

Review

Highly Efficient Robust and Stable M -Estimates of Location

Georgy Shevlyakov [†]

Department of Applied Mathematics, Peter the Great St. Petersburg Polytechnic University,
195251 St. Petersburg, Russia; georgy.shevlyakov@phmf.spbstu.ru

[†] Prof. Georgy Shevlyakov passed away during the revision cycle of the manuscript, comments have been addressed by Dr. Maya Shevlyakova.

Abstract: This article is partially a review and partially a contribution. The classical two approaches to robustness, Huber's minimax and Hampel's based on influence functions, are reviewed with the accent on distribution classes of a non-neighborhood nature. Mainly, attention is paid to the minimax Huber's M -estimates of location designed for the classes with bounded quantiles and Meshalkin-Shurygin's stable M -estimates. The contribution is focused on the comparative performance evaluation study of these estimates, together with the classical robust M -estimates under the normal, double-exponential (Laplace), Cauchy, and contaminated normal (Tukey gross error) distributions. The obtained results are as follows: (i) under the normal, double-exponential, Cauchy, and heavily-contaminated normal distributions, the proposed robust minimax M -estimates outperform the classical Huber's and Hampel's M -estimates in asymptotic efficiency; (ii) in the case of heavy-tailed double-exponential and Cauchy distributions, the Meshalkin-Shurygin's radical stable M -estimate also outperforms the classical robust M -estimates; (iii) for moderately contaminated normal, the classical robust estimates slightly outperform the proposed minimax M -estimates. Several directions of future works are enlisted.

Keywords: robustness; minimax approach; stable estimation



Citation: Shevlyakov, G. Highly Efficient Robust and Stable M -Estimates of Location. *Mathematics* **2021**, *9*, 105. <https://doi.org/10.3390/math9010105>

Received: 27 August 2020
Accepted: 30 December 2020
Published: 5 January 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Robust statistics, as a new field of mathematical statistics, originates from the pioneering works of John Tukey (1960) [1], Peter Huber (1964) [2], and Frank Hampel (1968) [3]. The term “robust” (Latin: strong, vigorous, sturdy, tough, powerful) was introduced into statistics by George Box (1953) [4].

The reasons of research in this field of statistics are of a general mathematical nature: the conceptions of “optimality” and “stability” are mutually complementary in performance evaluation for almost all mathematical procedures, and the trade-off between them is often a sought goal.

It is not rare that the performance of optimal solutions is rather sensitive to small violations of the assumed conditions of optimality. In statistics, the classical example of such unstable optimal procedure is given by the least squares estimates, in which performance under small deviations from normality can become disastrous [5].

Since the term “stability” is overloaded in mathematics, the term “robustness” being its synonym is at present conventionally used in statistics and in optimal control theory: in general, it means the stability of statistical inference under uncontrolled violations of accepted distribution models.

In present, there are two main approaches to robustness: historically, the first global minimax approach of Huber (quantitative robustness) [5] and the local approach of Hampel based on influence functions (qualitative robustness) [6]. Within the first approach, the least informative (favorable) distribution minimizing Fisher information over a certain distribution class is obtained with the subsequent use of the asymptotically optimal maximum likelihood parameter estimate for this distribution. In this case, the minimax approach

gives the guaranteed accuracy of robust estimates, that is, the asymptotic variance of the optimal parameter estimate is upper-bounded for distributions from the aforementioned class.

Within the second approach, a parameter estimate is defined by its desired influence function, which determines the qualitative robustness properties of an estimate, such as its low sensitivity to the presence of gross outliers in the data, to the data rounding-off, to the missing data, etc.

In what follows, we consider these methodologies in detail focusing on the optimization and variational calculus tools used in both aforementioned approaches. Within Huber's minimax approach, we review the conventional least informative (favorable) distributions obtained for the neighborhoods of a Gaussian [5], as well as those designed for a variety of the non-standard distribution classes of a non-neighborhood nature [7]. Within Hampel's local approach [6], we mostly emphasize its recently developed stable estimation branch with the originally posed variational calculus problems and rather prospective results on their application to robust statistics [8].

While this paper focuses on particular topics in the field of robust statistics, it is worth noting a few comprehensive reviews also covering the present state of art in this field, namely Reference [9–14].

An outline of the remainder of the article is as follows. In Section 2, a general problem setting for the design of minimax variance M -estimates of location is recalled. In Section 3, the globally stable Meshalkin-Shurygin's M -estimates are described. In Section 4, a comparative performance evaluation of the conventional robust M -estimates of location with the several novel proposed M -estimates is examined (univariate setting is considered throughout the paper), and several unforeseen and unexpected results have been obtained. In Section 5, some conclusions are given.

2. Huber's Minimax Variance Robust M -Estimates of Location

2.1. Preliminaries

The minimax principle aims at the worst case suggesting for it the best solution [2]; thus, this approach provides a guaranteed result [5]. However, being applied to the problem of estimation of location, it yields a robust version of the principle of the maximum likelihood [2]. Usually, estimation of location is of a primary interest, and, in this study, we focus on it.

Let x_1, \dots, x_n be a sample from a distribution with density $p(x - \theta)$ from a convex class \mathcal{P} , where θ is a location parameter. Further, we assume that p is a symmetric distribution density; hence, θ is the center of symmetry to be estimated.

An M -estimate T_n of θ is a solution to the following minimization problem:

$$T_n = \arg \min_{\theta} \sum_{i=1}^n \rho(x_i - \theta), \quad (1)$$

where $\rho(u)$ is called the *function of contrast* [15]: $\rho(x_i - \theta)$ is a measure of difference between the observation x_i and the estimated center of symmetry. The following particular cases of (1) are of a particular interest: (i) for $\rho(u) = u^2$, we have the least squares method with the sample mean \bar{x} as the estimate of location; (ii) for $\rho(u) = |u|$, we arrive at the least absolute values method with the sample median estimate $med x$; and (iii), mostly importantly, for a given density p , the choice $\rho(u) = -\log p(u)$ yields the maximum likelihood (ML) estimate of location.

It is more convenient to formulate the properties of M -estimates in terms of the derivative of the function of contrast $\psi(u) = \rho'(u)$ called a *score function*. In this case, an M -estimate is defined as a solution to the following implicit estimating equation

$$\sum_{i=1}^n \psi(x_i - T_n) = 0. \quad (2)$$

Under rather general conditions of regularity imposed on the class Ψ of score functions ψ and on the class \mathcal{P} of densities p (their various forms can be found in Reference [2,5,6]), M -estimates are consistent and asymptotically normal with the asymptotic variance AV :

$$Var(n^{1/2}T_n) = AV(\psi, p) = \frac{A}{B^2} = \frac{\int \psi^2(x)p(x) dx}{(\int \psi'(x)p(x) dx)^2}, \tag{3}$$

where

$$A(\psi, p) = \int \psi^2(x)p(x) dx,$$

$$B(\psi, p) = \int \psi'(x)p(x) dx.$$

For M -estimates (2), the following result holds [5].

Theorem 1. (Huber, 1964) Under regularity conditions, M -estimates satisfy the minimax property

$$AV(\psi^*, p) \leq AV(\psi^*, p^*) = \sup_{p \in \mathcal{P}} \inf_{\psi \in \Psi} AV(\psi, p) \leq AV(\psi, p^*), \tag{4}$$

where $p^*(x)$ is the least informative (favorable) density p^* minimizing Fisher information for location $I(p)$ over the class \mathcal{P} :

$$p^* = \arg \min_{p \in \mathcal{P}} I(p), \quad I(p) = \int \left[\frac{p'(x)}{p(x)} \right]^2 p(x) dx. \tag{5}$$

From (4) and (5), it follows that the minimax function of contrast and score function are given by the maximum likelihood method for the least informative density p^* :

$$\rho^*(x) = -\log p^*(x), \quad \psi^*(x) = -p^*(x)' / p^*(x). \tag{6}$$

Thus, the pair (ψ^*, p^*) is the saddle-point of the functional $AV(\psi, p)$. The right-hand part of inequality (4) is the Rao–Cramér inequality:

$$AV(\psi, p^*) \geq AV(-p^{*'} / p^*, p^*) = 1 / \int (p^*(x)')^2 / p^*(x) dx = 1 / I(p^*),$$

whereas its left-hand part guarantees the asymptotic accuracy of robust minimax estimation with the following upper bound upon the asymptotic variance of the minimax variance robust M -estimate of location: $AV(\psi^*, p) \leq 1 / I(p^*)$.

The key point of the minimax approach is the solution of the variational problem (5): further, various classes \mathcal{P} with the corresponding least informative densities p^* and minimax estimates are enlisted.

Now, we recall the Huber’s classical solution for ε -contaminated normal distributions (Tukey’s gross-error model):

$$\mathcal{P}_\varepsilon = \{p: p(x) \geq (1 - \varepsilon)\varphi(x), 0 \leq \varepsilon < 1\}, \tag{7}$$

where $\varphi(x) = (2\pi)^{-1/2} \exp(-x^2/2)$ is the standard normal distribution density.

Theorem 2. (Huber, 1964) In the class \mathcal{P}_ε , the least informative density p^* and the optimal score function has the following form [2]:

$$p^*(x) = p_{Huber}(x) = \begin{cases} (1 - \varepsilon)\varphi(x), & |x| \leq k, \\ C \exp(-D|x|), & |x| > k, \end{cases} \tag{8}$$

$$\psi^*(x) = \psi_{Huber}(x) = \begin{cases} x, & |x| \leq k, \\ CD \operatorname{sgn}(x), & |x| > k, \end{cases} \tag{9}$$

where the parameters $C, D,$ and k satisfy the conditions of norming, continuity, and continuous differentiability of the solution at $x = k$:

$$\int p^*(x) dx = 1, \quad p^*(k-0) = p^*(k+0), \quad p^{*'}(k-0) = p^{*'}(k+0).$$

Finally, we get the linear bounded score $\psi_{Huber}(x) = \max\{-k, \min\{x, k\}\}$, where k depends on the value of the contamination parameter ε , as follows:

$$\frac{2\varphi(k)}{k} - 2\Phi(-k) = \frac{\varepsilon}{1-\varepsilon}, \quad \Phi(x) = \int_{-\infty}^x \varphi(t) dt;$$

the values of the parameter $k = k(\varepsilon)$ are tabulated in Reference [5].

The particular cases of this solution for $\varepsilon = 0$ and $\varepsilon \rightarrow 1$ are given by $k \rightarrow \infty$ (the sample mean) and $k = 0$ (the sample median), respectively.

The Huber’s score function $\psi_{Huber}(x)$ is a robust version of the *ML* estimation: in the center $|x_i - \theta| \leq k$, the data are processed by the *ML* method, and they are trimmed within the exponential distribution tails $|x_i - \theta| > k$. In the limiting case of a completely unknown density as $\varepsilon \rightarrow 1$, the minimax variance *M*-estimate of location tends to the sample median.

2.2. Free Extremals of the Basic Variational Problem

Consider the problem of minimization of Fisher information for location (5) under two basic side conditions of non-negativeness and norming: $p(x) \geq 0, \int_{-\infty}^{\infty} p(x) dx = 1$.

Set $\sqrt{p(x)} = q(x)$, and rewrite this minimization problem as

$$\text{minimize } I(p) = 4 \int_{-\infty}^{\infty} (q'(x))^2 dx \quad \text{under} \quad \int_{-\infty}^{\infty} q^2(x) dx = 1.$$

Introducing the Lagrange multiplier λ related to the norming condition, we obtain the following differential equation for the function $q(x)$: $4q''(x) + \lambda q(x) = 0$.

The general solutions of this harmonic oscillator equation have the well-known forms depending on the sign of $\lambda = 4k^2$: (i) the exponential $q(x) = C_1 e^{kx} + C_2 e^{-kx}$, the cosine $q(x) = C_1 \sin kx + C_2 \cos kx$, and the linear $q(x) = C_1 + C_2 x$, where $k = \pm\sqrt{\lambda}/2$.

Further, we show that all these forms work in the structures of least informative distribution densities.

2.3. Least Informative Distributions

The neighborhoods of normal, generally, are not the only models of interest. In real-life applications, the information about the distribution central part tails, its moments, and/or subranges is rather often available. The empirical distribution and quantile functions, histograms, and kernel estimates, together with their tolerance limits, provide other examples. To enhance the efficiency of robust estimates, this information can be used in minimax settings.

Further, we deal with symmetric distribution densities $p(-x) = p(x)$. Evidently, distribution densities must also satisfy the non-negativeness and norming conditions common for all classes. For brevity, we do not write out all these conditions any time we define a distribution density class.

Now, we enlist several examples of distribution classes convenient for the description of a prior knowledge about data distributions [7].

- (1) The class of non-degenerate densities [15]: $\mathcal{P}_1 = \{p: p(0) \geq 1/(2a) > 0\}$. The parameter a of this class characterizes the dispersion of the central part of a distribution. The least informative density in this class is given by the double-exponential or Laplace density [16]: $p_1^*(x) = L(x; 0, a) = 2a^{-1} \exp(-|x|/a)$. This result is quite

natural if one recalls the exponential form of free extremals for the basic variational problem.

- (2) The class \mathcal{P}_2 of distribution densities with a bounded variance:

$$\mathcal{P}_2 = \left\{ p: \sigma^2(p) = \int_{-\infty}^{\infty} x^2 p(x) dx \leq \bar{\sigma}^2 \right\}.$$

All distribution densities with bounded variances are the members of this class. Evidently, the Cauchy-type distributions do not belong to it. The least informative density in this class is normal [17]:

$$p_2^*(x) = N(x; 0, \bar{\sigma}) = \frac{1}{\sqrt{2\pi\bar{\sigma}}} \exp\left(-\frac{x^2}{2\bar{\sigma}^2}\right).$$

- (3) The class \mathcal{P}_3 of approximately normal distribution densities is defined by Equation (7): $\{p: p(x) \geq (1 - \epsilon)\varphi(x), 0 \leq \epsilon < 1\}$.

- (4) The class of finite distributions: $\mathcal{P}_4 = \left\{ p: \int_{-l}^l p(x) dx = 1 \right\}$. This class defines the boundaries of the data (i.e., the inequality $|X| \leq l$ holds with probability one), and there is no more information about this distribution besides the boundary conditions of smoothness: $p(\pm l) = p'(\pm l) = 0$.

The least informative density in this class has the cosine-squared form [15]:

$$p_4^*(x) = \begin{cases} \frac{1}{l} \cos^2\left(\frac{\pi x}{2l}\right) & \text{for } |x| \leq l, \\ 0 & \text{for } |x| > l. \end{cases}$$

- (5) The class of distributions with a bounded interquantile distribution mass:

$$\mathcal{P}_5 = \left\{ p: \int_{-l}^l p(x) dx \geq 1 - \beta \right\}, \quad 0 \leq \beta < 1.$$

The parameters l and β are assumed given. The restriction upon the interquantile mass means that $P(|X| \leq l) \geq 1 - \beta$. We can redefine this class in a different way as the class with an upper-bounded interquantile range $IQR_\beta = P^{-1}\left(\frac{1+\beta}{2}\right) - P^{-1}\left(\frac{1-\beta}{2}\right) \leq 2l$, where $P(x)$ is a probability density function.

The least informative density in this class has both the cosine and exponential parts working at the center and tail areas, respectively [7],

$$p_5^*(x) = \begin{cases} A_1 \cos^2(B_1 x) & \text{for } |x| \leq l, \\ A_2 \exp(-B_2 |x|) & \text{for } |x| > l, \end{cases}$$

where the constants $A_1, A_2, B_1,$ and B_2 are determined from the simultaneous equations (restrictions) of the class \mathcal{P}_5 , namely the conditions of normalization and upon the distribution interquantile range, and the conditions of continuity and continuous differentiability at $x = l$ (for details, see Reference [7]). It is worth noting that the classes \mathcal{P}_1 and \mathcal{P}_4 are the particular cases of the class \mathcal{P}_5 when $\beta \rightarrow 0$ and $\beta = 1$, respectively.

3. Hampel’s Robust and Shurygin’s Stable Estimates of Location

Robust methods have lower sensitivity to possible departures from the accepted distribution models as compared to conventional statistical methods. To analyze the sensitivity of estimation, it is natural to have its specific indicator. In what follows, we introduce these indicators, namely the influence function and related to it measures.

3.1. Hampel's Local Approach to Robustness

Let P be a distribution function corresponding to $p \in \mathcal{P}$, the class of distribution densities, and let $T(P)$ be a functional defined in a subset of all distribution functions. The natural estimate defined by T is $T_n = T(P_n)$, i.e., the functional computed at the sample distribution function P_n .

The influence function $IF(x; T, p)$ of this functional is defined as

$$IF(x; T, p) = \lim_{t \rightarrow 0} \frac{T((1-t)P + t\Delta_x) - T(P)}{t}, \tag{10}$$

where Δ_x is the degenerate distribution taking mass 1 at x [6].

The influence function measures the impact of an infinitesimal contamination at x on the value of an estimate, formally being the Gâteaux derivative of the functional $T(P)$. For an M -estimate with a score function ψ , the influence function is proportional to it: $IF(x; \psi, p) = \psi(x)/B(\psi, p)$, where the term $B(\psi, p)$ stands in Equation (3) for the asymptotic variance of M -estimates.

Based on the influence function, several local measures of robustness are defined [6], including the gross-error sensitivity of T at p :

$$\gamma^*(T, p) = \sup_x |IF(x; T, p)|.$$

This indicator of sensitivity gives an upper bound upon the asymptotic estimate bias and measures the worst influence of an infinitesimal contamination on the value of an estimate. Maximizing the efficiency of an M -estimate of location under the condition of a bounded gross-error sensitivity at the normal distribution

$$\max_{\psi} \text{eff}(\psi, \varphi) \quad \text{under} \quad \gamma^* \leq \bar{\gamma},$$

where $\text{eff}(\psi, p) = \frac{1}{AV(\psi, p)I(p)}$ formally leads to the Huber's minimax linear bounded score $\psi_{Huber}(x) = \max\{-k, \min\{x, k\}\}$ in the class of contaminated normal distributions [6]. This particular result confirms the following general observation: the best estimates within both approaches, Huber's minimax and Hampel's local, are rather close in their performances.

3.2. Meshalkin-Shurygin's Stable Estimates of Location

This topic is partially reversal to the conventional setting: the maximum of some measure of sensitivity is minimized under the guaranteed value of the estimate variance or efficiency.

The conventional point-wise local measures of sensitivity, such as the influence and change-of-variance functions [6] are not appropriate here—a global indicator of sensitivity is desirable. We show that a novel global indicator of robustness proposed by Shurygin [18], the estimate stability, is closely related to the classical variation of the functional of the estimate asymptotic variance. Although the related theory has been developed for stable estimation of an arbitrary parameter of the underlying distribution, here, we focus on stable estimation of location.

A measure of the M -estimate sensitivity called variance sensitivity (VS) is introduced in Reference [19]. Formally, it is defined as the Lagrange functional derivative of the asymptotic variance (3):

$$VS(\psi, p) = \frac{\partial AV(\psi, p)}{\partial p} = \frac{\partial}{\partial p} \left(\frac{\int \psi^2(x)p(x) dx}{(\int \psi'(x)p(x) dx)^2} \right) = \frac{\int \psi^2(x) dx}{(\int \psi'(x)p(x) dx)^2}. \tag{11}$$

Equation (11) gives a global measure of the stability of an M -estimate in a model, where the outliers occur uniformly anywhere on the real line. The boundness of the

Lagrange derivative (11) holds under the condition of square integrability of ψ with the corresponding *redescending* scores when $\psi(x) \rightarrow 0$ for $|x| \rightarrow \infty$.

In Reference [7], it is shown that the principal part of the variation $\delta AV(\psi, p)$ of the asymptotic variance $AV(\psi, p)$ with respect to $\|\delta p\|$ is proportional to the variance sensitivity (11) or to the Lagrange derivative of the asymptotic variance: $\delta AV(\psi, p) \propto \partial AV(\psi, p) / \partial p$.

Further, consider the following optimization problem: what is the minimum variance sensitive score function for a given distribution density p ? The solution of this optimization problem is given by the minimum variance sensitive (MVS) score function:

$$\psi_{MVS}(x) = \arg \min_{\psi} VS(\psi, p) = -p'(x). \tag{12}$$

The estimate with this optimal score function (12) is called as the estimate of *minimum variance sensitivity* with $VS_{min} = VS(\psi_{MVS}, p)$. We define a global measure of the *stability* of any M -estimate comparing an estimate variance sensitivity with its minimum

$$0 \leq stb(\psi, p) = VS_{min}(p) / VS(\psi, p) \leq 1.$$

A number of optimization criteria settings with different weights for efficiency and stability of M -estimates of location have been proposed and solved; in other examples, efficiency is maximized under guaranteed stability, and vice versa [18]. Practically in all these cases, we deal with the score functions $\psi_{opt}(x)$ with the following limiting forms: the maximum likelihood case $\psi_{opt}(x) \rightarrow \psi_{ML}(x) = -p'(x) / p(x)$ when the requirement of high efficiency mostly matters and the opposite *redescending* case $\psi_{opt}(x) \rightarrow \psi_{MVS}(x) = -p'(x)$ when the requirement of high stability is important.

A compromise case is given by a stable estimate called *radical* with equal efficiency and stability, $eff(\psi, p) = stb(\psi, p)$, desirably both highly efficient and stable: its score function is given by

$$\psi_{rad}(x) = \psi_{ML}(x) \sqrt{p(x)} = -p'(x) / \sqrt{p(x)}. \tag{13}$$

Finally, it should be noted that the minimum sensitivity and radical estimates belong to the class of M -estimates with the exponentially weighted maximum likelihood score functions previously proposed by Meshalkin [20].

4. Asymptotic Efficiency of M -Estimates: A Comparative Study

Here, we compare the asymptotic efficiency performance of various robust and stable estimates of location at the conventional in robustness studies distributions: some particular results can be found in Reference [19,21]. We mainly focus on the minimax variance estimates for distributions with bounded interquartile ranges, as until present, their performance has not been thoroughly studied. It is important that some obtained results are unexpected and surprising.

4.1. Robust and Stable M -Estimates of a Location Parameter

We test the following M -estimates of location: (i) the sample mean and median, (ii) the Huber’s minimax variance M -estimate with the linear bounded score $\psi_{Huber}(x) = \max[-1.14, \min(x, 1.14)]$ optimal for ϵ -contaminated standard normal distributions with the parameter of contamination $\epsilon = 0.1$, (iii) the Hampel’s M -estimate with the *redescending* three-part score

$$\psi_{Hampel}(x) = \begin{cases} x & \text{for } 0 \leq |x| \leq a, \\ a \operatorname{sgn}(x) & \text{for } a \leq |x| \leq b, \\ a \frac{r - |x|}{r - b} \operatorname{sgn}(x) & \text{for } a \leq |x| \leq b, \\ 0 & \text{for } r \leq |x|, \end{cases}$$

where the parameters $a = 1.31$, $b = 2.039$, and $r = 4$ (see Reference [6], pp. 166–167), (iv) the minimax variance M -estimates with the scores $\psi_\beta(x)$ for various values of the parameter β : 0.01, 0.05, 0.1, 0.5, 0.9, 0.95, 0.99, and (v) the stable Shurygin's minimum variance sensitivity and radical M -estimates with the scores $\psi_{MVS}(x) = -p'(x)$ and $\psi_{rad}(x) = -p'(x)/\sqrt{p(x)}$, respectively.

4.2. Data Distributions

In our study, the following distribution densities are used:

- (i) the standard normal $p(x) = N(x; 0, 1) = (2\pi)^{-1/2} \exp(-x^2/2)$,
- (ii) the standard Laplace $p(x) = L(x; 0, 1/\sqrt{2}) = 2^{-1/2} \exp(-\sqrt{2}|x|)$,
- (iii) the standard Cauchy $p(x) = C(x; 0, 1) = \pi^{-1}(1+x^2)^{-1}$, and
- (iv) the heavy-tailed Tukey gross-error model as the Cauchy contaminated normal density $p(x) = 0.9N(x; 0, 1) + 0.1C(x; 0, 1)$.

4.3. Asymptotic Efficiency

The asymptotic efficiency of M -estimates is numerically computed as follows:

$$eff(T_n) = \frac{Var(T_{ML})}{Var(T_n)} = \frac{(\int \psi'(x)p(x) dx)^2}{I(p) \int \psi^2(x)p(x) dx}.$$

5. Conclusions

From Table 1, it follows:

- (1) As usual, the performance of the sample mean under heavy-tailed Cauchy and contaminated normal distributions is awful. Designed for these models, Huber's and Hampel's M -estimates perform well except the Laplace distribution case. This distribution with moderately heavy tails against a sharp peak at the center is a rather tough test for the asymptotic performance of M -estimates of location, especially as compared to the Cauchy distribution case. Recall that the Laplace and distributions close to it are the least informative ones in wide classes of distributions, for instance, in the class \mathcal{P}_β with the parameter β close to unit (the corresponding minimax variance M -estimates perform quite well in these cases). For a statistical user, the version with a bounded IQR (interquartile range, $\beta = 1/2$) seems a reasonable choice.
- (2) Surprisingly, the proposed minimax variance M -estimates with the scores ψ_β outperform the classical robust Huber's and Hampel's M -estimates at the normal, although the shape of the least informative distribution is not at all normal: note that the Taylor expansion of the *cosine*²-bell shape is close to the exponential one with small values of β . Moreover, these M -estimates are better than the classical robust estimates in heavy-tailed distribution models. We explain this effect by the nature of the distribution class \mathcal{P}_β —it is one of the most wide possible distribution classes. Finally, these M -estimates and their statistical properties can be obtained in a closed analytical form.
- (3) The globally stable Meshalkin-Shurygin's radical M -estimate also outperforms the classical robust Huber's and Hampel's M -estimates at the heavy-tailed Laplace and Cauchy distributions.

Table 1. Asymptotic efficiency (two best values are boldfaced).

<i>M</i> -Estimate	Normal	Laplace	Cauchy	Tukey Gross Error
mean	1	0.500	0	0
median	0.636	1	0.811	0.678
Huber's linear bounded	0.924	0.669	0.759	0.953
Hampel's three-part	0.911	0.644	0.869	0.946
$\psi_{0.01}$	0.895	0.310	0.250	0.747
$\psi_{0.1}$	0.976	0.503	0.484	0.948
$\psi_{0.5}$	0.844	0.799	0.765	0.962
$\psi_{0.9}$	0.679	0.965	0.898	0.825
$\psi_{0.99}$	0.641	0.995	0.859	0.727
MVS	0.650	0.750	0.800	0.710
radical	0.840	0.890	0.920	0.890

Finally, we outline the prospective future works: (i) an extension of the proposed minimax variance *M*-estimates to the multivariate case and the classes with simultaneously bounded subranges of different parameter β values—in the latter case, we may expect a uniformly better performance; (ii) a generalization of the Meshalkin-Shurygin's stable estimates to the multivariate case; and (iii) a thorough comparative study of the small sample performance of *M*-estimates of location—Monte Carlo experiments show a slightly better performance of the classical Huber's and Hampel's estimates in this case.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Tukey, J.W. A Survey of Sampling from Contaminated Distributions. *Contributions to Probability and Statistics*; Olkin, I., Ed.; Stanford Univ. Press: Redwood City, CA, USA, 1960; pp. 448–485.
2. Huber, P.J. Robust estimation of a location parameter. *Ann. Math. Stat.* **1964**, *35*, 73–101. [[CrossRef](#)]
3. Hampel, F.R. Contributions to the Theory of Robust Estimation. Ph.D. Thesis, University of California, Berkeley, CA, USA, 1968.
4. Box, G.E.P. Non-normality and tests on variances. *Biometrika* **1953**, *40*, 318–335. [[CrossRef](#)]
5. Huber, P.J.; Ronchetti, E. *Robust Statistics*; Wiley: New York, NY, USA, 2009.
6. Hampel, F.R.; Ronchetti, E.; Rousseeuw, P.J.; Stahel, W.A. *Robust Statistics. The Approach Based on Influence Functions*; Wiley: New York, NY, USA, 2005.
7. Shevlyakov, G.L.; Oja, H. *Robust Correlation. Theory and Applications*; Wiley, John Wiley & Sons Ltd.: Chichester, UK, 2016.
8. Shevlyakov, G.L.; Morgenthaler, S.; Shurygin, A.M. Redescending *M*-estimators. *J. Stat. Plann. Inference* **2008**, *138*, 2906–2916. [[CrossRef](#)]
9. Daszykowski, M.; Kaczmarek, K.; Vander Heyden, Y.; Walczak, B. Robust statistics in data analysis—A review: Basic concepts. *Chemom. Intell. Lab. Syst.* **2007**, *85*, 203–219. [[CrossRef](#)]
10. Augustin, T.; Hable, R. On the impact of robust statistics on imprecise probability models: A review. *Struct. Saf.* **2010**, *32*, 358–365. [[CrossRef](#)]
11. Rousseeuw, P. J.; Hubert, M. Robust statistics for outlier detection. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2011**, *1*, 73–79. [[CrossRef](#)]
12. Maronna, R.A.; Martin, R.D.; Yohai, V.J. *Robust Statistics: Theory and Methods (with R)*; Wiley: John Wiley & Sons Ltd.: Chichester, UK, 2019.
13. Rousseeuw, P.J.; Hubert, M. Anomaly detection by robust statistics. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2018**, *8*, e1236. [[CrossRef](#)]
14. Ronchetti, E. The main contributions of robust statistics to statistical science and a new challenge. *METRON* **2020**, 1–9. [[CrossRef](#)]
15. Polyak, B.T.; Tsytkin, Y.Z. Robust identification. In *Identification of Systems and Parameter Estimation, Part 1, Proceedings of the 4th IFAC Symposium, Tbilisi, GA, USA, 21–27 September 1976*; North-Holland Pub. Co.: Amsterdam, The Netherlands, 1976; pp. 203–224.
16. Shevlyakov, G.L.; Vilchevski, N.O. *Robustness in Data Analysis: Criteria and Methods*; VSP BV: Zeist, The Netherlands, 2002.
17. Kagan, A.M.; Linnik, Y.V.; Rao, S.R. *Characterization Problems in Mathematical Statistics*; Wiley: New York, NY, USA, 1972.

18. Shurygin, A.M. New approach to optimization of stable estimation. In *Proceedings of the 1 US/Japan Conference on Frontiers of Statist. Modeling*; Bozdogan, H., Ed.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1994; pp. 315–340.
19. Shevlyakov, G.L.; Shagal, A.A.; Shin, V.I. A comparative study of robust and stable estimates of multivariate location. *J. Math. Sci.* **2019**, *237*, 831–845. [[CrossRef](#)]
20. Meshalkin, L.D. Some mathematical methods for the study of non-communicable diseases. In *Proceedings of the 6th International Meeting of Uses of Epidemiology in Planning Health Services*, Primosten, Yugoslavia, 29 August–3 September 1971; Volume 1, pp. 250–256.
21. Shevlyakov, G.L.; Tkhakushinova, R.V.; Snin, V.I. Robust minimax variance estimation of location under bounded interquartile ranges. *J. Math. Sci.* **2020**, *248*, 25–32. [[CrossRef](#)]