The truncated mean of an asymmetric distribution

A. Marazzi, C. Ruffieux

Institut universitaire de médecine sociale et préventive,
University of Lausanne, Bugnon 17, CH 1005 Lausanne, Switzerland

---

Abstract

This paper investigates a simple procedure to estimate robustly the mean of an asymmetric distribution. The procedure removes the observations which are larger or smaller than certain limits and takes the arithmetic mean of the remaining observations, the limits being determined with the help of a parametric model, e.g., the Gamma, the Weibull or the Lognormal distribution. The breakdown point, the influence function, the (asymptotic) variance, and the contamination bias of this estimator are explored and compared numerically with those of competing estimates.

Keywords. Asymmetric distributions, robust estimates, trimming, M-estimates.

---

1. Introduction

Positive random variables with asymmetric distributions arise in many statistical applications and often the distribution mean μ (i.e., the expected value of the random variable) is the parameter of interest. Unfortunately, the mean is a difficult parameter to estimate well: the sample mean, which is the natural estimate, is very nonrobust. Various rules for removing discrepant values are used in practical applications in order to improve the stability of the sample mean.

Often, observations which are larger or smaller than certain limits are removed, and the arithmetic mean of the remaining observations is computed. However, the questions of how to define the limits and of how good these estimators are, have not yet been answered. For example, rules of the following kind are routinely used for the analysis of hospital length of stay (Beguin et al., 1991):

(a) the mean after “trimming” observations beyond a certain quantile of the empirical distribution;
(b) the mean after removing observations beyond k_1 interquartile ranges from the median (e.g., $k_1 = 1.5$) on log data;
(c) the mean after removing observations beyond k_2 median absolute deviations from the median (e.g., $k_2 = 3$) on log data.

Length of stay (LOS) is an important indicator of hospital resource consumption and LOS means of “medically homogeneous groups of patients” are used for budgeting, the cost of such a group patients being proportional to its mean LOS. Unfortunately, due to the definition of the limits, rule (a) is a biased estimate of the population mean, and rules (b) and (c) are biased estimates of the hypothetical lognormal mean. By increasing the proportion of removed data (e.g., decreasing k_1 and k_2), the three rules tend to the median which is a very robust but inadequate estimate for the purpose of budgeting.

This paper addresses the question of defining the limits using an approach based on four steps. First, a parametric model is fitted to the data, so that the mean of the fitted model is a robust and Fisher consistent estimate of μ (i.e., an asympotically unbiased estimate, if the model is the true distribution). Second, an upper truncation limit T_u is defined as the u-quantile of the fitted model, where $u \in (0.5, 1)$ is a user chosen number (i.e., a "tuning constant"). Third, a lower limit T_l is determined so that the mean of the truncated model coincides with the mean of the entire model. Finally, the arithmetic mean of the observations which are larger than T_l and smaller than T_u is computed. The result is a Fisher consistent robust estimate of μ. We call it a *truncated mean* (to distinguish it from the "trimmed mean" defined in case (a), above, cf. Hampel et al. 1986, p. 99).

In the first step we suppose that a monotone transformation of the random variable can be described by means of a location-scale or a shape-scale model. This situation often occurs in practice and many popular models are of this kind. For example, the Lognormal and the Weibull distributions are shape-scale or location-scale models depending on the transformation and the parametrization; the Gamma distribution and the Pareto distribution of first and second kind (Johnson et al., 1994, p. 574) are shape-scale models.

We base the fitting procedure of the first step on two equations relating the empirical and the model expressions of a measure of location and a measure of dispersion with a high breakdown point, e.g., the median and the median absolute deviation, and call their solution a *LD-estimate*. The truncated mean inherits the breakdown point of the LD-estimate but can have a much higher efficiency. In important special cases, an approximate solution of the two equations can be computed without iterations.

More sophisticated estimators (M-estimators) for general parametric models have been proposed in Hampel et al. (1986); recently some of them (the B_s- and the B_s^p-estimates) have been applied to the Lognormal, the Gamma, and the Weibull distributions (Victoria-Feser and Ronchetti, 1994; Marazzi and Randriamiharisoa, 1997a-c). With appropriate tuning constants, M-estimates are extremely efficient; their breakdown point can be made very high by means of redescending weight systems for the likelihood scores (redescending M-estimates, Hampel et al., 1986). The computation of M-estimates requires, however, rather complex iterative algorithms, the convergence of which cannot be guaranteed in some important cases (Marazzi and Ruffieux, 1996). Although iterations can be reduced to a single step (one-step M-estimates), the implementation of M-estimates is very laborious; we are not aware of any implementation of redescending M-estimates for genuinely asymmetric distributions. Thus, despite the optimality properties of M-estimates, many users prefer simpler trimming rules, like those mentioned above, although they have clear flaws and their performance has never been assessed.

By means of the truncated mean, this paper improves rules (b) and (c) in the Lognormal case and extends them to other models. It then compares the truncated mean to the B_s^p- and to a redescending one-step M-estimate of μ based on Tukey's biweight. Using the influence functions, the asymptotic relative efficiencies of all estimates with respect to the maximum likelihood estimates of μ is computed and used to determine the tuning constants. With the help of simulations, it is shown that the truncated mean is comparable to the competing estimates with respect to standard error, breakdown point, and contamination bias over a wide range of situations of practical interest.

Section 2 defines the LD-, the truncated mean-, the B_s^p-, and the one-step M-estimates used in the comparisons. The influence functions are given in Section 3. Section 4 discusses some computational aspects. Section 5 reports numerical results that are discussed in Section 6. S-plus functions are made available at Internet address www.hospvd.ch/iumsp.

REMARK. Transformations are often used to make the distribution (closely) symmetric; many robust procedures (based on the Gaussian model) are then available to estimate the transformed mean. Unfortunately the original mean cannot usually be estimated by transforming back the transformed mean. (Moreover, symmetrizing transformations do not always exist.) For example, to estimate a lognormal mean both the estimates of normal mean and scale are required; in addition, the distinction between main (mean) and nuisance (scale) parameter – that characterizes most location and scale procedures – is not appropriate anymore. Therefore, robust mean estimates for genuinely asymmetric models are necessary.

2. Estimates

Let $X > 0$ be a random variable with unknown cumulative distribution function G and (asymmetric) density g. We write $X \sim G$. We are interested in estimating $\mu(G)$, the expected value of X, using n independent observations $x_1, \ldots, x_n$. We suppose that, as it often occurs in practice, we are willing to use a parametric model $F_{\alpha,\sigma}$ for the distribution $F = G \circ h^{-1}$ (with density f) of some monotone increasing transformation $Y = h(X)$. The corresponding model for G is $G_{\alpha,\sigma}$ and define $\mu = \mu(G_{\alpha,\sigma})$. We assume that:

(a) σ is a scale parameter of $F_{\alpha,\sigma}$;

(b) either α is a location parameter or α is a shape parameter of $F_{\alpha,\sigma}$;

(c) α and σ can be characterized by means of the location and dispersion measures $m(F)$ and $s(F)$ defined below (Section 2.1), i.e., for given values of $m(F)$ and $s(F)$, α and σ can be obtained by solving

$$m(F_{\alpha,\sigma}) = m(F) \qquad \text{and} \qquad s(F_{\alpha,\sigma}) = s(F).$$

In the following, $F_n(y) = (1/n) \sum \Delta_{y_i}(y)$ denotes the empirical distribution function of Y, Δ_y stands for the pointmass at y (i.e., $\Delta_y(x) = 0$ if $x < y$, 1 otherwise), and $y_{[1]} \leq y_{[2]} \leq \ldots \leq y_{[n]}$ denote the Y-order statistics. Similar notations (G_n, $x_{[i]}$, etc.) are used in relation with X. We use the abbreviation $\theta = (\alpha, \sigma)^{\mathrm{T}}$ (column vector) and denote f_θ and g_θ the model densities. The inverse of any distribution function is defined in the usual way as $F^{-1}(s) = \inf\{y|F(y) \geq s\}$ and $q_u(F) = F^{-1}(u)$ denotes the u-quantile of F. The notation $\lfloor r \rfloor$ is used for the largest integer smaller than r.

This section gives brief definitions of the estimates, including the B_s^p- and a one-step M-estimate for the reader's convenience.

2.1 The LD-estimates

We assume that $m(F_n)$ and $s(F_n)$ are robust asymptotically unbiased estimates of $m(F)$ and $s(F)$. If α is a location parameter, let $(\tilde{\alpha}, \tilde{\sigma})$ be a solution of:

$$\tilde{\sigma} m(F_{0,1}) + \tilde{\alpha} = m(F_n), \qquad \tilde{\sigma} s(F_{0,1}) = s(F_n), \tag{2.1.1}$$

If α is a shape parameter, let $(\tilde{\alpha}, \tilde{\sigma})$ be a solution of:

$$\tilde{\sigma} m(F_{\tilde{\alpha},1}) = m(F_n), \qquad \tilde{\sigma} s(F_{\tilde{\alpha},1}) = s(F_n). \tag{2.1.2}$$

In both cases, we call $\tilde{\theta} = (\tilde{\alpha}, \tilde{\sigma})^{\mathrm{T}}$ the *LD-estimate* of θ ("Location-Dispersion" based estimate). The corresponding estimate of $\mu(G)$ is $\tilde{\mu} = \mu(G_{\tilde{\alpha},\tilde{\sigma}})$. We will consider the following special cases:

- *The* LD_R*-estimate* defined by

$$m(F_n) = \mathrm{med}(y_i), \qquad s(F_n) = F_n^{-1}(3/4) - F_n^{-1}(1/4), \tag{2.1.3}$$

  i.e., s is the interquartile range estimate of scale.

- *The* LD_D*-estimate* defined by

$$m(F_n) = \frac{1}{i_u + r_u - i_\ell - r_\ell} \sum_{i=i_\ell+2}^{i_u} y_{[i]} + r_u y_{[i_u+1]} + (1 - r_\ell) y_{[i_\ell+1]}, \tag{2.1.5}$$

$$s(F_n) = \frac{1}{j_u + s_u - j_\ell - s_\ell} \sum_{j=j_\ell+2}^{j_u} z_{[j]} + s_u z_{[j_u+1]} + (1 - s_\ell) z_{[j_\ell+1]}, \tag{2.1.6}$$

  where $i_u = \lfloor n(1-\beta) \rfloor$, $i_\ell = \lfloor n\beta \rfloor$, $r_u = n(1-\beta) - i_u$, $r_\ell = n\beta - i_\ell$, $j_u = \lfloor n(1-\gamma) \rfloor$, $j_\ell = \lfloor n\gamma \rfloor$, $s_u = n(1-\gamma) - i_u$, $s_\ell = n\gamma - i_\ell$, and $z_i = |y_i - m(F_n)|$. In words, m is the β-trimmed mean and s the γ*-trimmed (mean) absolute deviation*. The constants $\beta \in [0, 0.5]$ and $\gamma \in [0, 0.5]$ are user chosen (tuning constants). For $\beta = \gamma = 0.5$, m is the median and s the median absolute deviation. For β and γ close to 0.5, m and s are "smoothed" versions of these estimates. In our examples, we will use $\beta = \gamma = 0.4$.

REMARK. Obviously, there are other choices of m and s. An interesting one is, for example, $m(F_n) = \mathrm{med}(y_i)$ and $s(F_n) = \{|y_i - y_j|,\ i < j\}_{(k)}$ where $k = \binom{h}{2} \approx \binom{n}{2}/4$ and $h = \lceil \frac{n}{2} \rceil + 1$ is roughly half the number of observations, i.e., s is the Q_n scale estimate defined in Rousseeuw and Croux (1993). For the sake of brevity, we do not explore this option here.

2.2 The truncated mean estimate

Assume that $\bar{\theta}$ is a given Fisher consistent estimate of θ. We take $\bar{\theta} = \tilde{\theta}$, the LD-estimate defined above (although other choices are possible). Let $u \in (0.5, 1)$ be a user chosen number and let $T_u = q_u(G_{\bar{\theta}})$. Define T_l so that

$$\frac{1}{u - G_{\bar{\theta}}(T_l)} \int_{T_l}^{T_u} x dG_{\bar{\theta}}(x) = \int x dG_{\bar{\theta}}(x). \tag{2.2.1}$$

The *truncated mean estimate* of μ is then defined as:

$$\check{\mu}(G_n) = \mathrm{ave}\{x_i \mid T_l < x_i \le T_u\}, \tag{2.2.2}$$

i.e., the arithmetic mean of the x_i such that $T_l < x_i \le T_u$.

REMARK 1. The truncated mean estimates can be generalized as follows. Let a and b $(a < b)$ be two user chosen constants and let $w(x, a, b)$ be a weight function such that $w(x, a, b) > 0$ for $x \in [a, b]$ and 0 otherwise. For example, $w(x, a, b) = \psi_{b-a}(x - a) I_{\{x > a\}}$, where

$$\psi_k(x) = \left(\frac{6}{k}\right) \left(\frac{x}{k}\right) \max\left(1 - \left(\frac{x}{k}\right)^2, 0\right)^2 \tag{2.2.3}$$

is the rescaled Tukey's biweight. Let $u \in (0.5, 1)$ be a user chosen number and let $T_u = q_u(G_{\bar{\theta}})$. Define T_l so that

$$\frac{1}{\int w(x, T_l, T_u) dG_{\bar{\theta}}(x)} \int x w(x, T_l, T_u) dG_{\bar{\theta}}(x) = \int x dF_{\bar{\theta}}(x).$$

The *weighted mean estimate* of μ is then defined as:

$$\breve{\mu}(G_n) = \frac{1}{\sum_i w(x_i, T_l, T_u)} \sum_i x_i w(x_i, T_l, T_u).$$

Although the weighted mean improves certain properties (e.g., the local shift sensitivity, Hampel et al., 1986) of the truncated mean, it will not be considered further because it increases the complexity of computations.

REMARK 2. The "Huber-type skipped mean" (reject everything which is more than 5.2 median deviations away from the median and take the mean of the reminder, cf. Hampel et al., 1986, p.65) is a special case of truncated mean for the Gaussian distribution.

2.3 The B^p_s-estimate

Let $\mathcal{S}(y, \theta)$ denote the score vector function of F_θ, i.e., $\mathcal{S}(y, \theta) = \partial \ln f_\theta(y) / \partial \theta$ and let $\underline{b} = (b_1, b_2) \in \mathbb{R}^2$ a user chosen tuning constant. The *B^p_s-estimate* $\hat{\theta}$ of θ is then defined as a solution of

$$\sum_{i=1}^{n} \Psi_{\underline{b}} \left[A_{\underline{b}}(\theta)(\mathcal{S}(y_i, \theta) - C_{\underline{b}}(\theta)) \right] = 0, \tag{2.3.1}$$

where the function $\Psi_{\underline{b}}(z)$, $z = (z_1, z_2) \in \mathbb{R}^2$, is defined by $\Psi_{\underline{b}}(z) = (H_{b_1}(z_1), H_{b_2}(z_2))^{\mathrm{T}}$, and $H_b(y) = \min[b, \max(y, -b)]$ denotes the Huber function. Moreover, $A_{\underline{b}}(\theta)$ is a 2×2 nonsingular lower triangular matrix and $C_{\underline{b}}(\theta)$ is a 2-component vector; they are both functions of θ and are defined jointly and implicitly by the following equations:

$$\int \Psi_{\underline{b}} \left[A_{\underline{b}}(\theta)(\mathcal{S}(y, \theta) - C_{\underline{b}}(\theta)) \right] \Psi_{\underline{b}} \left[A_{\underline{b}}(\theta)(\mathcal{S}(y, \theta) - C_{\underline{b}}(\theta)) \right]^{\mathrm{T}} f_\theta(y) dy = I, \tag{2.3.2}$$

$$\int \Psi_{\underline{b}} \left[A_{\underline{b}}(\theta)(\mathcal{S}(y, \theta) - C_{\underline{b}}(\theta)) \right] f_\theta(y) dy = 0. \tag{2.3.3}$$

The B^p_s-estimate is an M-estimate, according to the usual definition, with

$$\psi(y, \theta) = \Psi_{\underline{b}}[A_{\underline{b}}(\theta)(\mathcal{S}(y, \theta) - C_{\underline{b}}(\theta))]. \tag{2.3.4}$$

The corresponding estimate of μ is $\hat{\mu} = \mu(G_{\hat{\theta}})$. (We will use $b_1 = b_2$ in the examples.) The rationale for the B^p_s-estimate is given in Hampel et al. (1986, pp. 238–257).

2.4 The one-step M-estimate

In this section, we assume that the model is parametrized with $\theta = (\mu, \sigma)^{\mathrm{T}}$ and denote by $\mathcal{S}(x, \mu, \sigma)$ the score function $\partial \ln g_\theta(x) / \partial \mu$. (In some special cases it may be more convenient to use $\theta = (\mu, \alpha)^{\mathrm{T}}$ with obvious changes in what follows.) Assume that

$\bar{\mu}$ is a given Fisher consistent estimate of μ, e.g., $\bar{\mu} = \mu(G_{\tilde{\theta}})$. We consider the one-step M-estimate $\breve{\mu}$ (Huber, 1981) defined by

$$\breve{\mu}(G_n) = \bar{\mu} - \frac{\sum_i \psi(x_i, \bar{\mu}, \bar{\sigma})}{\sum_i \dot{\psi}(x_i, \bar{\mu}, \bar{\sigma})}, \tag{2.4.1}$$

where

$$\psi(x, \mu, \sigma) = \psi_k \left(\mathcal{S}(x, \mu, \sigma) - B_k(\mu, \sigma) \right), \tag{2.4.2}$$

the function $B_k(\mu, \sigma)$ satisfies

$$\int \psi_k \left(\mathcal{S}(x, \mu, \sigma) - B_k(\mu, \sigma) \right) dG_{\mu,\sigma}(x) = 0, \tag{2.4.3}$$

$\psi_k(y)$ is the rescaled Tukey's biweight (2.2.3), k is a user defined tuning constant, and $\dot{\psi}_k(x, \mu, \sigma) = \partial\psi_k(x, \mu, \sigma)/\partial\mu$.

3. Influence functions, asymptotic variances, and breakdown points

Suppose that $\bar{\theta}$ and $\bar{\mu}$ are some of the estimates described in Section 2, i.e., $\bar{\theta} = \tilde{\theta}$ and $\bar{\mu} = \tilde{\mu}$ (the LD-estimates), or $\bar{\theta} = \hat{\theta}$ and $\bar{\mu} = \hat{\mu}$ (the B$_s^p$-estimates), or $\bar{\mu} = \breve{\mu}$ (the one-step M-estimate), or $\bar{\mu} = \check{\mu}$ (the truncated mean estimate). Then, under mild regularity conditions, $\bar{\theta}$ and $\bar{\mu}$ are asymptotically unbiased, compact differentiable (Rieder, 1994), and therefore, asymptotically normally distributed. The asymptotic covariance matrix $V(\bar{\theta}, F)$ of $\bar{\theta}$ and the asymptotic variance $v(\bar{\mu}, G)$ of $\bar{\mu}$ are

$$V(\bar{\theta}, F) = \int IF(y, \bar{\theta}, F) IF(y, \bar{\theta}, F)^{\mathrm{T}} dF(y), \tag{3.0.1}$$

$$v(\bar{\mu}, G) = \int IF(x, \bar{\mu}, G)^2 dG(x), \tag{3.0.2}$$

where $IF(y, \bar{\theta}, F) = (IF(y, \bar{\alpha}, F), IF(y, \bar{\sigma}, F))^{\mathrm{T}}$ and $IF(x, \bar{\mu}, G)$ denote the influence functions (Hampel et al., 1986) of $\bar{\theta}$ and $\bar{\mu}$. In applications, $V(\bar{\theta}, F_n)/n$ and $V(\bar{\theta}, F_{\bar{\theta}})/n$ are used as approximations of the covariance matrix of $\bar{\theta}$. Similarly, $v(\bar{\mu}, G_n)$ and $v(\bar{\mu}, G_{\bar{\theta}})$ are used as approximations of the variance of $\bar{\mu}$.

In the following subsections expressions for $IF(y, \bar{\alpha}, F)$, $IF(y, \bar{\sigma}, F)$, and $IF(x, \bar{\mu}, G)$ are given for the most typical cases of Section 2. These IFs have been derived by means of standard differentiation techniques. Note that, as $F = G \circ h^{-1}$, any functional p of F may also be considered as a functional $p^*(G) = p(G \circ h^{-1})$ of G and

$$IF(x, p^*, G) = IF(y, p, F) \quad \text{with } y = h(x). \tag{3.0.3}$$

Moreover, we find it convenient to use the following notations:

$$I(t, \alpha) = \int_{-\infty}^{t} dF_{\alpha,1}(y), \quad I_1(t, \alpha) = \partial I(t, \alpha)/\partial t, \quad I_2(t, \alpha) = \partial I(t, \alpha)/\partial \alpha,$$

$$J(t, \alpha) = \int_{-\infty}^{t} y dF_{\alpha,1}(y), \quad J_1(t, \alpha) = \partial J(t, \alpha)/\partial t, \quad J_2(t, \alpha) = \partial J(t, \alpha)/\partial \alpha.$$

Thus, $I(t, \alpha) = F_{\alpha,1}(t)$, $I_1(t, \alpha) = f_{\alpha,1}(t)$, and $J_1(t, \alpha) = t f_{\alpha,1}(t)$. Finally, the following auxiliary lemma is used on several occasions.

Lemma. Let $p(F)$ be a functional with influence function $IF(y, p, F)$. Let

$$F \circ p(F) = F(p(F)) \quad \text{and} \quad i \circ p(F) = \int_0^{p(F)} y dF(y).$$

Then

$$\begin{aligned} IF(y, F \circ p, F) &= f(p(F)) IF(y, p, F) + \Delta_y(p(F)) - F \circ p(F), \\ IF(y, \ i \circ p, F) &= f(p(F)) p(F) IF(y, p, F) + y \Delta_y(p(F)) - i \circ p(F). \end{aligned}$$

3.1 Influence function of the LD-estimate

If α is a location parameter, then

$$\tilde{\alpha}(F) = m(F) - [m(F_{0,1})/s(F_{0,1})]s(F) \qquad \text{and} \qquad \tilde{\sigma}(F) = [1/s(F_{0,1})]s(F). \tag{3.1.1}$$

Therefore, the IFs of $\tilde{\alpha}$ and $\tilde{\sigma}$ can be calculated straightforwardly, if the IFs of $m(F)$, $s(F)$ are known:

$$IF(y, \tilde{\alpha}, F) = IF(y, m, F) - [m(F_{0,1})/s(F_{0,1})] IF(y, s, F). \tag{3.1.2}$$

$$IF(y, \tilde{\sigma}, F) = [1/s(F_{0,1})] IF(y, s, F). \tag{3.1.3}$$

If α is a shape parameter, the functionals $\tilde{\alpha}(F)$ and $\tilde{\sigma}(F)$ of α and σ are defined by:

$$m(F)/s(F) = M(\tilde{\alpha}(F))/S(\tilde{\alpha}(F)) \qquad \text{and} \qquad \tilde{\sigma}(F) = S(F)/S(\tilde{\alpha}(F)), \tag{3.1.4}$$

where $M(\alpha) = m(F_{\alpha,1})$ and $S(\alpha) = s(F_{\alpha,1})$. In this case, the IFs of $\tilde{\alpha}$ and $\tilde{\sigma}$ depends on the IFs of m and s as well as on the functions M, S, and their derivatives M' and S':

$IF(y, \tilde{\alpha}, F) = a/b$, where

$$\begin{aligned} a &= [s(F) IF(y, m, F) - m(F) IF(y, s, F)]/s(F)^2, \\ b &= [M'(\tilde{\alpha}(F)) S(\tilde{\alpha}(F)) - S'(\tilde{\alpha}(F)) M(\tilde{\alpha}(F))]/S(\tilde{\alpha}(F))^2, \end{aligned}$$

$$IF(y, \tilde{\sigma}, F) = [S(\tilde{\alpha}(F)) IF(y, s, F) - s(F) S'(\tilde{\alpha}(F)) IF(y, \tilde{\alpha}, F)]/S(\tilde{\alpha}(F))^2.$$

We consider the LD$_D$-estimate as a special case. Then, $m(F)$ denotes the β-trimmed mean and $s(F)$ the γ-trimmed absolute deviation functionals:

- for $\beta < 0.5$ and $\gamma < 0.5$,

  $$m(F) = \frac{1}{1 - 2\beta} \int_{F^{-1}(\beta)}^{F^{-1}(1-\beta)} y dF(y), \qquad s(F) = \frac{1}{1 - 2\gamma} \int_{\tilde{F}^{-1}(\gamma)}^{\tilde{F}^{-1}(1-\gamma)} y d\tilde{F}(y),$$

  where

  $$\tilde{F}(z) = F(m(F) + z) - F(m(F) - z)$$

  denotes the distribution of $Z = |Y - m(F)|$, when $Y \sim F$;
- for $\beta = 0.5$, $m(F)$ is the median functional and, for $\gamma = 0.5$, $s(F)$ is the median absolute deviation functional (see, e.g., Rieder 1994).

The IF of m can be found in Huber (1981). In order to compute the IF of s, define $d_\gamma(F) = \tilde{F}^{-1}(\gamma)$ and note that $s = d_\gamma$ ($\gamma = 0.5$) and $s = (s_{1-\gamma} - s_\gamma)/(1-2\gamma)$ ($\gamma < 0.5$), where

$$s_\gamma = i \circ (m + d_\gamma) + i \circ (m - d_\gamma) - 2i \circ m - m \cdot [F \circ (m + d_\gamma) + F \circ (m - d_\gamma) - 2F \circ m].$$

Thus, use the auxiliary Lemma and the following result:

$IF(y, d_\gamma, F) = a/b$ where,

$$\begin{aligned} a &= -g(y) + IF(y, m, F)[f(m(F) - d_\gamma(F)) - f(m(F) + d_\gamma(F))], \\ b &= f(m(F) - d_\gamma(F)) + f(m(F) + d_\gamma(F)), \\ g(y) &= \Delta_y(m(F) + d_\gamma(F)) - \Delta_y(m(F) - d_\gamma(F)) - \gamma. \end{aligned}$$

Finally, the derivatives M' and S' can be expressed by means of

$$Q_\beta(\alpha) = F_{\alpha,1}^{-1}(\beta), \qquad D_\gamma(\alpha) = \tilde{F}_{\alpha,1}^{-1}(\gamma), \qquad S_\gamma(\alpha) = \int_0^{D_{1-\gamma}(\alpha)} y d\tilde{F}_{\alpha,1}(y),$$

so that $S(\alpha) = (S_{1-\gamma}(\alpha) - S_\gamma(\alpha))/(1-2\gamma)$ for $\gamma < 0.5$ and $S(\alpha) = D_\gamma(\alpha)$ for $\gamma = 0.5$. One obtains (suppressing the argument α in the notation):

$$\begin{aligned} Q'_\beta &= -I_2(Q_\beta)/I_1(Q_\beta), \\ M' &= \begin{cases} [J_1(Q_{1-\beta})Q'_{1-\beta} + J_2(Q_{1-\beta}) - J_1(Q_\beta)Q'_\beta - J_1(Q_\beta)]/(1-2\beta) & \text{for } \beta < 0.5, \\ -I_2(M)/I_1(M) & \text{for } \beta = 0.5, \end{cases} \\ D'_\gamma &= I_2(B) - I_2(A) + M' \cdot [I_1(A) - I_1(B)]/I_1(A) + I_1(B), \\ S'_\gamma &= J_1(A)A' + J_2(A) + J_1(B)B' + J_2(B), \\ &\quad - M \cdot [I_1(A)A' + I_2(A) + I_1(B)B' + I_2(B)] - M' \cdot [I(A) + I(B)] + O \\ S' &= \begin{cases} (S'_{1-\gamma} - S'_\gamma)/(1-2\gamma) & \text{for } \gamma < 0.5, \\ D'_\gamma & \text{for } \gamma = 0.5, \end{cases} \end{aligned}$$

where $A = M + D_\gamma$, $B = M - D_\gamma$, $A' = M' + D'_\gamma$, $B' = M' - D'_\gamma$, and O cancels out in S'.

3.2 Influence function of the truncated mean

Here we use the notation

$$H(t, \alpha, \sigma) = G_{\alpha,\sigma}(t), \qquad K(t, \alpha, \sigma) = \int_0^t x dG_{\alpha,\sigma}(x),$$

$H_1 = \partial H/\partial t$, $H_2 = \partial H/\partial \alpha$, $H_3 = \partial H/\partial \sigma$, $K_1 = \partial K/\partial t$, $K_2 = \partial K/\partial \alpha$, $K_3 = \partial K/\partial \sigma$. Let $\bar{\alpha}^*(G) = \bar{\alpha}(F)$, $\bar{\sigma}^*(G) = \bar{\sigma}(F)$, $\bar{\mu}(G) = \mu(G_{\bar{\alpha}^*(G),\bar{\sigma}^*(G)})$. Let $q(G)$ be the solution of

$$H(q, \bar{\alpha}^*, \bar{\sigma}^*) = u \tag{3.2.1}$$

and $r(G)$ the solution of

$$\frac{1}{u - H(r, \bar{\alpha}^*, \bar{\sigma}^*)}[K(q, \bar{\alpha}^*, \bar{\sigma}^*) - K(r, \bar{\alpha}^*, \bar{\sigma}^*)] = \bar{\mu}, \tag{3.2.2}$$

where $\bar{\alpha}^*$, $\bar{\sigma}^*$, $\bar{\mu}$, r, q have the argument G. Note that $T_u = q(G_n)$, $T_l = r(G_n)$, and

$$\breve{\mu}(G) = (i \circ q - i \circ r)(G)/(G \circ q - G \circ r)(G).$$

Therefore,

$$\begin{aligned} IF(x, \breve{\mu}, G) = &- \frac{(i \circ q - i \circ r)(G)}{(G \circ q - G \circ r)^2(G)} \cdot [IF(x, G \circ q, G) - IF(x, G \circ r, G)] \\ &+ \frac{1}{(G \circ q - G \circ r)(G)} \cdot [IF(x, i \circ q, G) - IF(x, i \circ r, G)], \end{aligned}$$

which can be computed with the help of the auxiliary Lemma. The IFs of q and r are required and can be obtained by differentiating (3.2.1) and (3.2.2). We obtain:

$$\begin{aligned} IF(x, q, G) &= -[1/H_1(q, \bar{\alpha}, \bar{\sigma})][H_2(q, \bar{\alpha}, \bar{\sigma}) IF(x, \bar{\alpha}^*, G) + H_3(q, \bar{\alpha}, \bar{\sigma}) IF(x, \bar{\sigma}^*, G)] \\ &= -[1/H_1(q, \bar{\alpha}, \bar{\sigma})][H_2(q, \bar{\alpha}, \bar{\sigma}) IF(h(x), \bar{\alpha}, F) + H_3(q, \bar{\alpha}, \bar{\sigma}) IF(h(x), \bar{\sigma}, F)], \end{aligned}$$

and

$IF(x, r, G) = a/b$, where

$$\begin{aligned} a = &\ [K_2(r, \bar{\alpha}, \bar{\sigma}) - K_2(q, \bar{\alpha}, \bar{\sigma}) - H_2(r, \bar{\alpha}, \bar{\sigma}) \bar{\mu}(G)] IF(h(x), \bar{\alpha}, F) \\ &+ [K_3(r, \bar{\alpha}, \bar{\sigma}) - K_3(q, \bar{\alpha}, \bar{\sigma}) - H_3(r, \bar{\alpha}, \bar{\sigma}) \bar{\mu}(G)] IF(h(x), \bar{\sigma}, F) \\ &+ [u - H(r, \bar{\alpha}, \bar{\sigma})] IF(x, \bar{\mu}, G) - IF(x, q, G) K_1(q, \bar{\alpha}, \bar{\sigma}), \\ b = &\ H_1(r, \bar{\alpha}, \bar{\sigma}) \bar{\mu}(G) - K_1(r, \bar{\alpha}, \bar{\sigma}). \end{aligned}$$

Here, $IF(x, \bar{\mu}, G)$ must be derived from the expression of $\mu(G_{\bar{\alpha},\bar{\sigma}})$. (Note: expressions can be somewhat simplified if $G_{\alpha,\sigma}$ can be re-parametrized using a scale parameter.)

3.3 Influence function of the B_s^p-estimate

The influence function of the B_s^p-estimate is given in Marazzi and Ruffieux (1996).

3.4 Influence function of the one-step M-estimate

We use the notation of Sections 2.3 and 3.2. Further, the derivatives of any function $\xi(y, \mu, \sigma)$ are denoted by $\xi' = \partial H/\partial y$, $\dot{\xi} = \partial \xi/\partial \mu$, $\ddot{\xi} = \partial \dot{\xi}/\partial \mu$, $\acute{\xi} = \partial \xi/\partial \sigma$ $\grave{\dot{\xi}} = \partial \dot{\xi}/\partial \sigma$. The functional expression of the one-step M-estimate defined in Section 2.3 is

$$\breve{\mu}(G) = \bar{\mu}(G) - h_1(G)/h_2(G), \tag{3.4.1}$$

where

$$h_1(G) = \int \psi(x, \bar{\mu}(G), \bar{\sigma}^*(G)) dG(x), \qquad h_2(G) = \int \dot{\psi}(x, \bar{\mu}(G), \bar{\sigma}^*(G)) dG(x),$$

$$\psi(x, \mu, \sigma) = \psi_k \left[\mathcal{S}(x, \mu, \sigma) - B_k(\mu, \sigma) \right],$$

$\psi_k(x)$ is the rescaled Tukey's biweight (2.2.3) and the function $B_k(\mu, \sigma)$ satisfies

$$\int \psi_k \left[\mathcal{S}(x, \mu, \sigma) - B_k(\mu, \sigma) \right] dG_{\mu,\sigma}(x) = 0. \tag{3.4.2}$$

One obtains (suppressing obvious arguments in the notation):

$$\begin{aligned}
IF(x, h_1, G) &= \psi - h_1 + \left[\int \dot{\psi}\right] IF(x, \bar{\mu}, G) + \left[\int \acute{\psi}\right] IF(h(x), \bar{\sigma}, G), \\
IF(x, h_2, G) &= \dot{\psi} - h_2 + \left[\int \ddot{\psi}\right] IF(x, \bar{\mu}, G) + \left[\int \grave{\psi}\right] IF(h(x), \bar{\sigma}, G), \\
IF(x, \breve{\mu}, G) &= (1/h_2) IF(x, h_1, G) - (h_1/h_2^2) IF(x, h_2, G).
\end{aligned}$$

3.5 Breakdown points

The breakdown point (BDP) of the LD$_D$-estimate is $\min(\beta, \gamma)$ and The BDP of the LD$_R$-estimate is 25%. The truncated mean estimate and the (redescending) one-step M-estimate maintain the breakdown point of the initial estimate.

The problem of analytically computing the breakdown point of the B^p_s-estimate is still unsolved in the general case. Some examples can be found in Marazzi and Ruffieux (1996), where, depending on the model and on the tuning constants, it ranges from 20% to 40%.

4. Notes on computation

The computation of the B^p_s-estimator is discussed in Marazzi and Ruffieux (1996). The iterative procedures for solving equations (2.3.1)-(2.3.3) (as well as the programs available in Marazzi and Randriamiharisoa, 1997a-c) are rather complex.

The one-step M-estimate avoids iterations. In its simplest form (2.4.1), it may however assume negative values and some modifications should be added. Moreover, for the computation of the influence function, beside ψ, ψ', and ψ'', the functions $\mathcal{S}$, $\dot{\mathcal{S}}$, $\ddot{\mathcal{S}}$, $\acute{\mathcal{S}}$, $\grave{\mathcal{S}}$, as well as the values of B_k, $\dot{B}_k$, $\ddot{B}_k$, $\acute{B}_k$, $\grave{B}_k$ for $\mu = \bar{\mu}(G)$ and $\sigma = \bar{\sigma}(F)$ are required. Note that the function $B_k(\mu, \sigma)$ is implicitly defined by (3.4.2); thus, the derivatives of B_k must be computed by differentiating these equations with respect to μ and σ. (Details may be found in the S-plus functions made available by the authors). The computational complexity is relevant.

If α is a location parameter, the LD-estimator $\tilde{\alpha}$ and $\tilde{\sigma}$ is explicitly obtained from (2.1.1). If α is a shape parameter, $\tilde{\alpha}$ must be computed numerically by solving (3.1.4), where:

- for the LD$_R$-estimate,

$$M(\alpha) = F^{-1}_{\alpha,1}(1/2), \qquad S(\alpha) = F^{-1}_{\alpha,1}(3/4) - F^{-1}_{\alpha,1}(1/4); \tag{4.1.1}$$

- for the LD$_D$-estimate,

$$M(\alpha) = \frac{1}{1-2\beta} \int_{F^{-1}_{\alpha,1}(\beta)}^{F^{-1}_{\alpha,1}(1-\beta)} y dF_{\alpha,1}(y), \quad S(\alpha) = \frac{1}{1-2\gamma} \int_{\tilde{F}^{-1}_{\alpha,1}(\gamma)}^{\tilde{F}^{-1}_{\alpha,1}(1-\gamma)} z d\tilde{F}_{\alpha,1}(z), \tag{4.1.2}$$

and

$$\tilde{F}_{\alpha,1}(z) = F_{\alpha,1}(M(\alpha) + z)) - F_{\alpha,1}(M(\alpha) - z)$$

denotes the distribution of $Z = |Y - M(\alpha)|$ when $Y \sim F_{\alpha,1}$.

Fast accurate programs for computing $F^{-1}_{\alpha,1}(\cdot)$ in (4.1.1) are readily available in standard statistical environments and spreadsheets; yet the computation of $M(\alpha)$ and $S(\alpha)$ according to (4.1.2) requires ad hoc programs. Moreover, the solution of (3.1.4) may slow down the iterations (especially within a resampling scheme). The following types of approximations can however be used to reduce the amount of computations:

(a) Choose an integer k (e.g., $k = 100$) and an interval $[\alpha_1, \alpha_2]$ that hopefully contains the estimate of α. Store $M(\alpha)$ and $S(\alpha)$ for $\alpha = \alpha_1 + (i-1)(\alpha_2 - \alpha_1)/(k-1)$ and $i = 1, \ldots, k$ in a preliminary computation. When required, compute $M(\alpha)$ and $S(\alpha)$ by means of an interpolation scheme.

(b) Use "simple approximations" for $M(\alpha)$, $S(\alpha)$, or $M(\alpha)/S(\alpha)$.

Example. If $F_{\alpha,1}$ is a Gamma distribution, $\beta = 0.4$, and $\gamma = 0.4$ then, for $\alpha \in [1, 20]$,

$$-0.020 + 0.052(M(\alpha)/S(\alpha)) + 0.0454(M(\alpha)/S(\alpha))^2 \approx \alpha$$

with an absolute error smaller than 0.016. Therefore, we may set $\tilde{\alpha} = -0.020 + 0.055(m/s) + 0.0454(m/s)^2$, where $m = m(F_n)$ and $s = s(F_n)$, provided the result belongs to $[1, 20]$. Moreover, $M(\alpha) \approx -0.314 + \alpha$ with an error smaller than 0.014 over the same range. Thus, we may set $\tilde{\sigma} = m/(\tilde{\alpha} - 0.314)$. Other approximations are under investigation and will be described elsewhere.

For the computation of the truncated mean, the truncation limits $T_u = G^{-1}_{\bar{\alpha},\bar{\sigma}}(u)$ (for given u) and $T_l = G^{-1}_{\bar{\alpha},\bar{\sigma}}(l)$ are required. Here, l solves

$$K(T_u, \bar{\alpha}, \bar{\sigma}) - K(G^{-1}_{\bar{\alpha},\bar{\sigma}}(l), \bar{\alpha}, \bar{\sigma}) = (u - l)\bar{\mu}, \tag{4.1.3}$$

$\bar{\mu} = \int G^{-1}_{\bar{\alpha},\bar{\sigma}}(x)dx$ and $K(t, \alpha, \sigma) = \int_0^t x dG_{\alpha,\sigma}(x)$. Again, if programs for computing K and solving (4.1.3) are not available, approximations can be used (e.g., an interpolation of Table 1, below).

5. Numerical results

Here we consider ϵ-contamination models of the form $G = (1-\epsilon)G_{\alpha,\sigma} + \epsilon\tilde{G}$, where the contamination $\tilde{G}$ is a uniform distribution over the interval $[0, a]$ (see the remark, below), $\epsilon \in (0, 1)$ is the contamination proportion, and the central model $G_{\alpha,\sigma}$ is one of the following distributions:

- the Weibull distribution with shape parameter α and scale σ;
- the Gamma distribution with shape parameter α and scale σ.
- the Lognormal distribution with normal mean α and normal scale σ;

In the Lognormal and Weibull cases, the transformation $h(X) = \ln(X)$ (that provides location-scale parametrizations for $\ln(X)$) is used; in the Gamma case $h(X) = X$ (that provides the usual shape-scale parametrization). Moreover, we consider the following estimators of $\mu(G)$:

- the classical maximum likelihood estimator of μ (for $\epsilon = 0$), (abbreviation MLE);
- the LD$_D$-estimator with $\beta = \gamma = 0.4$, (abbreviation D);
- the LD$_R$-estimator, (abbreviation R);
- the truncated mean estimator starting with LD$_D$, (abbreviation TD);

- the truncated mean estimator starting with LD_R, (abbreviation TR);
- the B_s^p-estimator, (abbreviation B);
- the one-step M-estimator based on Tukey's biweight, (abbreviation O).

In order to indicate the model on which the estimator is based, a suffix is sometimes added to the estimator notation: 1 for Weibull, 2 for Gamma, 3 for Lognormal. Thus, MLE_1 is the maximum likelihood estimator for the Weibull model, TD_2 is the truncated mean estimator for the Gamma model (starting with D_2), etc.

Figure 1 shows the asymptotic relative efficiency (ARE) of TD with respect to MLE, as a function of the upper truncation limit u (the lower truncation limit l being determined according to (2.2.1)), for the three models and a variety of values of α (Weibull and Gamma cases) and σ (Lognormal case). Except for the Lognormal case, these functions are monotonely increasing. Table 1 reports the values of u (and l) that solve ARE $= 0.80$, 0.85, and 0.90.

The IFs (at the models) of the 80% efficient estimators TD_j, TR_j, O_j, and B_j ($j = 1, 2, 3$) are shown in Figure 2. Here, moderately asymmetric models with mean = 2 and variance = 2 have been used: Gamma $\alpha = 2$, $\sigma = 1$; Weibull $\alpha = 1.435$, $\sigma = 2.203$; Lognormal $\lambda = 0.490$, $\sigma = 0.637$. (The IFs of the MLEs, not shown in Figure 2, are unbounded.)

A simulation has been performed with the 21 estimators MLE_j, D_j, R_j, TD_j, TR_j, B_j, and O_j (for $j = 1, 2, 3$). TD_j, TR_j, B_j, and O_j have been tuned so that ARE$= 0.8$. 2000 samples have been generated (using S-plus) according to the three central models $G_{\alpha,\sigma}$ mentioned above and two different contaminations $\tilde{G}$, $a = 10$ (short tailed) and $a = 50$ (long tailed), both with $\epsilon = 0.1$. The mean values of the contamination bias and variances over the 2000 samples of size $n = 200$ are reported in Table 2. (With a confidence of 90%, the relative error of the variances does not exceed 5%.) Some of the results are graphically represented in Figure 3. Other sample sizes have been investigated but the detailed results are not shown here. For example, Table 3 gives the average variances of TD_j (over 2000 samples) for $n = 200$ and $n = 50$ as well as the asymptotic approximations $V(\hat{\theta}, G_{\alpha,\sigma})/n$ provided by (3.0.1).

Finally, a real example is considered. The lengths of 32 stays (LOS, in days) in a Swiss hospital during 1988 for certain "disorders of the nervous system" are given in Table 4. LOS is an important indicator of hospital costs (Lave and Leinhardt, 1976) and the "LOS means"of medically homogeneous groups of patients within a hospital are used for budgeting purposes. LOS distributions often contain outliers whose value and frequency fluctuate from sample to sample, e.g., from year to year; for example, there are three clear outliers in Table 5. In order to obtain stable summaries, LOS means are usually replaced by the crude truncated means mentioned in the introduction. As a probability plot would clearly show, the Weibull model is a reasonable description. (This model can indeed be used for typical LOS distributions, according to Marazzi et al., 1998). The values of TD_1, TR_1, and B_1, as well as their bootstrap and asymptotic variance approximations, are reported in Table 5.

REMARK. The uniform distribution has been chosen because it best reproduces the kind of contamination we have observed in a variety of LOS distributions (see real example). Other contamination distributions, e.g., Gamma and Weibull, have been experimented. The results were, however, slightly confusing because it was more difficult to identify the central model.

6. Discussion

Few robust estimators for the parameters of asymmetric models have been studied in the literature. Most of these procedures are rather complex: for example, in order to compute a certain type of M-estimate for popular models like the Gamma and Weibull, seven scalar equations must be solved numerically. Thus, although programs are made available, many practical statisticians are reluctant to use them. On the other hand, the disturbing effects of outliers are well known among practitioners and very crude rules for outlier trimming are used as a remedy. Yet these rules have obvious flaws and insufficient theoretical support. The truncated mean estimator aims at improving these rules with the help of the basic robustness concepts and tools, while maintaining their computational simplicity.

In a large variety of situations that are common in practice, the ARE of TD and TR attains satisfactory levels. In our examples, except for extremely asymmetric models (e.g. Weibull with $\alpha = 0.5$ or Lognormal with $\sigma > 1$), the maximum values of ARE are very close to 1. Thus, efficiency loss at the model, with respect to the more sophisticated M-estimates, is not a matter of concern. Yet the truncation limits necessary to attain the usual values of ARE (e.g, 0.90 or 0.95) are much higher than the truncation limits that are currently used on the basis of experience with real data. For example, rules (b) and (c) mentioned in the introduction reject observations beyond 3 median absolute deviations or beyond 1.5 interquartile ranges from the median. If a Gaussian distribution on the log scale can be assumed, these limits correspond to a (2-sided) truncation proportion of about 5% and to an efficiency loss of 15%-25% (for σ ranging from 0.2 to 1). Thus, an ARE of 0.80 (the value used in the examples) seems a reasonable compromise.

According to the IFs, TD and TR are more sensitive than B and O to contaminating observations on the flanks of the central distributions, but far less sensitive to large outliers. As the example with real data shows, this can be a desirable feature. Not surprisingly, we also note that the IFs (and hence the behaviour) of TD and TR are more stable to model changes than the IFs of O and M (which depend on the score functions). In particular, the truncation points provided by the three models are close.

The simulation results in Table 2 (and other results not reported here) show that:

1. The one-step M-estimators cannot be used in their naive form and some refinement must be added in order to prevent erratic (e.g. negative) values. This will, however, complicate computation, an aspect we are trying to keep simple in this paper. The behaviour of these estimators will not be further commented on.
2. When $\epsilon \neq 0$, the MLE breaks down.
3. For $\epsilon = 0$, the estimators TD, TR, and B based on the appropriate models (Table 2, results in italic characters) perform according to the designed characteristics: the bias is very small and there is about 20% loss of variance with respect to MLE.
4. When $\epsilon \geq 0$ but the central model is still appropriate, TD and TR have comparable values of bias and variance. However, B is slightly better than TD and TR for $a = 10$ (short tailed contamination), whereas TD and TR are definitely better than B for $a = 50$ (long tailed contamination) – see also Figure 3. This is perfectly explained by the shape of the IFs: TD and TR – but not B – almost reject all observations beyond the truncation point (about 7); they are, however, more sensitive than B over a short interval of middle observations (between 4 and 7).
5. When $\epsilon \geq 0$ and the estimators are based on a wrong model, TD and TR outperform

B for $a = 50$. The performances of these three estimators are similar for $a = 10$.
6. The truncated means based on the Lognormal model are less stable than those based on Weibull or Gamma, even if the Lognormal model is correct. This is not surprising, since the Lognormal model provides the highest truncation points.
7. For $\epsilon \geq 0$ the variance of TD is often smaller than the variance of TR, whereas the bias is similar. Also note that, as expected, TD and TR have smaller bias and variances than the respective initial values D and R. See again the diagrams in Figure 3.

In the real example, only TD_1 has satisfactory behaviour, the value of its bootstrap variance being very close to the asymptotic approximation. Despite the small number of original outliers ($\epsilon \approx 10\%$), some highly contaminated ($\epsilon > 25\%$) bootstrap samples provided very discrepant simulated values of TR_1 and B_1 – but not TD_1 – that inflated the bootstrap variances.

We note that the asymptotic approximation for the variace of TD_j (Table 3) is very satisfactory for $n = 200$, especially with $\epsilon = 0$ or with a long tailed contamination; it is, however, less satisfactory when the sample size is small and when the tails of the contamination are short. Alternative procedures should be developed for these cases.

We conclude that the truncated means form a class of computationally and statistically attractive estimators. Within this class, estimators based on models with moderate tails (e.g., Weibull) offer better protection than those based on long tailed models (e.g., Lognormal). Although the interquartile range is used in practice as a simple estimate of dispersion for current trimming rules, other estimates with a high breakdown point are to be preferred. As a final choice, we recommend TD_1 or TD_2 tuned at ARE= 0.8. Further developments will include their extension to regression models with asymmetrically distributed responses.

Aknowledgements

We would like to thank Alex Randriamiharisoa for the generous help in programming. Thanks are also due to the referees for their valuable comments. Alfio Marazzi was partially supported by the FOMEC Project, Mathematics Department, Faculty of Exact Sciences, National University of La Plata, Argentina.

References

Beguin, C., Closon, M.C., Roger, F.H., 1991. Advances in DRGs data pooling in Europe: results from the Hoscom project in relation to outliers. Paper presented at the second EURODRG Workshop, Dublin, 24-25 April 1991.

Hampel F.R., Ronchetti E.M., Rousseeuw P.J., Stahel W.A., 1986. Robust statistics: The approach based on influence functions. Wiley, New York.

Huber, P.J., 1981. Robust statistics. Wiley, New York.

Johnson N.L, Kotz S., Balakrishnan, N., 1994. Continuous univariate distributions. Volume 1, 2nd ed. Wiley, New York.

Lave JR., Leinhardt S., 1976. The cost and length of a hospital stay, Inquiry, 13, pp. 327-343.

Marazzi A., Ruffieux C., 1996. Implementing M-estimators of the Gamma distribution. In: H. Rieder (Ed.), Robust statistics, data analysis, and computer intensive methods, In honor of Peter Huber's 60th birthday, Lecture notes in statistics, 109, Springer Verlag, Heidelberg.

Marazzi A., Randriamiharisoa A., 1997a. S-plus functions for robust estimators of the parameters of the gaussian and the lognormal distributions. Technical report. Institut universitaire de médecine sociale et préventive, Bugnon 17, CH-1005 Lausanne. (Report and programs available at www.hospvd.ch/iumsp).

Marazzi A., Randriamiharisoa A., 1997b. S-plus functions for robust estimators of the parameters of the gamma distribution. Technical report. Institut universitaire de médecine sociale et préventive, Bugnon 17, CH-1005 Lausanne. (Report and programs available at www.hospvd.ch/iumsp).

Marazzi A., Randriamiharisoa A., 1997c. S-plus functions for robust estimators of the parameters of the Weibull distribution. Technical report. Institut universitaire de médecine sociale et préventive, Bugnon 17, CH-1005 Lausanne. (Report and programs available at www.hospvd.ch/iumsp).

Marazzi A., Paccaud F., Ruffieux C., Beguin C., 1998. Fitting the distributions of length of stay by parametric models. Medical Care, 36, 915-927.

Rieder, H., 1994. Robust asymptotic statistics. Springer-Verlag, New York.

Rousseeuw P.J., Croux C., 1993. Alternatives to the median absolute deviation. J. Amer. Statist. Assoc., 88, 1273-1283.

Victoria-Feser M.P., Ronchetti E., 1994. Robust methods for personal-income distribution models. Canad. J. Statist., 22, 247-258.

Table 1. Upper and lower limits u, l, so that the truncated mean estimate (based on LD_D, with $\beta = \gamma = 0.4$) for the Weibull, the Gamma, and the Lognormal models have a given asymptotic relative efficiency (ARE) with respect to the maximum likelihood estimate. If two values exist (Lognormal case), only the smallest one is reported. In the Weibull and Gamma cases, the ARE depends only on α; in the Lognormal case, the ARE depends only on σ.

		Weibull		Gamma		Lognormal		
ARE	α	u	l	u	l	u	l	σ
0.80	1	0.994	0.031	0.993	0.034	0.994	0.087	1.15
	2	0.989	0.020	0.991	0.030	0.993	0.074	1.00
	3	0.984	0.020	0.989	0.028	0.991	0.058	0.80
	4	0.980	0.019	0.988	0.026	0.990	0.044	0.60
	5	0.976	0.019	0.988	0.026	0.988	0.033	0.40
	10	0.969	0.019	0.986	0.024	0.985	0.024	0.20
0.85	1	0.997	0.019	0.996	0.023	0.997	0.059	1.15
	2	0.996	0.008	0.994	0.020	0.996	0.052	1.00
	3	0.989	0.013	0.983	0.018	0.995	0.041	0.80
	4	0.987	0.013	0.983	0.018	0.993	0.030	0.60
	5	0.985	0.013	0.992	0.017	0.992	0.022	0.40
	10	0.979	0.012	0.991	0.015	0.991	0.016	0.20
0.90	1	0.998	0.012	0.998	0.014	—	—	1.15
	2	0.996	0.008	0.997	0.012	0.998	0.027	1.00
	3	0.994	0.007	0.996	0.011	0.997	0.024	0.80
	4	0.992	0.007	0.996	0.010	0.996	0.018	0.60
	5	0.991	0.007	0.996	0.010	0.996	0.013	0.40
	10	0.988	0.007	0.995	0.009	0.995	0.009	0.20

Weibull ($\alpha = 1.435,\ \sigma = 2.203$)

	$\varepsilon = 0$		$\varepsilon = 0.1$ $a = 10$		$\varepsilon = 0.1$ $a = 50$	
	Bias	Var	Bias	Var	Bias	Var
MLE_1	*0.001*	*0.010*	*0.304*	*0.017*	*2.071*	*0.267*
D_1	*0.005*	*0.016*	*0.197*	*0.022*	*0.313*	*0.030*
R_1	*0.007*	*0.020*	*0.187*	*0.028*	*0.287*	*0.034*
TD_1	*-0.002*	*0.013*	*0.176*	*0.024*	*0.078*	*0.015*
TR_1	*-0.006*	*0.013*	*0.175*	*0.025*	*0.071*	*0.015*
B_1	*0.004*	*0.013*	*0.208*	*0.019*	*0.344*	*0.028*
O_1	*-0.008*	*0.013*	*0.192*	*0.030*	*-0.010*	*0.348*
MLE_2	0.001	0.010	0.299	0.017	2.301	0.341
D_2	0.081	0.018	0.265	0.024	0.365	0.030
R_2	0.068	0.016	0.259	0.023	0.373	0.030
TD_2	0.054	0.013	0.258	0.026	0.121	0.016
TR_2	0.057	0.013	0.256	0.026	0.124	0.016
B_2	0.058	0.013	0.261	0.020	0.389	0.028
O_2	0.067	0.013	0.273	0.020	0.355	0.046
MLE_3	0.196	0.017	0.522	0.028	1.823	0.256
D_3	0.336	0.024	0.593	0.039	0.760	0.056
R_3	0.384	0.037	0.628	0.054	0.773	0.069
TD_3	0.222	0.014	0.555	0.024	0.426	0.040
TR_3	0.214	0.013	0.548	0.024	0.437	0.044
B_3	0.275	0.017	0.542	0.030	0.739	0.050
O_3	0.071	0.012	0.321	0.022	0.746	0.075

Gamma ($\alpha = 2,\ \sigma = 1$)

	$\varepsilon = 0$		$\varepsilon = 0.1$ $a = 10$		$\varepsilon = 0.1$ $a = 50$	
	Bias	Var	Bias	Var	Bias	Var
MLE_1	0.008	0.010	0.312	0.017	2.061	0.269
D_1	-0.063	0.014	0.124	0.020	0.224	0.026
R_1	-0.073	0.017	0.105	0.023	0.194	0.028
TD_1	-0.063	0.012	0.104	0.022	0.028	0.014
TR_1	-0.068	0.013	0.097	0.022	0.020	0.014
B_1	-0.046	0.012	0.156	0.017	0.278	0.024
O_1	-0.063	0.015	0.131	0.034	-0.027	0.018
MLE_2	*0.002*	*0.010*	*0.302*	*0.016*	*2.287*	*0.352*
D_2	*0.005*	*0.016*	*0.183*	*0.021*	*0.269*	*0.025*
R_2	*0.004*	*0.015*	*0.195*	*0.020*	*0.300*	*0.026*
TD_2	*-0.001*	*0.012*	*0.180*	*0.023*	*0.070*	*0.014*
TR_2	*0.000*	*0.012*	*0.190*	*0.023*	*0.076*	*0.015*
B_2	*0.005*	*0.012*	*0.205*	*0.018*	*0.320*	*0.024*
O_2	*0.003*	*0.012*	*0.208*	*0.019*	*0.231*	*0.039*
MLE_3	0.108	0.012	0.425	0.021	1.638	0.219
D_3	0.234	0.022	0.484	0.034	0.627	0.047
R_3	0.252	0.030	0.496	0.043	0.624	0.054
TD_3	0.168	0.013	0.496	0.024	0.328	0.029
TR_3	0.163	0.013	0.491	0.024	0.329	0.030
B_3	0.193	0.016	0.455	0.027	0.627	0.042
O_3	0.077	0.012	0.340	0.023	0.714	0.072

Lognormal ($\lambda = 0.490,\ \sigma = 0.637$))

	$\varepsilon = 0$		$\varepsilon = 0.1$ $a = 10$		$\varepsilon = 0.1$ $a = 50$	
	Bias	Var	Bias	Var	Bias	Var
MLE_1	0.017	0.010	0.324	0.017	2.085	0.263
D_1	-0.211	0.010	-0.051	0.016	0.026	0.020
R_1	-0.226	0.011	-0.072	0.017	0.004	0.021
TD_1	-0.207	0.011	-0.074	0.017	-0.097	0.014
TR_1	-0.214	0.011	-0.081	0.017	-0.101	0.014
B_1	-0.164	0.009	0.019	0.015	0.122	0.021
O_1	-0.247	0.016	-0.115	0.030	-0.160	0.019
MLE_2	0.000	0.010	0.303	0.017	2.300	0.346
D_2	-0.150	0.011	0.002	0.016	0.070	0.019
R_2	-0.134	0.011	0.040	0.017	0.133	0.023
TD_2	-0.149	0.010	-0.016	0.017	-0.064	0.013
TR_2	-0.137	0.010	0.021	0.020	-0.034	0.014
B_2	-0.119	0.009	0.061	0.015	0.159	0.021
O_2	-0.184	0.013	-0.004	0.024	-0.042	0.025
MLE_3	*0.001*	*0.010*	*0.305*	*0.018*	*1.438*	*0.184*
D_3	*0.001*	*0.015*	*0.212*	*0.025*	*0.317*	*0.034*
R_3	*0.001*	*0.017*	*0.210*	*0.028*	*0.318*	*0.037*
TD_3	*-0.001*	*0.012*	*0.268*	*0.033*	*0.118*	*0.020*
TR_3	*-0.002*	*0.012*	*0.270*	*0.033*	*0.121*	*0.021*
B_3	*0.001*	*0.012*	*0.231*	*0.022*	*0.366*	*0.033*
O_3	*0.000*	*0.012*	*0.257*	*0.023*	*0.504*	*0.077*

Table 2. Average bias and variance of various robust estimators of the mean for samples of size 200. 2000 samples were simulated according to three central models (Weibull, Gamma, Lognormal) and two contaminations (Uniform $[0, a]$, $a = 10$ and $a = 50$). The contamination proportion is ϵ.

Table 3. Average variances (Var) of TD_j and asymptotic approximations (V). 2000 samples were simulated according to three the central models – Weibull ($\alpha = 1.435$, $\sigma = 2.203$), Gamma ($\alpha = 2$, $\sigma = 1$), Lognormal ($\lambda = 0.490$, $\sigma = 0.637$) – and two contaminations (Uniform $[0, a]$, $a = 10$ and $a = 50$). The contamination proportion is ϵ.

Estimate	Central Model	n	Var $\varepsilon = 0$	Var $\varepsilon = 0.1$ $a = 10$	Var $\varepsilon = 0.1$ $a = 50$	V
TD_1	Weibull	200	0.013	0.024	0.015	0.012
		50	0.051	0.093	0.075	0.050
TD_2	Gamma	200	0.012	0.023	0.014	0.012
		50	0.050	0.095	0.066	0.050
TD_3	Lognormal	200	0.012	0.033	0.020	0.012
		50	0.048	0.106	0.115	0.049

Table 4. Frequency distribution (Freq.) of lengths of stay (LOS) in days of patients hospitalized in Switzerland during 1988 for certain “disorders of the nervous system”.

LOS	1	2	3	4	5	6	7	8	9	16	115	198	374
Freq.	2	6	5	5	4	2	2	1	1	1	1	1	1

Table 5. Values and variance approximations of the estimates TD_1, TR_1, and B_1 obtained with the data of Table 3.

Estimate	Value	Variance asympt. approx.	Variance bootstr. approx.
Truncated mean TD_1	4.00	0.32	0.32
Truncated mean TR_1	4.00	0.20	7.02
M-estimate B_1	4.59	0.29	0.64

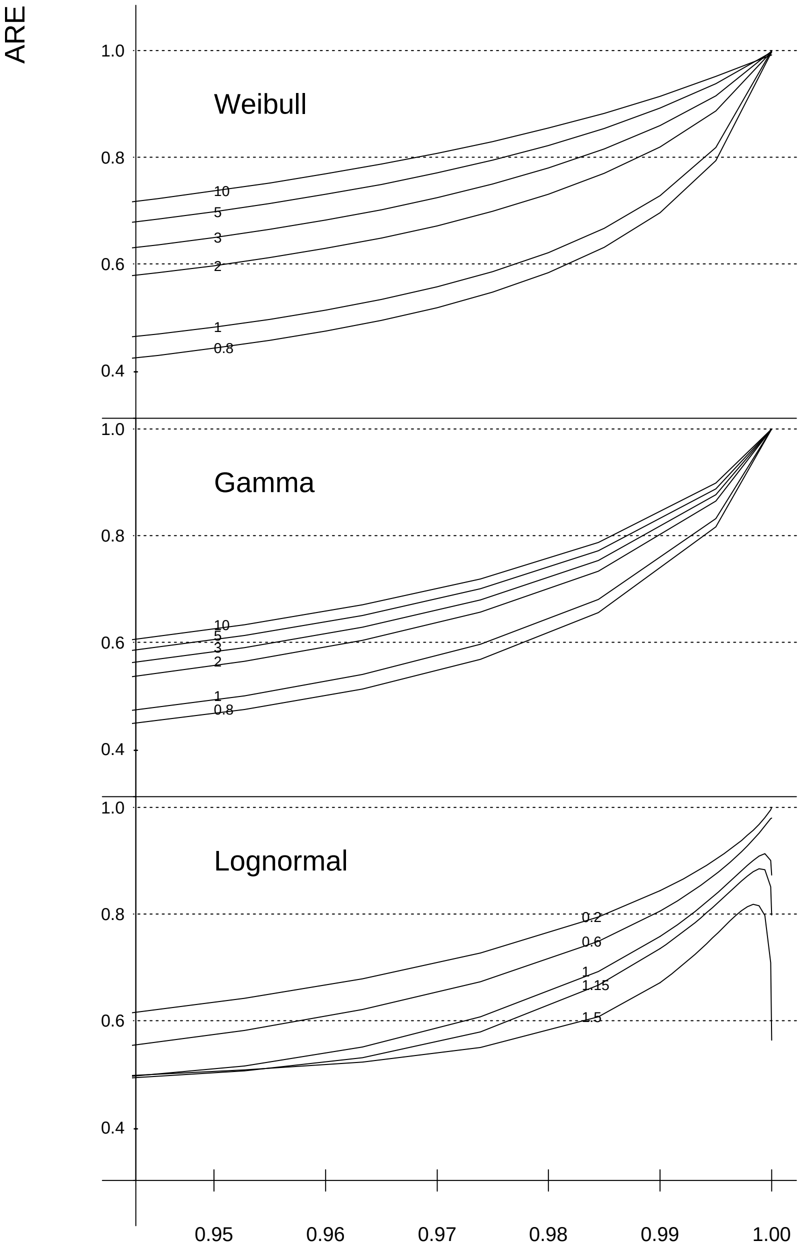

Figure 1. Asymptotic relative efficiency (ARE) of the truncated mean estimate (based on LD_D, with $\beta = \gamma = 0.4$) as a function of the upper truncation point, for a variety of values of α (Weibull and Gamma cases) and σ (Lognormal case). In the Weibull and Gamma cases, the ARE depends only on α; in the Lognormal case, the ARE depends only on σ.

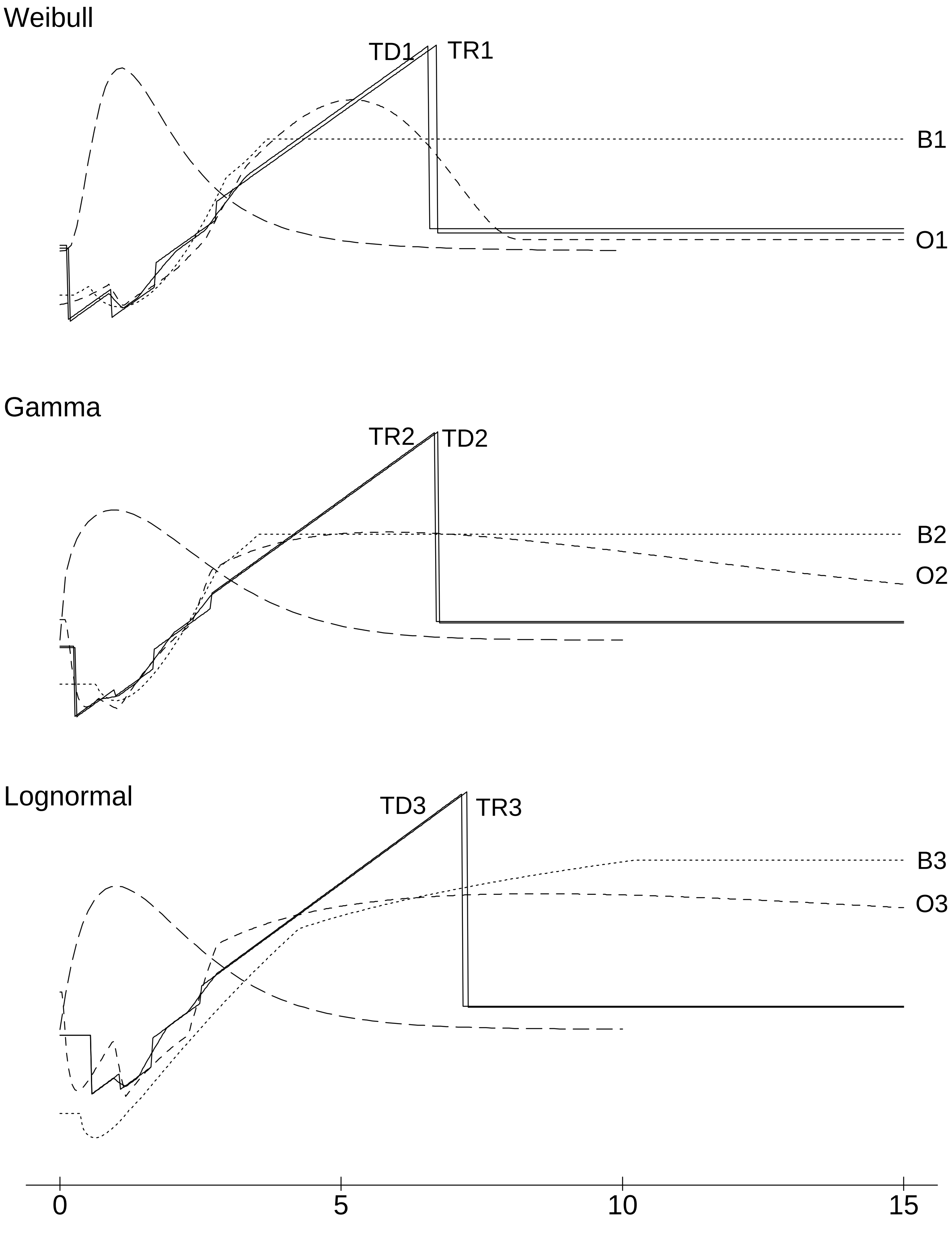

Figure 2. Influence functions (at the models) of truncated mean estimates (TD_j, TR_j), M-estimates (B_j), and one-step M-estimates O_j. Models: Weibull ($j = 1$, $\alpha = 1.435$, $\sigma = 2.203$), Gamma ($j = 2$, $\alpha = 2, \sigma = 1$), Lognormal ($j = 3, \lambda = 0.490, \sigma = 0.637$)). Densities are shown on the diagrams. Mean values = 2; variances = 2.

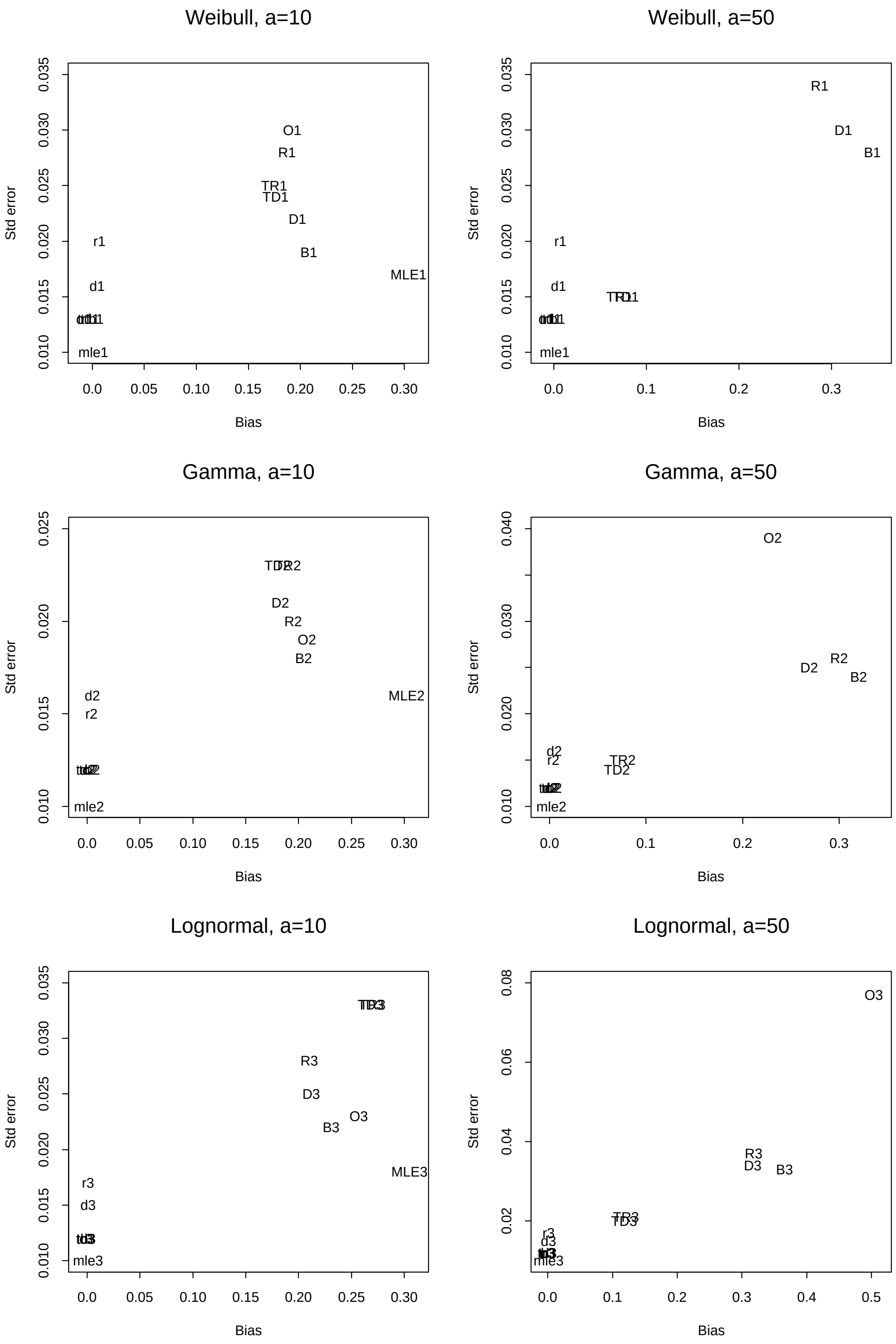

Figure 3. Scatter diagrams of average contamination bias versus standard error of various robust mean estimates. The numerical values of bias and variance are given in Table 2. Low cases indicate results obtained with $\epsilon = 0$; upper cases indicate results obtained with $\epsilon > 0$. Some of the results fall outside the frames.