The Annals of Statistics
1981, Vol. 9, No. 3, 578–585

TAIL-BEHAVIOR OF LOCATION ESTIMATORS[1]

BY JANA JUREČKOVÁ

Charles University, Prague

Let $X_1, \cdots, X_n$ be a sample from a population with density $f(x - \theta)$ such that f is symmetric and positive. It is proved that the tails of the distribution of a translation-invariant estimator of θ tend to 0 at most n times faster than the tails of the basic distribution. The sample mean is shown to be good in this sense for exponentially-tailed distributions while it becomes poor if there is contamination by a heavy-tailed distribution. The rates of convergence of the tails of robust estimators are shown to be bounded away from the lower as well as from the upper bound.

1. Introduction. Let $X_1, \cdots, X_n$ be a sequence of independent random variables identically distributed according to an absolutely continuous distribution function $F(x - \theta)$ with the density $f(x - \theta)$ such that $f(-x) = f(x) > 0$, $x \in R^1$; otherwise f is unspecified. The problem is that of estimating θ as a center of symmetry of an unknown symmetric absolutely continuous distribution. For each fixed n let $T_n = T_n(X_1, \cdots, X_n)$ be an estimator of θ based on $X_1, \cdots, X_n$.

Different measures of performance of T_n have been suggested and investigated. Besides the classical mean-square-error approach, the probability

$$P_\theta(|T_n - \theta| > a) \tag{1.1}$$

of the absolute error not exceeding a fixed number $a > 0$ has been considered by several authors. If the sequence $\{T_n\}$ is consistent for θ, then the inaccuracy (1.1) tends to 0 as $n \to \infty$. Bahadur [1], [2] proposed the limit

$$\lim_{n\to\infty}\left\{-\frac{1}{n}\ln P_\theta(|T_n - \theta| > a)\right\} = e \tag{1.2}$$

for a fixed $a > 0$ as a measure of performance of T_n, if the limit exists. Bahadur [2] and Fu [4] gave an upper bound for e for consistent sequences of estimators. Sievers [6] evaluated the limits e and their upper bounds for several estimators and several distribution shapes. From this point of view he found the sample median less efficient than the sample mean not only for normal but also for logistic distribution. He observed a similar feature even in the case of double-exponential distribution unless the values of a were small.

We shall consider estimators based on a finite sample $X_1, \cdots, X_n$. One intuitively expects from a good estimator T_n that the inaccuracy (1.1) will tend to 0 as fast as possible i.e., that the distribution of T_n will have the least possible tails. The tails of an estimator cannot be made arbitrarily small; for instance, if the sample comes from the Cauchy distribution one cannot find an estimator with exponentially decreasing tails.

We shall prove that the tails of a translation-invariant estimator could decrease to 0 at most n times faster than the tails of the basic distribution and that, on the other hand, there are estimators which behave from this point of view in the same way as one single observation (Theorem 2.1). Moreover, we shall show that both extreme cases may happen

Received November, 1979; revised April, 1980.

[1] This work was completed while the author was visiting Mathematical Centre in Amsterdam.

AMS 1970 *subject classifications.* Primary 62F10; secondary 62G05, 62G35.

Key words and phrases. Tails of the distribution, sample mean, L-estimator, trimmed mean, M-estimator, median, Hodges-Lehmann's estimator.

for the sample mean $\bar{X}_n$; $\bar{X}_n$ attains the upper bound if the basic distribution has exponentially decreasing tails of the type $\exp[-ba^r]$, $b > 0$, $r \geq 1$ and $\bar{X}_n$ attains only the lower bound if the basic distribution is heavy-tailed with the tails of the type ba^{-m}, $b > 0$, $m > 0$.

Estimating the centre of symmetry of an unknown symmetric distribution, we want to find an estimator which has small tails for as large a family of distributions as possible. Since an exponentially-tailed distribution contaminated by a heavy-tailed distribution becomes heavy-tailed, the sample mean $\bar{X}_n$ is not too good for such families of distributions. On the other hand, $\bar{X}_n$ remains good for such cases as a mixture of two normal distributions, for the normal distributions contaminated by the double-exponential distribution, etc.

If we trim off some extreme observations, then the rate of convergence of the tails of any resulting L-estimator attains neither the upper nor the lower bound (Theorem 3.1). The situation is similar for the estimators based on the ranks, e.g., for Hodges-Lehmann's estimator (Theorem 3.4). The tails of the sample median decrease exactly $(n + 1)/2$ times faster (for n odd) than the tails of the basic distribution, for both exponentially-tailed as well as for heavy-tailed distributions. The same holds for the Huber M-estimator generated by a bounded monotone odd function ψ (Theorem 3.3).

2. Behavior of the sample mean. Let us consider the model satisfying the following assumption:

ASSUMPTION A. $X_1, \cdots, X_n$ are random variables identically distributed according to the distribution function $F(x - \theta)$ with the density $f(x - \theta)$ such that $f(-x) = f(x) > 0$, $x \in R^1$; $\theta \in R^1$ is the parameter to be estimated.

All estimators we consider are translation-invariant, i.e., they satisfy the condition

$$T_n(X_1 + c, \cdots, X_n + c) = T_n(X_1, \cdots, X_n) + c \tag{2.1}$$

for any $c \in R^1$. If T_n is translation-invariant, then $P_\theta(|T_n - \theta| > a) = P_0(|T_n| > a)$ for any $\theta \in R^1$ and the inaccuracy (1.1) does not depend on θ.

The following theorem gives upper and lower bounds for the rate of convergence of the tails of a translation-invariant estimator.

THEOREM 2.1. *Let $T_n = T_n(X_1, \ldots, X_n)$ be a translation-invariant estimator of θ such that*

$$X^{(1)} > 0 \Rightarrow T_n(X_1, \cdots, X_n) > 0 \tag{2.2}$$
$$X^{(n)} < 0 \Rightarrow T_n(X_1, \cdots, X_n) < 0$$

where $X^{(1)} \leq X^{(2)} \leq \cdots \leq X^{(n)}$ are the order statistics corresponding to $X_1, \cdots, X_n$. Then, under Assumption A,

$$1 \leq \liminf_{a\to\infty} B(a, T_n) \leq \limsup_{a\to\infty} B(a, T_n) \leq n \tag{2.3}$$

where

$$B(a; T_n) = \frac{-\ln P_0(|T_n| > a)}{-\ln P_0(|X_1| > a)} \tag{2.4}$$

and P_0 is the probability distribution corresponding to F.

PROOF. We have

$$\begin{aligned} P_0(|T_n| > a) &= P_0(T_n > a) + P_0(T_n < -a) \\ &= P_0(T_n(X_1 - a, \cdots, X_n - a) > 0) + P_0(T_n(X_1 + a, \cdots, X_n + a) < 0) \end{aligned}$$

$$\geq P_0(X^{(1)} > a) + P_0(X^{(n)} < -a) = 2^{-n+1}[P_0(|X_1| > a)]^n$$

which implies the second inequality in (2.3). Similarly,

$$P_0(|T_n| > a) \leq P_0(X^{(n)} \geq a) + P_0(X^{(1)} \leq -a)$$
$$= 2\{1 - [1 - \tfrac{1}{2}P_0(|X_1| > a)]^n\}$$

and this implies the first inequality in (2.3). □

In the subsequent text, we shall investigate which estimators attain the upper bound in (2.3), which estimators are so poor that they attain the lower bound only and generally, what is the position of some well-known estimators from this point of view. We shall first consider the sample mean $\bar{X}_n$. The next theorem shows that the $\bar{X}_n$ attains the uper bound if the basic distribution has exponentially decreasing tails while it is a poor estimator for a heavy-tailed basic distribution.

THEOREM 2.2 *Let* $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ *be the sample mean, let* $X_1, \cdots, X_n$ *satisfy Assumption* A.

(i) *If*

$$\lim_{a\to\infty}\frac{-\ln(1-F(x))}{ba^r} = 1 \qquad \textit{for some} \quad b > 0, r \geq 1 \tag{2.5}$$

then

$$\lim_{a\to\infty} B(a; \bar{X}_n) = n. \tag{2.6}$$

(ii) *If*

$$\lim_{a\to\infty}\frac{-\ln(1-F(a))}{m\ln a} = 1, \qquad m > 0 \tag{2.7}$$

then

$$\lim_{a\to\infty} B(a; \bar{X}_n) = 1. \tag{2.8}$$

PROOF. Part (i) was proved by the author in [5]. Considering part (ii), we have

$$P_0(|\bar{X}_n| > a) = P_0(\bar{X}_n > a) + P(\bar{X}_n < -a)$$
$$\geq P_0(X_1 > -a, \cdots, X_{n-1} > -a, X_n > (2n-1)a)$$
$$+ P_0(X_1 < a, \cdots, X_n < a, X_n < -(2n-1)a)$$
$$= 2(F(a))^{n-1}[1 - F((2n-1)a)]$$

so that

$$\limsup_{a\to\infty} B(a; \bar{X}_n) \leq \limsup_{a\to\infty}\frac{-\ln[1 - F((2n-1)a)]}{m\ln[(2n-1)a]} = 1. \qquad \square$$

Part (i) concerns not only the normal ($r = 2$) but also the logistic and double-exponential distributions ($r = 1$); part (ii) covers Cauchy distribution ($m = 1$) and t-distribution with m degrees of freedom ($m > 1$). Theorem 2.2 says that $\bar{X}_n$ is a good estimator for the case (i) while it is a poor estimator for the case (ii). Now, what is the situation for $\bar{X}_n$ if F is a mixture of two distributions, one from each group?

The following lemma shows that if a distribution is contaminated by a heavy-tailed distribution then the resulting distribution is heavy-tailed. The sample mean $\bar{X}_n$ is a poor estimator in such a case.

LEMMA 2.1. *Let $F(x) = (1-\varepsilon)G(x) + \varepsilon H(x)$ where G and H are absolutely continuous distribution functions with the respective densities g and h such that $g(-x) = g(x) > 0$, $h(-x) = h(x) > 0$, $x \in R^1$; $0 < \varepsilon < 1$. If*

$$\lim_{x\to\infty} \frac{1-G(x)}{1-H(x)} = 0 \tag{2.9}$$

and

$$\lim_{x\to\infty} \frac{g(x)}{h(x)} = 0 \tag{2.10}$$

then

$$\lim_{x\to\infty} \frac{\ln(1-F(x))}{\ln(1-H(x))} = 1. \tag{2.11}$$

PROOF.

$$\lim_{x\to\infty} \frac{\ln(1-F(x))}{\ln(1-H(x))} = \lim_{x\to\infty} \frac{(1-H(x))f(x)}{(1-F(x))h(x)}$$

$$= \lim_{x\to\infty} \frac{(1-\varepsilon)\dfrac{g(x)}{h(x)} + \varepsilon}{(1-\varepsilon)\dfrac{1-G(x)}{1-H(x)} + \varepsilon} = 1. \qquad \square$$

3. Behavior of some robust estimators. If it is possible that the distribution of $X_1, \cdots, X_n$ is contaminated by a heavy-tailed distribution we must look for some more robust estimators of location. Let us consider what is the position of three basic types of robust estimators: L-estimators, M-estimators and R-estimators.

We shall show that the rate of convergence of the tails of such estimators is more or less bounded away from the lower as well as from the upper bound in (2.3). It means that the estimators are not optimal but, on the other hand, they may not be very poor.

3.1. L-*estimators*

THEOREM 3.1. *Let T_n be an* L-*estimator of the form*

$$T_n = \textstyle\sum_{i=1}^n c_i X^{(i)} \tag{3.1}$$

where $X^{(1)} \le \cdots \le X^{(n)}$ are the order statistics corresponding to $X_1, \cdots, X_n$ and $c_i \ge 0$, $i = 1, \cdots, n$ and $\sum_{i=1}^n c_i = 1$. Put $c_0 = c_{n+1} = 0$ and assume that $c_i = c_{n-i+1} = 0$ for $i = 0, 1, \cdots, k$ where $0 \le k < n/2$. Then, under Assumption A,

$$k + 1 \le \liminf_{a\to\infty} B(a; T_n) \le \limsup_{a\to\infty} B(a; T_n) \le n - k. \tag{3.2}$$

PROOF. The theorem was proved by the author in [5]. $\square$

COROLLARY. *Let T_n be the sample median corresponding to $X_1, \cdots, X_n$. Then, under the Assumption* A,

$$\frac{n}{2} \le \liminf_{a\to\infty} B(a; T_n) \le \limsup_{a\to\infty} B(a; T_n) \le \frac{n}{2} + 1 \tag{3.3}$$

for n even, and

$$\lim_{a\to\infty} B(a; T_n) = \frac{n+1}{2} \tag{3.4}$$

for n odd.

3.2. *Trimmed mean.* As a special case of L-estimators, consider the *trimmed mean* in the form

$$T_n = \frac{1}{n-2k} \sum_{i=k+1}^{n-k} X^{(i)}, \qquad k < \frac{n}{2}, n \geqq 3. \tag{3.5}$$

The following theorem shows that the behavior of the trimmed mean is similar to that of the sample mean: while it is near to the upper bound in (3.2) in the case of an exponentially-tailed distribution, it just attains the lower bound in the case of heavy-tailed distribution.

THEOREM 3.2. *Let T_n be the trimmed mean defined in* (3.5), *let $X_1, \cdots, X_n$ satisfy assumption* A.

(i) *If*

$$\lim_{a\to\infty} \frac{-\ln(1-F(a))}{ba^r} = 1 \qquad \textit{for some } b > 0, r \geqq 1 \tag{3.6}$$

then

$$n - 2k \leqq \liminf_{a\to\infty} B(T_n, a) \leqq \limsup_{a\to\infty} B(T_n, a) \leqq n - k. \tag{3.7}$$

(ii) *If* $k < \frac{n-1}{2}$ *and*

$$\lim_{a\to\infty} \frac{-\ln(1-F(a))}{m \ln a} = 1 \tag{3.8}$$

then

$$\lim_{a\to\infty} B(T_n, a) = k + 1. \tag{3.9}$$

PROOF. (i) According to Lemma 3.1 in [5], it suffices to show that

$$E_0 \exp\{(1-\varepsilon)(n-2k)b\,|\,T_n\,|^r\} < \infty \tag{3.10}$$

for any $\varepsilon \in (0, 1)$.

Put $d_n = (1-\varepsilon)(n-2k)$. Then, using the Hölder's inequality, we get

$$\begin{aligned} E_0[\exp\{d_n \cdot b\,|\,T_n\,|^r\}] &\leq E_0[\exp\{(1-\varepsilon)b \textstyle\sum_{i=k+1}^{n-k} |X^{(i)}|^r\}] \\ &\leq E_0[\exp\{(1-\varepsilon)b \textstyle\sum_{i=1}^{n} |X_i|^r\}] \\ &\leq (E_0 \exp\{(1-\varepsilon)b\,|X_1|^r\})^n < \infty \end{aligned} \tag{3.11}$$

where the last inequality follows from (3.6). It then follows from Lemma 3.1 in [5] (which is an immediate consequence of Markov's inequality) that

$$\liminf_{a\to\infty} B(T_n, a) \geq (1-\varepsilon)(n-2k) \tag{3.12}$$

and this implies the first inequality in (3.7). The second inequality follows from (3.2).

(ii) We have

$$P_0\{|\,T_n\,| > a\} = P_0\{T_n > a\} + P_0\{T_n < -a\} \tag{3.13}$$

and

$$\begin{aligned} &P_0\{T_n > a\} \\ &\qquad \geq P_0\{X^{(i)} > -a,\ i = k+1, \cdots, n-k-1,\ X^{(n-k)} > (2(n-2k)-1)a\} \end{aligned}$$

$$\geq P_0\{X_1 > -a, \cdots, X_{n-k-1} > -a, X_{n-k} > (2(n-2k)-1) a, \cdots, X_n > (2(n-2k)-1)a\} = (F(a))^{n-k-1}(1 - F((2n-4k-1)a))^{k+1}. \tag{3.14}$$

An analogous proof gives the same inequality for $P_0\{T_n < -a\}$. Hence,

$$-\ln P_0\{|T_n| > a\} \leq -\ln 2 + (k+1)\ln(1 - F((2n-4k-1)a)) -(n-k-1)\ln F(a). \tag{3.15}$$

Thus,

$$\limsup_{a\to\infty} B(T_n, a) \leq k + 1 \tag{3.16}$$

and (3.9) follows from (3.2). □

3.3. M-*estimators.* An M-estimator T_n is defined as any solution of the equation

$$\textstyle\sum_{i=1}^n \psi(X_i - t) = 0 \tag{3.17}$$

with respect to t; ψ is an appropriate nondecreasing odd function. We shall show that T_n behaves similarly as the sample median, at least for the distributions with exponentially decreasing and slowly decreasing tails.

THEOREM 3.3. *Let T_n be an* M-*estimator corresponding to the nondecreasing odd function ψ such that $\psi(x) = \psi(k)$ for $x \geq k$, $k > 0$. Suppose that the common distribution of $X_1, \cdots, X_n$ satisfies Assumption* A *and either of the following conditions*:

$$\lim_{a\to\infty}\frac{-\ln P_0(|X_1| > a)}{ba^r} = 1, \qquad b > 0, r > 0, \tag{3.18}$$

$$\lim_{a\to\infty}\frac{-\ln P_0(|X_1| > a)}{m \ln a} = 1, \qquad m > 0. \tag{3.19}$$

Then T_n satisfies (3.3) *and* (3.4).

PROOF.

(a) Suppose that n is even and denote $s = \frac{n}{2}$. Then

$$\begin{aligned} P_0(|T_n| > a) &= P_0(T_n > a) + P_0(T_n < -a) \\ &\geq P_0(\textstyle\sum_{i=1}^n \psi(X_i - a) > 0) + P_0(\sum_{i=1}^n \psi(X_i + a) < 0) \\ &\geq P_0(X^{(s)} - a > k) + P_0(X^{(s+1)} + a < -k) \\ &\geq 2\binom{n}{s-1}(F(a+k))^{s-1}(1 - F(a+k))^{s+1}; \end{aligned}$$

thus

$$\limsup_{a\to\infty} B(a; T_n) \leq (s+1) \limsup_{a\to\infty}\frac{\ln(1 - F(a+k))}{\ln(1 - F(a))} = s + 1.$$

Analogously,

$$\begin{aligned} P_0(|T_n| > a) &\leq P_0(X^{(s+1)} \geq a - k) + P_0(X^{(s)} \leq -a + k) \\ &= 2n\binom{n-1}{s}\int_{F(a-k)}^1 t^s(1-t)^{s-1}dt \leq 4\binom{n-1}{s}(1 - F(a-k))^s \end{aligned}$$

so that

$$\liminf_{a\to\infty} B(a; T_n) \geq s \liminf_{a\to\infty} \frac{(1 - F(a - k))}{(1 - F(a))} = s.$$

(b) The proof for n odd is analogous. □

3.4. R-*estimators.* We shall consider in detail only the Hodges-Lehmann estimator which has the form

$$T_n = \mathrm{med}_{1\leq i\leq j\leq n} \frac{X_i + X_j}{2}. \tag{3.20}$$

Other R-estimators could be investigated by the same method but it provides only the numerical values of the lower and upper bounds for the rate of convergence of the tails; we do not yet have an analytical formula expressing the bounds through the score-generating function of the underlying signed-rank test.

THEOREM 3.4. *Let T_n be the Hodges-Lehmann estimator* (3.20). *Then, under Assumption* A,

$$k_n + 1 \leq \liminf_{a\to\infty} B(a; T_n) \leq \limsup_{a\to\infty} B(a; T_n) \leq n - k_n \tag{3.21}$$

where k_n is the largest integer not exceeding $0.2n$.

PROOF. We shall first prove a simple lemma.

LEMMA 3.1. *Let $y_1, \cdots, y_n$ be integers satisfying $|y_i| = i$, $i = 1, \cdots, n$. If at least* $0.8n$ *of those numbers are negative, then* $\sum_{i=1}^n y_i < 0$.

PROOF OF LEMMA 3.1. If $0.8n$ is an integer, then

$$\sum_{i=1}^n y_i \leq -\sum_{i=1}^{0.8n} i + \sum_{i=0.8n+1}^{n} i = -0.14n^2 - 0.3n < 0;$$

$\sum_{i=1}^n y_i$ is still less in the case that $0.8n$ is not an integer. □

PROOF OF THEOREM 3.4. For any $t \in R^1$, *let* $R^+(|X_i - t|)$ be the rank of $|X_i - t|$ among $|X_1 - t|, \cdots, |X_n - t|$. T_n is an inversion of the Wilcoxon signed rank test, i.e., $T_n = \frac{1}{2}(T_n^* + T_n^{**})$ where

$$\begin{aligned} T_n^* &= \sup\{t: \textstyle\sum_{i=1}^n \mathrm{sign}(X_i - t)R^+(|X_i - t|) > 0\} \\ T_n^{**} &= \inf\{t: \textstyle\sum_{i=1}^n \mathrm{sign}(X_i - t)R^+(X_i - t) < 0\}. \end{aligned} \tag{3.22}$$

Then

$$\begin{aligned} P_0(|T_n| > a) &\leq 2P_0\{\textstyle\sum_{i=1}^n \mathrm{sign}(X_i - a)R^+(|X_i - a|) \geq 0\} \\ &\leq 2P_0(X^{n-k_n)} \geq a) \leq 2\binom{n-1}{k_n}\frac{(1 - F(a))^{k_n+1}}{k_n + 1}, \end{aligned}$$

thus

$$\liminf_{a\to\infty} B(a; T_n) \geq k_n + 1.$$

Similarly,

$$P_0(|T_n| > a) \geq 2P_0(\textstyle\sum_{i=1}^n \mathrm{sign}(X_i - a)R^+(|X_i - a|) > 0)$$

$$\geq 2P_0(X^{(k_n+1)} > a) \geq 2\binom{n}{k_n}(F(a))^{k_n}(1 - F(a))^{n-k_n}$$

so that

$$\limsup_{a\to\infty} B(a; T_n) \leq n - k_n.$$

REFERENCES

[1] BAHADUR, R. R. (1960). Asymptotic efficiency of tests and estimates. *Sankhyā* **22** 229–252.

[2] BAHADUR, R. R. (1967). Rates of convergence of estimates and test statistics. *Ann. Math. Statist.* **38** 303–324.

[3] BAHADUR, R. R. (1971). *Some Limit Theorems in Statistics.* SIAM, Philadelphia.

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

[5] JUREČKOVÁ, J. (1979). Finite-sample comparison of L-estimators of location, *Commentationes mathematicae Universitatis Carolinae* **20** 509–518.

[6] SIEVERS, G. L. (1978). Estimation of location; A large deviation comparison. *Ann. Statist.* **6** 610–618.

DEPARTMENT OF PROBABILITY AND STATISTICS
CHARLES UNIVERSITY
SOKOLOVSKA 83
18600 PRAGUE 8, CZECHOSLOVAKIA