Asymptotic and Finite-Sample Properties in Statistical Estimation

Jana Jurečková

1 Introduction

Consider first the problem of estimating the shift parameter θ based on observations $X_1, \ldots, X_n$, distributed according to distribution function $F(x-\theta)$. Parallel problem consists of estimating the regression parameter in model $Y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + e_i, i = 1, \ldots, n$. Many estimators of θ are asymptotically normally distributed, which is proven with the aid of the central limit theorem. The word "central" is suitable, because it approximates well the central part, but less accurately the tails of the true distribution of the estimator. The leading idea of robust estimators was their assumed resistance to heavy-tailed distributions and to the gross errors. However, while they are often asymptotically normal, we can show that they themselves can be heavy-tailed for any finite n.

Another interesting fact is that though many estimators are asymptotically admissible with respect to quadratic or generally to convex risk functions, some of them are not finite-sample admissible for any distribution at all, and cannot be even Bayesian. This is true mainly for trimmed estimators, as the median, trimmed mean or the trimmed least squares estimator. Generally this is true for many estimators with bounded influence functions; cf. [6, 7].

If we do not know F exactly, we usually take recourse to robust estimators, less sensitive to the outlying observations and to the gross errors. Well-known are the classes of M-, L- and R-estimators, each of which containing elements, asymptotically normal and efficient for specific distributions. In the family of symmetric contaminated distributions, $\mathscr{F} = (1-\varepsilon)F + \varepsilon H, \ H \in \mathscr{H}$ with unimodal central distribution F, any of these classes contains an element with the mini-maximally optimal asymptotic variance over $\mathscr{F}$. Under a fixed F, we can

J. Jurečková (✉)
Faculty of Mathematics and Physics, Charles University in Prague, Prague, Czech Republic
e-mail: jurecko@karlin.mff.cuni.cz

D. Dawson et al. (eds.), *Asymptotic Laws and Methods in Stochastics*,
Fields Institute Communications 76, DOI 10.1007/978-1-4939-3076-0_20

obtain the M-, L- and R-estimators with identical influence functions by a suitable transformation (dependent on F) of the respective score (weight) function. However, the influence function characterizes the statistical functional rather than its finite-sample estimator, and the M-, L- and R-estimators can behave differently for finite n.

The asymptotic approach often stretches the truth; when the number of observations is finite, the distribution of a robust estimator is far from normal, and it inherits the tails from the parent distribution F. From this point of view, the estimator is non-robust. Our purpose in the present paper is to illustrate some distinctive differences between the asymptotic and finite-sample properties of robust estimators. We shall devote attention to the tail-behavior of M-estimators and of their one-step versions, and generally to the tail-behavior of equivariant estimators. Concerning the one-step version $T_n^{(1)}$ of estimator T_n, starting with an initial estimator $T_n^{(0)}$, it is interesting though not well known that while asymptotic properties of $T_n^{(1)}$ depend on those of non-iterated T_n, its finite-sample properties rather depend on the initial $T_n^{(0)}$. The finite-sample properties of an estimator depend on its finite sample distribution; we shall illustrate the exact finite-sample densities of some equivariant estimators. However, to calculate the density numerically requires a multiple numerical integration, for which a very good approximation is needed. We recommend the saddle-point approximation, which is very precise even for a very small n.

2 Tail-Behavior of Equivariant Estimators

2.1 Estimation of Shift Parameter, i.i.d. Observations

Let $X_1, \ldots, X_n$ be a random sample from an unknown distribution function $F(x-\theta)$, where F is absolutely continuous with positive density f. For the sake of identifiability of θ, assume that f is symmetric around 0, or another condition guaranteeing the identifiability. Suppose that F is heavy-tailed in the sense that

$$\lim_{x\to\infty} \frac{-\ln(1-F(x))}{m \ln x} = 1, \quad \text{for some } \ m > 0. \tag{1}$$

Then, for $x > 0$,

$$1 - F(x) = x^{-m} L(x) \tag{2}$$

where $L(x)$ is slowly varying at infinity, i.e. $\lim_{x\to\infty} \frac{L(ax)}{L(x)} = 1 \ \forall a > 0$.

For that, we should verify that $L_m(x) = x^m(1-F(x))$ is slowly varying at infinity. Indeed, for $x > 0$ and any $a > 0$ fixed, under (1)

jurecko@karlin.mff.cuni.cz

$$\ln\left(\frac{L_m(ax)}{L_m(x)}\right) = m\ \ln a + \ln(1-F(ax)) - \ln(1-F(x))$$
$$= m\ \ln a + \left(\frac{\ln(1-F(ax))}{m\ \ln(ax)}\right)\cdot m\ \ln(ax) - \left(\frac{\ln(1-F(x))}{m\ \ln x}\right)\cdot m\ \ln x \to 0$$

as $x \to \infty$, and it confirms (2). In that case F belongs to domain of attraction of the Fréchet distribution. Conversely, (2) implies (1).

Let $T_n = T_n(X_1, \ldots, X_n)$ be a translation equivariant estimator of θ, further satisfying the following natural condition:

$$\min_{1\le i\le n} X_i > 0 \Rightarrow T_n(\mathbf{X}) > 0, \quad \max_{1\le i\le n} X_i < 0 \Rightarrow T_n(\mathbf{X}) < 0. \tag{3}$$

Tail-behavior of T_n can be characterized by means of a measure proposed in [3]:

$$B(a, T_n) = \frac{-\ln P_\theta(|T_n - \theta| \ge a)}{-\ln(1-F(a))} = \frac{-\ln P_0(|T_n| \ge a)}{-\ln(1-F(a))} \tag{4}$$

and its values for $a \gg 0$. If T_n satisfies (3), then under any fixed n

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

(see [3] for the proof). Particularly, if $\lim_{a\to\infty} B(a, T_n) = \lambda_n > 0$ and F is heavy-tailed with tail index m, then

$$P_0(T_n \ge a) = a^{-m\lambda_n} L_1(a),\ L_1 \text{ slowly varying at infinity,}$$

hence T_n is also heavy-tailed. Specifically, it applies also to median $\tilde{X}_n$ and to the M-estimator M_n with bounded ψ-function, where $\lambda_n = \frac{n}{2}$. It means that $\tilde{X}_n$ and M_n are heavy-tailed with the tail index $\frac{mn}{2}$. It is finite for every n, though increasing with n, which classifies the distribution of these estimates as heavy-tailed for any finite n. The distribution of estimates is light-tailed (normally, exponentially tailed) only under $n = \infty$. The sample mean $\bar{X}_n$ has $\lambda_n \equiv 1$; thus $\bar{X}_n$ is heavy-tailed with the tail index m for any $n < \infty$.

2.2 *Estimation of Shift Parameter, Non-identically Distributed Observations*

Let us now consider the case where the $X_i,\ i = 1, \ldots, n$ are independent, but non-identically distributed, X_i having continuous distribution function $F_i(x-\theta)$, symmetric around θ, and heavy-tailed in the sense that

$$1 - F_i(x) = x^{-m_i} L_i(x),\ 0 < m_i < \infty,\ L_i \text{ slowly varying at infinity, } i = 1, \ldots, n.$$

Denote

$$m_* = \min\{m_i,\ 1 \le i \le n\} \quad m^* = \max\{m_i,\ 1 \le i \le n\}.$$

If we are not aware of the difference between $F_1, \dots, F_n$, we automatically use an equivariant estimate T_n satisfying (3) as before. Then even its tail behavior cannot be exponentially-tailed. In fact, as proven in [8],

$$a^{-m^*} L(a) \le P_\theta(T_n - \theta > a) \le a^{-m_*} L(a) \ \text{ for } \ a > a_0,$$

where $L(\cdot)$ is slowly varying at infinity. Particularly, if $X_1, \dots, X_n$ are heteroscedastic in the sense that $F_i(x) \equiv F(x/\sigma_i),\ i = 1, \dots, n$, then $m_1, \dots, m_n$ coincide. Hence, the heteroscedasticity does not affect the tail index of T_n, which is always equal to m.

2.3 Estimation of Regression Parameter

Consider the linear model $\mathbf{Y}_n = \mathbf{X}_n \boldsymbol{\beta} + \mathbf{e}_n$ with a fixed (nonrandom) design matrix $\mathbf{X}_n$ of order $n \times p$ and of rank p, with the rows $\mathbf{x}_i^\top,\ i = 1, \dots, n$. The vector of errors $\mathbf{e}_n$ consists of n independent components, identically distributed with a symmetric distribution function F such that $0 < F(z) < 1,\ z \in \mathbb{R}^1$. Let $\mathbf{T}_n$ be an estimator of $\boldsymbol{\beta}$, regression equivariant in the sense

$$\mathbf{T}_n(\mathbf{Y} + \mathbf{X}\mathbf{b}) = \mathbf{T}_n(\mathbf{Y}) + \mathbf{b}, \quad \forall \mathbf{b} \in \mathbb{R}^p.$$

He et al. [2] extended the tail measure (4) to $\mathbf{T}_n$ in the linear model in the following way:

$$B(a, \mathbf{T}_n) = \frac{-\ln\ P\left(\max_i |\mathbf{x}_i^\top(\mathbf{T}_n - \boldsymbol{\beta})| > a\right)}{-\ln\ (1 - F(a))},\ a \gg 0. \tag{5}$$

The same authors showed that if there exists at least one non-positive and one non-negative residual $r_i = Y_i - \mathbf{x}_i^\top \mathbf{T}_n$, then $\limsup_{a\to\infty} B(a, \mathbf{T}_n) \le n$. The properties of this measure were further studied by Mizera and Müller [12] and Portnoy and Jurečková [13], and this measure was extended to multivariate models by Zuo ([15, 16] and [17]). Jurečková, Koenker and Portnoy [11] studied the tail behavior of the least-squares estimator with random (possibly heavy-tailed) matrix $\mathbf{X}$.

It is traditionally claimed that robust estimators are insensitive to outliers in $\mathbf{Y}$ and to heavy-tailed distributions of model errors. However, we can show that an equivariant estimator $\mathbf{T}_n$ in the linear model is still heavy-tailed for any finite n provided the distribution function F is heavy-tailed, even if $\mathbf{X}$ is non-random. More precisely, if $\mathbf{T}_n$ is a regression equivariant estimator of $\boldsymbol{\beta}$ such that there exists at least one non-negative and one non-positive residual $r_i = Y_i - \mathbf{x}_i^\top \mathbf{T}_n,\ i = 1, \ldots, n,$ then

$$P_\beta\left(\|\mathbf{T}_n - \boldsymbol{\beta}\| > a\right) \geq a^{-m(n+1)} L(a)$$

where $L(\cdot)$ is slowly varying at infinity. Hence, the distribution of $\|\mathbf{T}_n - \boldsymbol{\beta}\|$ is heavy-tailed under every finite n (see [8] for the proof).

2.4 Tail-Behavior of M-Estimator of Regression Parameter

The class of M-estimators defined as

$$\mathbf{T}_n = \arg\min_{\mathbf{b} \in \mathbb{R}^p} \left\{ \sum_{i=1}^n \rho(Y_i - \mathbf{x}_i^\top \mathbf{b}) \right\}$$

covers the Huber estimator and some redescending M-estimators. Assume that F is symmetric with nondegenerate tails (heavy or light) and such that

$$\lim_{a \to \infty} \frac{-\ln(1 - F(a + c))}{-\ln(1 - F(a))} = 1 \quad \text{for } \forall c > 0.$$

Following [12], we suppose that ρ satisfies the conditions (discussed in [12] in detail):

(i) ρ is absolutely continuous, nondecreasing on $[0, \infty),\ \rho(z) \geq 0,$ $\rho(z) = \rho(-z),\ z \in \mathbb{R}^1.$
(ii) $\rho(z)$ is unbounded and its derivative $\psi(z)$ is bounded for $z \in \mathbb{R}^1.$
(iii) ρ is subadditive in the sense that there exists $L > 0$ such that $\rho(z_1 + z_2) \leq \rho(z_1) + \rho(z_2) + L$ for $z_1, z_2 \geq 0.$

Define

$$\begin{aligned} m_* &= m_*(n, \mathbf{X}, \rho) \\ &= \min\left\{ \text{card}\, \mathscr{M} :\ \sum_{i \in \mathscr{M}} \rho(\mathbf{x}_i^\top \mathbf{b}) \geq \sum_{i \notin \mathscr{M}} \rho(\mathbf{x}_i^\top \mathbf{b}) \text{ for some } \mathbf{b} \neq \mathbf{0} \right\} \end{aligned}$$

where $\mathscr{M}$ runs over subsets of $\mathscr{N} = \{1, 2, \ldots, n\}$. Then it is proven in [5] that

$$\liminf_{a \to \infty} B(a, \mathbf{T}_n) \geq m_*.$$

It means that m_* is the lower bound for the tail behavior of M-estimator generated by ρ and it coincides with the lower bound derived in [12] for the finite-sample breakdown point of the M-estimator $\mathbf{T}_n$.

3 One-Step Version of an Estimator, Its Tail-Behavior and Breakdown Point

A broad class of estimators $\mathbf{T}_n$ of $\boldsymbol{\beta}$ admit a representation

$$\mathbf{T}_n(\mathbf{Y}) = \boldsymbol{\beta} + \frac{1}{\gamma}(\mathbf{X}_n^\top \mathbf{X}_n)^{-1} \sum_{i=1}^{n} \mathbf{x}_i \psi(Y_i - \mathbf{x}_i^\top \boldsymbol{\beta}) + \mathbf{R}_n,$$
$$\|\mathbf{R}_n\| = o_p(\|\mathbf{X}_n^\top \mathbf{X}_n\|^{-1/2}) \tag{6}$$

with a suitable function ψ and a functional $\gamma = \gamma(\psi, F)$.

The one-step version of $\mathbf{T}_n$ is defined as the one-step Newton-Raphson iteration of the system of equations $\sum_{i=1}^n \mathbf{x}_i \psi(Y_i - \mathbf{x}_i^\top \mathbf{b}) = \mathbf{0}$, even when the estimator is not a root of this system (as in the case of L_1-estimator or of other M-estimators with discontinuous ψ).

Let us start with a consistent initial estimator $\mathbf{T}_n^{(0)}$ of $\boldsymbol{\beta}$, satisfying $n^{1/2}(\mathbf{T}_n^{(0)} - \boldsymbol{\beta}) = O_p(1)$. The one-step version of $\mathbf{T}_n$ is defined as

$$\mathbf{T}_n^{(1)} = \begin{cases} \mathbf{T}_n^{(0)} + \frac{1}{n\hat{\gamma}_n}(\mathbf{Q}_n^*)^{-1} \sum_{i=1}^n \mathbf{x}_i \psi(Y_i - \mathbf{x}_i^\top \mathbf{T}_n^{(0)}) \ \ldots \ \text{if } \hat{\gamma}_n \neq 0 \\ \mathbf{T}_n^{(0)} \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \ldots \ \text{otherwise} \end{cases}$$

where $\mathbf{Q}_n^* = n^{-1}\mathbf{X}_n^\top \mathbf{X}_n$. The two-step or the k-step versions of $\mathbf{T}_n$ are defined analogously for $k = 2, 3, \ldots$. Here we assume that $\gamma \neq 0$ and that $\hat{\gamma}_n$ is a consistent estimator of γ such that $1 - (\gamma/\hat{\gamma}_n) = O_p(n^{-1/2})$. For possible regression invariant estimates of γ we refer the reader to [9].

While the asymptotic properties of $\mathbf{T}_n^{(1)}$ depend on those of the non-iterated estimator $\mathbf{T}_n$, its finite-sample breakdown point depends on that of initial $\mathbf{T}_n^{(0)}$ (see [13]). There is a conjecture that even more finite sample properties of $\mathbf{T}_n^{(1)}$ depend solely on the initial estimator. We shall illustrate this phenomenon at least in the special case of location model:

3.1 *One-Step Version in the Location Model*

Let T_n be an equivariant estimator of a location parameter and $T_n^{(0)}$ be an equivariant initial estimator. Consider a modified one-step version of T_n :

$$T_n^{(1)} = \begin{cases} T_n^{(0)} + \hat{\gamma}_n^{-1} W_n \ \ldots \ \text{if } |\hat{\gamma}_n^{-1} W_n| \leq c, \ 0 < c < \infty \\ T_n^{(0)} \qquad\qquad\quad \ldots \ \text{otherwise} \end{cases}$$

where $W_n = n^{-1}\sum_{i=1}^{n}\psi(Y_i - T_n^{(0)}) = O_p(n^{-1/2})$. Then $T_n^{(1)} - T_n = o_p(n^{-1/2})$ and $T_n^{(1)}$ is also equivariant. Surprisingly, the tail behavior of $T_n^{(1)}$ and of $T_n^{(k)}$ depends more on that of $T_n^{(0)}$ than on the tail-behavior of non-iterative T_n. The following theorem is proven in [5]:

Theorem 1. *Let $Y_1, \ldots, Y_n$ be a sample from a population with distribution function $F(y - \theta)$, F symmetric and increasing on the set $\{x : 0 < F(x) < 1\}$. Let T_n be an equivariant estimator of θ admitting the representation*

$$T_n(\mathbf{Y}) = \theta + \frac{1}{n\gamma}\sum_{i=1}^{n}\psi(Y_i - \theta) + R_n,\ R_n = o_p(n^{-1/2})$$

with a bounded skew-symmetric non-decreasing ψ. Then, for $k = 1, 2, \ldots$

$$\liminf_{a\to\infty} B(T_n^{(0)}, a) \le \liminf_{a\to\infty} B(T_n^{(k)}, a) \le \limsup_{a\to\infty} B(T_n^{(k)}, a) \le \limsup_{a\to\infty} B(T_n^{(0)}, a).$$

Example 1. (i) Let $T_n^{(0)} = \tilde{X}_n$ be the sample median, n odd. Let T_n be an equivariant estimator and $T_n^{(k)}$ its k-step version starting with $\tilde{X}_n$. Then, under the conditions of Theorem 1,

$$\lim_{a\to\infty} B(T_n^{(k)}, a) = \frac{n+1}{2} \qquad \text{for} \quad k = 1, 2, \ldots.$$

(ii) Let $T_n^{(0)} = \bar{X}_n$ be the sample mean. Let T_n be an equivariant estimator and $T_n^{(k)}$ its k-step version starting with $\bar{X}_n$. Then, under the conditions of Theorem 1,

$$\lim_{a\to\infty} B(T_n^{(k)}, a) = \begin{cases} n & \text{if } F \text{ is of type I (exponentially tailed)} \\ 1 & \text{if } F \text{ is of type II (heavy tailed)} \end{cases}$$

for $k = 1, 2, \ldots$, where the types I or II of F mean that its tails satisfy

$$\lim_{a\to\infty} \frac{-\ln(1 - F(a))}{ba^r} = 1, \quad b > 0, \quad r \ge 1$$
$$\lim_{a\to\infty} \frac{-\ln(1 - F(a))}{m\ln a} = 1, \quad m > 0,$$

respectively (see [3] for more details).

4 Finite-Sample Density of Equivariant Estimators

The finite-sample properties of estimator T_n, including the moments, depend on its entire scope, not only on its central part. The finite sample density can be sometimes derived, though it does not have a simple form. For instance, let $X_1, \ldots, X_n$ be a sample from the distribution with distribution function $F(x - \theta)$ where F has a continuously differentiable density f and finite Fisher information. Denote by $g_\theta(t)$ the density of a translation equivariant estimator T_n of θ. Then (see [10])

$$\begin{aligned} g_\theta(t) &= \int_{T(x_1,\ldots,x_n)\le t} \cdots \int \sum_{i=1}^{n} \frac{f'(x_i-\theta)}{f(x_i-\theta)} \prod_{k=1}^{n} f(x_k-\theta)dx_1\ldots dx_n \\ &= \mathrm{E}_0\left\{\sum_{i=1}^{n} \frac{f'(X_i)}{f(X_i)}\, I\Big[T(X_1,\ldots,X_n) \le t-\theta\Big]\right\}. \end{aligned}$$

If T_n is a solution of the equation $\sum_{i=1}^{n} \psi(X_i - t) = 0$ with monotone ψ, then $g_\theta(t)$ can be rewritten as

$$g_\theta(t) = \mathrm{E}_0\left\{\sum_{i=1}^{n} \frac{f'(X_i)}{f(X_i)}\, I\Big[\sum_{j=1}^{n} \psi(X_j - (t-\theta)) \le 0\Big]\right\}.$$

To calculate it numerically means an n-fold integration, and we recommend to use a saddle point approximation as it is more precise.

This density is numerically compared in [10] with its saddle-point approximation, developed in [1], for the Huber and maximum likelihood estimators, and for various parent distributions, including the Cauchy. The numerical comparisons demonstrate that the saddle-point approximations are very precise even for small sample sizes, and thus can be recommended in applications. A similar approach applies to the density of a regression quantile, derived in [4], and its saddle-point approximation, computed in [14].

Acknowledgements Research was supported by the Grant GAČR 15-00243S. The author would like to thank the Editor for organizing the volume, and the Referee for his/her valuable comments.

References

1. Field, C.A., Ronchetti, E.: Small Sample Asymptotics. Lecture Notes–Monograph Series. Institute of Mathematical Statistics, Hayward (1990)
2. He, X., Jurečková, J., Koenker, R., Portnoy, S.: Tail behavior of regression estimators and their breakdown points. Econometrica **58**, 1195–1214 (1990)
3. Jurečková, J.: Tail behavior of location estimators. Ann. Stat. **9**, 578–585 (1981)

4. Jurečková, J.: Finite sample distribution of regression quantiles. Stat. Probab. Lett. **80**, 1940–1946 (2010)
5. Jurečková, J.: Tail-behavior of estimators and of their one-step versions. Journal de la Société Francaise de Statistique **153/1**, 44–51 (2012)
6. Jurečková, J., Klebanov, L.B.: Inadmissibility of robust estimators with respect to L1 norm. In: Dodge, Y. (ed.) L_1-Statistical Procedures and Related Topics. Lecture Notes–Monographs Series, vol. 31, pp. 71–78. Institute of Mathematical Statistics, Hayward (1997)
7. Jurečková, J., Klebanov, L.B.: Trimmed, Bayesian and admissible estimators. Stat. Probab. Lett. **42**, 47–51 (1998)
8. Jurečková, J., Picek, J.: Finite-sample behavior of robust estimators. In: Chen, S., Mastorakis, N., Rivas-Echeverria, F., Mladenov, V. (eds.) Recent Researches in Instrumentation, Measurement, Circuits and Systems pp. 15–20. ISBN: 978-960-474-282-0. ISSN: 1792-8575
9. Jurečková, J., Portnoy, S.: Asymptotics for one-step M-estimators in regression with application to combining efficiency and high breakdown point. Commun. Stat. A **16**, 2187–2199 (1987)
10. Jurečková, J., Sabolová, R.: Finite-sample density and its small sample asymptotic approximation. Stat. Probab. Lett. **81**, 1311–1318 (2011)
11. Jurečková, J., Koenker, R., Portnoy, S.: Tail behavior of the least-squares estimator. Stat. Probab. Lett. **55**, 377–384 (2001)
12. Mizera, I., Mueller, C.H.: Breakdown points and variation exponents of robust M-estimators in linear models. Ann. Stat. **27**, 1164–1177 (1999)
13. Portnoy, S., Jurečková, J.: On extreme regression quantiles. Extremes **2**(3), 227–243 (1999)
14. Sabolová, R.: Small sample inference for regression quantiles. KPMS Preprint 70, Charles University in Prague (2012)
15. Zuo, Y.: Finite sample tail behavior of the multivariate trimmed mean based on Tukey-Donoho halfspace depth. Metrika **52**, 69–75 (2000)
16. Zuo, Y.: Finite sample tail behavior of Hodges-Lehmann type estimators. Statistics **35**, 557–568 (2001)
17. Zuo, Y.: Finite sample tail behavior of multivariate location estimators. J. Multivar. Anal. **85**, 91–105 (2003)

jurecko@karlin.mff.cuni.cz