Estimating the minimal value of a function in global random search: Comparison of estimation procedures

E. Hamilton, V. Savani, A. Zhigljavsky

School of Mathematics, Cardiff University

In the problem of estimating the lower bound of a function over a continuous set, a variety of linear estimators are used, as well as the maximum likelihood estimator. The asymptotic mean square errors (MSE) of several linear estimators asymptotically coincide with the asymptotic MSE of the maximum likelihood estimator. In this paper we consider the non-asymptotic behaviour of different estimators and show that the MSE of the best linear estimators is superior to the MSE of the the maximum likelihood estimator.

1 Introduction

Let $f : A \to \mathbb{R}$ be an objective function defined in feasible region A and $m = \min_{x \in A} f(x)$ be its global minimum. We always assume that A is a compact subset of $\mathbb{R}^d$ for some $d \geq 1$, $\text{vol}(X) > 0$ (where $\text{vol}(\cdot)$ stands for 'volume'), $m > -\infty$ and there is at least one global minimizer; that is, the point $x^* \in A$ such that $f(x^*) = m$. We shall also assume that the objective function f is continuous in the neighbourhood of this minimizer x^*. For simplicity we also assume that the objective function f is bounded from above (this last condition is made purely for technical reasons and can be relaxed).

We consider several estimators of m that are based on taking random samples of n points from the feasible region A and computing the corresponding values of the objective function. The set of n points, $X_n = \{x_1, \ldots, x_n\}$, will be independent and identically distributed random variables/vectors (i.i.d.r.v.) with common distribution P, where P is a probability measure defined on A. We assume that there is a positive probability that a random point x_j will be in the vicinity of x^*.

Let $Y_n = \{f(x_1), \ldots, f(x_n)\} = \{y_1, \ldots, y_n\}$ be the i.i.d. r.v. obtained by computing $f(\cdot)$ at the elements of the sample X_n. The y_j will have common cumulative distribution function (c.d.f.).

$$F(t) = \mathbb{P}\{x \in A : f(x) \leq t\} = \int_{f(x) \leq t} P(dx) . \tag{1}$$

The minimum value of $f(\cdot)$ is the essential infimum of the r.v. y with the c.d.f. (1):

$$m = \min_{x \in A} f(x) = \operatorname{ess\,inf} y = \inf\{a : F(a) > 0\}\,.$$

2 Asymptotic distribution of the minimum order statistic

Let the sample size n be fixed and $y_{1,n} \leq \ldots \leq y_{n,n}$ be the order statistics corresponding to the independent random sample Y_n. Here $y_{i,n}$ represents the i-th smallest member within the sample $Y_n = \{y_1, \ldots, y_n\}$.

Consider first the asymptotic distribution of the sequence of minimum order statistic $y_{1,n}$, as $n \to \infty$. Generally, in the case $m = \operatorname{ess\,inf} y > -\infty$ (where y has c.d.f. $F(t)$) there are two limiting distributions possible; however, in global random search applications, when $F(\cdot)$ has the form (1), only one asymptotic distribution arises; specifically, the Weibull distribution with the c.d.f.

$$\Psi_\alpha(z) = \begin{cases} 0 & \text{for } z < 0 \\ 1 - \exp\{-z^\alpha\} & \text{for } z \geq 0\,. \end{cases} \tag{2}$$

This c.d.f. has only one parameter, α, which is called 'tail index'. The mean of the Weibull distribution with tail index α is $\Gamma(1 + 1/\alpha)$; the density corresponding to the c.d.f. (2) is

$$\psi_\alpha(t) = (\Psi_\alpha(t))' = \alpha\, t^{\alpha-1} \exp\{-t^\alpha\}\,, \quad t > 0\,. \tag{3}$$

Let κ_n be the $\left(\frac{1}{n}\right)$-quantile of a c.d.f. $F(\cdot)$; that is, $\kappa_n = \inf\{u | F(u) \geq 1/n\}$. Note that since the objective function $f(\cdot)$ is continuous in a neighbouhood of x^*, the c.d.f. $F(\cdot)$ is continuous in the vicinity of m and for n large enough we have $F(\kappa_n) = 1/n$.

The following classical result from the theory of extreme order statistics is of primary importance for us.

Theorem 1. *Assume* $\operatorname{ess\,inf} \eta = m > -\infty$*, where* η *has c.d.f.* $F(t)$*, and the function*

$$V(v) = F\left(m + \frac{1}{v}\right), \qquad v > 0,$$

regularly varies at infinity with some exponent $(-\alpha)$*,* $0 < \alpha < \infty$*; that is,*

$$\lim_{v \to \infty} \frac{V(tv)}{V(v)} = t^{-\alpha}, \qquad \text{for each } t > 0\,. \tag{4}$$

Then

$$\lim_{n\to\infty} F_{1,n}(m + (\kappa_n - m)z) = \Psi_\alpha(z)\,, \tag{5}$$

where $F_{1,n}$ is the c.d.f. of the minimum order statistics $y_{1,n}$, the c.d.f. $\Psi_\alpha(z)$ is defined in (2) and κ_n is the $\left(\frac{1}{n}\right)$-quantile of $F(\cdot)$.

The asymptotic relation (5) means that the distribution of the sequence of random variables $(y_{1,n} - m)/(\kappa_n - m)$ converges (as $n \to \infty$) to the random variable with c.d.f. $\Psi_\alpha(z)$.

The c.d.f. $\Psi_\alpha(z)$, along with its limiting case $\Psi_\infty(z) = \exp\{-\exp(z)\}$, $z>0$, are the only nondegenerate limits of the c.d.f.'s of the sequences $(y_{1,n}-a_n)/b_n$, where $\{a_n\}$ and $\{b_n\}$ are arbitrary sequences of positive numbers.

If there exist numerical sequences $\{a_n\}$ and $\{b_n\}$ such that the c.d.f.'s of $(y_{1,n}-a_n)/b_n$ converge to Ψ_α, then we say that $F(\cdot)$ belongs to the domain of attraction of $\Psi_\alpha(\cdot)$ and express this as $F \in D(\Psi_\alpha)$. The conditions stated in Theorem 1 are necessary and sufficient for $F \in D(\Psi_\alpha)$. There are two conditions: $m = \operatorname{ess\,sup} \eta < \infty$ and the condition (4). The first one is always valid in global random search applications. The condition (4) demands more attention. For example, it is never valid in discrete optimization problems as in this problems the c.d.f. $F(\cdot)$ is not continuous (but $F(\cdot)$ has to be continuous in the vicinity of $m = \operatorname{ess\,inf} y$). In fact, for a c.d.f. with a jump at its lower end-point no non-degenerate asymptotic distribution for $y_{1,n}$ exists, whatever the normalization (that is, sequences $\{a_n\}$ and $\{b_n\}$).

The condition (4) can be written as

$$F(t) = c_0(t-m)^\alpha + \mathrm{o}((t-m)^\alpha) \qquad \text{as } t \downarrow m\,, \tag{6}$$

where c_0 is a function of $v = 1/(t-m)$, slowly varying at infinity as $v \to \infty$. Of course, any positive constant is a slowly varying function, but the actual range of eligible functions c_0 is much wider.

The following sufficient condition (so-called von Mises condition) for (4) and (5) is often used: if $F(t)$ has a positive derivative $F'(t)$ for all $t \in (m, m+\varepsilon)$ for some $\varepsilon > 0$ and

$$\lim_{t\downarrow m} \frac{(t-m)F'(t)}{F(t)} = \alpha\,,$$

then (4) holds.

The following condition is stronger that the condition (6) and often used for justifying properties of the maximum likelihood estimators:

$$F(t) = c_0(t-m)^\alpha \left(1 + \mathrm{O}((t-m)^\beta)\right) \qquad \text{as } t \downarrow m \tag{7}$$

for some positive constants c_0, α and β.

In problems of global optimization we can establish a direct link between the form (1) of the c.d.f. $F(\cdot)$, the condition (6) and the value of the tail index α. The basic result is as follows.

Assume that the objective function f is such that its global minimizer x^* is unique and

$$f(x)-m = w(\|x-x_*\|)H(x-x_*) + \mathrm{O}(\|x-x_*\|^{\beta}), \quad \|x-x_*\| \to 0, \tag{8}$$

for some homogeneous function $H : \mathbb{R}^d\backslash\{0\} \to (0,\infty)$ of order $\beta > 0$ (for H the relation $H(\lambda z) = \lambda^{\beta} H(z)$ holds for all $\lambda > 0$ and $z \in \mathbb{R}^d$) and function $w : \mathbb{R} \to \mathbb{R}$ is positive and continuous. Then the condition (6) for the c.d.f. (1) holds and the value of the tail index α is equal to $\alpha = d/\beta$.

Proof of this result and many generalizations can be found in []. Two important particular cases of (8) are:

- let $f(\cdot)$ be twice continuously differentiable in the vicinity of x_*, $\nabla f(x_*) = 0$ (here $\nabla f(x_*)$ is the gradient of $f(\cdot)$ in x_*) and the Hessian $\nabla^2 f(x_*)$ of $f(\cdot)$ at x_* is nondegenerate; in this case, we can take

  $$w(\cdot) = 1, \quad H(z) = -z'[\nabla^2 f(x_*)]z\,,$$

  which implies $\beta = 2$ and $\alpha = d/2$;
- let all components of $\nabla f(x_*)$ are finite and non-zero which often happens if the global minimum of $f(\cdot)$ is achieved at the boundary of A. Then we may take $H(z) = z'\nabla f(x_*)$, $w(\cdot) = 1$; this gives $\beta = 1$ and $\alpha = d$.

The quantity $\kappa_n - m$, where $m = \operatorname{ess\,inf} \eta$ and κ_n is the $(1/n)$-quantile of $F(\cdot)$, enters many formulae below and therefore its asymptotic behaviour is very important. Fortunately, the asymptotic behaviour of $\kappa_n - m$ is clear. Indeed, as soon as (6) holds with some c_0, we have

$$\frac{1}{n} = F(\kappa_n) \sim c_0\,(\kappa_n - m)^{\alpha} \qquad \text{as } n \to \infty$$

implying

$$(\kappa_n - m) \sim (c_0 n)^{-1/\alpha} \qquad \text{as } n \to \infty\,. \tag{9}$$

3 Defining the Estimators

Let $Y_n = \{y_1, \ldots, y_n\}$ be an independent sample of values from the c.d.f. (1) and $y_{1,n} \le \ldots \le y_{n,n}$ be the corresponding order statistics. In this section we define six estimators of m that will be numerically studied below.

For constructing the estimators of m, only the first k order statistics $\{y_{1,n}, \ldots, y_{k,n}\}$ will be used (here k is much smaller than n). There are two major reasons for this: (a) the higher order statistics contain very little information about m; and (b) we can use the limit theorem for extreme order statistics (Theorem 1) only if k is such that $k/n \to 0$ as $n \to \infty$.

The estimators rely on the assumption that the c.d.f. $F(\cdot)$ satisfies the condition (6) with known value of the tail index α.

3.1 Maximum Likelihood Estimator

In defining the maximum likelihood estimator (MLE) we have to assume the condition (7) which is stronger than (6). Additionally, we have to assume $\alpha \geq 2$ (numerical study shows that the MLE as defined below coincides with the minimum order statistic $y_{1,n}$ for $\alpha < 2$).

Taking the asymptotic form of the likelihood function as exact (for details see Hall , Zh), we obtain that the maximum likelihood estimator of the minimum, $\hat{m^*}$, is the solution in z to the following likelihood equation:

$$(\alpha - 1) \sum_{j=1}^{k-1} \frac{(y_{k,n} - y_{j,n})}{(y_{j,n} - z)} = k \tag{10}$$

conditionally $z < y_{1,n}$; if there are no solution to this equation for $z < \{y_{1,n}$, then we set $\hat{m^*} = y_{1,n}$; if there are more than one solution in the region $z \in (-\infty, y_{1,n})$, then we take the smallest of these solutions.

If the conditions (7), $\alpha \geq 2$, $k \to \infty$, $k/n \to 0$ (as $n \to \infty$) are satisfied then the maximum likelihood estimators of m are asymptotically normal and asymptotically efficient in the class of asymptotically normal estimators and their mean square error $E(\hat{m} - m)^2$ is asymptotically

$$E(\hat{m} - m)^2 \sim \begin{cases} (1 - \frac{2}{\alpha})(\kappa_n - m)^2 \, k^{-1+2/\alpha} & \text{for } \alpha > 2, \\ (\kappa_n - m)^2 \log k & \text{for } \alpha = 2\,. \end{cases} \tag{11}$$

3.2 Linear Estimators

We will now define five different linear estimators.

A general linear estimator of m can be written as

$$\hat{m}_{n,k}(a) = \sum_{i=1}^{k} a_i y_{i,n}\,, \tag{12}$$

where $a = (a_1, \ldots, a_k)' \in \mathbb{R}^k$ is a vector of coefficients. It can be shown (see Zh ....) that as $n \to \infty$ we have

$$E\hat{m}_{n,k}(a) = m \sum_{i=1}^{k} a_i - (\kappa_n - m)a'b + \mathrm{o}(\kappa_n - m) = m \sum_{i=1}^{k} a_i + \mathrm{o}(1)\,. \tag{13}$$

Here $b = (b_1, \ldots b_k)' \in \mathbb{R}^k$, where $b_i = \Gamma(i + 1/\alpha) \,/\, \Gamma(i)$.

From (13) it is clear that (as the objective function f is bounded and therefore the variances of all $y_{i,n}$ are finite), a necessary and sufficient condition for an estimator with vector of coefficients a to be consistent is:

$$\sum_{i=1}^{k} a_i = 1\,. \tag{14}$$

Additionally, to ensure a small bias we may require

$$\sum_{i=1}^{k} a_i b_i = 0\,. \tag{15}$$

The main criteria is to ensure minimization of the mean square error given by

$$E(\hat{m}_{n,k}(a) - m)^2 \sim (\kappa_n - m)^2\, a'\Lambda a, \quad n \to \infty\,, \tag{16}$$

where $\Lambda = \|\lambda_{ij}\|_{i,j=1}^{k}$ is a symmetric $k \times k$-matrix with elements λ_{ij} defined for $i \geq j$ by the formula

$$\lambda_{ij} = \frac{\Gamma(i+2/\alpha)\,\Gamma(j+1/\alpha)}{\Gamma(i+1/\alpha)\,\Gamma(j)}\,.$$

Optimal linear estimator

The r.h.s. of (16) is a natural optimality criterion for selecting the vector a. The optimal consistent estimator $m^\circ = \hat{m}_{n,k}(a^\circ)$, we shall call it *the optimal linear estimator*, is determined by the vector of coefficients

$$a^\circ = \arg\min_{a:a'\mathbf{1}=1} a'\Lambda a = \frac{\Lambda^{-1}\mathbf{1}}{\mathbf{1}'\Lambda^{-1}\mathbf{1}}\,. \tag{17}$$

The estimator m° has been suggested in Cooke (1979), where the form (17) for the vector of coefficients was obtained. Solving the quadratic programming problem in (17) is straightforward. In the process of doing that, we obtain

$$\min_{a:a'\mathbf{1}=1} a'\Lambda a = (a^*)'\Lambda a^* = 1/\mathbf{1}'\Lambda^{-1}\mathbf{1}\,. \tag{18}$$

Lemma 7.3.4 in Zhigljavsky (1991) gives the following expression for the r.h.s. of (18):

$$\mathbf{1}'\Lambda^{-1}\mathbf{1} \;=\; \begin{cases} \frac{1}{\alpha-2}\left(\frac{\alpha\Gamma(k+1)}{\Gamma(k+2/\alpha)} - \frac{2}{\Gamma(1+2/\alpha)}\right) & \text{for } \alpha \neq 2, \\ \sum_{i=1}^{k} 1/i & \text{for } \alpha = 2\,; \end{cases} \tag{19}$$

this expression is valid for all $\alpha > 0$ and $k = 1, 2, \ldots$

The components a_i° $(i = 1, \ldots, k)$ of the vector a° can be evaluated explicitely: $a_i^\circ = u_i/\mathbf{1}'\Lambda^{-1}\mathbf{1}$ for $i = 1, \ldots, k$ with

$$\begin{aligned} u_1 &= (\alpha+1)\,/\,\Gamma(1+2/\alpha), \\ u_i &= (\alpha-1)\,\Gamma(i)/\Gamma(i+2/\alpha) \qquad \text{for } i = 2, \ldots, k-1, \\ u_k &= -(\alpha k - \alpha + 1)\Gamma(k)\,/\,\Gamma(k+2/\alpha). \end{aligned}$$

Deriving this expression for the coefficients of the vector a° is far from trivial, see Zhigljavsky (1991), Section 7.3.3.

The asymptotic properties of the optimal linear estimators coincide with the properties of the maximum likelihood estimators and hold under the same regularity conditions (we again refer to Zhigljavsky (1991), Section 7.3.3). In particular, the optimal linear estimators $m^\circ = \hat{m}_{n,k}(a^\circ)$ of m are asymptotically normal and their mean square error $E(m^\circ - m)^2$ asymptotically behaves like the r.h.s. of (11). Unlike the MLE, this estimator is defined for all $\alpha > 0$ and indeed behaves well for small α, see below.

Second optimal linear estimator

The second linear estimator (suggested in Cooke (1980)) satisfies the conditions (14) and (15): that is, the consistency and unbiasedness conditions. The coefficients of this estimator are:

$$a^{\triangle} = \arg \min_{\substack{a:\, a'\mathbf{1}=1,\\ a'b=0}} a'\Lambda a = \frac{\Lambda^{-1}\mathbf{1} - (b'\Lambda^{-1}\mathbf{1})\Lambda^{-1}b/(b'\Lambda^{-1}b)}{\mathbf{1}'\Lambda^{-1}\mathbf{1} - (b'\Lambda^{-1}\mathbf{1})^2/(b'\Lambda^{-1}b)} \tag{20}$$

Asymptotic properties of the estimator $m^{\triangle} = \hat{m}(a^{\triangle})$ (when $\alpha \geq 2$, $n \to \infty, k \to \infty$ and $k/n \to 0$) are the same as the asymptotic properties of the MLE and the estimator m°.

Csörgö-Mason estimator

The Csörgö-Mason estimator is linear consistent and has similar asymptotic properties to the maximum likelihood estimator and the first two linear estimators. It is defined by the vector of coefficients $a^{\square} = \left(a_1^{\square}, \ldots, a_n^{\square}\right)'$, where

$$a_i^{\square} = \begin{cases} v_i & \text{for } \alpha > 2, \quad i = 1, \ldots, k-1 \\ v_k + 2 - \alpha & \text{for } \alpha > 2, \quad i = k \\ 2/\log(k) & \text{for } \alpha = 2, \quad i = 1 \\ \log(1+1/i)/\log(k) & \text{for } \alpha = 2, \quad i = 2, \ldots, k-1 \\ \left(\log(1+1/k) - 2\right)/\log(k) & \text{for } \alpha = 2, \quad i = k \end{cases}$$

with

$$v_j = (\alpha - 1)k^{2/\alpha - 1}\left(j^{1-2/\alpha} - (j-1)^{1-2/\alpha}\right).$$

For all $\alpha \geq 2$, the asymptotic properties of the estimators $m^{\triangle} = \hat{m}(a^{\triangle})$ and $m^{\square} = \hat{m}(a^{\square})$, when $n \to \infty, k \to \infty$ and $k/n \to 0$, coincide with the asymptotic properties of the MLE and the estimator m°.

A practical linear estimator

For practical purposes the following estimator $m^{\odot} = \hat{m}(a^{\odot})$ is useful:

$$a^{\odot} = ((1 + C_k), 0, \ldots, 0, -C_k)' \tag{21}$$

where $C_k = b_1/(b_k - b_1)$ is found from the unbiasedness condition $a'b = 0$. A similar estimator was proposed by Van der Waat (1980).

Minimum order statistic

The simplest estimator that can be used is the minimum order statistic, $m^{\bullet} = \hat{m}(a^{\bullet}) = y_{1,n}$. This means that

$$a^{\bullet} = (1, 0, \ldots, 0)'.$$

4 Comparison of estimators in finite samples

In this section we make a comparison of the efficiency and bias for the maximum likelihood and linear estimators of the finite global minimum m given finite samples of size $n = 500$ drawn from the Weibull distribution with probability distribution function

$$f(x) = \alpha x^{\alpha-1} \exp\{-x^{\alpha}\}, \quad x \geq 0. \tag{22}$$

The values of α considered when sampling from the Weibull distribution will be $\alpha \in \{2, 3, 5\}$ and the parameter k for the number of order statistics used in the computation of estimators will be $k \leq 20$. Note that the c.d.f. of the Weibull distribution can be represented in the form of (7) since $F(t) = 1 - \exp\{-t^{\alpha}\} = (t-0)^{\alpha} + \mathrm{O}((t-0)^{\alpha})$.

Efficiency of Linear Estimators

By definition, the optimal linear estimator $\hat{m}_{n,k}(a^{\circ})$, with a° given by (**??**), provides the lowest mean square error in the class of all linear consistent estimators as $n \to \infty$. The asymptotic mean square error for m° is

$$\lim_{N\to\infty} \mathrm{Var}(m^{\circ}) = \frac{(\kappa_n - m)^2}{\mathbf{1}'\Lambda^{-1}\mathbf{1}}, \quad n \to \infty. \tag{23}$$

Fig. 1 shows the efficiency, $\left[\kappa_n^2/(\mathbf{1}'\Lambda^{-1}\mathbf{1})\right] / \left[\frac{1}{R}\sum_{i=1}^{R}(\hat{m}_i - m)^2\right]$, of the maximum likelihood estimator, the consistent linear estimators and the minimum order statistic for α=1, 2, 5 and 10 against different values of k. The efficiency is based on taking $R = 10,000$ estimators of $\hat{m}_i$, where each $\hat{m}_i$ is estimated from a sample of size $n = 100$. Since we consider finite samples, it is possible for the efficiency to be greater than 1. Fig. 1 clearly shows that the mean square error of the $\hat{m}_{100,k}(a^{\circ})$ estimator converges to the asymptotic mean square error given by (23) for α=1, 2, 5 and 10. The estimator $\hat{m}_{100,k}(a^{\circ})$ clearly provides the lowest mean square error in the class of estimators considered. The efficiency of the minimum order statistic decreases monotonically as $k \to \infty$, this is because the mean square error of the minimum order statistic is independent of k, where as the mean square error of the consistent linear estimator $\hat{m}_{100,k}(a^{\circ})$ decreases as k increases. The efficiency of the $\hat{m}_{500,k}(a^{\triangle})$ estimator increases monotonically as k increases until k becomes too large, when the estimator begins to decrease. When the value of n is increased this decreasing in $\hat{m}_{500,k}(a^{\triangle})$ does not occur until $k > 20$. The efficiencies of the minimum order statistic and the $\hat{m}_{n,k}(a^{\triangle})$ estimators are equal for $k = 2$. This can be verified by considering the asymptotic mean square errors (as $n \to \infty$) of these two estimators at this point. Both of which are approximately $(\kappa_n - m)^2 \Gamma(1+2/\alpha))$ as $n \to \infty$. Estimators m° and $\hat{m}_{n,k}(a^{\triangle})$ become more similar in efficiency as $\alpha \to \infty$.

Comparison of estimators for fixed sample size

Efficiency $n = 100$

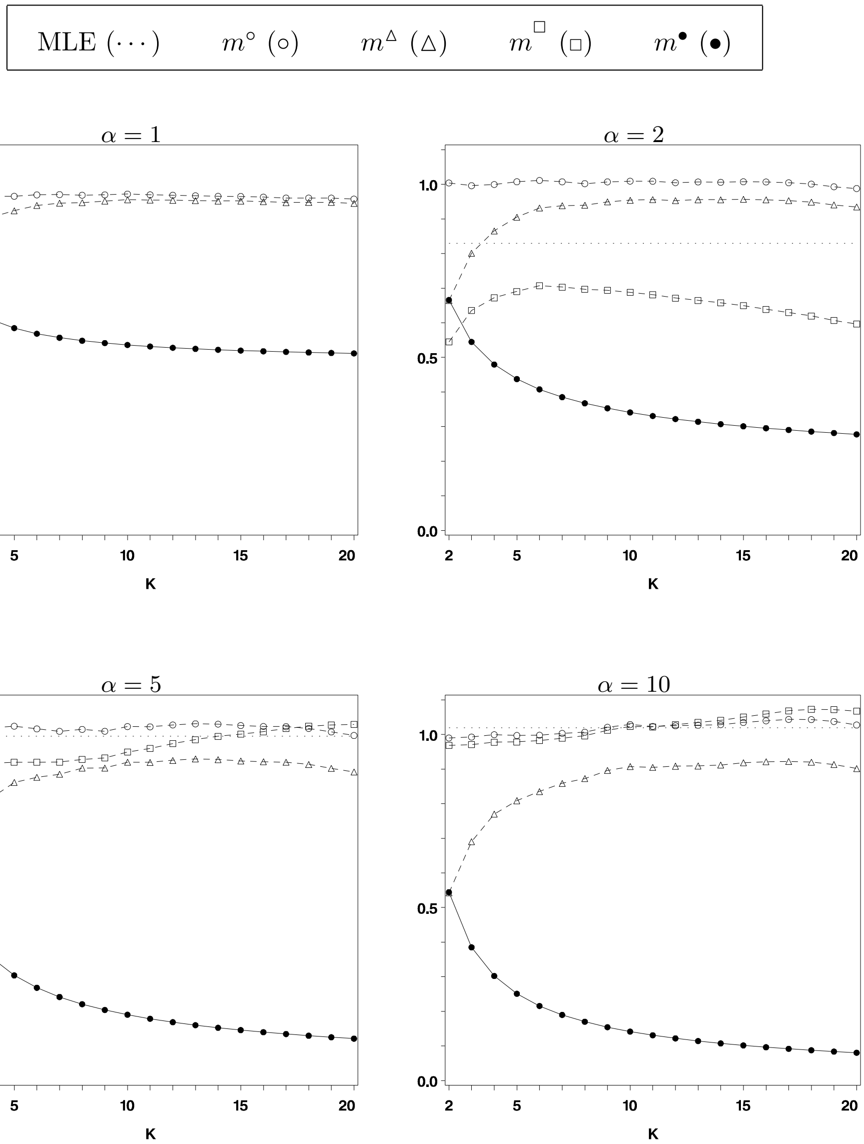

Fig. 1. Efficiency, $\left[\kappa_n^2/(\mathbf{1}'\Lambda^{-1}\mathbf{1})\right] / \left[\frac{1}{R}\sum_{i=1}^{R}(\hat{m}_i - m)^2\right]$, of different estimators for sample size $n = 100$ and $\alpha = 1, 2$, 5 and 10 against k.

Bias $n = 100$

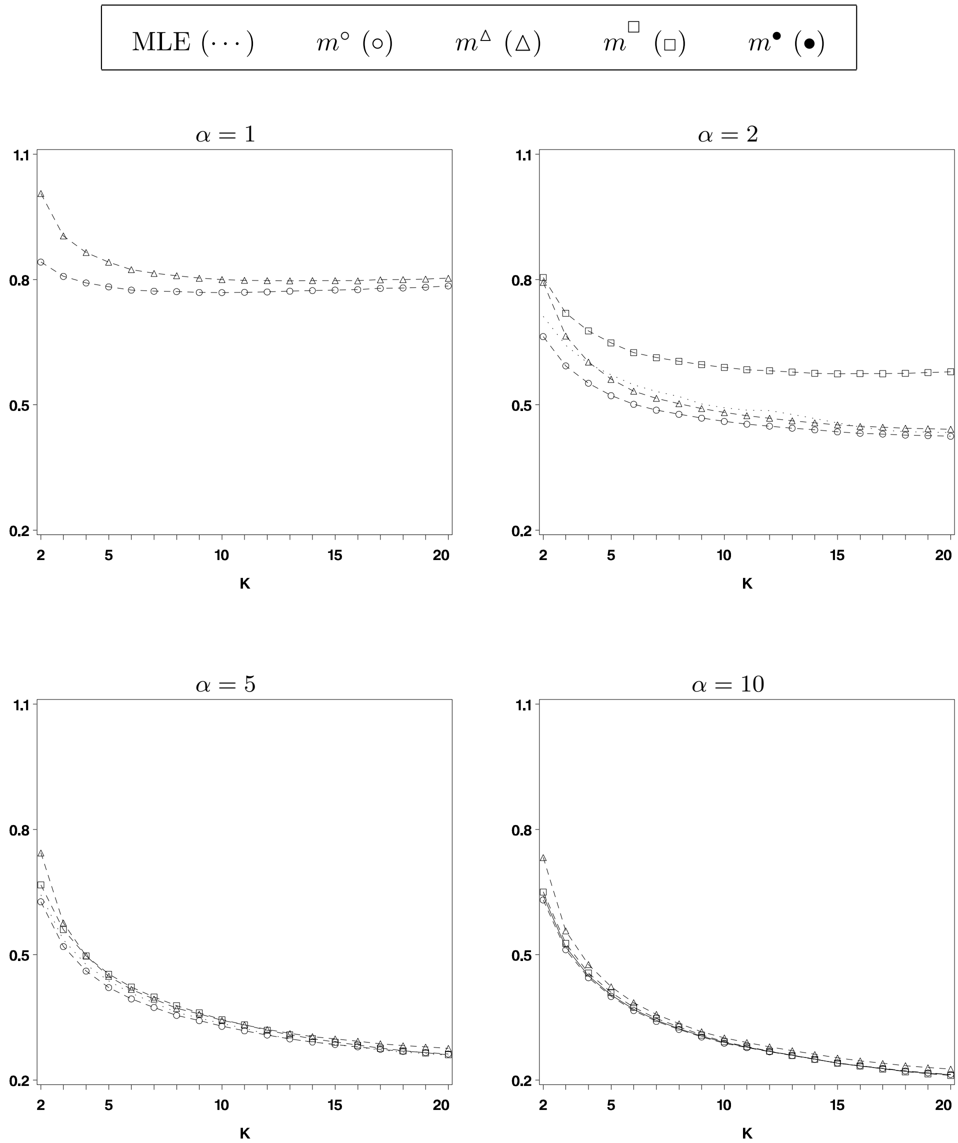

Fig. 2. Bias, $\frac{1}{R}\sum_{i=1}^{R} |n^{(1/\alpha)}(\hat{m}_i - m)|$, of different estimators for sample size $n =$ 100 and $\alpha = 1, 2,$ 5 and 10 against k.

Comparison of estimators for varying sample sizes

— $n = 100$ ··· $n = 500$ - - - $n = 1000$

Efficiency $\alpha = 1$

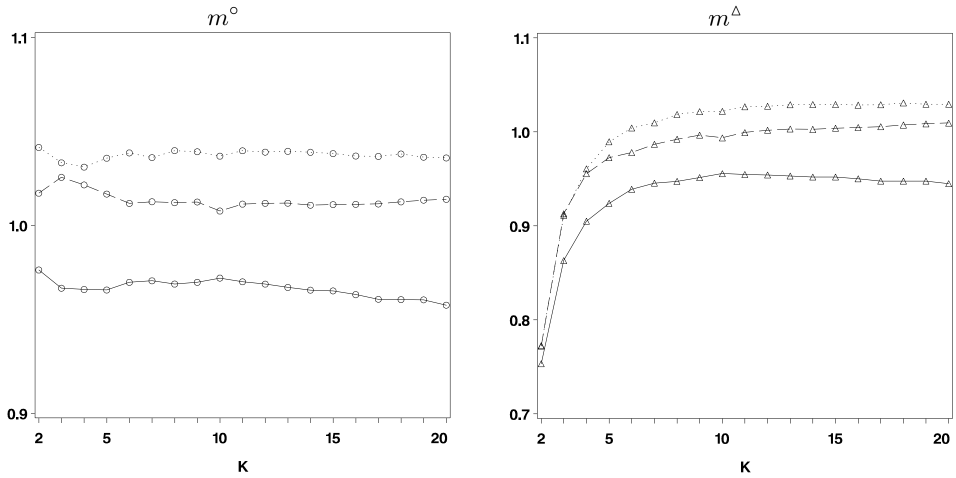

Fig. 3. Efficiency, $\left[\kappa_n^2/(\mathbf{1}'\Lambda^{-1}\mathbf{1})\right] / \left[\frac{1}{R}\sum_{i=1}^{R}(\hat{m}_i - m)^2\right]$, of different estimators for $\alpha = 1$ and sample sizes $n = 100$, 500, 1000 against k.

Bias $\alpha = 1$

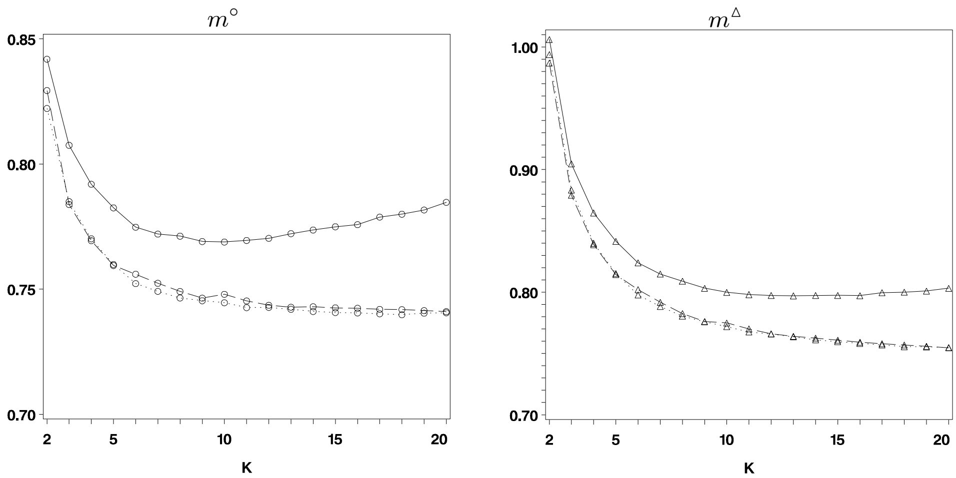

Fig. 4. Bias, $\frac{1}{R}\sum_{i=1}^{R}\left|n^{1/\alpha}(\hat{m}_i - m)\right|$, of different estimators for $\alpha = 1$ and sample sizes $n = 100$, 500, 1000 against k.

Efficiency $\alpha = 2$

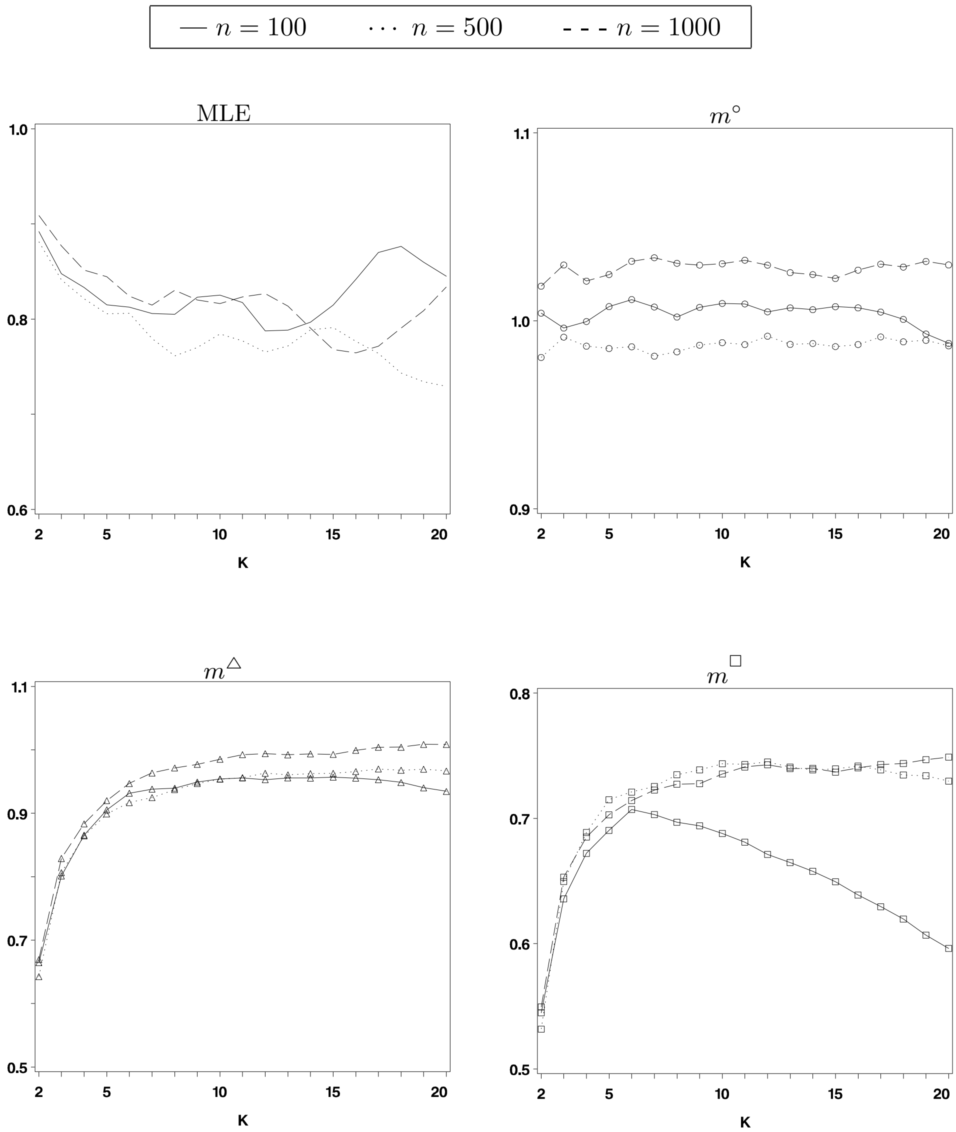

Fig. 5. Efficiency, $\left[\kappa_n^2/(\mathbf{1}'\Lambda^{-1}\mathbf{1})\right] / \left[\frac{1}{R}\sum_{i=1}^{R}(\hat{m}_i - m)^2\right]$, of different estimators for $\alpha = 2$ and sample sizes n = 100, 500, 1000 against k.

Bias $\alpha = 2$

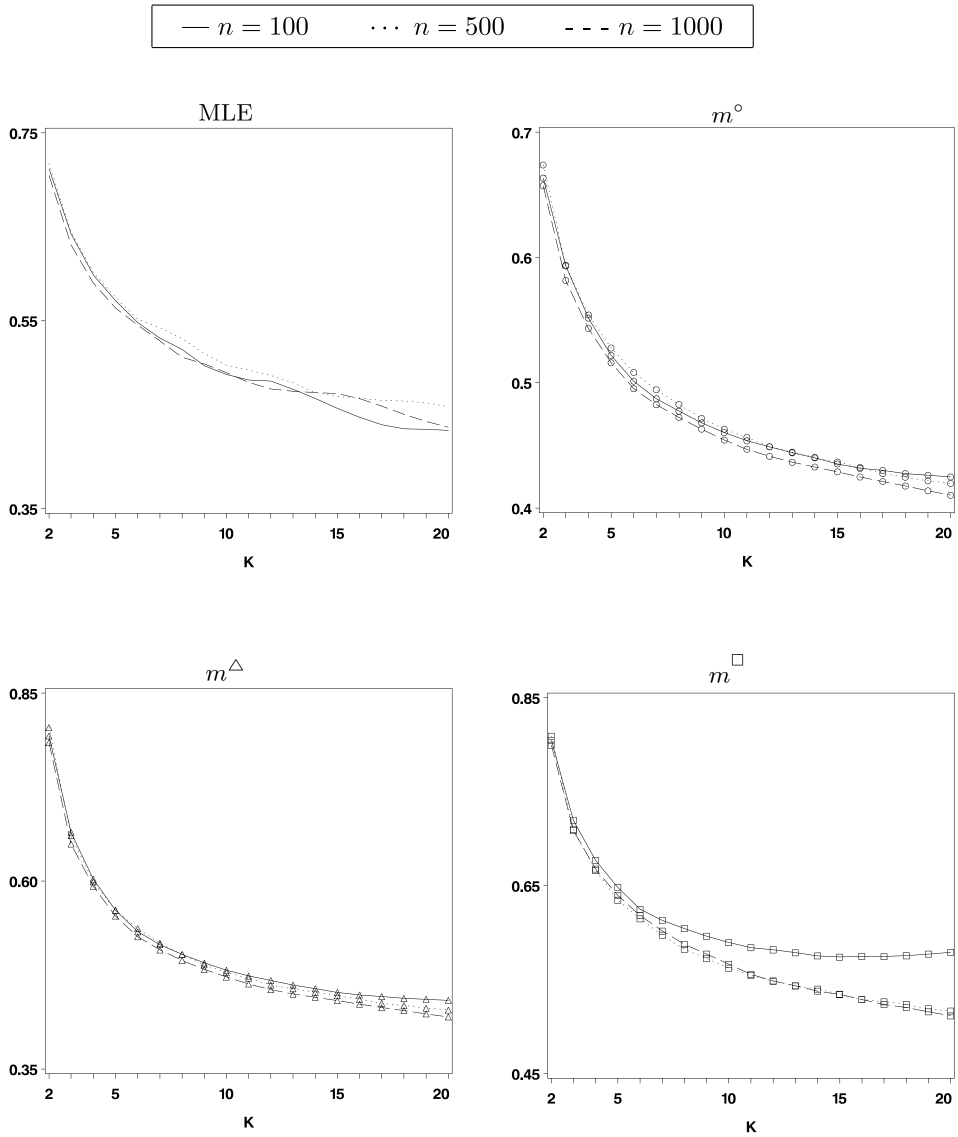

Fig. 6. Bias, $\frac{1}{R}\sum_{i=1}^{R}|n^{1/\alpha}(\hat{m}_i - m)|$, of different estimators for $\alpha = 2$ and sample sizes $n = 100$, 500, 1000 against k.

Histograms $n = 500$ $\alpha = 1$

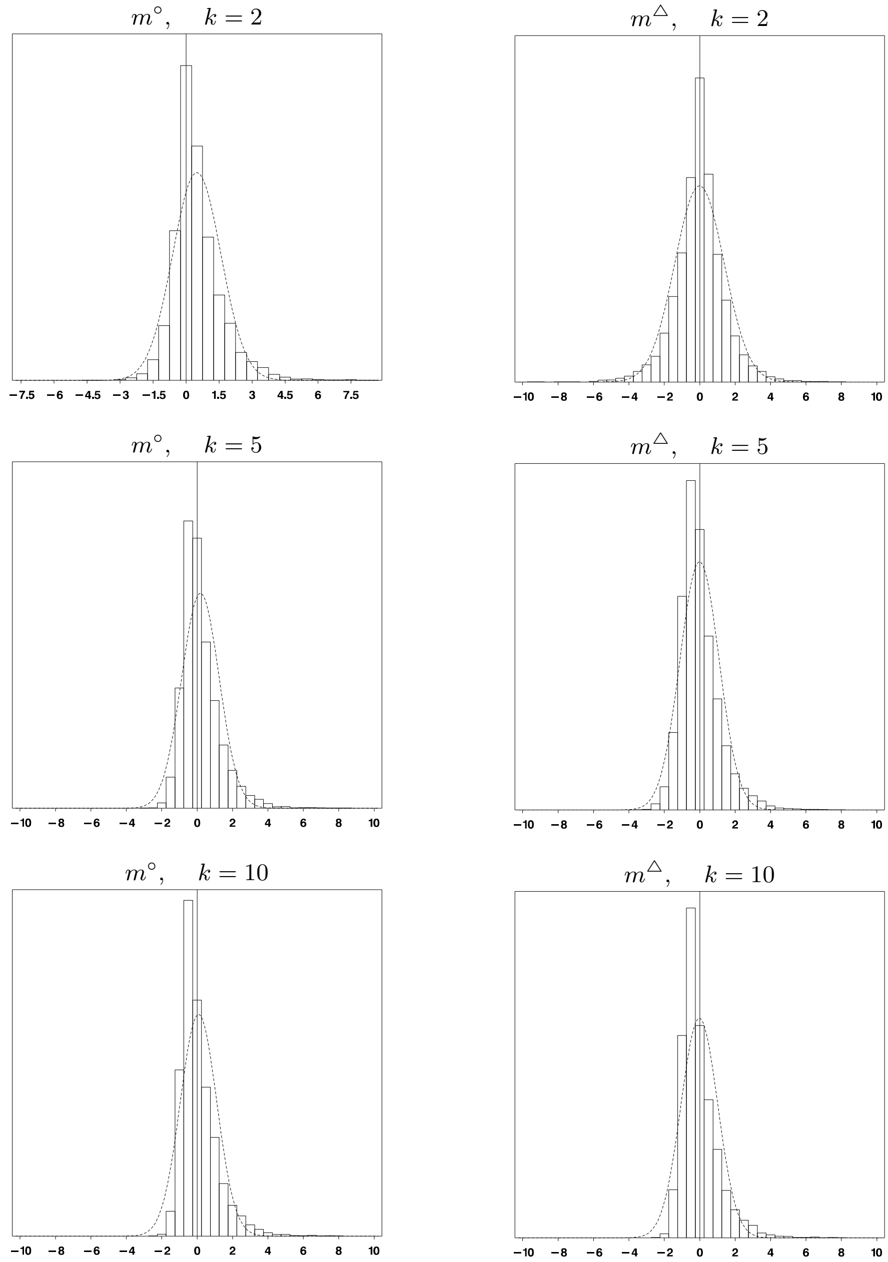

Fig. 7. Histograms of the normalized estimators $n^{(1/\alpha)}(\hat{m}_i - m)$, $i = 1 \ldots R$, $R =$ 10,000 , plotted with a normal probability density function whose parameters are taken from the sample.

Histograms n = 500 α = 2 k = 2

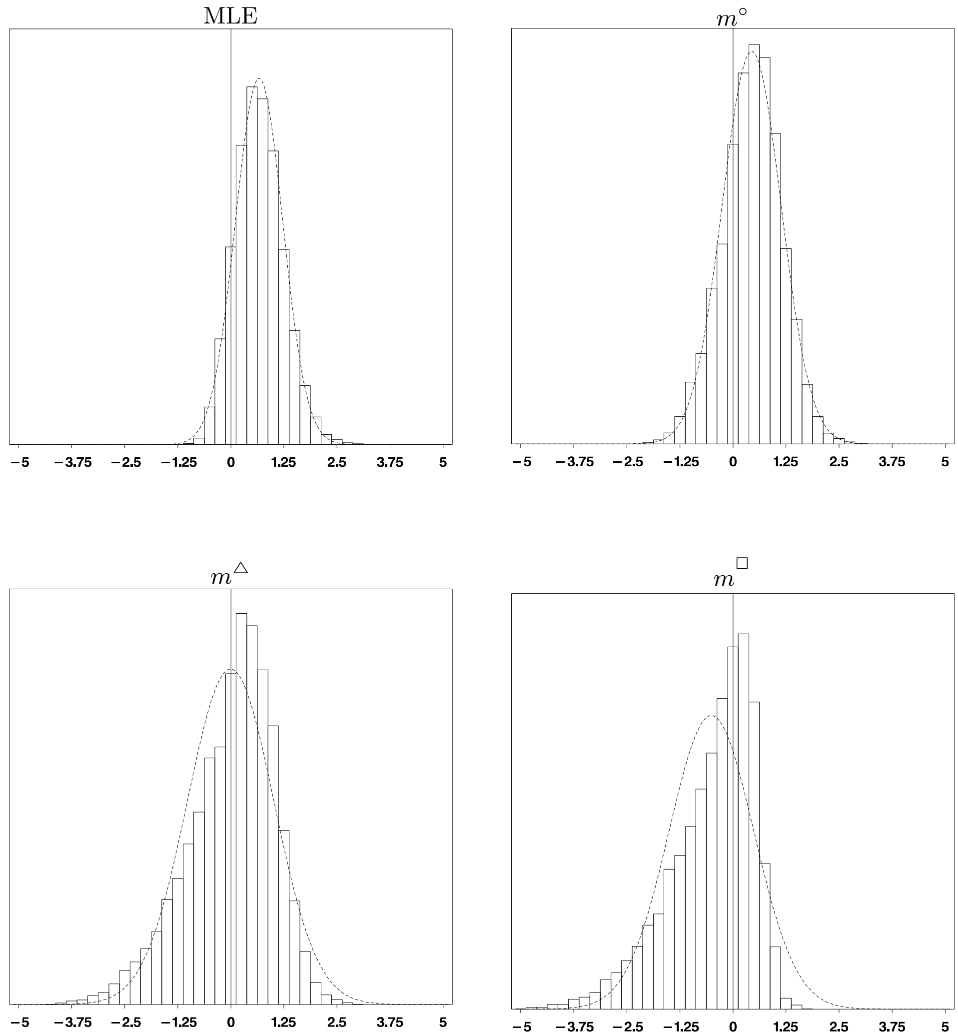

Fig. 8. Histograms of the normalized estimators $n^{(1/\alpha)}(\hat{m}_i - m)$, $i = 1 \ldots R$, $R =$ 10,000 , plotted with a normal probability density function whose parameters are taken from the sample.

Histograms n = 500 $\alpha = 2$ $k = 5$

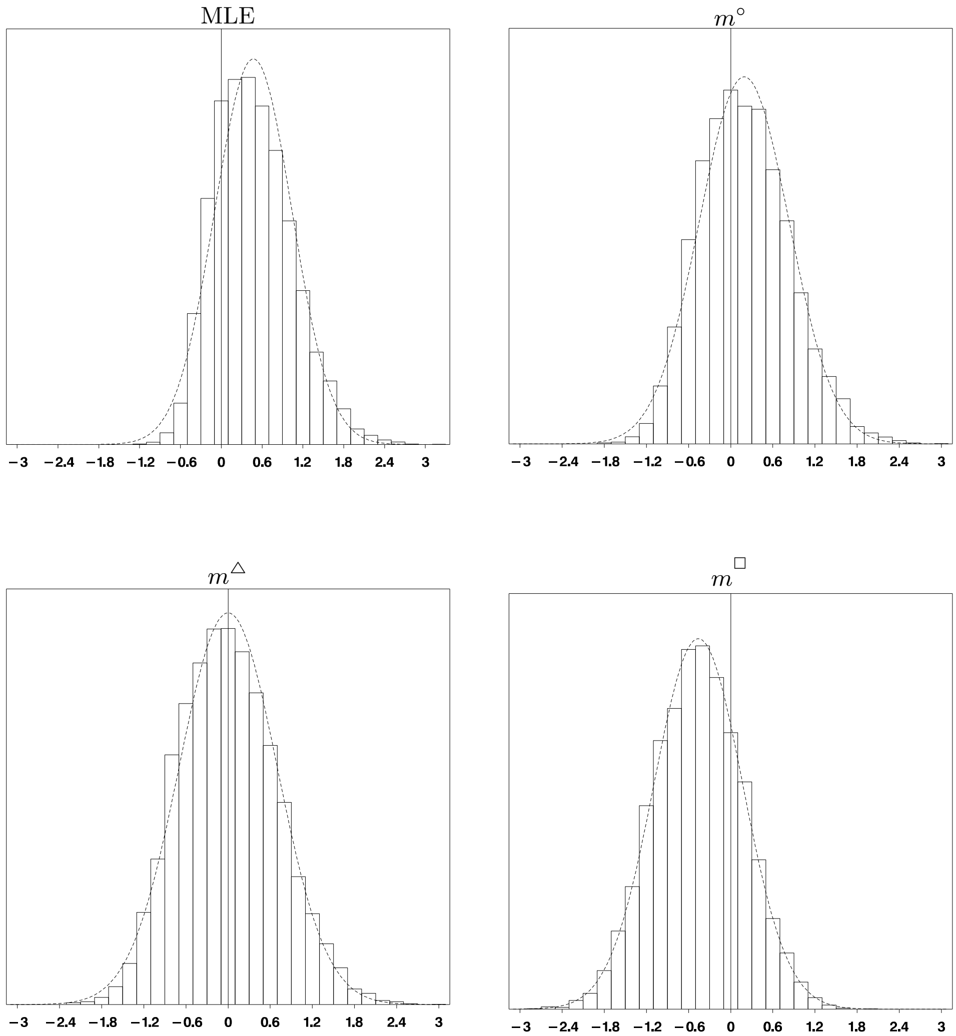

Fig. 9. Histograms of the normalized estimators $n^{(1/\alpha)}(\hat{m}_i - m)$, $i = 1 \ldots R$, $R =$ 10,000 , plotted with a normal probability density function whose parameters are taken from the sample.

$\alpha = 2$

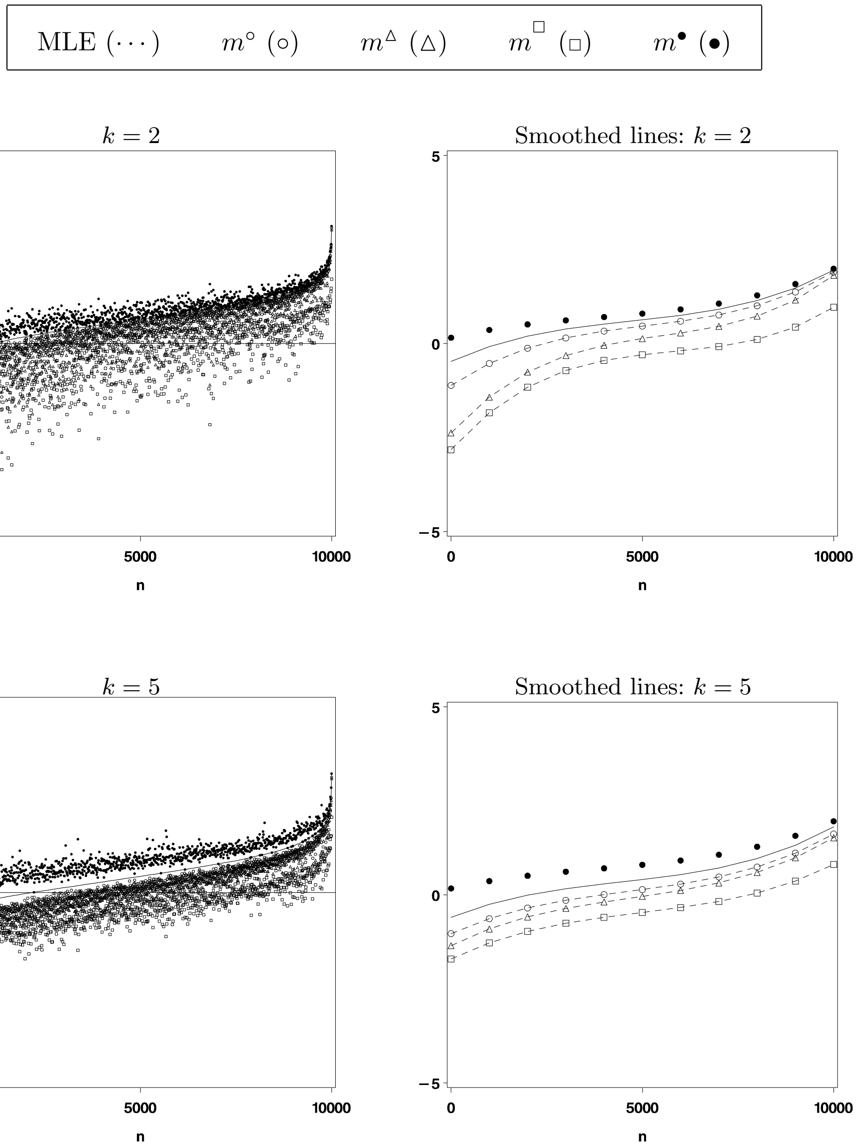

Fig. 10. Scatter plots showing $R = 10,000$ points of $n^{(1/\alpha)}(\hat{m}_i - m)$ for each estimator. For every fixed n, values of the maximum likelihood estimator and linear estimators are plotted. Graphs on the right hand side show the cubic regression lines for each estimator, calculated from the 10,000 points.

$\alpha = 5$

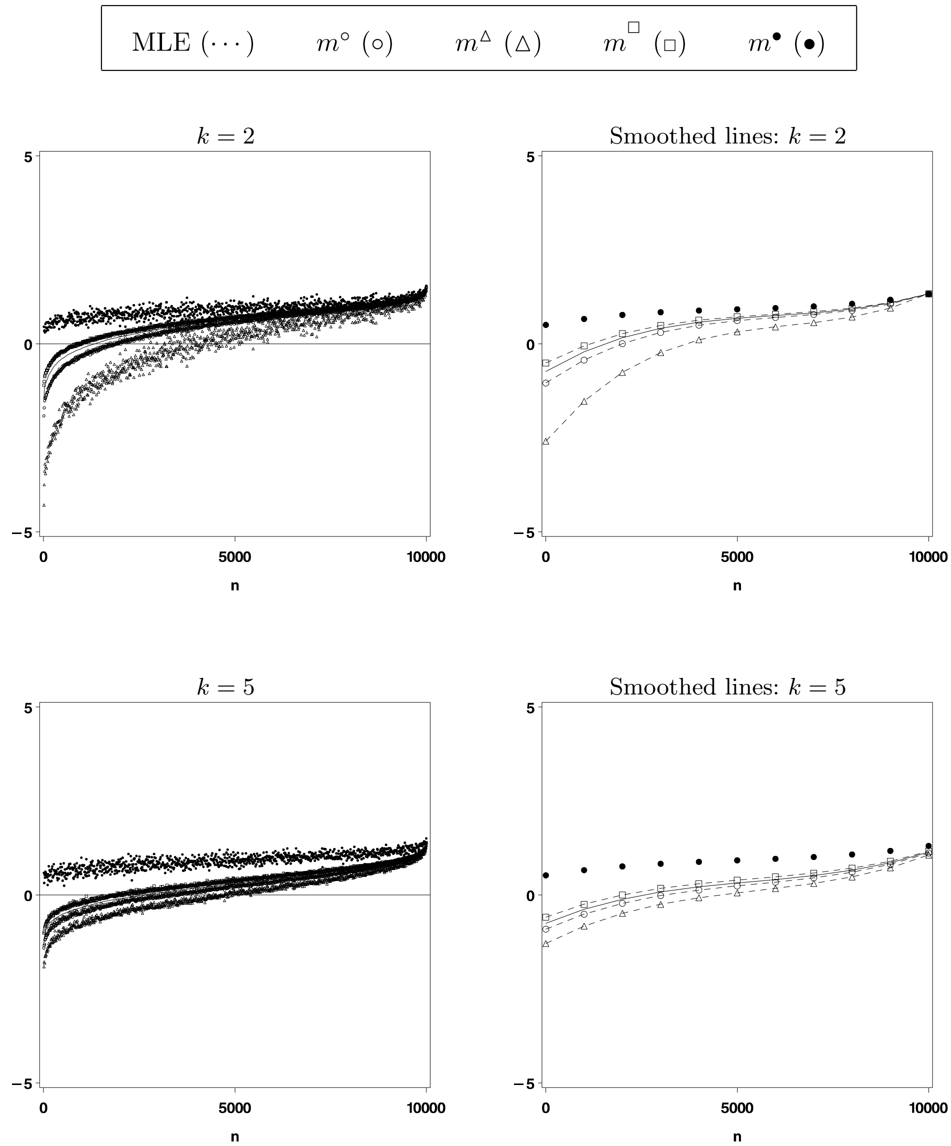

Fig. 11. Scatter plots showing $R = 10,000$ points of $n^{(1/\alpha)}(\hat{m}_i - m)$ for each estimator. For every fixed n, values of the maximum likelihood estimator and linear estimators are plotted. Graphs on the right hand side show the cubic regression lines for each estimator, calculated from the 10,000 points.

$\alpha = 10$

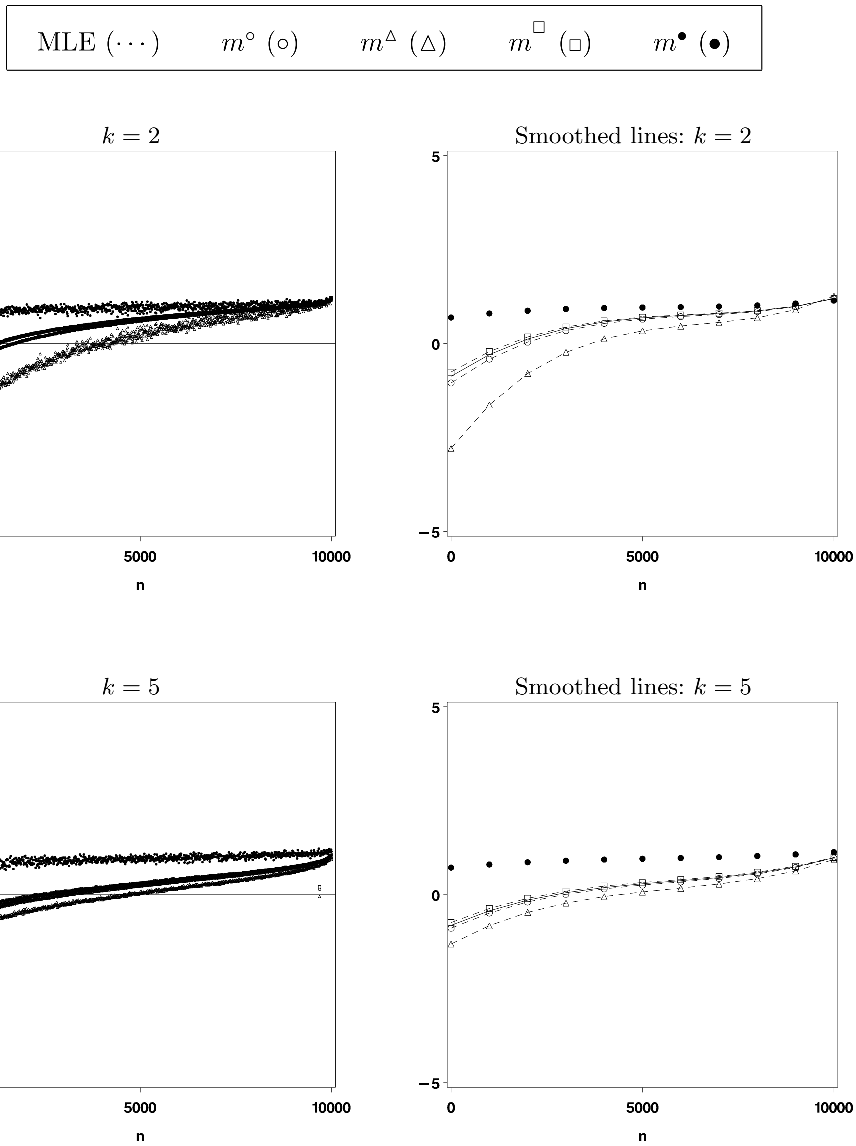

Fig. 12. Scatter plots showing $R = 10,000$ points of $n^{(1/\alpha)}(\hat{m}_i - m)$ for each estimator. For every fixed n, values of the maximum likelihood estimator and linear estimators are plotted. Graphs on the right hand side show the cubic regression lines for each estimator, calculated from the 10,000 points.

Scatter plots

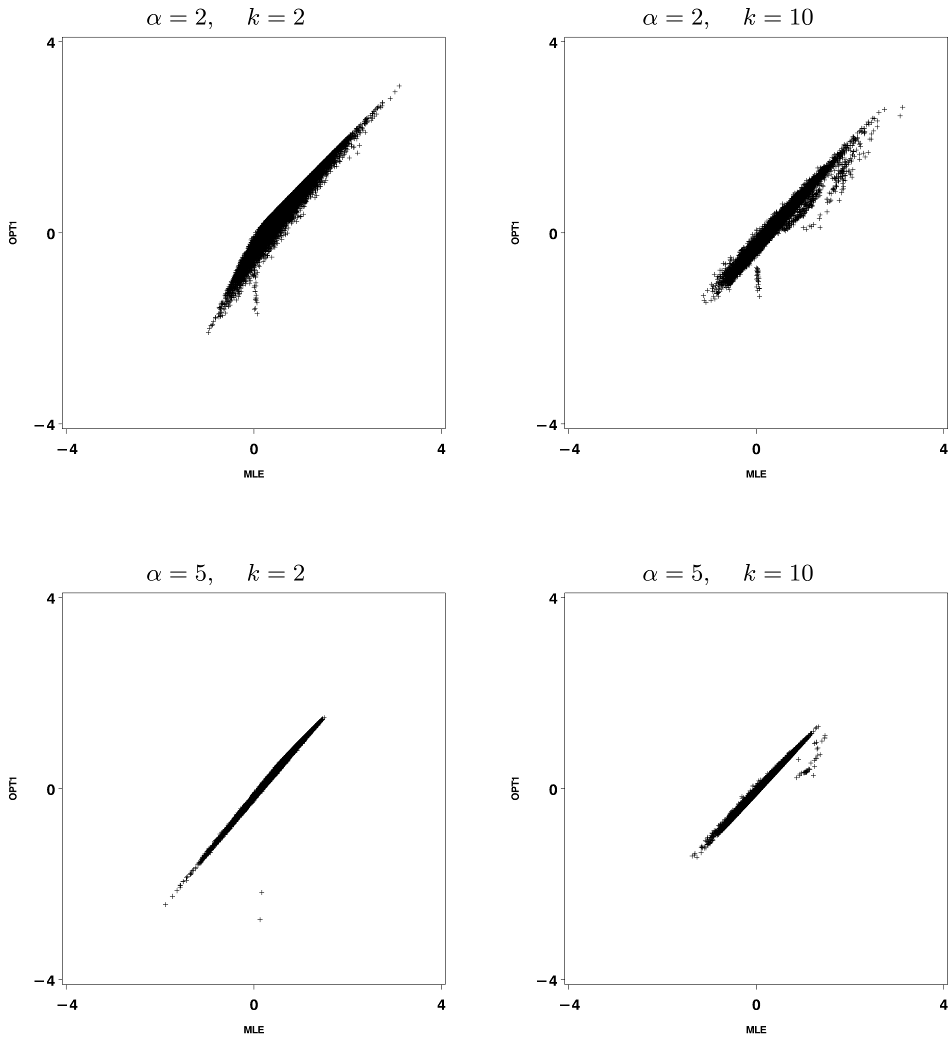

Fig. 13. Scatter plots showing $R = 10,000$ points of $n^{(1/\alpha)}(\hat{m}_i - m)$ for (MLE) plotted against the optimal linear estimator $\hat{m}_{n,k}(a^\circ)$, (OPT1).

References