# ON ESTIMATION OF PARAMETERS BY THE MINIMUM DISTANCE METHOD

V. P. SHULENIN
•
Tomsk State University, Tomsk, RUSSIA
shulenin-vp@rambler.ru

## ABSTRACT

Parameter estimates, constructed by the minimum distance method, are briefly called the *MD*-estimates. The minimum distance method has been proposed by Wolfowitz (1957). An extensive bibliography was compiled and published by Parr (1981). In this paper the effectiveness of the shift parameter estimation based on the use of Cramer - von Mises weighted distance is discussed. The robustness of this kind of *MD*-estimates under various supermodels describing deviations from the Gaussian model is considered. Numerical results are given for the case of contaminated normal distributions.

### Statement of the problem

Let us consider first a case when the statistical model $(\mathrm{X}, \Im_\theta)$ is given in parametric form. $\mathrm{X} = \{\vec{x}\}$ denotes the sample space, the elements of which are realizations $\vec{x} = (x_1,...,x_n)$ of a random vector $\vec{X} = (X_1,...,X_n)$; $\Im_\theta = \{F : F(x,\theta), \theta \in \Theta\}$ is a parametric set of admissible probability distributions for the experiment considered; $X_1,...,X_n$ is a sequence of i.i.d. random variables with the distribution function $F(x,\theta)$ and the density $f(x,\theta)$, $x \in R^1$, $\theta \in \Theta$. The functional form of the distribution is defined up to an unknown parameter (scalar or vector), which belongs to a given parameter set $\Theta$. It is required to construct the estimate of an unknown parameter $\theta \in \Theta$ based on a sample $X_1,...,X_n$ from a distribution $F(x,\theta)$.

.

### The essence of the minimum distance method

If a distance $\rho(F,G)$ between any two distributions, $F, G \in \Im$, is given, then parameter $\theta$ may be estimated by minimization of the distance between the empirical distribution function $F_n(x)$, constructed from a sample $X_1,...,X_n$, and the distribution function $F_\theta(x) = F_X(x,\theta)$ adopted in the model $(\mathrm{X}, \Im_\theta)$. Thus, for a chosen distance $\rho(F,G)$ $MD$-estimator for $\theta$ is defined as $\hat{\theta} = \arg\min_\theta \{\rho(F_n, F_\theta)\}$. Various distances could be used for constructing $MD$-estimates (see Parr, and Schucany (1980)). For instance, the maximum likelihood method is based on a distance

$$\rho(F_n, F_\theta) = -\int \ln f(x,\theta) dF_n(x).$$

In this paper, we consider the estimates that are based on the weighted Cramer - von Mises distance

$$\rho_W(F_n, F_\theta) = \int [F_n(x) - F_\theta(x)]^2 W_\theta(x, F_\theta)\, dF_\theta(x) \tag{1}$$

where $W_\theta = W(x, F_\theta)$ is a certain weight function, which may depend on d.f. $F_\theta$ (or on density $f_\theta$).

Assuming that $\rho_W(F_n, F_\theta)$ a differentiable function of the parameter $\theta$, its derivative is $\tilde{\lambda}_{F_n}(\theta) = \partial\rho_W(F_n, F_\theta)/\partial\theta$. With this notations, the estimation $\theta_n$ for parameter $\theta$ based on the use of weighted Cramer-von Mises distance (1) is a solution of the equation

$$\tilde{\lambda}_{F_n}(\theta) = -2\int[F_n(x) - F_\theta(x)]\frac{\partial F_\theta(x)}{\partial\theta}W_\theta(x)dF_\theta(x) + \int[F_n(x) - F_\theta(x)]^2\frac{\partial}{\partial\theta}[W_\theta f_\theta(x)]dx \qquad (2)$$

In this paper we consider the $MD$-estimation of the location parameter; in this case $F_\theta(x) = F(x-\theta)$. Let a family of reference distributions be designated as $\Im_0 = \{F : F_\theta(x) = F_0(x-\theta),\ \theta \in R^1\}$, where $F_0$ is a distribution with density $f_0$. Rewrite (1) as

$$\rho_{F_n,F_0}(\theta, W) = \int[F_n(x) - F_0(x-\theta)]^2 W(x-\theta)dx. \qquad (3)$$

Note that the choice of the weight function $W$ in the form of the density of reference distribution, i.e., in the form $W(x) = f_0(x)$, corresponds to the Cramer-von Mises distance; the choice of the weighting function $W(x) = f_0(x)/F_0(x)(1-F_0(x))$ gives the distance of Anderson-Darling (see for example, Boos (1981), Shulenin (1993a)). Assuming that $\rho_{F_n,F_0}(\theta, W)$ is a differentiable function of the parameter $\theta$, its derivative is $\lambda_{F_n}(\theta) = \partial\rho_{F_n,F_0}(\theta, W)/\partial\theta$. Then the equation $\lambda_{F_n}(\theta) = 0$ for the obtaining the $MD$-estimation, may be written in the form

$$\frac{2}{n}\sum_{i=1}^{n}\left[\frac{2i-1}{2n} - F_0(X_{(i)} - \theta)\right]W(X_{(i)} - \theta) = 0, \qquad (4)$$

where $X_{(1)},..., X_{(n)}$ the ordered statistics of the sample $X_1,..., X_n$.

**Asymptotic normality of the $MD$-estimators**

The asymptotic properties of $MD$-estimators were studied by several authors (see, for example, Boos (1981), Wiens (1987), Shulenin (1992)). In this paper, we discuss the asymptotic properties of estimators $\theta_n$ of the parameter of location $\theta$, which, for a given reference d.f. $F_0$, and given weight function $W$, is a solution of equation (4). There are two variants of parameter estimating:

**Version 1**. The distribution function $F$ of the observations $X_1,..., X_n$ is *known* and it coincides with the reference distribution function $F_0$, that is $F = F_0$ (or $F \in \Im_0$).

**Version 2**. The distribution function of the observations is *not known* and it is not necessarily the same as the reference distribution function, that is $F \neq F_0$ (or $F \notin \Im_0$).

Note that the $MD$-estimator $\theta_n$ of the location parameter $\theta$, which is the solution of equation (4), can be written as a functional of the empirical distribution function, in the form of $\theta_n = \theta(F_n)$. Here the functional $\theta(F)$ is defined either by relation

$$\min_\theta \rho_{F,F_0}(\theta, W) = \rho_{F,F_0}(\theta(F), W),$$

or may be given implicitly (as functional $T(F) = \theta(F)$) by expression

$$2\int[F(x+T(F)) - F_0(x)]f_0(x)W(x)dx - \int[F(x+T(F)) - F_0(x)]^2 W'(x)dx = 0. \qquad (5)$$

For studying the asymptotic properties of the $MD$-estimators $\theta_n = \theta(F_n)$ for the location parameter $\theta$, we use the approach of Mises (see Serfling, R. J. (1980), Shulenin (2012)). Let us consider the expansion of the form

$$\theta(F_n) = \theta(F) + V_{1n} + R_{1n}, \qquad (6)$$

where $V_{1n}$ is approximation statistics, and $R_{1n} = \theta(F_n) - \theta(F) - V_{1n}$ is the remainder of the expansion (6). Let us start from defining approximation statistics $V_{1n}$ and the remainder $R_{1n}$. It is necessary to compute the Gateaux differential of the first order $d_1T(F;G-F)$ for functional $T(F)$ defined by (5). Let $F_\lambda = F + \lambda(G-F)$, $0 \le \lambda \le 1$. Replacing the distribution function $F$ in (5) by the d.f. $F_\lambda$, we obtain the expression

$$2\int \{F(x+T(F_\lambda)) + \lambda[G(x+T(F_\lambda)) - F(x+T(F_\lambda))] - F_0(x)\} f_0(x)W(x)dx -$$

$$-\int \{F(x+T(F_\lambda)) + \lambda[G(x+T(F_\lambda)) - F(x+T(F_\lambda))] - F_0(x)\}^2 f_0(x)W^{/}(x)dx = 0.$$

Differentiating the expression on $\lambda$, setting $\lambda = 0$, and taking into account that $d_1T(F;G-F) = \partial T(F_\lambda)/\partial\lambda\,|_{\lambda=0}$, $T(F_\lambda)\,|_{\lambda=0} = T(F) = \theta$, we get

$$d_1T(F;G-F) = \frac{\int [G(x)-F(x)]\{[F(x)-F_0(x-\theta)]W^{/}(x-\theta) - f_0(x-\theta)W(x-\theta)\}dx}{\int f(x)f_0(x-\theta)W(x-\theta)dx - \int [F(x)-F_0(x-\theta)]f(x)W^{/}(x-\theta)dx}$$

From this expression, after replacing $G$ by the empirical d.f. $F_n$ , we get an approximation for statistics $V_{1n}$ :

$$V_{1n} = d_1T(F;F_n - F) = n^{-1}\sum IF(X_i;F,F_0,W).$$

Here $IF(u;F,F_0,W) = d_1T(F;\Delta_u - F)$, $0 \le u < \infty$, is the Hampel influence function for the $MD$ -estimator $\theta_n = \theta(F_n)$ of the location parameter $\theta$ , which for a given reference d.f. $F_0$ and given weight function $W$ is a solution of equation (4). Note that the expression for the influence function also follows from the above formula by replacing d.f. $G$ by degenerated at the point $u$ distribution function $\Delta_u$. The resulting formulas, together with the expansion (6), are the basis for the proof of asymptotic normality of the $MD$ -estimators, which are solutions of the equation (4).

Note that the general conditions of regularity (which impose restrictions on the behavior of the tails of d.f. $F$ and the weight function $W$) under which the expression $\sqrt{n}R_{1n} \to^p 0, n \to \infty$, and for which $MD$ - estimator is consistent and asymptotically normal, given in Boos (1981). In addition, the considered here $MD$ - estimates belong to the family of $MD_\alpha$ - estimates whose asymptotic properties are described in Shulenin (1992).

To facilitate formulating further results, let us denote by $\Im_S$ a family of absolutely continuous symmetric distributions. Let the class of weight functions $W_S$ consists of differentiable and even functions, that is $W(-x) = W(x)$ and

$$\int \{F(x)(1-F(x))\}^p W(x+c)dx < \infty, p > 0, c \in (-\infty, +\infty).$$

**Theorem**. Let $(F, F_0) \in \Im_S, W \in W_S$. Then, under fulfillment of the inequalities $0 < \sigma^2(F;F_0,W) = \int IF^2(x;F,F_0,W)dF(x) < \infty$ ,

the asymptotic expression can be written in the form of

$$L\{\sqrt{n}[\theta(F_n) - \theta(F)]/\sigma(F;F_0,W)\} = N(0,1), n \to \infty.$$

The asymptotic variance of $MD$ -estimate with the reference d.f. $F_0$ and the weight function $W$ under the distribution $F$ of observations $X_1,...,X_n$, is equal to $D(F;F_0,W) = \sigma^2(F;F_0,W)/n$; the Hampel influence function $IF(u;F,F_0,W) = -IF(-u;F,F_0,W)$ for the $MD$ -estimates is calculated by formulas

$$IF(u;F,F_0,W)=A_{F,F_0}(u;W)/B_{F,F_0}(W),\ 0\le u<\infty, \quad (7)$$

$$A_{F,F_0}(u;W)=\int_0^u W(x)dF(x)-W(u)[F(u)-F_0(u)], \quad (8)$$

$$B_{F,F_0}(W)=\int_{-\infty}^{\infty} f_0(x)W(x)dF(x)-\int_{-\infty}^{\infty}[F(x)-F_0(x)]W'(x)dF(x). \quad (9)$$

The proof can be found in Boos (1981), Wiens (1987), Parr and de Wet (1981).

Note that for the first version of parameter estimation $\theta$, when $F\in\Im_0$ the influence function $IF(u;F,W)$, $0\le u<\infty$ is given by

$$IF(u;F,W)=\frac{\int_{-\infty}^{+\infty}\{F(x)-I[u\le x]\}W(x)dF(x)}{\int_{-\infty}^{+\infty}f^2(x)W(x)dx}=\frac{\int_0^u W(x)dF(x)}{\int_{-\infty}^{+\infty}f(x)W(x)dF(x)}=$$

$$=J^{-1}(F,W)\int_0^u f(x)W(x)dx\ ,0\le u<\infty\ , \quad (10)$$

and the asymptotic variance of $\sqrt{n}\,MD$-estimate is given by

$$\sigma^2(F,W)=\frac{\int_{-\infty}^{+\infty}\left(\int_{-\infty}^{+\infty}\{F(y)-I[u\le y]\}W(y)dF(y)\right)^2dF(u)}{\left(\int_{-\infty}^{+\infty}f(x)W(x)dF(x)\right)^2}=\frac{\int_{-\infty}^{+\infty}\left(\int_0^x W(y)dF(y)\right)^2dF(x)}{\left(\int_{-\infty}^{+\infty}f^2(x)W(x)dx\right)^2}. \quad (11)$$

**Efficient *MD* - estimators**

For the first version of parameter $\theta$ estimation (when the distribution function $F$ of the observations $X_1,...,X_n$ is known and coincides with the reference function of a symmetric distribution $F_0$) there is an effective parameter estimate in the class of $MD$ - estimators. Its asymptotic variance is equal to the inverse of the Fisher information $I(f_0)$ about $\theta$ in distribution $F_0(x-\theta)$ with the density $f_0$. This score is determined by the effective weight function of the form

$$W^*(x)=a\,\frac{d^2\{-\ln f_0(x)\}}{dx^2}\cdot\frac{1}{f_0(x)}. \quad (12)$$

This effect was observed earlier in Boos (1981), Parr, De Wet (1981). Correctness of this fact can be seen from the following. Let us denote $\psi(x)=-f'(x)/f(x)$; then $\psi'(x)=d^2\{-\ln f(x)\}/dx^2$, and the expression (12) can be rewritten, taking into account that $F=F_0$, as $W(x)=a\,\psi'(x)/f(x)$. Substituting this weight function $W\in W_S$ in (11), and taking into account that $F\in\Im_S$, $\psi(0)=0$, we obtain

$$\sigma^2(F,W)=\frac{\int_{-\infty}^{+\infty}\left(\int_0^x W(y)dF(y)\right)^2dF(x)}{\left(\int_{-\infty}^{+\infty}f^2(x)W(x)dx\right)^2}=\frac{a^2\int_{-\infty}^{+\infty}\psi^2(x)dF(x)}{a^2\left(\int_{-\infty}^{+\infty}\psi'(x)dF(x)\right)^2}=\frac{I(f)}{I^2(f)}=\frac{1}{I(f)}.$$

**Example 1**. Note that the use of (12) allows to find the distribution function $F_0$, under which the Cramer - von Mises $MD$-estimator with the weighting function $W(x)=f_0(x)$ produces asymptotically efficient parameter estimates. In fact, solving the differential equation $d^2\{-\ln f_0(x)\}/dx^2=a\cdot f_0^2(x)$ under $W(x)=f_0(x)$, we obtain the density of the form

$$f_0(x) = 2/[\pi(e^x + e^{-x})] = (1/\pi)\sec h(x)\,,\, x \in R^1,$$

with the distribution function

$$F_0(x) = (2/\pi)\, arctg(e^x)\,,\, x \in R^1,$$

which is called the hyperbolic secant. Note that the Fisher information for the parameter $\theta$ in the density $f_0(x) = (1/\pi)\sec h(x)$ is hyperbolic secant as for the Cauchy distribution, and is equal $I(f_0) = 1/2$. Hence $\sigma^2(F_0, W = f_0) = 2$. Note, in addition, that the influence function for $MD$ - estimation with the weighting function $W \equiv 1$, with $F = F_0$ is *limited* and defined as

$$IF(x; F_0, W \equiv 1) = \frac{F_0 - (1/2)}{\int_0^1 f_0(F_0^{-1}(t))dt} = \frac{(2/\pi)arctg(e^x) - (1/2)}{(2/\pi^2)} = \pi\, arctg(e^x) - (\pi^2/4)\ ,\, x \in R^1.$$

The asymptotic variance of the $MD$ - estimate with weight function $W \equiv 1$ and $F = F_0$ is the same as the asymptotic variance of Hodges - Lehmann estimate $HL$, and for distribution $F_0(x) = (2/\pi)\, arctg(e^x)$ is given by

$$\sigma^2(F_0, W = 1) = \frac{1}{12\left(\int_0^1 f_0(F_0^{-1}(t))dt\right)^2} =$$

$$= \frac{1}{12\left((2/\pi)\int_0^1 \sin(\pi t/2)\cos(\pi t/2)dt\right)^2} = \frac{\pi^4}{48} \approx 2{,}029 = \sigma^2(F_0, HL)\ .$$

**Example 2**. Let the supermodel $\Im_S^* = \{F_{(1)}, F_{(2)}, F_{(3)}, F_{(4)}, F_{(5)}\}$ be a finite set of distributions, where $F_{(1)} = \Phi$ is the standard normal distribution, Fisher information $I(f_{(1)}) = 1$; $F_{(2)}$ is logistic, $I(f_{(2)}) = 1/3$; $F_{(3)}$ is Laplace, $I(f_{(3)}) = 1$; $F_{(4)}$ is Cauchy, $I(f_{(2)}) = 1/2$; $F_{(5)}$ is hyperbolic secant, $I(f_{(5)}) = 1/2$. Optimal weight functions of the form (12) for these distributions are given in Table 1 and in Figure 1.

Table 1. Optimal weight functions of the form $W^*(x) = a \cdot \psi'(x)/f(x)$

| $F_{(1)}$ | $F_{(2)}$ | $F_{(3)}$ | $F_{(4)}$ | $F_{(5)}$ |
|---|---|---|---|---|
| $W_{(1)}^*(x) = 1/\phi(x)$ | $W_{(2)}^*(x) \equiv 1$ | $W_{(3)}^*(x) = 2e^{\lvert x\rvert}\delta(x-0)$ | $W_{(4)}^*(x) = (1-x^2)/(1+x^2)$ | $W_{(5)}^*(x) = (2/\pi)(e^x + e^{-x})^{-1}$ |

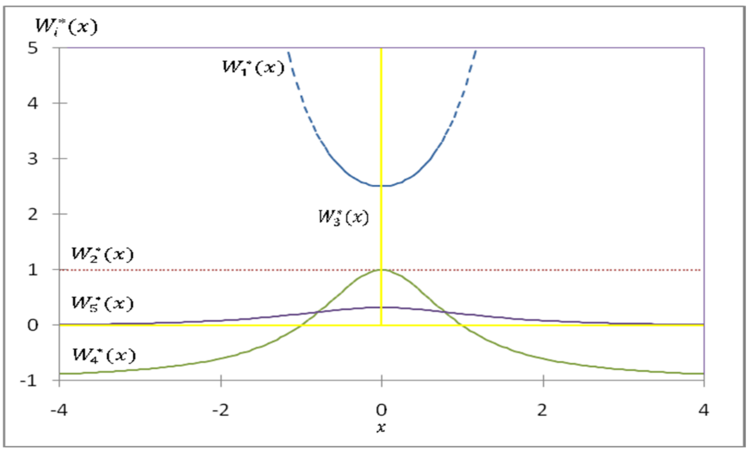


**Fig.1. Optimal weight functions-estimates for $F \in \Im_S^*$**

Note that the asymptotic variance of $MD$- estimate with the reference distribution $F_0(x) = F(x)$ and the weight function $W(x) = 1/f(x)$ coincides with the asymptotic variance of the sample mean $\overline{X}$, and is calculated by the formula

$$\sigma^2(F, W = 1/f) = \frac{\int_{-\infty}^{+\infty}\left(\int_0^x W(y)dF(y)\right)^2 dF(x)}{\left(\int_{-\infty}^{+\infty} f^2(x)W(x)dx\right)^2} = \frac{\int_{-\infty}^{+\infty}\left(\int_0^x (1/f(y))f(y)dy\right)^2 dF(x)}{\left(\int_{-\infty}^{+\infty} f^2(x)(1/f(x))dx\right)^2} = \int_{-\infty}^{\infty} x^2 dF(x).$$

For the weight function $W(x) = 1/\phi(x)$, where $\phi(x)$ is the standard normal density, $MD$- estimator is an efficient estimate of the location parameter $\theta$ of the normal distribution, but it has, like the sample mean $\overline{X}$, the *unlimited* influence function $IF(x;\Phi, W = 1/\phi) = x$, $x \in R^1$ and its sensitivity to gross errors is not limited, that is $\gamma^*(\Phi, W = 1/\phi) = \infty$. Note also that the choice of the weighting function $W(x) \equiv 1$ leads to asymptotically efficient $MD$- estimator for the logistic cdf $F_{(2)}$ (the variance in this case coincides with the variance of $HL$- estimator), and the absolute efficiency of the $MD$- estimator with weight function $W(x) = f_{(2)}(x)$ is equal to $AE(F_{(2)}, W = f_{(2)}) = [3{,}036(1/3)]^{-1} = 0{,}988$. Recall that for the logistic distribution $F_{(2)}$ with density $f_{(2)}$, the equality $f_{(2)} = F_{(2)}(1 - F_{(2)})$ holds, and therefore, the choice of the weighting function in the form inherent in $MD$- estimation based on the use of the Anderson-Darling distance, $W(x) = f_0 / F_0(1 - F_0)$, also leads to an effective $MD$-estimation for the logistic distribution. For the Laplace distribution with density $f_{(3)}(x) = (1/2)\exp(-|x|)$, $x \in R^1$ function $\psi(x) = -f_{(3)}^{/}(x)/f_{(3)}(x) = sign(x)$ and, therefore, the optimal weight function $W^*(x) = a \cdot \psi^{/}(x)/f(x)$ defined by (12), takes the form $W_{(3)}^*(x) = \{sign(x)\}^{/} / f_{(3)}(x) = \delta(x-0)/f_{(3)}(x) = 2e^{|x|}\delta(x-0)$. Using this expression for the optimal weight function, and (11), one may see that the asymptotic variance of $MD$- estimate coincides with the asymptotic variance of the sample median $\overline{X}_{1/2}$, which is asymptotically efficient estimate of parameter $\theta$ for the Laplace distribution. In fact, from (11) with the weighting function $W(x) = \delta(x-0)/f(x)$, we obtain:

$$\sigma^2(F, W) = \frac{\int_{-\infty}^{+\infty}\left(\int_{-\infty}^{+\infty}\{F(y) - I[u \le y]\}W(y)dF(y)\right)^2 dF(u)}{\left(\int_{-\infty}^{+\infty} f(x)W(x)dF(x)\right)^2} =$$

$$= \frac{\int_{-\infty}^{+\infty}\left(\int_{-\infty}^{+\infty}\{F(y) - I[u \le y]\}\delta(y-0)dy\right)^2 dF(u)}{\left(\int_{-\infty}^{+\infty} f(x)\delta(x-0)dx\right)^2} =$$

$$= \frac{\int_{-\infty}^{+\infty}\{F(0) - I[u \le 0]\}^2 dF(u)}{f^2(0)} =$$

$$= \frac{(1/4) - \int_{-\infty}^{+\infty} I[u \le 0]dF(u) + \int_{-\infty}^{+\infty} I^2[u \le 0]dF(u)}{f^2(0)} = \frac{1}{4f^2(0)} = \sigma^2(F, \overline{X}_{1/2}).$$

Note that for the Cauchy distribution the optimal weight function $W^*_{(4)}(x) = a(1-x^2)/(1+x^2)$ is negative outside the interval $[-1, 1]$. This fact can be explained as follows. From (10) it follows that the weight function $W$ is expressed through the derivative of the influence function in the form $W(u) = J(F,W)\, IF^{/}(u;F,W)/f(u)$, $0 \le u < \infty$. So, to "reduce" the influence outliers on the $MD$-estimation, it is necessary its influence function to decrease for large values of the argument and, consequently, the weight function should be *negative*, as is observed for the optimal weight function $W^*_{(4)}(x) = a(1-x^2)/(1+x^2)$ for the Cauchy distribution.

**Example 3**. Consider the family of t-distributions $\Im_r \in \Im_S$, for which the density distribution $f_r(x)$ with degrees of freedom $r$ can be written as

$$f_r(x) = A(r)(1+(x^2/r))^{-(r+1)/2}, x \in R^1, A(r) = \Gamma((r+1)/2)/\sqrt{r\pi}\,\Gamma(r/2).$$

Using (11), we can see that the optimal weight function for this family of distributions is calculated by the formula

$$W^*_r(x) = a \cdot r^{-(r+1)/2}(r+1)A^{-1}(r)(r-x^2)(r+x^2)^{(r-3)/2} \ .$$

Hence, under $r=1$ we obtain the optimal weight function for Cauchy distributions as $W^*_r(x)|_{r=1} = a \cdot 2\pi(1-x^2)/(1+x^2) = W^*_{(4)}(x)$. The case of $r \to \infty$ corresponds to the normal distribution. Given that under $r \to \infty$, the expressions $A(r) \to 1/\sqrt{2\pi}$ and $(1+(x^2/r))^{-(r+1)/2} \to e^{-x^2/2}$ are hold, from the general formula, we obtain:

$$\lim_{r\to\infty} W^*_r(x) = a \cdot \sqrt{2\pi}\, \exp(x^2/2) = a \cdot 1/\phi(x) = W^*_{(1)}(x).$$

**Robustness of the MD-estimators**

To study the properties of robustness, we consider two types of supermodels that describe deviations from the Gaussian model of observations. The first supermodel $\Im^*_S$, which was used in Example 2, is defined as a finite set of given distributions, that is, $\Im^*_S = \{F_{(1)}, F_{(2)}, F_{(3)}, F_{(4)}, F_{(5)}\}$.

Second supermodel $\Im_{\varepsilon,\tau}(\Phi)$ called Gaussian model with scale contamination, is determined as

$$\Im_{\varepsilon,\tau}(\Phi) = \{F : F_{\varepsilon,\tau}(x) = (1-\varepsilon)\Phi(x) + \varepsilon\Phi(x/\tau)\} \ , 0 \le \varepsilon \le 1, \tau \ge 1,$$

where $\Phi(x)$ is the standard normal distribution function with density $\phi(x)$, $\varepsilon$ - the proportion of sample contamination, and $\tau$ is a parameter of the scale contamination.

**Example 4. The first option**. First, we consider the properties of $MD$-estimators within a supermodel under different types of reference cdf $F_0$ and weighting functions $W$. For the first version of parameter $\theta$ estimation (when the distribution function $F$ is known and equals the reference distribution function $F_0$, that is $F \in \Im_0$), the influence function of $MD$-estimation and its asymptotic variance are given by (10) and (11). Let us consider various types of the weighting function $W \in W_S$.

A) Let $W(x) \equiv 1$, $F(x) = F_0(x)$. Under these conditions the $MD$-estimators with the weight function $W(x) \equiv 1$ are $B$-robust, that is, they have *limited* influence functions, which are defined as $IF(x;F,W\equiv 1) = \{2F(x)-1\}/2\int f^2(x)dx$. In the Gaussian case $F = \Phi$, the influence function is given by $IF(x;\Phi,W\equiv 1) = \sqrt{\pi}[2\Phi(x)-1]$. The sensitivity to gross errors

$\gamma^*(F,T)=\sup_x|IF(x;F,T)|$ of $MD$-estimators with the weighting function $W(x)\equiv 1$ is equal to $\gamma^*(\Phi, W\equiv 1)=\sqrt{\pi}\approx 1{,}77$.

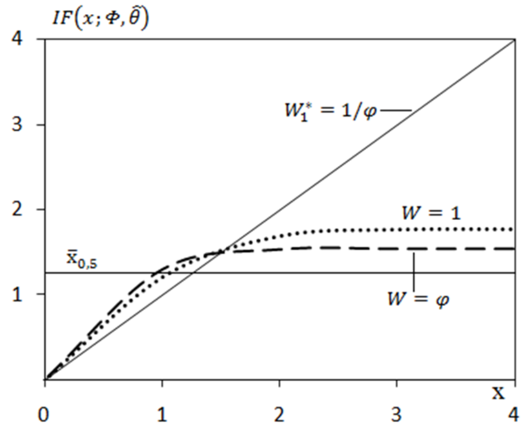


**Fig. 2. Influence function of $MD$-estimators for the normal distribution**

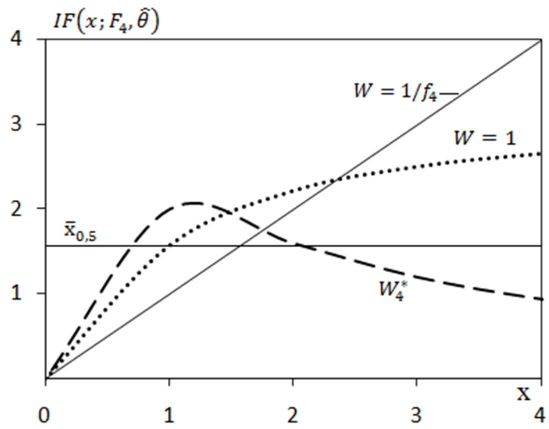


**Fig. 3. Influence function of $MD$-estimators for the Cauchy distribution**

**B)** Let the weight function coincides with the reference density, $W(x)=f_0(x)$, and $F(x)=F_0(x)$. Under these assumptions the asymptotic variance of the $MD$ - estimation is given by

$$\sigma^2(F,W=f)=\frac{\int_{-\infty}^{+\infty}\left(\int_0^x f^2(y)dy\right)^2 dF(x)}{\left(\int_{-\infty}^{+\infty} f^3(x)dx\right)^2}.$$

Note that for a Gaussian distribution $F(x)=\Phi(x)$ and the weight function $W(x)=\phi(x)=(1/\sqrt{2\pi})\exp\{-x^2/2\}$ we obtain from (10) the *limited* influence function

$$IF(x;\Phi,W=\phi)=(\sqrt{3\pi}/2)\tilde{\Phi}(x)=(\sqrt{3\pi}/2)[2\Phi(x\sqrt{2})-1],\, x\in R^1,$$

where $\tilde{\Phi}(x)$ is the Laplace function given by

$$\tilde{\Phi}(x)=(2/\sqrt{\pi})\int_0^x \exp\{-x^2\}dx,\, \tilde{\Phi}(x)=2\Phi(x\sqrt{2})-1,\, x\ge 0, \qquad \Phi(x)=(1/\sqrt{2\pi})\int_{-\infty}^x \exp\{-x^2/2\}dx.$$

Sensitivity to gross errors $\gamma^*(F,T)$ of $MD$ - estimation, with the weighting function $W(x)=\phi(x)$, is equal to $\gamma^*(\Phi,W=\phi)=\sqrt{3\pi}/2=1{,}53$. In this case, the asymptotic variance of $\sqrt{n}\,MD$ - estimation is

$$\sigma^2(\Phi,W=\phi)=2\int_0^\infty IF^2(x;\Phi,W=\phi)\,d\Phi(x)=\frac{3\pi}{2}\cdot\frac{1}{\sqrt{2\pi}}\int_0^\infty \tilde{\Phi}^2(x)e^{-x^2/2}dx=$$

$$=(3/2)\,arctg(2/\sqrt{5})=1{,}095.$$

The asymptotic variance of the $MD$ - estimators for the cases (A) and (B) were calculated for the following distributions: $F_{(1)}$ - normal, $F_{(2)}$ - logistic, $F_{(3)}$ - Laplace, $F_{(4)}$ - Cauchy, $F_{(5)}$ -hyperbolic secant. Numerical calculations derived from formulas for $F_0=F_{(i)}, i=1,\dots,5$ and with different weight functions are shown in Table 2.

Table 2. The asymptotic variance of $\sqrt{n}\,MD$-estimators for the supermodel $\Im_S^*$ at $F_0 = F_{(i)}$, $i = 1,...,5$

| The weight function | $F_{(1)} = \Phi$ | $F_{(2)}$ | $F_{(3)}$ | $F_{(4)}$ | $F_{(5)}$ |
|---|---|---|---|---|---|
| $W \equiv 1$ | 1,047 (0,96) | 3,000 **(1,00)** | 1,333 (0,75) | 3,287 (0,61) | 2,029 (0,98) |
| $W_{(i)}(x) = f_{(i)}(x)$ | 1,095 (0,91) | 3,036 (0,99) | 1,200 (0,83) | 2,573 (0,78) | 2,000 **(1,00)** |
| $\widetilde{W}_{(i)}(x) = f_{(i)} / F_{(i)}(1-F_{(i)})$ | 1,035 (0,97) | 3,000 **(1,00)** | 1,262 (0,79) | 2,317 (0,86) | 2,020 (0,99) |
| $W_{(i)}(x) = 1/ f_{(i)}(x)$ | 1,000 **(1,00)** | 3,290 (0,91) | 2,000 (0,50) | $\infty$ (0,00) | 2,467 (0,81) |
| $W_{(4)}^*(x) = (1-x^2)/(1+x^2)$ | 1,109 (0,90) | 4,204 (0,71) | 1,230 (0,81) | 2,000 **(1,00)** | 2,103 (0,95) |

The absolute values of efficiency of $MD$-estimates are given in parentheses, they were calculated according to the formula $AE(F,W) = [\sigma^2(F,W)I(f)]^{-1}$. Note that for distributions with "heavy tails" (Cauchy and Laplace), the absolute efficiency of $MD$-estimators depends mainly on the choice of the weighting function $W$. For normal distribution, the optimal weight function is $W_{(1)}(x) = 1/ f_{(1)}(x)$. Weight functions $W \equiv 1$ and $W_{(2)}(x) = f_{(2)} / F_{(2)}(1-F_{(2)})$ are optimal for the logistic distribution $F_{(2)}$. Weight function $W_{(4)}^*(x) = (1-x^2)/(1+x^2)$ is optimal for the Cauchy distribution. Weight function $W_{(5)}(x) = f_{(5)}(x)$ is optimal for distribution $F_{(5)}$ - hyperbolic secant.

**Example 5. The second option**. Consider the case when $F \neq F_0$, and the supermodel $\Im_S^* = \{F_{(1)}, F_{(2)}, F_{(3)}, F_{(4)}, F_{(5)}\}$ is the finite set of distributions, $F \in \Im_S^*$. In this case, the asymptotic variance of $\sqrt{n}MD$-estimators under the weight function $W = 1$ is given by

$$\sigma^2(F, F_0, W \equiv 1) = \frac{2\int_0^{\infty} [F_0(u) - (1/2)]^2 dF(u)}{\left(\int f_0(x) f(x) dx\right)^2}, F \in \Im_S^*. \quad (13)$$

The numerical values of the asymptotic variance of $\sqrt{n}MD$-estimators for $F \in \Im_S^*$ and the weight function $W = 1$, calculated using the formula (13). are shown in Table 3.

Table 3. Asymptotic variance of $\sqrt{n}MD$-estimators, for $\hat{\theta}_{(i)} = \hat{\theta}(F_0 = F_{(i)}, W \equiv 1)$, $i = 1,...,5$, $F \in \Im_S^*$

| $\hat{\theta} \setminus F$ | $F_{(1)}$ | $F_{(2)}$ | $F_{(3)}$ | $F_{(4)}$ | $F_{(5)}$ | $d(\hat{\theta}, \Im_S^*)$ |
|---|---|---|---|---|---|---|
| $\hat{\theta}_{(1)}$ | 1,047 (0,96) | 3,051 (0,98) | 1,383 (0,72) | 2,911 (0,69) | 2,008 (0,99) | 0,42 |
| $\hat{\theta}_{(2)}$ | 1,016 (0,98) | 3,000 (1,00) | 1,524 (0,66) | 3,679 (0,54) | 2,069 (0,97) | 0,57 |
| $\hat{\theta}_{(3)}$ | 1,059 (0,94) | 3,048 (0,98) | 1,333 (0,75) | 2,957 (0,68) | 2,006 (0,99) | **0,41** |
| $\hat{\theta}_{(4)}$ | 1,046 (0,96) | 3,025 (0,99) | 1,385 (0,72) | 3,290 (0,61) | 2,017 (0,99) | 0,48 |
| $\hat{\theta}_{(5)}$ | 1,031 (0,97) | 3,011 (0,99) | 1,439 (0,70) | 3,276 (0,61) | 2,029 (0,98) | 0,49 |

Note that in the table (3) in parentheses the absolute efficiency estimates are presented, calculated by the formula $AE(F, \hat{\theta}) = \{\sigma^2(F, F_0, W \equiv 1) I(f)\}^{-1}$. In the last column of the table, the defects of the estimates in the supermodel $\Im_S^*$, calculated from (19), are given.

**Note 1**. One of convenient means for comparing qualities of estimates $\hat{\theta}_1,...,\hat{\theta}_k$ of a given parameter $\theta$ of a symmetric distribution $F$ is a concept of defect of the estimator (see, for example, Andrews *at all*. (972), Shulenin (2012)). Let $\hat{\theta}_1,...,\hat{\theta}_k$ be a finite set of asymptotically normal and unbiased estimates of the location parameter $\theta$, based on a sample $X_1,...,X_n$ from the distribution $F$, obeying the expression

$$L\left\{\frac{\sqrt{n}(\hat{\theta}_i-\theta)}{\sigma_F(\hat{\theta}_i)}\right\}=N(0,\ 1),\quad n\to\infty,\quad i=1,...,k.$$

Defect of estimator $\hat{\theta}_i, i=1,...,k$ among the compared parameter estimates $\hat{\theta}_1,...,\hat{\theta}_k$ for a symmetrical distribution $F$ is defined as

$$DE_F(\hat{\theta}_i)=1-\min\{\sigma_F^2(\hat{\theta}_1),...,\sigma_F^2(\hat{\theta}_k)\}/\sigma_F^2(\hat{\theta}_i),\ i=1,...,k. \quad (14)$$

Note that if among the estimators $\hat{\theta}_1,...,\hat{\theta}_k$ there is an effective estimate, for which $\sigma_F^2(\hat{\theta}^*)=1/I(f)$ and, therefore, $\min\{\sigma_F^2(\hat{\theta}_1),...,\sigma_F^2(\hat{\theta}_k)\}=1/I(f)$, then the absolute defect of the estimator $\hat{\theta}_i$ is equal to one minus its absolute efficiency, i.e.,

$$ADE_F(\hat{\theta}_i)=1-A\text{Э}_F(\hat{\theta}_i),\ i=1,...,k. \quad (15)$$

**Note 2**. Studying robustness of compared estimates $\hat{\theta}_1,...,\hat{\theta}_k$ of the location parameter $\theta$ in the supermodel $\mathfrak{I}$ consisting of a finite set of symmetric distributions, $\mathfrak{I}=\{F_1,...,F_r\}$, usually is made by observing the disposition of estimates' defects on the plane of two distributions. The defect for basic (ideal, usually a Gaussian) model is laid along the horizontal axis, and along vertical axis the defects for an alternative model, which is a part of a supermodel $\mathfrak{I}=\{F_1,...,F_r\}$, is laid. With this visual representation of the defects count on the plane of the two distributions, the preference is given to the estimate, which is closest to the origin. As examples, the absolute defects of estimates are presented on the plane of distributions "Gauss-Laplace" and "Gauss-Cauchy", see Figures (4) and (5).

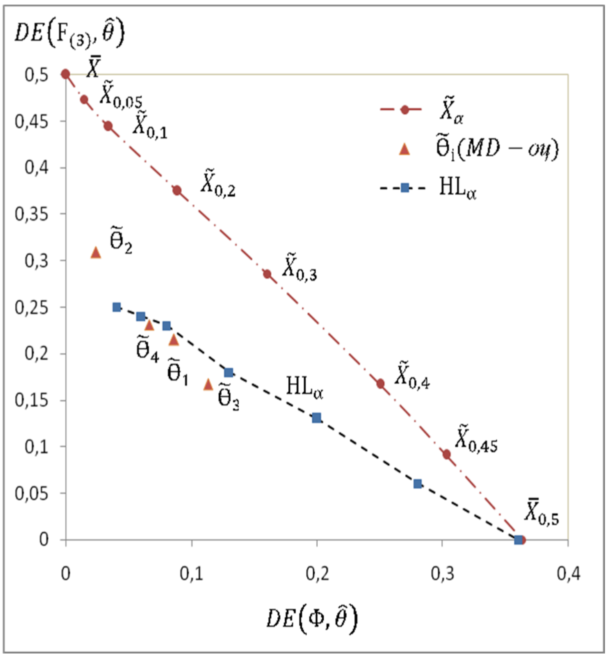


**Fig. 4. Defects estimates in the plane "Gauss-Cauchy"**

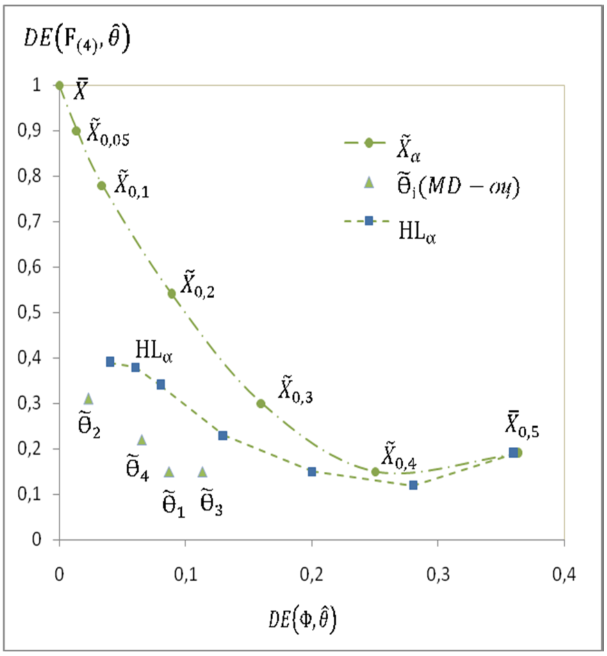


**Fig. 5. Defects estimates in the plane "Gauss-Laplace"**

The advantages of the $MD$-estimates $\hat{\theta}_{(i)} = \hat{\theta}(F_0 = F_{(i)}, W = f_{(i)})$, $i = 1,...,5$ for $F \in \Im_S^*$ before the family $\widetilde{X}_\alpha$ - Winzor-means and family $HL_\alpha$-estimates Hodges-Lehmann $0 \le \alpha \le 1/2$.
are clearly seen in these figures (they are placed closer to the origin).

**Note 3**. If we want to draw a conclusion on the preferenced estimator among compared estimates $\hat{\theta}_1,...,\hat{\theta}_k$ of the parameter $\theta$ within the entire supermodel $\Im = \{F_1,...,F_r\}$, we can use the Euclidean metric using the above notations:

$$d(\hat{\theta}_i; \Im) = \left\{ \sum_{j=1}^{r} [DE_{F_j}(\hat{\theta}_i)]^2 \right\}^{1/2}, \tag{16}$$

or

$$Ad(\hat{\theta}_i; \Im) = \left\{ \sum_{j=1}^{r} [ADE_{F_j}(\hat{\theta}_i)]^2 \right\}^{1/2}, \quad i = 1,...,k.$$

(17)

The preference is given to the estimator $\hat{\theta}_i$ with the minimal value of $d(\hat{\theta}_i; \Im)$, that is

$$d(\hat{\theta}_i; \Im) = \min\{d(\hat{\theta}_1; \Im),...,d(\hat{\theta}_k; \Im)\}. \tag{18}$$

For the supermodel $\Im_S^* = \{F_{(1)}, F_{(2)}, F_{(3)}, F_{(4)}, F_{(5)}\}$, the formula (16) can be written as

$$d(\widetilde{\theta}_{(i)}, \Im_S^*) = \left( \sum_{j=1}^{5} [1 - \{\sigma^2(F_{(j)}, \widetilde{\theta}_{(i)}) I(f_{(j)})\}^{-1}]^2 \right)^{1/2} = \left( \sum_{j=1}^{5} [1 - A\text{Э}(F_{(j)}, \widetilde{\theta}_{(i)})]^2 \right)^{1/2}, i = 1,..., 5. \tag{19}$$

According to the criterion (18), the preference among estimators $\widetilde{\theta}_{(1)},...,\widetilde{\theta}_{(5)}$ in the supermodel $\Im_S^*$, should be given to the $MD$ - estimator for $F_0 = F_{(3)}$ with reference Laplace distribution, and with weight function $W \equiv 1$, since this estimator has the minimum value of

$$d(\hat{\theta}_{(3)}, \Im_S^*) = \min\{d(\hat{\theta}_{(i)}, \Im_S^*),\ i = 1,...,5\} = 0{,}41$$

(see the last column of Table 3). Compare it with that of Hodges-Lehmann $d(HL, \Im_S^*) = 0{,}47$, of $\widetilde{X}_\alpha$ - Winzor-mean $d(\widetilde{X}_{0,45}, \Im_S^*) = 0{,}41$; of the sample median $d(\overline{X}_{1/2}, \Im_S^*) = 0{,}51$; of the sample mean $d(\overline{X}, \Im_S^*) = 1{,}14$, Shulenin (2012, p.256).

**Example 6. The second option**. Consider the Gaussian model with a scale contamination $\Im_{\varepsilon,\tau}(\Phi)$. Let the reference distribution be a normal distribution $F_0 = \Phi$, and the distribution of the observations is characterized by normal distribution with a scale contamination, $F \in \Im_{\varepsilon,\tau}(\Phi)$. Under these assumptions, the asymptotic variance of $\sqrt{n}\,MD$-estimation for $W = 1$ is calculated by the formula

$$\sigma^2(F_{\varepsilon,\tau}, \Phi, W \equiv 1) = \frac{2\int_0^{+\infty} [\Phi(x) - (1/2)]^2 [(1-\varepsilon)\phi(x) + (\varepsilon/\tau)\phi(x/\tau)]dx}{\left( \int_{-\infty}^{\infty} \phi(x)[(1-\varepsilon)\phi(x) + (\varepsilon/\tau)\phi(x/\tau)]dx \right)^2} =$$

$$= \frac{[\pi(1-\varepsilon)/6] + [\varepsilon\, arctg(\tau^2/\sqrt{2\tau^2+1})]}{\{[(1-\varepsilon)/\sqrt{2}] + (\varepsilon/\sqrt{\tau^2+1})\}^2}.$$

For the weight function $W(x)=f_0(x)=\phi(x)$ the asymptotic variance of $\sqrt{n}\,MD$ -estimator is given by

$$\sigma^2(F_{\varepsilon,\tau},\Phi,W=\phi)=\frac{2\int_0^{\infty}\left(\int_0^{u}\phi(x)f_{\varepsilon,\tau}(x)dx-\phi(u)[F_{\varepsilon,\tau}(u)-\Phi(u)]\right)^2 dF_{\varepsilon,\tau}(u)}{\left(\int_{-\infty}^{\infty}\phi^2(x)dF_{\varepsilon,\tau}(x)-\int_{-\infty}^{\infty}\phi'(x)[F_{\varepsilon,\tau}(x)-\Phi(x)]dF_{\varepsilon,\tau}(x)\right)^2}=$$

$$=\frac{1}{4\pi^2\cdot\widetilde{B}^2(\varepsilon,\tau)}\sum_{i=1}^{20}A_i(\varepsilon,\tau),$$

where $\widetilde{B}(\varepsilon,\tau)$ and $A_i(\varepsilon,\tau), i=1,\ldots,20$ are certain functions of the parameters $\varepsilon$ and $\tau$. The numerical values of the asymptotic variance of $\sqrt{n}\,MD$ -estimators for $F\in\Im_{\varepsilon,\tau}(\Phi)$ at different weight functions are given in the Table. 5.

Table 5. The asymptotic variance of $\sqrt{n}\,MD$ -estimators for $F\notin\Im_0$, $F=F_{\varepsilon,\tau}$, $F_0=\Phi$

| $W,\tau \setminus \varepsilon$ | 0,00 | 0,01 | 0,05 | 0,10 | 0,15 | 0,20 | 0,25 | 0,30 |
|---|---|---|---|---|---|---|---|---|
| $\tau=3$ | 1,047(0,95) | 1,071(0,96) | 1,171(0,97) | 1,307(0,97) | 1,458(0,95) | 1,625(0,94) | 1,811(0,93) | 2,019(0,93) |
| $W\equiv 1$, 5 | 1,047(0,95) | 1,078(0,95) | 1,210(0,93) | 1,395(0,90) | 1,607(0,86) | 1,851(0,83) | 2,132(0,80) | 2,459(0,78) |
| $\tau=3$ | 1,095(0,91) | 1,117(0,92) | 1,209(0,93) | 1,333(0,94) | 1,470(0,95) | 1,620(0,95) | 1,786(0,95) | 1,972(0,96) |
| $W=\phi$, 5 | 1,095(0,91) | 1,122(0,92) | 1,237(0,91) | 1,393(0,90) | 1,562(0,89) | 1,749(0,88) | 1,956(0,87) | 2,187(0,87) |

The absolute efficiency of $MD$ -estimates calculated using the formula $AE(F_{\varepsilon,\tau},\hat{\theta})=\{\sigma^2(F_{\varepsilon,\tau},W)I(f_{\varepsilon,\tau})\}^{-1}$, where $I(f_{\varepsilon,\tau})$ is the Fisher information about the location parameter of distributions from the supermodel $\Im_{\varepsilon,\tau}(\Phi)$, are given in the table in parentheses.

Fig. (6) shows the absolute efficiency of estimates for $F\in\Im_{\varepsilon,\tau}(\Phi)$. It is clearly seen that $MD$ -estimates with the reference function $F_0=\Phi$ and the weight function $W(x)=\phi(x)$, as well as the weight function $W(x)=1$, provide high absolute efficiency when $0\le\varepsilon\le 0{,}3$. The absolute efficiency of the sample mean $\overline{X}$ decreases sharply, and the median for the sample $\overline{X}_{1/2}$ is slowly growing, remaining at low levels.

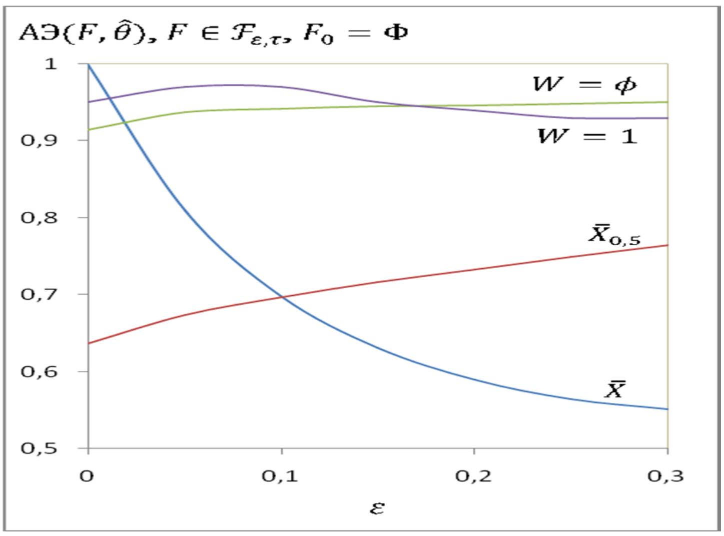


**Fig. 6. Absolute efficiency estimates for** $F\in\Im_{\varepsilon,\tau}(\Phi)$, $\tau=3$

**Example 7. Adaptive version**. Properties of the $MD$ - estimates depend strongly on the choice of the weighting function $W$ for distributions with "heavy tails". Therefore, the study of the properties of the efficiency and robustness of $MD$ -estimates (for the case $F \notin \Im_0$) opens the possibility of an adaptive approach to the choice of the reference distribution $F_0$ and weighting function $W$ within the given supermodel, based on the sample estimates of functionals that determine the "degree of heaviness of tails" of distributions (see Shulenin (1993a)). Adaptive selection of the weighting function can provide the required quality of $MD$ -estimates for a given supermodel.

Let us consider an example of the supermodel $\Im_{\varepsilon,\tau}(\Phi) = \{F : F(x) = \Phi_{\varepsilon,\tau}(x)\}$. We assume that the proportion of contamination $\varepsilon$ may vary in limits $0 \le \varepsilon \le 0{,}3$, and the scale parameter $\tau$ is $\tau = 3$. For this supermodel with the reference function $F_0 = \Phi$, let us define an adaptive weighting function $\hat{W}$ as

$$\hat{W}(x;\, X_1,\dots,X_n) = \begin{cases} 1/\phi(x), & 1{,}71 < Q(F_n) \le 1{,}76 \\ 1, & 1{,}76 < Q(F_n) \le 1{,}86\,, \\ \phi(x), & 1{,}86 < Q(F_n) \le 1{,}91 \end{cases} \tag{20}$$

where $Q(F_n)$ is the sample estimate of the functional $Q(F\,;\nu,\mu)$ which characterizes the "degree of heaviness of the distribution tails" and is defined in Shulenin (1993a). Sample estimate of $Q(F_n)$ is based on a sample $X_1,\dots,X_n$ and may be written as

$$Q(F_n;\,\nu,\mu) = \frac{m}{k}\left(\sum_{i=n-k+1}^{n} X_{(i)} - \sum_{i=1}^{k} X_{(i)}\right) \Big/ \left(\sum_{i=n-m+1}^{n} X_{(i)} - \sum_{i=1}^{m} X_{(i)}\right), k = [\nu n],\; m = [\mu n]. \tag{21}$$

Here the parameters $\nu$ and $\mu$ satisfy inequalities $0 < \nu < \mu \le 0{,}5$, $\nu = 0{,}2$, $\mu = 0{,}5$ and $X_{(1)},\dots,X_{(n)}$ are the order statistics of the sample $X_1,\dots,X_n$.

Note that the choice of the weighting function in the form of (20), the absolute efficiency of adaptive $MD$ - estimates do not fall below the level of 0.95 when the proportion $\varepsilon$ of contamination is $0 \le \varepsilon \le 0{,}3$. It means that within a given supermodel the absolute efficiency satisfies inequalities $0{,}95 \le A\text{Э}(\Phi_{\varepsilon,\tau}, \hat{W}) \le 1$ if $\tau = 3$, $0 \le \varepsilon \le 0{,}3$, $n \ge 40$ (see Figure 6). If we choose not to adapt the weighting function, and use, for example, the Anderson - Darling weight function in form of $\tilde{W}(x,\phi) = \phi(x)/\Phi(x)(1-\Phi(x))$, then the absolute efficiency of $MD$ - estimates with such a weight function in the framework of the supermodel $\Im_{\varepsilon,\tau}(\Phi)$ could fall to the level of 0.47.

**Conclusion**

We studied the asymptotic properties of the $MD$ - estimators of the location parameter $\theta$, based on the use of a weighted Cramer - Mises distance. It is shown that these estimates are $B$ - robust, that is, their influence functions are limited, and therefore, they are "protected" against outliers in the sample. For the case $F \in \Im_0$, the optimal weight functions are given that make $MD$ - estimates asymptotically efficient. For the Gaussian model with a scale contamination (for $F \in \Im_{\varepsilon,\tau}(\Phi)$, $\tau = 3$) the absolute efficiency of $MD$ - estimates with the weight function $W \equiv 1$ does not fall below 0.93 at $0 \le \varepsilon \le 0{,}3$, and it increases from 0.91 to 0.96 for the weight function $W \equiv \phi$.

Summarizing, we note that there is a close connection of $MD$ -estimators of parameter $\theta$ with the other robust $M$ -, $L$ -,and $R$ - estimators (see Shulenin and Tarasenko (1994), Shulenin and Serykh (1993), Shulenin(1995)). Properties of $MD$ - estimators in some cases coincide with those of many well-known estimates of the location parameter $\theta$; for example, with the properties of the Hodges - Lehmann estimates, the sample mean and median. Note also that the abovementioned asymptotic results is quite good approximation for properties of $MD$-estimators for sample sizes

$n \geq 20$. This is confirmed by the numerous computer simulation results. Studied properties of the efficiency and robustness of $MD$-estimates open (for the case $F \notin \Im_0$) the possibility to use an adaptive approach to the choice of the reference distribution function $F_0$ and the weighting function $W$ within the given supermodel, based on sample estimates of functionals that determine the "degree of heaviness of tails" of distributions (see Example 7 and Shulenin (2010), Shulenin(2010a)).

**Acknowledgement**

The author expresses gratitude's to Prof. F.P. Tarasenko for valuable comments and help in preparing English version of the paper.

## REFERENCES

Andrews D.F., Bickel P.J., Hampel F.R., Huber P. J., Rogers W.H., Tukey J.W. (1972). Robust estimation of location: survey and advances. Princeton. N.Y.: Princeton Univ. Press. 1972, 375 p.

Bickel, P. J. (1976). Another look at robustness: a review of reviews and some new development. *Scand. J. Statist. Theory and App1*. **3**, 145-168.

Boos, D. D. (1981). Minimum distance estimators for location and goodness of fit. *J. Amer. Statist. Assoc*. **76**, N.375, 663-670.

Parr, W.S. and Schucany, W.R. (1980). Minimum distance and robust estimation. J. Amer. Statist. Assoc. **75,** No. 371, 616 – 624.

Parr, W. C. (1981). Minimum distance estimation: a bibliography. Comm. Statist. **A10**, 1205-1224.

Parr, W.C., De Wet. (1981). On minimum weighted Cramer-von Mists statistical estimation. *Comm. Statist*. **A10(12),** 1149 – 1166.

Serfling, R. J. (1980). Approximation Theorems of Mathematical Statistics. N. – Y.: Wiley, 371p.

Shulenin, V. P. (1992). ( Шуленин В. П. Асимптотические свойства и робастность MD-оценок. Теория вероятностей и её применение. 1992 т. 37, в. 4, с. 816-818).

Shulenin, V. P., Serykh, A.P. (1993). (Шуленин В. П. , Серых А. П. Робастные и непараметрические алгоритмы обработки данных физических экспериментов. Известия вузов Физика. 1993, № 10, с.128-136).

Shulenin, V. P.(1993a). (Шуленин В. П. Введение в робастную статистику. Томск: Изд-во Том. ун-та,1993.-227с.)

Shulenin, V. P. and Tarasenko, F. P. (1994). Connections of MD-estimates with classes of robust estimates of location parameter. 12th Prague Conf. on Inform. Theory. August 29 – September 2, 220-223.

Shulenin, V. P. (1995). ( Шуленин В. П. Границы эффективности оценок, построенных методом минимума расстояния Крамера-Мизеса. Известия вузов Физика. 1995, № 9, с.84-89).

Shulenin, V. P. (2010). ( Шуленин В.П. Свойства адаптивных оценок Ходжеса – Лемана в асимптотике и при конечных объемах выборки. Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2010, № 2(11), с. 96 – 112).

Shulenin, V. P. (2010a). (Шуленин В.П. Эффективные и робастные MD-оценки Крамера–Мизеса. ВестникТомского государственного университета. Управление, вычислительная техника и информатика. 2010, № 3(12), с. 107 – 121).

Shulenin, V. P. (2012). (Шуленин В.П. Математическая статистика. Ч. 3. Робастная статистика: учебник. – Томск: Изд-во НТЛ, 2012. – 520 с.)

Wiens, D.P. (1987). Robust weighted Cramer-von Mises estimators of location, with minimax variance in $\varepsilon$ - contamination neighbourhoods. *The Canadian Journal of Statistics*. **15,** N. 3, 269-278.
Wolfowitz, J. (1957). The minimum distance method. *Ann. Math. Statist*. **28**, 75-88.