$L_1$-*Statistical Procedures and Related Topics*
IMS Lecture Notes – Monograph Series (1997) Volume 31

# Inadmissibility of robust estimators with respect to $L_1$ norm

**Jana Jurečkovà**

*Charles University, Prague, Czech Republic*

**L.B. Klebanov**

*Institute of Mathematical Geology, St. Petersburg, Russia*

*Abstract:* We show that robust M-estimators as well as equivariant estimators which do not depend on the extreme observations are inadmissible estimators of the location with respect to the $L_1$ loss function for a broad class of distributions. As a consequence, it implies that the sample median is inadmissible as an estimator of the location of the double-exponential distribution.

*Key words:* Admissibility, equivariant estimator, Pitman estimator, $M$-estimator, $L_1$ norm.

AMS subject classification: 62C15, 62F11, 62F35.

## 1 Introduction

Let $X_1, \ldots, X_n$ be a random sample from a distribution with the absolutely continuous distribution function $F(x-\theta)$, $\theta \in \mathbb{R}^1$. The problem is that of estimating the parameter $\theta$. We shall assume that the loss $L(t,\theta)$ incurred when estimating $\theta$ by $t$ depends only on $|t-\theta|$, *i.e.*

$$L(t,\theta) = L(|t-\theta|). \tag{1}$$

Then it is natural to restrict considerations to the estimators equivariant with respect to the shift in location, *i.e.* satisfying

$$T_n(X_1+c, \ldots, X_n+c) = T_n(X_1, \ldots, X_n) + c \quad \forall c \in \mathbb{R}^1 \quad \text{and} \quad \forall \mathbf{X} \in \mathbb{R}^n. \tag{2}$$

Let $\mathcal{T}$ denote the family of all equivariant estimators.

Different measures of performance of equivariant estimators were investigated. Among them, the probability

$$P_\theta(|T_n - \theta| > a) \tag{3}$$

that the absolute error exceeds $a > 0$, was considered by several authors, either for $n \to \infty$ and $a > 0$ fixed or $n$ fixed and $a \to \infty$. Bahadur [2], [3], Fu [4] and Sievers [12] studied the limit

$$\lim_{n\to\infty} \left\{ -\frac{1}{n} \log P_\theta(|T_n - \theta| > a) \right\} = e, \quad a \quad \text{fixed} \tag{4}$$

as a measure of performance of $T_n$. Sievers [12], who calculated the limits $e$ for several estimators and several distribution shapes, found the sample median less efficient than the sample mean not only for normal but also for logistic distribution and even for the double exponential distribution for sufficiently large $a$. Similar phenomenon was observed by Jurečkov´a [5] who considered the measure of performance

$$B(a; T_n) = \frac{-\log P_0(|T_n| > a)}{-\log P_0(|X_1| > a)} \tag{5}$$

under $n$ fixed and $a \to \infty$. The estimators which trimm-off the extreme observations were found more robust but less efficient than $\bar{X}_n$ for densities with exponential tails, including the double exponential. Akahira and Takeuchi [1] computed the loss of information caused by trimming the extreme order statistics in the double exponential population.

Denote

$$\mathbf{Y} = (Y_1, \ldots, Y_n) \text{ where } Y_i = X_i - X_1, \; i = 1, \ldots, n \tag{6}$$

the maximal invariant of $\mathbf{X}$ with respect to the group of translations of $X_1, \ldots, X_n$. Then the minimum risk equivariant estimator $T_n^*$ (*Pitman estimator*, MRE) exists provided there exists at least one $T_n \in \mathcal{T}$ with finite risk; then $T_n^*$ could be written in the form

$$T_n^*(\mathbf{X}) = T_n(\mathbf{X}) - v^*(\mathbf{Y}) \tag{7}$$

where $v^*(\mathbf{Y})$ satisfies

$$\mathbb{E}_0 L(|T_n(\mathbf{X}) - v^*(\mathbf{Y})|) = \min_{v(\mathbf{Y})} \mathbb{E}_0 L(|T_n(\mathbf{X}) - v(\mathbf{Y})|) \tag{8}$$

with the minimum taken over all functions $v(\mathbf{y})$.

If the loss function is quadratic, $L_2(t,\theta) = (t-\theta)^2$, the minimum risk estimator has the form

$$T_n^{(2)}(\mathbf{X}) = T_n(\mathbf{X}) - \mathbb{E}_0(T_n(\mathbf{X})|\mathbf{Y})$$

and it equals to the sample mean $\bar{X}$ provided $F$ is normal. Conversely, Kagan, Linnik and Rao [9] proved that, provided $\bar{X}_n$, $n \geq 3$, is the minimum risk estimator of $\theta$ for some $F$ with respect to the quadratic loss, then $F$ is normal. Otherwise speaking, the equality $\mathbb{E}_0(\bar{X}|\mathbf{Y}) = 0$ characterizes the normal distribution.

If $f$ is not normal, the minimum risk estimator of $\theta$ is typically nonlinear and only the sample mean is a sum of independent summands. However, many estimators admit asymptotic representations of the type

$$T_n(\mathbf{X}) = \theta + \frac{1}{n}\sum_{i=1}^{n}\psi(X_i-\theta) + o_p(n^{-1/2}) \quad \text{as} \quad n \to \infty \tag{9}$$

with appropriate functions $\psi$. Especially, if $T_n(\mathbf{X})$ is an asymptotically efficient estimator, then $\psi(x) = -\frac{1}{I(f)}\frac{f'(x)}{f(x)}$, with $f$ being the density of $F$ and $I(f)$ its Fisher information. A systematic study of the representations of this type could be found in [7]. Jurečkovà and Milhaud [6] recently proved that if the equality $\mathbb{E}_0(\sum_{i=1}^{n}\psi(X_i)|\mathbf{Y}) = 0$ holds for $n \geq 4$ and for a function $\psi$ satisfying some regularity conditions, then $\psi(x) = c\frac{f'(x)}{f(x)}$, $x \in \mathbb{R}^1$, where $c$ is a constant and $f$ is the density of $F$. This result indicates that not only asymptotically, but also in the finite sample case, many properties of $\frac{1}{nI_f}\sum_{i=1}^{n}\left(-\frac{f'(X_i)}{f(X_i)}\right)$ under $F$ are in correspondence with the respective properties of the sample mean under the normal distribution.

In the present paper, we shall consider the performance of some robust estimators with respect to the $L_1$ loss, *i.e.*,

$$L_1(t,\theta) = |t-\theta|. \tag{10}$$

Zinger, Kagan and Klebanov [13] and Kagan and Zinger [10] (see also [9], Section 7.9) proved that if $\bar{X}_n$ is the minimum risk estimator of $\theta$ with respect to $L_1$-loss for $f(x-\theta)$, $f$ unimodal and $n \geq 6$, then the underlying distribution is normal. This means that, for the normal distribution, $\bar{X}$ is the minimum risk estimator of $\mathcal{T}$ with respect to both $L_1$ and $L_2$ loss functions and in both cases its admissibility it is a characteristic property of the normal distribution.

If $f$ is unknown then we prefer robust estimators which are not connected with a fixed density shape. However, if we know $f$, then we are interested in admissible estimators whose risk cannot be uniformly improved. We shall show that the robust estimators are not admissible under the $L_1$ norm for a broad class of densities.

## 2 Inadmissibility of trimmed estimators

Let $X_1, \ldots, X_n$ be a random sample from a distribution with the continuous density $f(x-\theta)$ such that $f(x) > 0,\ x \in \mathbb{R}^1$. Let $X_{n:1} \leq \ldots \leq X_{n:n}$ be the order statistics corresponding to $X_1, \ldots, X_n$. Let $\mathcal{T}^* \subset \mathcal{T}$ denote the set of equivariant estimators satisfying the following condition:

**(A1)** If $T_n = T_n(X_{n:i_1}, \ldots, X_{n:i_k}),\quad i_1 < \ldots < i_k,\ 1 \leq k \leq n$, then $X_{n:i_1} \leq T_n \leq X_{n:i_k}$.

**(A2)** $T_n(0, \ldots, 0) = 0$.

The following theorem shows that the trimmed estimators as well as the M-estimators with a score function constant outside a bounded interval are inadmissible for unimodal densities.

**Theorem 1** *Let $X_1, \ldots, X_n,\quad n \geq 5$, be independent observations from a distribution with the density $f(x-\theta)$, where $f(x) > 0,\ x \in \mathbb{R}^1$, is unimodal, i.e. increasing for $x < 0$ and decreasing for $x > 0$.*
*(i) Let $T_n \in \mathcal{T}^*$ be an equivariant estimator, $T_n(X_1, \ldots, X_n)$, continuous in each argument, constant with respect to $X_{n:1}, X_{n:2}, X_{n:n-1}$ and $X_{n:n}$, but uniquely determined as a function of $X_{n:3}, \ldots, X_{n:n-2}$. Then $T_n$ is inadmissible estimator of $\theta$ with respect to the loss (10).*
*(ii) Let $M_n$ be an M-estimator generated by a continuous non-decreasing function $\psi$ such that $\psi(x) = \psi(c_1)$ for $x \leq c_1$ and $\psi(x) = \psi(c_2)$ for $x \geq c_2,\quad c_1 < 0 < c_2$, as*

$$M_n = \frac{1}{2}(M_n^+ + M_n^-) \quad \text{where} \tag{11}$$

$$M_n^- = \sup\{t : \sum_{i=1}^n \psi(X_i - t) > 0\}, \quad M_n^+ = \inf\{t : \sum_{i=1}^n \psi(X_i - t) < 0\}.$$

*Then $M_n$ is inadmissible as an estimator of $\theta$ with respect to the loss (10).*

**Proof:** In the case of $L_1$ norm, (7) specializes to

$$T_n^* = T_n - \text{med}_0(T_n|\mathbf{y}) \tag{12}$$

where $\text{med}_0(T_n|\mathbf{y})$ stands for any conditional median of $T_n$ given the maximal invariant $\mathbf{y}$ under $\theta = 0$. Hence, $T_n$ is the MRE provided

$$\text{med}_0(T_n|\mathbf{y}) = 0 \tag{13}$$

and the median is unique.

(i) Let $T_n \in \mathcal{T}$ be uniquely determined and do not depend on $X_{n:1}, X_{n:2}$, $X_{n:n-1}$ and $X_{n:n}$ Denote

$$\mathbf{Y} = (Y_1, \ldots, Y_n) = (X_1 - T_n, \ldots, X_n - T_n). \tag{14}$$

Then $\mathbf{Y}$ is the maximal invariant for the group of translations of $X_1, \ldots, X_n$. The conditional distribution of $T_n$ given $\mathbf{Y} = \mathbf{y}$ has the density

$$g(t|\mathbf{y}) = \frac{\prod_{i=1}^n f(t + y_i)}{\int_{-\infty}^{\infty} \prod_{i=1}^n f(z + y_i) dz}. \tag{15}$$

The condition (13) rewrites in the following way:

$$\int_{-\infty}^{0} \prod_{i=1}^{n} f(t + y_i) dt = \int_{0}^{\infty} \prod_{i=1}^{n} f(t + y_i) dt \quad \text{a.s. } [F]. \tag{16}$$

In view of continuity of the density $f$ and the estimator $T_n$ previous equation holds not only for almost all but for all $y_i$.

Denoting $w(t) = \operatorname{sign} t,\ t \in \mathbb{R}^1$, we rewrite (16) in the form

$$\int_{-\infty}^{\infty} w(t) \prod_{i=1}^{n} f(t + y_i) dt = 0. \tag{17}$$

Denote $A = \{\mathbf{y} : y_2 = \ldots = y_{n-2} = 0,\ y_1 \le y_2 \le 0,\ y_n \ge y_{n-1} \ge 0\}$. Then $T_n(\mathbf{y}) = 0$ for $\mathbf{y} \in A$ independently of the values of $y_1, y_2, y_{n-1}, y_n$ and (17) takes on the form

$$\int_{-\infty}^{\infty} w(t) f(t+y_1) f(t+y_2) f(t+y_{n-1}) f(t+y_n) f^{n-2}(t) dt = 0, \quad \mathbf{y} \in A. \tag{18}$$

Differentiating (18) in $y_\nu,\ \nu = 1, 2, n-1, n$ gives

$$\int_{-\infty}^{\infty} w(t) f'(t + y_\nu) \prod_{i \ne \nu} f(t + y_i) f^{n-4}(t) dt = 0, \quad \nu = 1, 2, n-1, n. \tag{19}$$

Integrating the left-hand side of (19) by parts for $\nu = 1$ and using (19) for $\nu = 2, n-1, n$, we obtain

$$(n-4) \int_{-\infty}^{\infty} w(t) f'(t) \prod_{i=1,2,n-1,n} \frac{f(t + y_i)}{f(y_i)} f^{n-5}(t) dt = 2 f^{n-4}(0) \tag{20}$$

If we especially take three following choices of $\mathbf{y} \in A$ :

$$\begin{aligned} y_1 &= y_2 = u < 0, y_{n-1} = y_n = 0 \\ y_1 &= u < 0, y_2 = 0, y_{n-1} = 0, y_n = v > 0 \\ y_1 &= y_2 = 0, y_{n-1} = y_n = v > 0, \end{aligned}$$

and subtract twice the second equality from the sum of the first and the third ones, we get

$$\int_{-\infty}^{\infty} w(t)f'(t)\left[\frac{f(t+u)}{f(u)} - \frac{f(t+v)}{f(v)}\right]^2 f^{n-3}(t)dt = 0. \tag{21}$$

Because $w(t)f'(t) > 0$ for $t \neq 0$, (21) implies that

$$\frac{f(t+u)}{f(u)} = \frac{f(t+v)}{f(v)} \tag{22}$$

holds for all $t \neq 0,\ u < 0,\ v > 0$, and this in turn implies that

$$\frac{f(t+u)}{f(u)} = \frac{f(t+v)}{f(v)} \tag{23}$$

holds for all $t \neq 0,\ u, v \in \mathbb{R}^1$. By the Cauchy equation, the only function satisfying (23) is either the exponential function or the constant. By the unimodality assumption on $f$, $\forall v > 0$ there exists $u < 0$ such that $f(u) = f(v)$, and then (22) leads to the constant $f$, what is a contradiction. Hence, there exists at least one $\mathbf{y}^* \in A$ such that either

$$\int_{-\infty}^{0} \prod_{i=1}^{n} f(t+y_i^*)dt < \int_{0}^{\infty} \prod_{i=1}^{n} f(t+y_i^*)dt \tag{24}$$

[or the opposite inequality] holds for $\mathbf{y}^*$. Then, because of the continuity, (24) [or the opposite inequality] holds in a neighborhood of $\mathbf{y}^*$, hence (16) is not true *a.s.* $[F]$ and $T_n$ is not admissible.

(ii) Let $M_n$ be the M-estimator defined in (11). Put $Y_i = X_i - M_n,\ i = 1, \ldots, n$. Then $\mathbf{Y}$ is the maximal invariant with respect to the group of translations and the conditional density of $M_n$ given $\mathbf{Y} = \mathbf{y}$ has the form (15). Analogously as in the part (i), $M_n$ would be the MRE under the condition (17). Let $B = \{\mathbf{y} : y_5 = \ldots = y_n = 0\}$; proceed analogously as in (19) - (21) and take successively the following choices of $y_1, \ldots, y_4$ :

$$\begin{aligned}
y_1 &= y_2 = c_1 - u,\ y_3 = y_4 = c_1 - u', \\
y_1 &= c_1 - u,\ y_2 = c_2 + v, \quad y_3 = c_1 - u',\ y_4 = c_2 + v', \\
y_1 &= y_2 = c_2 + v, \quad y_3 = y_4 = c_2 + v',
\end{aligned}$$

$u, u', v, v' \geq 0$. Analogously as in the part (i), we arrive at the equation

$$\int_{-\infty}^{\infty} w(t)f'(t)\frac{f(t+c_1-u)f(t+c_1-u')}{f(c_1-u)f(c_1-u')} -$$

$$\frac{f(t+c_2+v)f(t+c_2+v')}{f(c_2+v)f(c_2+v')}^2 f^{n-5}(t)dt = 0 \tag{25}$$

$\forall t \in \mathbb{R}^1,\ u, u', v, v' \geq 0$. Quite analogously we get

$$\int_{-\infty}^{\infty} w(t)f'(t)\frac{f(t+c_1-u)f(t+c_2+v')}{f(c_1-u)f(c_2+v')}-$$

$$\frac{f(t+c_2+v)f(t+c_1-u')}{f(c_2+v)f(c_1-u')}^2 f^{n-5}(t)dt \quad = \quad 0 \tag{26}$$

$\forall t \in \mathbb{R}^1,\ u, u', v, v' \geq 0$. Similarly as in part (i), we first conclude that the density f should be then either exponential or constant in the tails and (25) finally to leads to the constant tails, what is a contradiction with the conditions imposed on $f$. Thus, there exists $\mathbf{y}^* \in B$ and hence also its neighborhood satisfying either (24) or the opposite inequality; we conclude that $M_n$ cannot be an admissible estimator of $\theta$ with respect to $L_1$ loss. □

Notice that Theorem 1 covers the trimmed L-estimators, the sample median as well the linear combinations of several (non-extreme) sample quantiles; it also covers the Huber estimator and the related M-estimators. The results also imply that the sample median is nor admissible even for the double exponential distribution for which it is the maximum likelihood estimator. Similarly, while Huber's M-estimator is MLE for the contaminated normal distribution, it is not admissible for the same.

## Acknowledgements

For Jana Jurečková, the research was supported by the Charles University Grant GAUK 198/1996, and for L. B. Klebanov, the research was partially supported by RFFI Grants 95-01-01260,96-01-00852.

## References

[1] Akahira, M. and Takeuchi, K. (1995). *Non-Regular Statistical Estimation.* Lecture Notes in Statistics 107, Springer-Verlag.

[2] Bahadur, R.R. (1960). Asymptotic efficiency of tests and estimates. *Sankhyā* **22** 229-252.

[3] Bahadur, R.R. (1967). Rates of convergence of estimates and test statistics. *Ann. Math. Statist.* **38** 303-324.

[4] Fu, J.C. (1975). The rate of convergence of consistent point estimators. *Ann. Statist.* **3** 234-240.

[5] Jurečková, J. (1981). Tail-behavior of location estimators. *Ann. Statist.* **9** 578-586.

[6] Jurečková, J. and Milhaud, X. (1996). Characterization of distributions in invariant models. Submitted.

[7] Jurečková,J. and Sen, P.K. (1996). *Robust Statistical Procedures: Asymptotics and Interrelations.* New York: Wiley.

[8] Kagan, A.M., Linnik, Yu.V. and Rao, C.R. (1965). On a characterization of the normal law based on a property of the sample avarage. *Sankhyā* A **27** 405-406.

[9] Kagan, A.M., Linnik, Yu.V. and Rao, C.R. (1972). *Characterization Problems of Mathematical Statistics* (in Russian). Moscow: Nauka

[10] Kagan, A.M. and Zinger, A.A (1971). Sample mean as an estimator of location parameter. *Sankhyā* A **33** 351-358.

[11] Lehmann, E.L. (1983). *Theory of Point Estimation.* New York: Wiley.

[12] Sievers, G.L. (1978). Estimation of location; a large deviation comparison. *Ann. Statist.* **6** 610-618.

[13] Zinger, A.A., Kagan, A.M. and Klebanov, L.B. (1969). Sample mean as an estimator of the shift parameter with non-quadratic loss functions (in Russian). *Doklady AN SSSR* 186, 1.