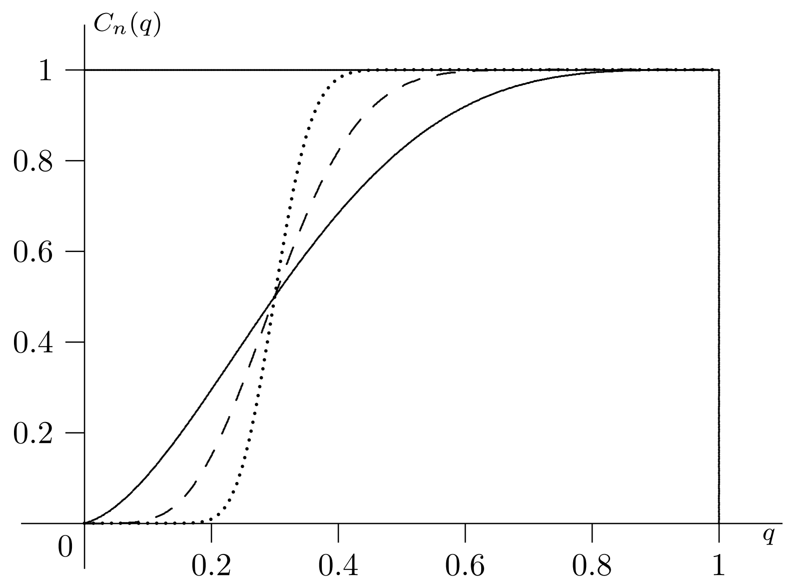


Fig. 6.1. Concentration function of the estimator $\hat{x}_{RZ}$, $q = 0.3$
$n = 5$ - solid, $n = 20$ - dashes, $n = 100$ - dots

Theorem 3 (Wieczorkowski et al. 1991). *For every* $\beta \in (0, q)$

$$C_n(\beta) \to 0, \quad \text{as } n \to +\infty$$

and for every $\beta \in (q, 1)$

$$C_n(\beta) \to 1, \quad \text{as } n \to +\infty$$

Proof. By the Berry-Esséen inequality we have

$$\left| \pi_k(q) - \bar{\Phi}\left(\frac{k-1-nq}{\sqrt{nq(1-q)}}\right) \right| \leq \frac{b(q)}{\sqrt{n}}$$

where $\bar{\Phi}(t) = 1 - \Phi(t)$, Φ is the distribution function of the standard normal distribution $N(0, 1)$, $b(q) = c[q^2 + (1-q)^2]/\sqrt{q(1-q)}$, and c is a Berry-Esséen constant. By the definition of k it follows that

$$\bar{\Phi}\left(\frac{k-nq}{\sqrt{nq(1-q)}}\right) - \frac{b(q)}{\sqrt{n}} \leq \pi_{k+1}(q) < \frac{1}{2} \leq \pi_k(q) \leq \bar{\Phi}\left(\frac{k-1-nq}{\sqrt{nq(1-q)}}\right) + \frac{b(q)}{\sqrt{n}}$$

The inequalities hold for all n which is possible only if $k-1-nq \leq 0 \leq k-nq$ and hence $k = nq + \delta$ for some $\delta \in (0, 1)$. The concentration function $C_n(\beta)$ can be written in the form

$$C_n(\beta) = \lambda \pi_k(\beta) + (1-\lambda)\pi_{k+1}(\beta)$$

and by the above estimates we obtain

$$W(n, q, \beta) - \frac{b(q)}{\sqrt{n}} \leq C_n(\beta) \leq W(n, q, \beta) + (1 + 2\lambda)\frac{b(q)}{\sqrt{n}}$$

where

$$W(n, q, \beta) = \lambda \left[\bar{\Phi}\left(\frac{n(q-\beta)+\delta-1}{\sqrt{nq(1-q)}}\right) - \bar{\Phi}\left(\frac{n(q-\beta)+\delta}{\sqrt{nq(1-q)}}\right) \right] + \bar{\Phi}\left(\frac{n(q-\beta)+\delta}{\sqrt{nq(1-q)}}\right)$$

To end the proof it is enough to observe that

$$W(n, q, \beta) \to 1, \quad \text{as } n \to \infty \quad \text{if } \beta < q$$

and

$$W(n, q, \beta) \to 0, \quad \text{as } n \to \infty \quad \text{if } \beta > q$$

□

References

Beran, R. and Hall, P. (1993), Interpolated Nonparametric Prediction Intervals and Confidence Intervals, *J.R.Statist.Soc.* B 55, No. 3, pp. 643–652

Bickel, P.J. and Lehmann, E.L. (1975–1979), Descriptive statistics for nonparametric models. I–III. *Ann. Statist.* 3, 1038-1045, 1045-1069; 4, 1139-1158; IV *Contribution to Statistics: Hájek Memorial Volume*, J.Jureckova, Ed. Academia, Prague, pp. 33–40

Boratyńska, A. and Zieliński, R. (1997): Asymptotic Behavior of Sample Median in a Parametric Model, *Annales Universitatis Mariae Curie-Sklodowska*, Lublin-Polonia, Sectio A, Vol. LI.1,2, 11-14

Borovkov, A.A. (1998), *Mathematical Statistics*, Gordon and Breach

Brown, B.M. (1995), Median estimates and sign tests, in *Encyclopedia of Statistical Sciences*, Vol. 5, Wiley

Brown, L.D., Cai, T.T. and DasGupta, A. (2001), Interval estimation for a Binomial Proportion, *Statistical Science* 16, 2, 101–133

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

David, C.E. and Steinberg, S.M. (1986), Quantile estimation, In *Encyclopedia of Statistical Sciences*, Vol. 7, Wiley, New York

David, H.A. (1981), *Order Statistics*, 2nd ed., Wiley, New York

David, H.A. and Nagaraja, H.N. (2003) *Order Statistics*, Third Edition, Wiley-Interscienc

Dielman, T., Lowry, C. and Pfaffenberger, R. (1994), A comparison of quantile estimators, *Commun. Statist.-Simula* 23(2), 355-371

Fama, E.F. and Roll, R. (1971), Parameter estimates for symmetric stable distributions, *JASA* 66, 331-338

Gass, S.I. (1958), *Linear Programming. Methods and Applications*, Mc Graw-Hill Book Company, Inc.

Harrell, F.E. and Davis, C.E. (1982), A new distribution-free quantile estimator, *Biometrika* 69, 635-640

Hoeffding, W. (1963), Probability inequalities for sums of bounded random variables. *J. Amer. Statis. Assos.* 58, 13-30

Huang, M.L. and Brill P. (1999): A level crossing quantile estimation method. *Statistics & Probability Letters* 45, 111-119

Huber, P.J. (1981), *Robust Statistics*, Wiley

Hutson, A.D. (1999), Calculating nonparametric confidence intervals for quantiles using fractional order statistics, *Journal of Applied Statistics*, Vol. 26, No. 3, pp. 343–353

Kaigh, W.D. and Lachenbruch, P.A. (1982), A generalized quantile estimator, *Commun. Statist. Theor. Meth.* 12, 2427-2443

Keating, J.P. and Tripathi, R.C. (1986): Percentiles, Estimation of. In *Encyclopedia of Statistical Sciences*, Vol. 7, Wiley, New York

Keating, J.P., Mason, R.L., Rao, C.R., Sen, P.K. ed's (1991) *Pitman's Measure of Closeness*. Special Issue of *Communications in Statistics - Theory and Methods*, Vol. 20, Number 11

Keating, J.P., Mason, R.L., and Sen, P.K. (1993), *Pitman's Measure of Closeness: A comparison of Statistical Estimators*, SIAM, Philadelphia.

Lehmann, E.L. (1983), *Theory of Point Estimation*, Wiley

Lehmann, E.L. (1986), *Testing Statistical Hypotheses*, Second Edition. Wiley

Neumann, P. (1965), Über den Median der Binomial- und Poissonverteilung, *Wiss. Z. Techn. Univers. Dresden*, 15, H.2, 223-226

Neumann, P. (1967), "Uber den Median der Binomial- und Poissonverteilung, *Wiss. Z. Humboldt-Univ. Berlin*, Math.-Nat.R. XVI, 62-64

Reiss, R.-D. (1989), *Approximate Distributions of Order Statistics with Applications to Nonparametric Statistics*, Springer-Verlag

Rychlik, T. and Zieliński, R. (1985), An asymptotically most bias-robust invariant estimator of location. In *Stability Problems for Stochastic Models*, Lecture Notes in Math. 1233, Springer-Verlag, 156-171

Serfling, R.J. (1980, re-edition XXXX), *Approximation Theorems of Mathematical Statistics*, Wiley

Shervish, M.J. (1995), *Theory of Statistics*, Springer

Uhlmann, W. (1963): Ranggrössen als Schätzfunktionen, *Metrika* 7(1), 23-40

Wieczorkowski, R. and Zieliński, R. (1991) Some properties and applications of a distribution-free quantile estimator, *Roczniki Polskiego Towarzystwa Matematycznego, Ser. III: Matematyka Stosowana*, XXXIV, 15-22

Zieliński, R. (1987), Robustness of sample mean and sample median under restrictions on outliers, *Appl. Math. (Warsaw).* 19,2, 239-240

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

Zieliński, R. (1988), Stable Estimation of Location Parameter - Nonasymptotic Approach, *Statistics* 19, 229-231

Zieliński, R.(1995), Estimating Median and Other Quantiles in Nonparametric Models. Applicationes Math. 23.3, 363-370. Correction: Applicationes Math. 23.4 (1996) p. 475

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. (2000a), A median-unbiased estimator of the characteristic exponent of a symmetric stable distribution, *Statistics* 34, 353-355

Zieliński, R. (2000b), High accuracy evaluation of cumulative distribution function of α-stable symmetric distribution, *Teor. Veroyatn. Prim.* 45, 4, 807-808

Zieliński, R. (2001a) High-accuracy evaluation of the cumulative distribution function of α-stable symmetric distributions, *Journal of Mathematical Sciences*, 105, 6, 2630-2632

Zieliński, R. (2001b), PMC-optimal nonparametric quantile estimator, *Statistics* 35, 453-462

Zieliński, R. (2002), The most stable estimator of location under integrable contaminants, *Appl. Math. (Warsaw)* 29,1, 1-6

Zieliński, R. (2003), Constructing median-unbiased estimators in one-parameter families of distributions via stochastic ordering, *Applicationes Math.* 30,3, 251-255

Zieliński, R. (2004), Effective WLLN, SLLN, and CLT in Statistical Models, *Applicationes Math.* 31,1, 117-125

Zieliński, R. and Zieliński, W. (2004), Best Exact Nonparametric Confidence Intervals for Quantiles, (*To appear*)