Ann. Inst. Statist. Math.
Vol. 53, No. 2, 244–261 (2001)

TAIL BEHAVIOR AND BREAKDOWN PROPERTIES OF EQUIVARIANT ESTIMATORS OF LOCATION

JOZEF KUŠNIER* AND IVAN MIZERA*, **

Department of Statistics, University of Illinois at Urbana-Champaign, 101 Illini Hall, 725 South Wright Street, Champaign, IL 61820, U.S.A.

(Received May 13, 1998; revised August 23, 1999)

Abstract. For translation and scale equivariant estimators of location, inequalities connecting tail behavior and the finite-sample breakdown point are proved, analogous to those established by He *et al.* (1990, *Econometrika*, **58**, 1195–1214) for monotone and translation equivariant estimators. Some other inequalities are given as well, enabling to establish refined bounds and in some cases exact values for the tail behavior under heavy- and light-tailed distributions. The inequalities cover translation and scale equivariant estimators in great generality, and they involve new breakdown-related quantities, whose relations to the breakdown point are discussed. The worth of tail-behavior considerations in robustness theory is demonstrated on examples, showing the impact of the basic two techniques in robust estimation: trimming and averaging. The mathematical language employs notions from regular variation theory.

Key words and phrases: Robustness, breakdown, tail behavior, equivariance, location estimator, regular variation.

1. Introduction and overview

Criteria for assessing robustness are often based on large-sample considerations. This raises some concern about the relevancy of the asymptotic results; as an alternative, tractable finite-sample criteria are sought. A successful and already well-established robustness measure of this kind is the breakdown point (sometimes called also breakdown value). Its data-analytic, non-probabilistic nature is particularly stressed by its most used formalization, the finite-sample breakdown of Donoho and Huber (1982). It should be said that the measure of robustness provided by the breakdown is somewhat crude: there are numerous estimators with the same breakdown point and there are even many estimators with the maximal possible breakdown.

A possible refinement of the breakdown concept was outlined in the pioneering work of Jurečková (1979, 1981*a*, 1981*b*, 1985). She introduced a robustness criterion depending on the tail behavior: large deviation probabilities of a location estimator are compared to tail probabilities of the underlying distribution.

Suppose that $X_1, X_2, \dots, X_n$ are independent observations, identically distributed according to a law P_ϑ with a distribution function $F(x-\vartheta)$. Let $T = T_n(X_1, X_2, \dots, X_n)$

*Now at Department of Probability and Statistics, Comenius University, Mlynská dolina, SK-84215 Bratislava, Slovakia.

**Research partially supported by the Slovak VEGA grant 1/4196/97.

be an estimator of the location parameter ϑ. The tail behavior of T is expressed through the quantity

$$B_{T,F}(a) = \frac{-\log P_\vartheta[|T-\vartheta| > a]}{H(a)}.$$

We restrict our attention to the permutation-invariant T which are translation equivariant:

$$T_n(x_1+c, x_2+c, \ldots, x_n+c) = T_n(x_1, x_2, \ldots, x_n) + c$$

for any c. For such T we may assume, without loss of generality, that $\vartheta = 0$ and

$$B_{T,F}(a) = \frac{-\log P[|T| > a]}{H(a)}.$$

In the sequel, we write F for the distribution function of $P = P_0$. Let $G(a) = F(-a) + 1 - F(a)$ be its (two-sided) *tail* and $H(a) = -\log G(a)$ its (two-sided) *cumulative hazard function*.

The asymptotics of $B_{T,F}(a)$ for $a \to \infty$ (note the fixed sample size) indicate how well the estimator performs in comparison with a single observation. In this respect, we study

$$\underline{B}_{T,F} = \liminf_{a\to\infty} B_{T,F}(a) \quad \text{and} \quad \overline{B}_{T,F} = \limsup_{a\to\infty} B_{T,F}(a).$$

Clearly, $\underline{B}_{T,F} \le \overline{B}_{T,F}$. He, Jurečková, Koenker, and Portnoy (1990) (HJKP) pointed out a connection to the breakdown point ε_T^*: for those estimators which are also monotone (non-decreasing in each argument),

$$\varepsilon_T^* \le \underline{B}_{T,F} \le \overline{B}_{T,F} \le n - \varepsilon_T^* + 1, \tag{1.1}$$

for any F satisfying a minor regularity condition. Inequalities (1.1) of HJKP provide a probabilistic interpretation of the breakdown point, and at the same time open a new possibility for a more refined assessment of robustness, in the breakdown vein.

The requirement of monotonicity is, however, more restrictive than it might appear. There are estimators which are not monotone: redescending M-estimators, shorth, location versions of high-breakdown procedures used in regression, like LMS, S-estimators, and others. For some of these, non-monotonicity follows from the result of Bassett (1991), which says that the only monotone, translation and scale equivariant estimator with 50% breakdown point is the sample median. This implies, for instance, that the studentized version of Huber's estimator is not monotone. A natural question in this respect is whether the inequalities of HJKP are just a specific virtue of monotone estimators, or whether they represent a more general "law of nature".

In this paper, we settle this problem for scale equivariant estimators, satisfying $T_n(cx_1, cx_2, \ldots, cx_n) = cT_n(x_1, x_2, \ldots, x_n)$, for any nonzero c. (Note that the definition of scale equivariance implies also symmetry with respect to the reflection of the data around zero. This excludes asymmetric location estimators, like sample quantiles.) Translation and scale equivariant estimators possess the following constant-fit property: $T_n(c0, c0, \ldots, c0) = cT_n(0, 0, \ldots, 0)$ for any c, hence $T_n(0, 0, \ldots, 0) = 0$ and $T_n(c, c, \ldots, c) = c$.

Section 2 contains prerequisities and all general theorems. Theorem 1 brings the extension of HJKP inequalities for translation and scale equivariant location estimators.

It is not exactly breakdown point but a related quantity which appears in our inequalities; nonetheless, both breakdown indicators coincide in most practical cases.

Jurečková (1981*a*) and HJKP considered two general classes of distributions:

• distributions with *exponential tails* (light-tailed, Type I): there are $C > 0$, $r > 0$ such that $H(a) \sim Ca^r$; this class covers the logistic and Laplace distribution ($r = 1$) as well as the normal ($r = 2$);

• distributions with *algebraic tails* (heavy-tailed, Type II): there is $M > 0$ such that $H(a) \sim M \log a$; a representative of this class is the family of t distributions (including the Cauchy distribution).

We express the tail-heaviness more generally, via the exponent ρ of regular variation of H (the definition is given below). It turns out that unlike for monotone estimators, the lower bound in our analog of HJKP inequalities depends on ρ. It is the same as in (1.1) for $\rho = 0$ (heavy-tailed distributions); but it is, in general, smaller for $\rho > 1$.

On the other hand, the upper bound in (1.1) holds in great generality; for heavy-tailed distributions, as well as for light-tailed ones. Theorem 1 does not cover, from technical point of view, some (very) heavy-tailed distributions, which are covered by Theorem 2. This theorem introduces another breakdown-related quantity, and gives not only the upper HJKP bound for all heavy-tailed distributions, but it yields also a somewhat sharper upper bound for certain estimators.

This opens another theme of the paper. The HJKP inequalities determine the tail behavior completely only for estimators with 50% breakdown; for others, they only delimit its possible range. This raises concern about possible refinements, and finally the ultimate determination of the tail behavior. The analyses of various examples, given in Section 3, in some cases complement the work of Jurečková, quoted above, and in some cases give final determination of the tail behavior. The results are summarized in Table 1 in Section 4, which brings also some concluding remarks.

Theorem 3 concerns the tail behavior of estimators which are obtained as the average of a subset of observations, at light-tailed distributions. Theorem 3 applies also to studentized M-estimators, showing that their tail behavior is the same under light- as well as heavy-tailed distributions.

The proofs of theorems, as well as of some propositions from Section 3 are collected in the Appendix, which contains also several auxiliary results on regular variation. These might be useful in verifying the regularity conditions from Section 2.

2. Breakdown characteristics and tail behavior

We use the notation $\sim$, o, O in the usual meaning [as defined on p. 1 of Serfling (1980)]. A function g is called *regularly varying* if and only if for any $\lambda > 0$ there is a function $\varphi(\lambda)$ such that

$$\lim_{x\to\infty} \frac{g(\lambda x)}{g(x)} = \varphi(\lambda).$$

It is known [see Proposition 0.4 of Resnick (1987)] that in this case actually $\varphi(\lambda) = \lambda^\rho$, where ρ is called the *exponent* (or index) of regular variation. If $\rho = 0$, then g is called *slowly varying*. For more background on regular variation, see Bingham *et al.* (1987) or Resnick (1987).

The first regularity condition imposed on P delimits the scope of our investigations.

(A1) F has a positive density: $f(a) > 0$ for all $a \in \mathbb{R}$.

The results of Jurečková (1981b) reveal considerably different features of the tail behavior for distributions with compact support, compared to the regular case.

The second condition expresses the amount of regularity needed to avoid symmetry. Jurečková (1981a), as well as HJKP, originally considered symmetric distributions only; instead of $G(a)$ and $H(a)$, $(1 - F(a))$ and $-\log(1 - F(a))$ were involved, respectively. However, symmetry is not essential here, provided some regularity in the relationship of the two tails is present. We formulate our results in the non-symmetric setting, to avoid possible false impression that the theory applies only to symmetric situations. The right tail of P is called *dominant*, if $F(-a) = O(1 - F(a))$ as $a \to \infty$; if, conversely, $1 - F(a) = O(F(-a))$, then the left tail is dominant.

(A2) At least one tail of P is dominant.

It is possible that both are: for any symmetric P, for instance. In all subsequent formulations and proofs, it is not important which of the dominant tails is considered—if only there is one. Thus, the words "a dominant tail" should be read in this sense. Note that (A2) is not satisfied only if $(1 - F(a))/F(-a)$ oscillates from 0 to ∞ while $a \to \infty$.

Next condition makes possible to express tail-heaviness through regular variation.

(A3) The cumulative hazard function H of P is regularly varying with exponent ρ.

Since $H(a) \to \infty$ as $a \to \infty$, we have always $\rho \geq 0$ in (A3), due to a property of regularly varying functions [see Proposition 1.5.1 in Bingham *et al.* (1987)].

If $\rho = 0$, an additional condition may be needed. If the right tail is dominant, its corresponding *one-sided tail function* is $\bar{G}(a) = G_2(a) = 1 - F(a)$; if the left tail is dominant, then $\bar{G}(a) = G_1(a) = F(a)$. The corresponding *one-sided cumulative hazard function* is then $\bar{H}(a) = -\log \bar{G}(a)$.

(A4) If $\bar{G}$ is the one-sided tail function of a dominant tail, then $\limsup_{a\to\infty} \bar{G}(va)/\bar{G}(ua) < 1$ for any $v > u > 0$.

Some further properties concerning assumptions (A2)–(A4) are given in the Appendix; in particular, Propositions A.1 and A.2 show that (A4) follows from (A3) if $\rho > 0$. It should be noted that conditions (A1)–(A4) are satisfied by all known specific distributions supported by the whole real line.

It is clear that distributions with exponential tails satisfy (A3) with $\rho = r > 0$; thus, they satisfy (A4) as well. It is also easily seen that distribution with algebraic tails satisfy (A3) with $\rho = 0$ (the cumulative hazard function is slowly varying). We were not able to establish (A4) for all distributions with algebraic tails; despite that, (A4) holds for all typical representatives of this class, including all t distributions. The properties of regular variation suggest that (A4) is violated only for very heavy-tailed distributions. Jurečková (1981a) showed that a mixture of a light-tailed and a heavy-tailed distribution inherits the heavy-tailed behavior. This fact follows (under slightly weaker assumptions) also from our Proposition A.1, which also implies that the heavier tail in asymmetric situations (when left and right tail differ in behavior) dominates. In the popular contamination model $(1 - \varepsilon)F_1 + \varepsilon F_2$ this means that the tail-heaviness of the mixture depends only on the smaller of ρ's, no matter the magnitude of ε. This is somewhat paradoxical, in the light of the fact that (as shown below) the behavior of $\underline{B}_{T,F}$ and $\overline{B}_{T,F}$ depends entirely on ρ. Note, however, that $B_{T,F}(a)$ would probably depend on ε; we leave this topic for the further investigation.

Let $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ be a sample. The set of all "contaminated-by-replacement" samples is denoted by $B(\boldsymbol{x}, m)$; it contains exactly those samples which can be obtained from $\boldsymbol{x}$ by replacing not more than $m \geq 0$ points with arbitrary values. Let

$$\beta_T(m, \boldsymbol{x}) = \sup\{|T_n(\boldsymbol{y})| : \boldsymbol{y} \in B(\boldsymbol{x}, m)\}.$$

The finite-sample (called also Donoho-Huber) replacement *breakdown point* of T at $\boldsymbol{x}$ is defined as

$$\varepsilon_T^*(\boldsymbol{x}) = \inf\{m : \beta_T(m, \boldsymbol{x}) = \infty\}.$$

The (overall) *breakdown point* of T can be then defined as $\varepsilon_T^* = \inf_{\boldsymbol{x}} \varepsilon_T^*(\boldsymbol{x})$, where $\boldsymbol{x}$ runs over all possible samples. In many cases, the breakdown point is *universal*: $\varepsilon_T^* = \varepsilon_T^*(\boldsymbol{x})$ for all $\boldsymbol{x}$.

In the study of tail performance of scale equivariant estimators, it is more useful to measure rather "the worst-case performance over all samples". Just taking the sup of β over all $\boldsymbol{x}$ would result in ∞; the solution is, in view of scale equivariance, to take

$$\gamma_T(m) = \sup\{\beta_T(m, \boldsymbol{x}) : \boldsymbol{x} \in [-1, 1]^n\}.$$

The constant-fit property of translation and scale equivariant estimators implies that $\gamma_T(m) \geq 1$ for all m. Elementary, but important observations are: γ_T is non-decreasing in m; and

$$\mu_T^* = \inf\{m : \gamma_T(m) = \infty\} \leq \varepsilon_T^*. \tag{2.1}$$

The equality holds, however, for the vast majority of practical instances—in fact, the exceptional cases can be regarded as pathological.

As a prolog to HJKP inequalities, Jurečková (1981*a*) proved that any estimator such that

(E) $T_n(X_1, X_2, \ldots, X_n)$ lies within a convex hull of $X_1, X_2, \ldots, X_n$

satisfies $1 \leq \underline{B}_{T,F} \leq \overline{B}_{T,F} \leq n$. The inequalities are sharp: the arithmetic mean attains n for distribution with exponential tails and 1 for algebraic tails—in this case, its tail performance does not improve over that of one observation. We mention this result because of the property (E), which is apparently a virtue of every reasonable location estimator.

THEOREM 1. *Suppose that* (A1), (A2) *hold and* T *is translation and scale equivariant estimator satisfying* (E). *If* (A3) *holds, then*

$$\max\{(m+1)[\gamma_T(m)]^{-\rho} : 0 \leq m < \mu_T^*\} \leq \underline{B}_{T,F}. \tag{2.2}$$

If (A3) *holds with* $\rho > 0$, *or if* (A3) *and* (A4) *hold, then*

$$\overline{B}_{T,F} \leq n - \mu_T^* + 1. \tag{2.3}$$

Since $\gamma_T(m) \geq 1$, the maximum in 2.2 does not exceed μ_T^*. In the majority of cases, the maximum is attained for $m = \mu_T^* - 1$. If H is slowly varying (that is, if $\rho = 0$), then

$$\mu_T^* \leq \underline{B}_{T,F} \leq \overline{B}_{T,F} \leq n - \mu_T^* + 1, \tag{2.4}$$

whenever (A1), (A2) and (A4) hold.

The form of (2.2) opens a possibility that under light-tailed distributions, a non-monotone estimator possesses worse tail behavior than a monotone estimator with the same breakdown point. Though we do not know whether inequality (2.2) is sharp, in the next section we give examples of estimators whose tail behavior is inferior to the lower bound in (1.1), valid for monotone estimators.

As already mentioned, though (A3) and (A4) hold for a large class of distributions, there is a possibility that (A4) is violated for distributions with very heavy tails (heavier than those of the Cauchy distribution, for instance). For this reason, we present an alternative way to assess (2.3) for distributions with slowly varying H. It employs another type of breakdown concept, suitable, like γ_T and μ_T^*, primarily for location and scale equivariant estimators. Moreover, in this vein we are able to obtain refined upper bounds for the tail behavior under heavy-tailed distributions. Let

$$\delta_T(m) = \sup\{T_n(\boldsymbol{y}) : \boldsymbol{y} \in B(\boldsymbol{x}, m) \cap (-\infty, 1]^n, \boldsymbol{x} \in (-\infty, 0]^n\}.$$

Obviously, $\delta_T(m)$ is non-decreasing in m; and $\delta_T(m) \leq 1$ whenever T satisfies (E). Of special interest are those m for which the equality holds; let $\eta_T^* = \inf\{m : \delta_T(m) = 1\}$. The relationship between η_T^* and ε_T^* or μ_T^* is more delicate than that between the latter two. Usually $\varepsilon_T^* \leq \eta_T^*$, but there are (again pathological) examples not satisfying this inequality. On the other hand, the inequality can be sharp: for trimmed means, for instance.

It can be shown by exact-fit considerations that for the estimators with maximal breakdown point, $\varepsilon_T^* = \mu_T^*$. (The exact-fit point of scale equivariant estimators is equal to the breakdown point at the sample with all points concentrated in c; and translation equivariance further makes the exact-fit point universal, that is, not dependent on c.) For odd n, ε_T^* and μ_T^* are both equal to η_T^*. For even n there is a small difference: while $\varepsilon_T^* = \mu_T^* \lfloor (n+1)/2 \rfloor$, $\eta_T^* = \lfloor n/2 \rfloor + 1$. Note also that due to symmetry,

$$-\delta_T(m) = \inf\{T_n(\boldsymbol{y}) : \boldsymbol{y} \in B(\boldsymbol{x}, m) \cap [-1, \infty)^n, \boldsymbol{x} \in [0, \infty)^n\}, \tag{2.5}$$

hence there is no real need for notions symmetric to $\delta_T(m)$ and η_T^*.

THEOREM 2. *Assume that* (A1), (A2), *and* (A3) *with* $\rho = 0$ *hold. If* T *is a translation and scale equivariant estimator satisfying* (E), *then*

$$\overline{B}_{T,F} \leq n - \eta_T^* + 1. \tag{2.6}$$

In the vein of Theorem 2, an inequality for $\rho > 0$ (involving η_T^*) might be proved; however, it would be inferior to that given by (2.2). On the other hand, Theorem 2 can yield stronger upper inequality than (1.1) for $\rho = 0$; note that in this case, (1.1) is implied by (2.6) if $\varepsilon_T^* \leq \eta_T^*$.

The quantity γ_T is very similar to the maximum deviation curve (MDC) introduced by Croux (1996). MDC is defined in the same way as our γ_T, only instead of β_T the sup of $|T_n(\boldsymbol{y}) - T_n(\boldsymbol{x})|$ (instead of $|T_n(\boldsymbol{y})|$) is taken. Thus,

$$\mathrm{MDC}(m) = \sup\{|T_n(\boldsymbol{y}) - T_n(\boldsymbol{x})| : \boldsymbol{y} \in B(\boldsymbol{x}, m), \boldsymbol{x} \in [-1, 1]^n\}.$$

Croux uses $[-1/2, 1/2]^n$ instead of $[-1, 1]^n$, but this is merely a matter of taste. MDC may be more natural for the study of the bias behavior; however, γ_T is better suited for our needs. For estimators satisfying (E),

$$\max\{1, \mathrm{MDC}(m) - 1\} \leq \gamma_T(m) \leq \mathrm{MDC}(m) + 1. \tag{2.7}$$

Consequently, the "MDC breakdown point", that is, the inf of those m for which $\mathrm{MDC}(m) = \infty$, is equal to μ_T^*. In certain cases, MDC is equal to γ_T.

Jurečková (1979, 1981*a*) proved that the mean behaves well under distributions with exponential tails. The following theorem shows that this property is shared to an extent by estimators obtained through averaging a subset of observations; for L-estimators, a version of this fact was proved by Jurečková (1981*a*). The decisive property is the following one; it depends on m.

(M) If $|T| > a$, then there is a subset $Y_1, Y_2, \ldots, Y_m$ of $X_1, X_2, \ldots, X_n$ such that $|\frac{1}{m}\sum_{i=1}^{m} Y_i| > a$.

THEOREM 3. *Suppose that F is continuous and $H(a) \sim h(a)$, where h is a non-negative, increasing, differentiable and convex function. If T is translation equivariant estimator satisfying* (M), *then* $\underline{B}_{T,F} \geq m$.

For distributions with algebraic tails, $h(a) = Ca^r$ and the assumptions of Theorem 3 are satisfied if $r \geq 1$, for the Laplace or normal distribution, for instance.

3. Some examples and their more detailed analysis

Monotone estimators. For monotone estimators which are translation and scale equivariant (scale equivariance was not required by HJKP) inequalities (1.1) follow from (2.4), Theorem 1, and the following proposition, which implies that $\varepsilon_T^* \leq \mu_T^*$ (then (2.1) yields equality, so that μ_T^* can be replaced by ε_T^* in the upper bound (2.3)). The proposition also asserts that $\gamma_T(\mu_T^* - 1) = 1$; hence the lower bound (2.2) reduces to ε_T^* under any type of distribution.

PROPOSITION 1. *For any translation and scale equivariant monotone estimator T satisfying* (E), (i) $\gamma_T(m) = 1$ *whenever* $m < \varepsilon_T^*$; (ii) $\varepsilon_T^* = \mu_T^* \leq \eta_T^*$.

PROOF. From Bassett (1991) it follows that any translation and scale equivariant monotone estimator with breakdown point equal to ε_T^* must lie between the ε_T^* and $1 - \varepsilon_T^*$ sample quantiles. This implies that $\gamma_T(m) \leq 1$ for $m < \varepsilon_T^*$. The constant-fit property of the estimators under consideration then implies that $\gamma_T(m) = 1$ for $m < \varepsilon_T^*$, and also that $\gamma_T(m) = \infty$ if $m \geq \varepsilon_T^*$. Consequently, $\mu_T^* = \varepsilon_T^*$ and the max in (2.2) is equal to ε_T^* for any ρ. Analogously, we obtain that $\delta_T(m) = 0$ for $m < \varepsilon_T^*$; thus, $\varepsilon_T^* \leq \eta_T^*$. □

All L-estimators with positive weights (including the median, all trimmed and Winsorized means) are monotone, translation and scale equivariant. Theorem 2 shows that their exact tail behavior is lower than the upper bound given by Theorem 1.

PROPOSITION 2. *For any L-estimator assigning a positive weight to the i-th and $(n - i + 1)$-th order statistic,* $\eta_T^* > n - i$.

PROOF. If all points are in $(-\infty, 0]$ and $n - i$ of them are allowed to be in $(-\infty, 1]$, then still the i-th order statistic does not exceed 0; hence T cannot be equal to 1. □

Note that, due to symmetry, an L-estimator assigns a positive weight to the i-th order statistic, if and only if it does so to the $(n - i + 1)$-th one. Proposition 2 and

Theorem 2 combined with Theorem 1 show—as established by Jurečková (1985)—that for all L-estimators, $\underline{B}_{T,F} = \overline{B}_{T,F} = \varepsilon_T^*$ if the underlying distribution is heavy-tailed (with $\rho = 0$).

On the other hand, Theorem 3 yields for trimmed means the highest possible (by inequalities of HJKP) tail behavior under distributions with exponential tails (satisfying assumptions of Theorem 3). To see this, just note that all trimmed means which trim k observations on each side satisfy (M) with $m = n - k$ (note that if the trimmed mean is greater than a, not only the average of the middle $n - 2k$ order statistics, but also the average of $n - k$ largest order statistics is greater than a).

Another example of monotone, translation and scale equivariant estimator is the Hodges-Lehmann estimator, whose properties slightly differ from those of trimmed means. Again $\mu_T^* = \varepsilon_T^*$, by Proposition 1; but here also $\eta_T^* = \varepsilon_T^*$ for odd n, thus Theorem 2 gives the same bound as Theorem 1 in this case.

LMS and shorth. According to Rousseeuw and Leroy (1987), the location LMS is the midpoint of the shortest "half" of the data—that is, the midpoint of the shortest interval containing $\lfloor n/2 \rfloor + 1$ data points. The shorth is the average of the observations lying in this interval. Clearly, both estimators are translation and scale equivariant. Since both have breakdown point equal to $\lfloor (n+1)/2 \rfloor$ for odd n, the result of Bassett (1991) implies they are not monotone (which can be also seen directly). If more than a half of the data points lie in $[-1, 1]$, then the shortest half of the data must contain at least one point from those lying in $[-1, 1]$. This yields that the shortest "half" of the data is contained in $[-3, 3]$. Thus, if $m < \lfloor (n+1)/2 \rfloor$, then $\gamma_T(m) \le 2$ for the LMS; in fact, $\gamma_T(m) = 2$ (allow the right endpoint of the shortest half tend to 3). Thus, $\varepsilon_T^* = \mu_T^* = \lfloor (n+1)/2 \rfloor + 1$. A similar argument shows that also $\eta_T^* = \varepsilon_T^*$ for odd n.

The maximum in (2.2) is equal to $\mu_T^* [\gamma_T(\mu_T^* - 1)]^{-\rho} = \lfloor \frac{1}{2}(n+1) \rfloor \, 2^{-\rho}$. For heavy-tailed distributions with $\rho = 0$, the lower bound is the same as that for monotone estimators. However, it can be much lower when ρ is equal to 1 or 2.

To get the better idea about the exact tail behavior of LMS, we derive the following upper bound.

PROPOSITION 3. *Assume* (A1)–(A4). *For the location LMS*,

$$\overline{B}_{T,F} \le \left(n - \left\lfloor \frac{1}{2}(n+1) \right\rfloor + 1\right) \left(\frac{2}{3}\right)^{\rho} + \left[\left(\frac{4}{3}\right)^{\rho} - \left(\frac{2}{3}\right)^{\rho}\right]. \tag{3.1}$$

PROOF. See the Appendix. □

For the shorth, we have again that $\varepsilon_T^* = \mu_T^* = \lfloor (n+1)/2 \rfloor + 1$; the same holds for η_T^* if n is odd. For $m < \varepsilon_T^*$,

$$\gamma_T(m) \le \frac{3m + \left\lfloor \frac{1}{2} n \right\rfloor + 1 - m}{\left\lfloor \frac{1}{2} n \right\rfloor + 1} = 1 + \frac{2m}{\left\lfloor \frac{1}{2} n \right\rfloor + 1}$$

(this bound is sharp, as can be easily seen). Now, the maximum in (2.2) does not exceed

$$\max\left\{\frac{(m+1)\left(\left\lfloor\frac{1}{2}n\right\rfloor+1\right)^{\rho}}{\left(2m+\left\lfloor\frac{1}{2}n\right\rfloor+1\right)^{\rho}} : m < \mu_T^*\right\}. \tag{3.2}$$

For $\rho = 1$, (3.2) is equal to

$$\mu_T^*[\gamma_T(\mu_T^*-1)]^{-\rho} = \frac{\left(\left\lfloor\frac{1}{2}(n+1)\right\rfloor\right)\left(\left\lfloor\frac{1}{2}n\right\rfloor+1\right)}{\left(2\left\lfloor\frac{1}{2}(n+1)\right\rfloor+\left\lfloor\frac{1}{2}n\right\rfloor-1\right)}.$$

However, for $\rho = 2$, the maximum in (3.2) is

$$\begin{aligned} &\frac{\left(\left\lfloor\frac{1}{2}n\right\rfloor+1\right)^2}{8\left(\left\lfloor\frac{1}{2}n\right\rfloor-1\right)} && \text{if } \left\lfloor\frac{1}{2}n\right\rfloor \text{ is odd,}\\ &\frac{\left\lfloor\frac{1}{2}n\right\rfloor\left(\left\lfloor\frac{1}{2}n\right\rfloor+1\right)^2}{2\left(2\left\lfloor\frac{1}{2}n\right\rfloor-1\right)^2} && \text{if } \left\lfloor\frac{1}{2}n\right\rfloor \text{ is even,}\\ &\geq \frac{1}{8}\left(\left\lfloor\frac{1}{2}n\right\rfloor+1\right) && \text{in both cases,} \end{aligned}$$

and is attained for $m < \mu_T^* - 1$ (which is a bit unusual, but not impossible).

However, the real tail behavior for distribution functions F satisfying the assumptions of Theorem 3 is much better: $\underline{B}_{T,F} \geq \lfloor n/2 \rfloor + 1$, since the shorth is computed by averaging the $\lfloor n/2 \rfloor + 1$ observations. For odd n, it is the same tail behavior as for heavy-tailed distributions (and for monotone estimators). We can see that tail-behavior considerations are able to distinguish between two very closely related estimators, which otherwise have the same breakdown point and the same cube-root rate of convergence. This shows the value of tail-behavior considerations.

M-estimators. Huber's estimator. Although simple M-estimators with monotone score functions are monotone, their studentized versions are generally not. In practical use of M-estimators, some kind of scale adjustment is highly necessary: the "plain" versions suffer not only from lack of scale equivariance (the mean and median are the only exceptions), but also from possible non-identifiability [Mizera (1994)], and bad change-of-variance and breakdown behavior [Donoho and Huber (1982)]. Consequently, some of the approaches recommended by Huber (1981) should be adopted: either studentization, or "Proposal 2" simultaneous estimation of location and scale.

Suppose that ψ is a monotone, even (that is, $\psi(-x) = -\psi(x)$) and bounded score function. Let $S = S_n(X_1, X_2, \ldots, X_n)$ be a scale estimator. An M-estimator with score function ψ studentized by S is defined as the solution T of the equation $\lambda(T, S) = 0$, where

$$\lambda(t, s) = \sum_{i=1}^{n} \psi\left(\frac{X_i - t}{s}\right),$$

for $t \in \mathbb{R}$ and $s > 0$. For $s = 0$, we set the value of $\psi((X_i - t)/s)$ to be the limiting value for $s > 0$: that is,

$$\lambda(t, 0) = \sum_{i=1}^{n} \text{sign}(X_i - t)$$

(we may suppose without loss of generality that $-\psi(-\infty) = \psi(\infty) = 1$). Note that monotonicity of ψ yields that

$$\lambda(t_1, s) \geq \lambda(t_2, s) \quad \text{whenever} \quad t_1 \leq t_2 \tag{3.3}$$

and also, if $s_1 \leq s_2$,

$$\begin{aligned} \psi\left(\frac{X_i - t}{s_1}\right) &\geq \psi\left(\frac{X_i - t}{s_2}\right) \quad \text{if} \quad X_i - t \geq 0, \\ &\leq \psi\left(\frac{X_i - t}{s_2}\right) \quad \text{if} \quad X_i - t \leq 0. \end{aligned} \tag{3.4}$$

As soon as S is location invariant, T keeps the location equivariance of the non-studentized version. Moreover, the scale equivariance of S implies that of T. The studentized estimator T inherits also the breakdown point of the non-studentized version, provided that the breakdown point of S is high enough.

The most common choice for S is MAD, the median absolute deviation from the median. It is well-known that its breakdown point, as well as that of an M-estimator studentized by it, is $\varepsilon_T^* = \lfloor (n+1)/2 \rfloor$ [see, for instance, Huber (1981)].

To compute $\gamma_T(m)$, note first that if $m \leq m^* = \varepsilon_T^* - 1$ and more than $n - \varepsilon_T^*$ of X_i's lie in $[-1, 1]$, then $S = \text{MAD} \leq 2$. If $\psi(c) > 1 - d$, then (3.4) yields that if $d \leq (n - 2m)/(n - m)$, then

$$\lambda(1 + 2c, S) < -(1 - d)(n - m) + m = -n + 2m + d(n - m) \leq 0,$$

and, analogously, also $\lambda(-1 - 2c, S) > 0$. Thus, by (3.3), $\gamma_T(m) \leq 1 + 2c_m$, where c_m is the smallest c satisfying $\psi(c) \geq m/(n - m)$. Note that $c_1 \leq c_2 \leq \cdots \leq c_{m^*}$. Consequently, $\mu_T^* = \varepsilon_T^*$.

A similar argument shows that also $\eta_T^* = \varepsilon_T^*$ for odd n: whatever is the value of the scale parameter s in $\lambda(t, s)$, the M-estimate can never be equal to 1 if all observation are in $(-\infty, 1]$ and more than half of them in $(-\infty, 0]$. Theorem 2 then establishes the upper bound $n - \lfloor (n+1)/2 \rfloor + 1$ for the tail behavior under heavy-tailed distributions. This bound applies also to other studentizing scale estimators, and also when location and scale are estimated simultaneously—provided that the scale part breakdown point is not less than $\lfloor (n+1)/2 \rfloor$.

If T is Huber's estimator with tuning constant k, that is, the score function $\psi(x)$ is equal to $\max\{-1, \min\{x/k, 1\}\}$, we have that $c_m = (km)/(n - m)$. Thus, the maximum in (2.2) is equal to

$$\mu_T^* [\gamma_T(\mu_T^* - 1)]^{-\rho} = \varepsilon_T^* \left(1 + 2k \frac{\varepsilon_T^* - 1}{n - \varepsilon_T^* + 1}\right)^{-\rho} \geq \varepsilon_T^* (1 + 2k)^{-\rho}.$$

This bound suggests that Huber's estimator might exhibit a similar loss in tail behavior as the location LMS estimator. However, it is plausible that its tail behavior under light-tailed distribution is rather that of shorth, since Huber's estimator also performs a kind of "averaging" in the middle of data. Let $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ denote the order statistics of the sample $X_1, X_2, \ldots, X_n$.

PROPOSITION 4. *Let ψ be the score function for Huber's estimator; let $m = \lfloor (n+1)/2 \rfloor$. If $\lambda(T_s, s) = 0$, then*

$$T_s \leq \frac{1}{m} \sum_{i=n-m+1}^{n} X_{(i)}$$

for any s.

PROOF. See the Appendix. □

As a consequence, Huber's estimator satisfies (M). Theorem 3 then yields that $\underline{B}_{T,F} \geq \lfloor (n+1)/2 \rfloor$ for distributions with exponential tails. Again, the result is valid for any type of studentization, and also for the location part of simultaneous ("Proposal 2") estimators of location and scale (now even without any assumptions on the breakdown point of the scale part).

4. Concluding remarks

As expressed by HJKP, robust estimators can be viewed as "minimax" in the tail-behavior sense: they sacrifice some tail performance (especially under light-tailed distributions), in order to maintain it stable under all types of distributions (particularly heavy-tailed ones).

Our findings relate this theoretical principle with the basic techniques for constructing robust estimators: trimming and averaging. In a sense, they act in opposite directions: trimming (note that not necessarily related to order statistics) ensures high breakdown and good tail behavior under heavy-tailed distributions. Averaging the rest of observations then rescues good tail behavior under light-tailed distributions.

Our examples also show that tail-behavior investigations are a worthy way of assessing robustness. For instance, they are able to distinguish between the location LMS

Table 1. Breakdown characteristics and bounds on the tail behavior.

Estimator	*Breakdown charact. (%)*			*Tail behavior (%)*					
				$\rho = 0$		$\rho = 1$		$\rho = 2$	
	ε^*_T	μ^*_T	η^*_T	$\geq$	$\leq$	$\geq$	$\leq$	$\geq$	$\leq$
Sample mean	0	0	100	0	0	100	100	100	100
Sample median	50	50	50	50	50	50	50	50	50
29% trimmed mean	29	29	71	29	29	71	71	71	71
Hodges-Lehmann	29	29	29	29	71	29	71	29	71
LMS	50	50	50	50	50	17	33	7	22
Shorth	50	50	50	50	50	50	50	50	50
Huber's M-estimator	50	50	50	50	50	50	50	50	50

estimator and the shorth, two closely related estimators with the same breakdown point and similar asymptotics. Also, tail-behavior considerations are of finite-sample, non-asymptotic nature.

Table 1 contains a summary of our results for various location estimators. Figures are simplified, in form of percentages of n, neglecting the additive terms vanishing with increasing n, and also the subtle differences between odd and even n. We hope that this simplification will make tail-behavior features more visible. Under heavy-tailed distributions we understand those satisfying the assumptions of Theorem 2 with $\rho = 0$: t family, for instance. Light-tailed distributions satisfy the assumptions of Theorem 3 with $\rho > 0$: examples are the Laplace or logistic ($\rho = 1$) or the normal distribution ($\rho = 2$). First four estimators are monotone: sample mean and median, trimmed mean (the trimming proportion was chosen to allow for a comparison with the Hodges-Lehmann estimator; analogous figures can be obtained for any other choice), and the Hodges-Lehmann estimator. The last three estimators are non-monotone: LMS, shorth and Huber's M-estimator. Huber's estimator can be either in its studentized, or simultaneous location-scale ("Proposal 2") form: the only requirement is that the scale part has 50% breakdown point.

For all but the Hodges-Lehmann and the LMS, Table 1 gives the definitive assessment of the tail behavior. The LMS is shown to be inferior to other 50% breakdown estimators. The behavior of the Hodges-Lehmann estimators remains unsettled, in the bounds given by HJKP.

Acknowledgements

We are grateful to Jana Jurečková, Steve Portnoy and an anonymous referee for pointing out numerous inaccuracies and valuable suggestions for improving the manuscript.

Appendix: proofs and some additional technical facts

We prove some auxiliary statements first. The first one of them shows that the tail behavior depends on a dominant tail. It is formulated slightly more generally, to cover also Lemma 2.1 from Jurečková (1981a) about mixtures, where similar behavior takes place: the dominant component, the component with heavier tails, prevails. Thus, when $c_1 + c_2 = 1$ below, the statement is about mixtures; on the other hand, for our proofs we need the version with $c_1 = c_2 = 1$.

PROPOSITION A.1. *Suppose that* $g(a) = c_1 g_1(a) + c_2 g_2(a)$. *If* $g_2(a) = O(g_1(a))$ *and* $g_1(a) \to 0$ *as* $a \to \infty$, *then* $-\log g(a) \sim -\log g_1(a)$. *Consequently, if* $\bar{H}$ *corresponds to a dominant tail, then* $\bar{H}(a) \sim H(a)$ *and* H *is regularly varying if and only if* $\bar{H}$ *is, with the same exponent.*

PROOF. Just note that

$$\frac{-\log g(a)}{-\log g_1(a)} = 1 + \frac{-\log \left[c_1 + c_2 \dfrac{g_2(a)}{g_1(a)}\right]}{-\log g_1(a)},$$

and that $-\log g_1(a) \to \infty$ as $a \to \infty$. The rest of the statement immediately follows. □

PROPOSITION A.2. *Suppose that $g(a)$ is a positive function such that $g(a) \to 0$ for $a \to \infty$. If $h(a) = -\log g(a)$ is regularly varying with exponent $\rho > 0$, then*

$$\limsup_{a\to\infty} \frac{g(va)}{g(ua)} < 1 \tag{A.1}$$

whenever $v > u > 0$.

PROOF. Just note that

$$\limsup_{a\to\infty} \frac{g(va)}{g(ua)} = \limsup_{a\to\infty} e^{-[h(va)-h(ua)]} = \lim_{a\to\infty} e^{-h(ua)[h(va)/h(ua)-1]} = 0,$$

since the limit in the exponent exists and is equal to $-\infty$. □

PROPOSITION A.3. *If g is a positive function satisfying* A.1, *and h is an arbitrary function such that $h(a) \to \infty$ for $a \to \infty$, then*

$$\lim_{a\to\infty} \frac{-\log[g(ua) - g(va)]}{h(a)} = \lim_{a\to\infty} \frac{-\log g(ua)}{h(a)},$$

whenever $v > u > 0$ and the limit on the right exists.

PROOF. The proposition follows from the equality

$$\frac{-\log[g(ua) - g(va)]}{h(a)} = \frac{-\log g(ua)}{h(a)} + \frac{-\log\left[1 - \dfrac{g(va)}{g(ua)}\right]}{h(a)}.$$

The limit of the rightmost term is 0 since the denominator is bounded. □

PROOF OF THEOREM 1. We first prove inequality (2.2). Let $m^* < \mu_T^*$. Note that if $m \geq n - m^* + 1$ of X_i's lie in $[-1, 1]$, then $|T| \leq K = \gamma_T(m^* - 1)$. Let $|X^*|$ be the $(n - m^* + 1)$-th order statistic among $|X_1|, |X_2|, \ldots, |X_n|$. By scale equivariance,

$$T_n(X_1, X_2, \ldots, X_n) = \left| |X^*| \, T_n\left(\frac{X_1}{|X^*|}, \frac{X_2}{|X^*|}, \ldots, \frac{X_1}{|X^*|}\right) \right| \leq K|X^*|.$$

Hence,

$$\begin{aligned} P[|T| > a] &\leq P[|X^*| > a/K] \\ &= \int_{a/K}^{\infty} n\binom{n-1}{n-m^*} G(t)^{n-m^*}(1 - G(t))^{m^*-1}\tilde{g}(t)\,dt \\ &\leq \int_{1-G(a/K)}^{1} n\binom{n-1}{n-m^*}(1-s)^{m^*-1}\,ds = \binom{n-1}{n-m^*} G\Big(\frac{a}{K}\Big)^{m^*}, \end{aligned}$$

where $1 - G(x)$ is the distribution function of the absolute value of the random variable with distribution function F, and $\tilde{g}$ is its density. Since $H(a) \to \infty$,

$$\underline{B}_{T,F} \geq m^* \lim_{a\to\infty} \frac{-\log G(a/K)}{H(a)} = m^* \lim_{a\to\infty} \frac{H(a/K)}{H(a)} = m^* K^{-\rho},$$

due to (A3). This proves (2.2).

By Proposition A.2, (A3) with $\rho > 0$ implies (A4). Hence, it is sufficient to prove (2.3) under (A3) and (A4). Suppose that $m \geq n - m^* + 1$. If m of X_i's lie in $[-1, 1]$, then $|T| \leq K$. By translation equivariance, $T \geq d - K$ whenever m or more of X_i's lie in $[d-1, d+1]$. If $d - K > 0$, scale equivariance yields for arbitrary $c > 0$ that $T \geq c(d-K)$ whenever m of X_i's lie in $[c(d-1), c(d+1)]$. Writing $a = c(d-K)$, $u = (d-1)/(d-K)$, and $v = (d+1)(d-K)$, we obtain the following: given any $d > K$, there are $v > u > 1$ such that $T \geq a$ if at least m of X_i's lie in $[ua, va]$. Thus, for any $a > 0$ and any $d > K$

$$\text{(A.2)} \qquad P[|T| > a] \geq P[X_{(m^*)} > ua \ \&\ X_{(n)} < va] \geq [F(va) - F(ua)]^{n-m^*+1}.$$

We can suppose, without loss of generality, that the right tail is dominant (otherwise the whole proof would be for a symmetrically reflected situation). By A.2,

$$\begin{aligned} \frac{-\log P[|T| > a]}{H(a)} &\leq \frac{-(n-m^*+1)\log[F(va) - F(ua)]}{H(a)} \\ &= \frac{-(n-m^*+1)\log[\bar{G}(ua) - \bar{G}(va)]}{H(a)}. \end{aligned}$$

By (A4) and Propositions A.3 and A.1,

$$\begin{aligned} \overline{B}_{T,F} &\leq (n-m^*+1) \lim_{a\to\infty} \frac{\bar{H}(ua)}{H(a)} = (n-m^*+1) \lim_{a\to\infty} \frac{\bar{H}(ua)}{\bar{H}(a)} \\ &= (n-m^*+1)u^\rho = (n-m^*+1)\left(\frac{d-1}{d-K}\right)^\rho. \end{aligned}$$

The last inequality holds for arbitrary $d > K$ chosen in advance; thus, allowing $d \to \infty$, we obtain 2.3. □

Proof of Theorem 2. Again, we can assume without loss of generality that the right tail is dominant. By symmetry, we have (2.5); by translation equivariance, it follows that $0 < 1 - \delta_T(m) \leq T$ whenever $m \leq \eta_T^* - 1$, $x_i \geq 0$ for all i, and $n - \eta_t^* + 1$ of x_i's lie in $[1, \infty)$. Finally, scale equivariance yields for $m \leq \eta_T^* - 1$ that $0 < c(1 - \delta_T(m)) \leq T$ whenever $X_{(1)} \geq 0$ and $X_{(m+1)} \geq c$. Hence, writing $m^* = \eta_T^*$, $a = c(1 - \delta_T(m^* - 1))$ and $u = (1 - \delta_T(m^* - 1))^{-1}$, we obtain

$$\begin{aligned} P[|T| > a] &\geq P[T > a] \geq P[X_{(1)} \geq 0 \ \&\ X_{(m^*)} \geq ua] \\ &\geq G_1(0)^{m^*-1} G_1(ua)^{n-m^*+1}. \end{aligned}$$

Hence, again by Proposition A.1,

$$\begin{aligned} \overline{B}_{T,F} &\leq \lim_{a\to\infty} \left[\frac{-\log \bar{G}(0)^{m^*-1}}{H(a)} + (n-m^*+1)\frac{\bar{H}(ua)}{H(a)}\right] \\ &= (n-m^*+1) \lim_{a\to\infty} \frac{\bar{H}(ua)}{\bar{H}(a)} = (n-m^*+1)u^\rho = (n-m^*+1). \end{aligned}$$ □

Next two propositions are needed for the proof of Theorem 3.

PROPOSITION A.4. *Suppose that $H(x)$ is the (two-sided) cumulative hazard function of the distribution of a random variable X. If X has a continuous distribution function and $H(x) \sim h(x)$ for $x \to \infty$, where $h(x)$ is a nonnegative, increasing function with derivative $h'(x)$, then $\mathbb{E}e^{(1-\varepsilon)h(|X|)} < \infty$ for any $\varepsilon > 0$.*

PROOF. Since $H(a) \sim h(a)$, there is U_ε such that

$$G(u) = e^{-H(u)} \leq e^{-(1-1/2\varepsilon)h(u)} \tag{A.3}$$

for any $u \geq U_\varepsilon$. We can choose U_ε large enough so that there is $u_\varepsilon \geq 0$ such that $e^{(1-\varepsilon)h(u_\varepsilon)} = U_\varepsilon$. Let $\bar{F}(x) = 1 - G(x)$ be the distribution function of $|X|$. Now we can write

$$\mathbb{E}e^{(1-\varepsilon)h(|X|)} = \int_0^{u_\varepsilon} e^{(1-\varepsilon)h(u)}\, d\bar{F}(u) + \int_{u_\varepsilon}^{\infty} e^{(1-\varepsilon)h(u)}\, d\bar{F}(u). \tag{A.4}$$

The first integrand in (A.4) is bounded by $e^{(1-\varepsilon)h(u_\varepsilon)} = U_\varepsilon$, since $h(x)$ is increasing. Using rules for computation with Stieltjes integrals, we obtain that the second integral is equal to

$$\begin{aligned}\int_{u_\varepsilon}^{\infty} e^{(1-\varepsilon)h(u)}\, d(-G(u)) &= G(u_\varepsilon)e^{(1-\varepsilon)h(u_\varepsilon)} - \lim_{u\to\infty} G(u)e^{(1-\varepsilon)h(u)} \\ &\quad + \int_{u_\varepsilon}^{\infty} G(u)e^{(1-\varepsilon)h(u)}(1-\varepsilon)h'(u)\, dx,\end{aligned} \tag{A.5}$$

integrating by parts—note that the continuity of F entails that of $\bar{F}$. For the last integral in (A.5),

$$\int_{u_\varepsilon}^{\infty} e^{-H(u)}e^{(1-\varepsilon)h(u)}(1-\varepsilon)h'(u)\, du \leq (1-\varepsilon)\int_{u_\varepsilon}^{\infty} e^{-1/2\varepsilon h(u)}h'(u)\, du < \infty,$$

by (A.3). The proof is finished by taking care of the limit in (A.5):

$$\lim_{u\to\infty} G(u)e^{(1-\varepsilon)h(u)} \leq \lim_{u\to\infty} e^{(1-\varepsilon)h(u)}e^{-(1-1/2\varepsilon)h(u)} = 0,$$

again by (A.3). □

PROPOSITION A.5. *Suppose that $H(x) \sim h(x)$, where $H(x)$ is the cumulative hazard function of the distribution of independent, identically distributed random variables $Y_1, Y_2, \ldots, Y_m$ and $h(x)$ satisfies the assumptions of Proposition 8 and is convex. If $\bar{Y} = \frac{1}{m}\sum_{i=1}^m Y_i$, then $\mathbb{E}e^{(1-\varepsilon)mh(|\bar{Y}|)} < \infty$ for any $\varepsilon > 0$.*

PROOF. By Jensen's inequality,

$$\begin{aligned}\mathbb{E}e^{(1-\varepsilon)mh(|\bar{Y}|)} &\leq \mathbb{E}e^{(1-\varepsilon)\sum_{i=1}^m h(|Y_i|)} \\ &= \mathbb{E}\prod_{i=1}^m e^{(1-\varepsilon)h(|Y_i|)} = (\mathbb{E}e^{(1-\varepsilon)h(|Y_1|)})^m < \infty,\end{aligned}$$

due to independence of the Y_i's and Proposition A.4. □

Next proposition is essentially Lemma 3.1 from Jurečková (1979). Since that contained some (for us) unnecessary assumptions, we rephrase it again.

PROPOSITION A.6. *Suppose that $h(a)$ is a non-decreasing function such that $H(a) \sim h(a)$. If the inequality $\mathbb{E}e^{bh(|T|)} < \infty$ holds for some $b > 0$, then $\underline{B}_{T,F} \geq b$.*

PROOF. By Markov's inequality,

$$P[|T| > a] \leq \frac{\mathbb{E}e^{bh(|T|)}}{e^{bh(a)}}.$$

Hence

$$-\log P[|T| > a] \geq -\log \mathbb{E}e^{bh(|T|)} + bh(a),$$

and therefore

$$\underline{B}_{T,F} \geq \lim_{a\to\infty} \frac{bh(a)}{H(a)} = b,$$

since the expectation is finite and $H(a) \to \infty$ as $a \to \infty$. □

PROOF OF THEOREM 3. Property (M) implies that

$$P[|T| > a] \leq \binom{n}{m} P[|\bar{Y}| > a],$$

where $\bar{Y} = \frac{1}{m}\sum_{i=1}^{m} Y_i$ is the average of m independent random variables $Y_1, Y_2, \ldots, Y_m$, with the distribution identical to that of X_i's. Hence by Propositions A.5 and A.6,

$$\underline{B}_{T,F} \geq -\lim_{a\to\infty}\left[\frac{\log\binom{n}{m}}{H(a)}\right] + \underline{B}_{\bar{Y},F} \geq m(1-\varepsilon),$$

for any $\varepsilon > 0$. □

PROOF OF PROPOSITION 3. Let $0 < \delta < 1/2$ be arbitrary. If $x_i \leq 0$ for $i = 1, 2, \ldots, m_0 = \lfloor (n-1)/2 \rfloor$, $A \leq x_i \leq (2-\delta)A$ for $i = m_0+1, m_0+2, \ldots, n-1$ and $(2-2\delta)A \leq x_n \leq (2-\delta)A$, then $T > (3-2\delta)A/2$. Writing $a = (3-2\delta)A/2$, $b = 2/(3-2\delta)$, $u = (4-4\delta)/(3-2\delta)$, and $v = (4-2\delta)/(3-2\delta)$, we have

$$P[|T| > a] \geq P[T > a] \geq F(0)^{m_0}(F(va) - F(ba))^{n-m_0-1}[F(va) - F(ua)].$$

Without loss of generality, we can assume that the right tail is dominant. By Propositions A.1 and A.3,

$$\begin{aligned}\overline{B}_{T,F} &\leq -\lim_{a\to\infty}\frac{m_0 \log F(0)}{H(a)} + (n-m_0-1)\lim_{a\to\infty}\frac{\bar{H}(ba)}{\bar{H}(a)} + \lim_{a\to\infty}\frac{\bar{H}(ua)}{\bar{H}(a)} \\ &= (n-m_0-1)b^\rho + u^\rho = \left\lfloor \frac{1}{2}n \right\rfloor b^\rho + u^\rho.\end{aligned}$$

Hence,

$$\overline{B}_{T,F} \leq \left\lfloor \frac{1}{2}n \right\rfloor \left(\frac{2}{3-\delta}\right)^\rho + \left(\frac{4-4\delta}{3-\delta}\right)^\rho.$$

Since δ was arbitrary, (3.1) follows. □

PROOF OF PROPOSITION 4. Let $m = \lfloor (n+1)/2 \rfloor$. Let $\tilde{T}_s = \tilde{T}_s(Y_1, Y_2, \ldots, Y_n)$ be an estimator defined through an equation

$$\sum_{i=1}^{n-m} \psi\left(\frac{Y_i - \tilde{T}_s}{s}\right) + \sum_{i=n-m+1}^{n} \tilde{\psi}\left(\frac{Y_i - \tilde{T}_s}{s}\right) = 0,$$

where $\tilde{\psi}(u) = \min\{x/k, 1\}$ (recall that $\psi(u) = \max\{-1, \tilde{\psi}(u)\}$). We prove first that

$$\tilde{T}_s(X_{(1)}, X_{(2)}, \ldots, X_{(n)}) = T_s. \tag{A.6}$$

To see this, note first that $\psi((X_{(n-m+1)} - T_s)/s) > -1$; if the contrary would hold, then $\psi((X_{(i)} - T_s)/s) = -1$ for all $i = 1, 2, \ldots, n-m+1$ (since ψ is non-decreasing)—but then $\lambda(T_s, s) < 0$. Thus, $\psi((X_{(n-m+1)} - T_s)/s) > -1$ and consequently (again by the monotonicity of ψ), $\psi((X_{(i)} - T_s)/s) > -1$ for all $i = n-m+1, \ldots, n$. Hence, $\tilde{T}_s$ is obtained through solving the same equation as T_s and A.6 holds.

Now, let $\bar{T}_s$ be an estimate satisfying an equation

$$\sum_{i=n-m+1}^{n} \tilde{\psi}\left(\frac{X_{(i)} - \bar{T}_s}{s}\right) = 0.$$

We claim that

$$\tilde{T}_s(X_{(1)}, X_{(2)}, \ldots, X_{(n)}) \le \tilde{T}_s(\bar{T}_s, \ldots, \bar{T}_s, X_{(n-m+1)}, \ldots, X_{(n)}) = \bar{T}_s. \tag{A.7}$$

The inequality follows from the fact that $\tilde{T}_s$ is non-decreasing in every variable, and that $\bar{T}_s$, as all M-estimators, satisfies (E): if none of the order statistics $X_{(1)}, X_{(2)}, \ldots, X_{(n-m)}$ exceeds $X_{(n-m+1)}$, then none of them exceeds $\bar{T}_s$ as well—hence the monotonicity of $\tilde{T}_s$ can be applied. To observe the equality in (A.7), just start with $X_{(n-m+1)}, \ldots, X_{(n)}$, compute $\bar{T}_s$, and then add m data points equal to $\bar{T}_s$: their contribution to the score function of $\tilde{T}_s$ is 0, hence both estimators are obtained through solving the same equation.

Finally, since $\tilde{\psi}(u) \le u/k$, the score function based on $\tilde{\psi}$ is less or equal to that based on u/k. Since both are non-increasing, the corresponding M-estimators satisfy an inequality

$$\bar{T}_s \le \frac{1}{m} \sum_{i=n-m+1}^{n} X_{(i)}. \tag{A.8}$$

The desired inequality then follows from (A.6), (A.7), and (A.8). □

The rest of this Appendix contains some additional facts connected with regular variation and tails of distributions. These can be useful in assessing and discussing the validity of various regularity conditions arising in tail-behavior investigations. We state them without proofs, which are available from the authors.

LEMMA A.1. *If $f(-a) = O(f(a))$ $[f(a) = O(f(-a))]$ as $a \to \infty$, then $G_2(a) = O(G_1(a))$ $[G_1(a) = O(G_2(a))$, respectively$]$.*

LEMMA A.2. (i) *If* $f(a)$ $[f(-a)]$ *is regularly varying with exponent* $\varkappa$ *for* $a \to \infty$, *then* $G_1(a)$ [*respectively*, $G_2(a)$] *is regularly varying with exponent* $\varkappa+1$. (ii) *If* $G(a)$ $[G_1(a), G_2(a)]$ *is regularly varying with exponent* $\varkappa$ *for* $a \to \infty$, *then* $H(a)$ $[H_1(a), H_2(a)$, *respectively*] *is regularly varying with exponent* $-\varkappa$.

The converse to Lemma A.2(ii) is in general not true (for the normal distribution, for instance).

LEMMA A.3. *If* $G(a)$ *is regularly varying for* $a \to \infty$ *with an exponent* $\varkappa < 0$, *then* (A4) *holds.*

It makes sense to say that (A4) holds for G: if $\limsup_{a\to\infty} G(va)/G(ua) < 1$ for any $v > u > 0$. We say that a tail is *strongly dominant* if $\lim_{a\to\infty} \bar{G}(a)/\tilde{G}(a)$ exists and is finite; here $\bar{G}$ is the one-sided tail function corresponding to a dominant tail and $\tilde{G}$ is the one-sided tail function corresponding to the other one. Strong dominance implies dominance. Lemma 2.1 of Jurečková (1981a) postulates the strong dominance with limit equal to 0 (compare with our Proposition A.1).

LEMMA A.4. *Suppose that* $v > u > 0$. (i) *If* (A4) *holds* (*for a dominant tail*), *then it holds for* G. (ii) *If* (A4) *holds for* G, *then* (A4) *holds if a dominant tail is strongly dominant.*

The last proposition shows that the regularity condition used in Theorem 2.1 of HJKP follows from (A3).

LEMMA A.5. *If* (A3) *holds, then* $H(a+c) \sim H(a)$ *for any* $c > 0$.

REFERENCES

Bassett, G. W., Jr. (1991). Equivariant, monotonic, 50% breakdown estimators, *Amer. Statist.*, **45**, 135–137.

Bingham, N. H., Goldie, C. M. and Teugels, J. L. (1987). *Regular Variation*, Cambridge University Press, Cambridge.

Croux, C. (1996). Maximum deviation curves for location estimators, *Statistics*, **28**, 295–305.

Donoho, D. L. and Huber, P. J. (1982). The notion of breakdown point, *Festschrift in Honor of Erich Lehmann*, (eds. K. Doksum and J. L. Hodges, Jr.), 157–183, Wadsworth, Belmont, California.

He, X., Jurečková, J., Koenker, R. and Portnoy, S. (1990). Tail-behavior of regression estimators and their breakdown points, *Econometrica*, **58**, 1195–1214.

Huber, P. J. (1981). *Robust Statistics*, Wiley, New York.

Jurečková, J. (1979). Finite-sample comparison of L-estimators of location, *Commentationes Mathematicae Universitatis Carolinae*, **20**, 509–518.

Jurečková, J. (1981a). Tail-behavior of location estimators, *Ann. Statist.*, **9**, 578–585.

Jurečková, J. (1981b). Tail-behavior of location estimators in nonregular cases, *Commentationes Mathematicae Universitatis Carolinae*, **22**, 365–375.

Jurečková, J. (1985). Tail-behavior of L-estimators and M-estimators, *Probability and Statistical Decision Theory, Vol. A (Bad Tatzmannsdorf, 1983)*, (eds. F. Konecny, J. Mogyoródi and W. Wertz), 205–217, Reidel, Dordrecht.

Mizera, I. (1994). On consistent M-estimators: tuning constants, unimodality and breakdown, *Kybernetika*, **30**, 289–300.

Resnick, S. I. (1987). *Extreme Values, Regular Variation and Point Processes*, Springer, New York.

Rousseeuw, P. J. and Leroy, A. (1987). *Robust Regression and Outlier Detection*, Wiley, New York.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*, Wiley, New York.