Robust Estimators for Transformed Location Scale Families

David J. Olive *

Southern Illinois University

May 5, 2006

Abstract

In analogy with the method of moments, the parameters of a location scale family can be estimated robustly by equating the population and sample medians and median absolute deviations. Asymptotically efficient robust estimators can be made using the cross checking technique.

A robust method is also given for estimating the parameters of distributions that are simple transformations of location scale families. The lognormal, Weibull and Pareto distributions are used to illustrate the method for the log transformation.

KEY WORDS: Censored Data; Median Absolute Deviation; Method of Moments. Outliers; Reliability; Survival Analysis.

*David J. Olive, Department of Mathematics, Southern Illinois University, Mailcode 4408, Carbondale, IL 62901-4408, USA. E-mail address: dolive@math.siu.edu. This work was supported by the National Science Foundation

1 Introduction

The population median MED(Y) is a measure of location and the population median absolute deviation MAD(Y) is a measure of scale. The population median is any value MED(Y) such that

$$P(Y \leq \text{MED}(Y)) \geq 0.5 \text{ and } P(Y \geq \text{MED}(Y)) \geq 0.5 \tag{1}$$

while

$$\text{MAD}(Y) = \text{MED}(|Y - \text{MED}(Y)|). \tag{2}$$

Since MAD(Y) is a median distance, MAD(Y) is any value such that

$$P(Y \in [\text{MED}(Y) - \text{MAD}(Y), \text{MED}(Y) + \text{MAD}(Y)]) \geq 0.5,$$

and

$$P(Y \in (\text{MED}(Y) - \text{MAD}(Y), \text{MED}(Y) + \text{MAD}(Y)) \leq 0.5.$$

These population quantities can be estimated from the sample $Y_1, \ldots, Y_n$. Let $Y_{(1)} \leq \cdots \leq Y_{(n)}$ be the order statistics. The *sample median*

$$\text{MED}(n) = Y_{((n+1)/2)} \text{ if n is odd,} \tag{3}$$

$$\text{MED}(n) = \frac{Y_{(n/2)} + Y_{((n/2)+1)}}{2} \text{ if n is even.}$$

The notation $\text{MED}(n) = \text{MED}(Y_1, ..., Y_n) = \text{MED}(Y_i,\ i = 1, \ldots, n)$ will be useful for the following definition: the *sample median absolute deviation*

$$\text{MAD}(n) = \text{MED}(|Y_i - \text{MED}(n)|,\ i = 1, \ldots, n). \tag{4}$$

In analogy with the method of moments, robust point estimators can be obtained by solving $\text{MED}(n) = \text{MED}(Y)$ and $\text{MAD}(n) = \text{MAD}(Y)$. This procedure is simple if the distribution of Y is a 1 parameter family, a 2 parameter symmetric family or a location scale family. The method is also useful if Y is a 1 or 2 parameter family such that $W = t(Y)$ has a distribution that is from a location scale family. Estimate the parameters of Y using functions of $\text{MED}(W_1, ..., W_n)$ and $\text{MAD}(W_1, ..., W_n)$. Similar (but nonrobust) estimators can usually be obtained from the method of moments estimators based on $W_1, ..., W_n$ as long as the moments exist.

The above "MAD method" has been suggested by several authors when $t(Y) = Y$ is the the identity transformation, e.g., see Marazzi and Ruffieux (1999); however, the theory given in Falk (1997) showing that these estimators are asymptotically normal for asymmetric distributions is recent. The He and Fung (1999) "method of medians" is actually a very different procedure. These estimators are found by equating the sample median and population median for the score function of the model.

The He and Fung (1999) "cross checking" technique can be used to make asymptotically efficient robust estimators. The basic idea is to compute a robust estimator $(\hat{\theta}_R, \hat{\lambda}_R)$ and an asymptotically efficient estimator $(\hat{\theta}, \hat{\lambda})$. Then the cross checking estimator $(\hat{\theta}_C, \hat{\lambda}_C)$ is the efficient estimator if the robust and efficient estimators are "close", otherwise, $(\hat{\theta}_C, \hat{\lambda}_C)$ is the robust estimator. The cross checking estimator is asymptotically efficient, and tends to have greater outlier resistance than competing highly efficient robust estimators (such as M-estimators) if $(\hat{\theta}_R, \hat{\lambda}_R)$ is good.

Section 2 suggests simple highly outlier resistant estimators $(\hat{\theta}_R, \hat{\lambda}_R)$ for 21 "brand name" distributions. Section 3 shows how to use the estimators for right or left censored data, and Section 4 presents a simple proof of $\sqrt{n}$ and almost sure convergence of MAD(n).

2 Examples

The following remark is useful for computing MED(Y) and MAD(Y). Assume that Y has a continuous distribution with cumulative distribution function (cdf) $F(y) = P(Y \leq y)$. Let y_α be the α percentile of Y so that $F(y_\alpha) = \alpha$ for $0 < \alpha < 1$. Notice that y_α is found by solving $F(y_\alpha) = \alpha$ for y_α and that MED(Y) $= y_{0.5}$.

Remark 1. a) If Y has a probability density function (pdf) that is continuous and positive on its support and symmetric about μ, then MED(Y) $= \mu$ and MAD(Y) $= y_{0.75} -$ MED(Y).

b) Suppose that Y is from a location scale family with standard pdf $f_Z(z)$ that is continuous and positive on its support. Then $Y = \mu + \sigma Z$ where $\sigma > 0$. Let $M =$ MED(Z) and $D =$ MAD(Z). Then find M by solving $F_Z(M) = 0.5$ for M. After finding M, find D by solving $F_Z(M + D) - F_Z(M - D) = 0.5$ for D (often numerically). Then MED$(Y) = \mu + \sigma M$ and MAD$(Y) = \sigma D$.

The following examples illustrate the MAD method for the identity transformation.

Example 1. Suppose the $Y \sim N(\mu, \sigma^2)$. Then $Y = \mu + \sigma Z$ where $Z \sim N(0, 1)$. The standard normal random variable Z has a pdf that is symmetric about 0. Hence MED$(Z) = 0$ and MED$(Y) = \mu + \sigma$MED$(Z) = \mu$. Let $D =$ MAD(Z) and let $P(Z \leq z) = \Phi(z)$ be the cdf of Z. Remark 1a) implies that $D = z_{0.75} - 0 = z_{0.75}$ where $P(Z \leq z_{0.75}) = 0.75$. Numerically, $D \approx 0.6745$. Since MED$(Y) = \mu$ and MAD$(Y) \approx 0.6745\sigma$, $\hat{\mu} =$ MED(n) and $\hat{\sigma} \approx$ MAD$(n)/0.6745 \approx 1.483$MAD(n).

Example 2. If Y is exponential (λ), then the cdf of Y is $F_Y(y) = 1 - \exp(-y/\lambda)$ for $y > 0$ and $F_Y(y) = 0$ otherwise. Since $\exp(\log(1/2)) = \exp(-\log(2)) = 0.5$, MED$(Y) = \log(2)\lambda$. Since the exponential distribution is a scale family with scale parameter λ, MAD$(Y) = D\lambda$ for some $D > 0$. Hence

$$0.5 = F_Y(\log(2)\lambda + D\lambda) - F_Y(\log(2)\lambda - D\lambda), \quad \text{or}$$

$$0.5 = 1 - \exp[-(\log(2) + D)] - (1 - \exp[-(\log(2) - D)]) = \exp(-\log(2))[e^D - e^{-D}].$$

Thus $1 = \exp(D) - \exp(-D)$ which needs to be solved numerically.

Table 1 provides the pdf, MED(Y) and MAD(Y) (except for the power and TEV distributions) for the Cauchy C(μ, σ), double exponential DE(θ, λ), exponential EXP(λ), two parameter exponential EXP(θ, λ), half Cauchy HC(μ, σ), half logistic HL(μ, σ), half normal HN(μ, σ), largest extreme value LEV(θ, σ), logistic L(μ, σ), Maxwell Boltzmann MB(μ, σ), normal N(μ, σ^2), power POW(λ), Rayleigh R(μ, σ), smallest extreme value SEV(θ, σ), truncated extreme value TEV(λ) and uniform U(θ_1, θ_2) distributions. Table 2 provides the robust point estimators for these distributions.

All of the above two parameter distributions are location scale families and only the C(μ, σ), DE(θ, λ), L(μ, σ), N(μ, σ^2) and U(θ_1, θ_2) distributions are symmetric. The Table 2 results are well known for the N(μ, σ^2) family and Rousseeuw and Croux (1993) was

Table 1: MED(Y) and MAD(Y) for some useful random variables.

Name	$f(y)$	MED(Y)	MAD(Y)
$C(\mu, \sigma)$	$1/(\pi\sigma[1+(\frac{y-\mu}{\sigma})^2])$	μ	σ
$DE(\theta, \lambda)$	$\frac{1}{2\lambda}\exp(-\frac{\|y-\theta\|}{\lambda})$	θ	0.6931λ
$EXP(\lambda)$	$\frac{1}{\lambda}\exp(-\frac{y}{\lambda})\ I(y \geq 0)$	0.6931λ	$\lambda/2.0781$
$EXP(\theta, \lambda)$	$\frac{1}{\lambda}\exp(-\frac{(y-\theta)}{\lambda})\ I(y \geq \theta)$	$\theta + 0.6931\lambda$	$\lambda/2.0781$
$HC(\mu, \sigma)$	$\frac{2}{\pi\sigma[1+(\frac{y-\mu}{\sigma})^2]} I(y \geq \mu)$	$\mu + \sigma$	0.73205σ
$HL(\mu, \sigma)$	$\frac{2\exp(-(y-\mu)/\sigma)}{\sigma[1+\exp(-(y-\mu)/\sigma)]^2} I(y \geq \mu)$	$\mu + \log(3)\sigma$	0.67346σ
$HN(\mu, \sigma)$	$\frac{2}{\sqrt{2\pi}\ \sigma}\exp(\frac{-(y-\mu)^2}{2\sigma^2})\ I(y \geq \mu)$	$\mu + 0.6745\sigma$	$0.3991\ \sigma$
$LEV(\theta, \sigma)$	$\frac{1}{\sigma}\exp(-(\frac{y-\theta}{\sigma}))\exp[-\exp(-(\frac{y-\theta}{\sigma}))]$	$\theta + 0.3665\sigma$	0.767049σ
$L(\mu, \sigma)$	$\frac{\exp(-(y-\mu)/\sigma)}{\sigma[1+\exp(-(y-\mu)/\sigma)]^2}$	μ	$1.0986\ \sigma$
$MB(\mu, \sigma)$	$\frac{\sqrt{2}(y-\mu)^2\exp(\frac{-1}{2\sigma^2}(y-\mu)^2)}{\sigma^3\sqrt{\pi}} I(y \geq \mu)$	$\mu + 1.5381722\sigma$	0.460244σ
$N(\mu, \sigma^2)$	$\frac{1}{\sqrt{2\pi\sigma^2}}\exp(\frac{-(y-\mu)^2}{2\sigma^2})$	μ	0.6745σ
$POW(\lambda)$	$\frac{1}{\lambda}y^{\frac{1}{\lambda}-1}I(0 < y < 1)$	$(0.5)^\lambda$	MAD(Y)
$R(\mu, \sigma)$	$\frac{y-\mu}{\sigma^2}\exp\left[-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right]\ I(y \geq \mu)$	$\mu + 1.1774\sigma$	0.4485σ
$SEV(\theta, \sigma)$	$\frac{1}{\sigma}\exp(\frac{w-\theta}{\sigma})\exp[-\exp(\frac{w-\theta}{\sigma})]$	$\theta - 0.3665\sigma$	0.767049σ
$TEV(\lambda)$	$\frac{1}{\lambda}\exp\left(y - \frac{e^y-1}{\lambda}\right) I(y > 0)$	$\log(1 + \lambda\log(2))$	MAD(Y)
$U(\theta_1, \theta_2)$	$\frac{1}{\theta_2-\theta_1} I(\theta_1 \leq y \leq \theta_2)$	$(\theta_1 + \theta_2)/2$	$(\theta_2 - \theta_1)/4$

Table 2: Robust point estimators for some useful random variables.

$C(\mu,\sigma)$	$\hat{\mu} = \text{MED}(n)$	$\hat{\sigma} = \text{MAD}(n)$
$DE(\theta,\lambda)$	$\hat{\theta} = \text{MED}(n)$	$\hat{\lambda} = 1.443\ \text{MAD}(n)$
$EXP(\lambda)$	$\hat{\lambda}_1 = 1.443\ \text{MED}(n)$	$\hat{\lambda}_2 = 2.0781\ \text{MAD}(n)$
$EXP(\theta,\lambda)$	$\hat{\theta} = \text{MED}(n) - 1.440\ \text{MAD}(n)$	$\hat{\lambda} = 2.0781\ \text{MAD}(n)$
$HC(\mu,\sigma)$	$\hat{\mu} = \text{MED}(n) -\ 1.3660\text{MAD}(n)$	$\hat{\sigma} =\ 1.3660\text{MAD}(n)$
$HL(\mu,\sigma)$	$\hat{\mu} = \text{MED}(n) -\ 1.6313\text{MAD}(n)$	$\hat{\sigma} =\ 1.4849\text{MAD}(n)$
$HN(\mu,\sigma)$	$\hat{\mu} = \text{MED}(n) -\ 1.6901\text{MAD}(n)$	$\hat{\sigma} = 2.5057\ \text{MAD}(n)$
$LEV(\theta,\sigma)$	$\hat{\theta} = \text{MED}(n) - 0.4778\ \text{MAD}(n)$	$\hat{\sigma} = 1.3037\ \text{MAD}(n)$
$L(\mu,\sigma)$	$\hat{\mu} = \text{MED}(n)$	$\hat{\sigma} = 0.9102\ \text{MAD}(n)$
$MB(\mu,\sigma)$	$\hat{\mu} = \text{MED}(n) - 3.3421\ \text{MAD}(n)$	$\hat{\sigma} = 2.17276\ \text{MAD}(n)$
$N(\mu,\sigma^2)$	$\hat{\mu} = \text{MED}(n)$	$\hat{\sigma} = 1.483\ \text{MAD}(n)$
$POW(\lambda)$	$\hat{\lambda} = \log(\text{MED}(n))/\log(0.5)$	
$R(\mu,\sigma)$	$\hat{\mu} = \text{MED}(n) - 2.6255\ \text{MAD}(n)$	$\hat{\sigma} = 2.230\ \text{MAD}(n)$
$SEV(\theta,\sigma)$	$\hat{\theta} = \text{MED}(n) + 0.4778\text{MAD}(n)$	$\hat{\sigma} = 1.3037\text{MAD}(n)$
$TEV(\lambda)$	$\hat{\lambda} = [\exp(\text{MED}(n)) - 1]/\log(2)$	
$U(\theta_1,\theta_2)$	$\hat{\theta}_1 = \text{MED}(n) - 2\ \text{MAD}(n)$	$\hat{\theta}_2 = \text{MED}(n) + 2\ \text{MAD}(n)$

used for the EXP(θ, λ) family. Pewsey (2002) provides maximum likelihood estimators for the HN(μ, σ) family.

Next, 5 examples of the MAD method are given when Y has a distribution such that $W = t(Y) = \log(Y)$ has a location scale family. McKane, Escobar, and Meeker (2005) say that Y has a log-location-scale distribution.

Example 3. If Y has a lognormal (μ, σ^2) distribution, then $W = \log(Y) \sim N(\mu, \sigma^2)$. Thus $\hat{\mu} = \text{MED}(W_1, ..., W_n)$ and $\hat{\sigma} = 1.483\text{MAD}(W_1, ..., W_n)$. This estimator is also given by He and Fung (1999). See Toma (2003) for related methods and Serfling (2002) for M-estimators.

Example 4. Suppose that Y has a Pareto(σ, λ) distribution with pdf

$$f(y) = \frac{\frac{1}{\lambda}\sigma^{1/\lambda}}{y^{1+1/\lambda}}$$

where $y \geq \sigma$, $\sigma > 0$, and $\lambda > 0$. Then $W = \log(Y) \sim EXP(\theta = \log(\sigma), \lambda)$. Let $\hat{\theta} = \text{MED}(W_1, ..., W_n) - 1.440\text{MAD}(W_1, ..., W_n)$. Then $\hat{\sigma} = \exp(\hat{\theta})$ and

$$\hat{\lambda} = 2.0781 \ \text{MAD}(W_1, ..., W_n).$$

See Brazauskas and Serfling (2000,2001) for alternative robust estimators.

Example 5. Suppose that Y has a Weibull (ϕ, λ) distribution with pdf

$$f(y) = \frac{\phi}{\lambda} y^{\phi - 1} e^{-\frac{y^\phi}{\lambda}}$$

where λ, y, and ϕ are all positive. Then $W = \log(Y)$ has a SEV$(\theta = \log(\lambda^{1/\phi}), \sigma = 1/\phi)$ distribution, also known as a log-Weibull distribution. Notice that $-W \sim LEV(-\theta, \sigma)$, also known as a Gumbel distribution. Let $\hat{\sigma} = \text{MAD}(W_1, ..., W_n)/0.767049$ and let $\hat{\theta} = \text{MED}(W_1, ..., W_n) - \log(\log(2))\hat{\sigma}$. Then $\hat{\phi} = 1/\hat{\sigma}$ and $\hat{\lambda} = \exp(\hat{\theta}/\hat{\sigma})$. See He and Fung (1999), Chandra and Chaudhuri (1990), Seki and Yokoyama (1996) and Smith (1977) for alternative simple or robust estimators.

Example 6. Suppose that Y has a log–Cauchy(μ, σ) distribution with pdf

$$f(y) = \frac{1}{\pi \sigma y [1 + (\frac{\log(y) - \mu}{\sigma})^2]}$$

where $y > 0$, $\sigma > 0$ and μ is a real number. See McDonald (1987). Then $W = \log(Y)$ has a Cauchy(μ, σ) distribution. Let $\hat{\mu} = \text{MED}(W_1, ..., W_n)$ and $\hat{\sigma} = \text{MAD}(W_1, ..., W_n)$.

Example 7. Suppose that Y has a log–logistic(ϕ, τ) distribution with pdf and cdf

$$f(y) = \frac{\phi\tau(\phi y)^{\tau-1}}{[1+(\phi y)^{\tau}]^2} \quad \text{and} \quad \mathrm{F(y)} = 1 - \frac{1}{1+(\phi \mathrm{y})^{\tau}}$$

where $y > 0$, $\phi > 0$ and $\tau > 0$. Then $W = \log(Y)$ has a logistic$(\mu = -\log(\phi), \sigma = 1/\tau)$ distribution. Hence $\phi = e^{-\mu}$ and $\tau = 1/\sigma$. See Kalbfleisch and Prentice (1980, pp. 27-28). Then $\hat{\tau} = \log(3)/\text{MAD}(W_1, ..., W_n)$ and $\hat{\phi} = 1/\text{MED}(Y_1, ..., Y_n)$ since $\text{MED}(Y) = 1/\phi$.

3 Robust Estimation for Censored Data

As noted in He and Fung (1999), estimators based on the median may not need to be adjusted if there are some possibly right or left censored observations present. Suppose that in a reliability study the Y_i are failure times and the study lasts for T hours. Let $Y_{(R)} < T$ but $T < Y_{(R+1)} < \cdots < Y_{(n)}$ so that only the first R failure times are known and the last $n - R$ failure times are unknown but greater than T (similar results hold if the first L failure times are less than T but unknown while the failure times $T < Y_{(L+1)} < \cdots < Y_{(n)}$ are known). Then create a pseudo sample $Z_{(i)} = Y_{(R)}$ for $i > R$ and $Z_{(i)} = Y_{(i)}$ for $i \leq R$. Then compute the robust estimators based on $Z_1, ..., Z_n$. These estimators will be identical to the estimators based on $Y_1, ..., Y_n$ (no censoring) if the amount of right censoring is moderate. For a one parameter family, nearly half of the data can be right censored if the estimator is based on the median. If the sample median and MAD are used for a two parameter family, the proportion of right censored data depends on the skewness of the distribution. Symmetric data can tolerate nearly 25% right censoring, right skewed data a larger percentage, and left skewed data a smaller percentage.

Tables 3 and 4 present the results from a small simulation study. For Table 3, samples of size $n = 100$ Weibull (ϕ, λ) observations $Y_1, ..., Y_n$ were generated. Then the pseudo data $Z_{(1)}, ..., Z_{(100)}$ was created by replacing $Y_{(86)}, ..., Y_{(100)}$ by $Y_{(85)}$. Then $\hat{\phi}$ and $\hat{\lambda}$ were estimated as in Example 5 using $W_i = \log(Z_i)$. Only 15% of the cases were right censored since the SEV distribution has a longer left tail than right. The means and standard deviations from 500 runs are given. Notice that $\hat{\phi} \approx \phi$ and that the bias and variability of $\hat{\lambda}$ increases as λ increases.

Table 3: Results for Right Censored Weibull Data

ϕ	λ	mean($\hat{\phi}$)	SD($\hat{\phi}$)	mean($\hat{\lambda}$)	SD($\hat{\lambda}$)
1	1	1.0130	0.1216	1.0070	0.1277
1	5	1.0235	0.1167	5.3301	1.2475
1	10	1.0211	0.1240	11.0228	3.6296
1	20	1.0240	0.1313	23.6023	11.2156
20	1	20.4128	2.3712	1.0091	0.1515
20	5	20.6121	2.7800	5.4354	1.5706
20	10	20.4656	2.5651	11.2070	4.5380
20	20	20.5479	2.7082	23.8544	13.3203

Table 4: Results for Right Censored Pareto Data

σ	λ	mean($\hat{\sigma}$)	SD($\hat{\sigma}$)	mean($\hat{\lambda}$)	SD($\hat{\lambda}$)
1	1	1.0088	0.0619	0.9903	0.1421
1	5	1.1220	0.4119	4.9413	0.6896
1	10	1.4348	1.3947	9.9577	1.4640
1	15	1.9633	3.7377	15.0809	2.2010
1	20	3.3504	8.0171	19.9893	2.8479

For Table 4, samples of size $n = 100$ Pareto (σ, λ) observations $Y_1, ..., Y_n$ were generated. Then the pseudo data $Z_{(1)}, ..., Z_{(100)}$ was created by replacing $Y_{(76)}, ..., Y_{(100)}$ by $Y_{(75)}$. Then $\hat{\sigma}$ and $\hat{\lambda}$ were estimated as in Example 4 using $W_i = \log(Z_i)$. The means and standard deviations from 500 runs are given. Notice that $\hat{\lambda} \approx \lambda$ and that the bias and variability of $\hat{\sigma}$ increases as λ increases.

4 Theory for the MAD

For theory, the following quantity will be useful:

$$\text{MD}(n) = \text{MED}(|Y_i - \text{MED}(Y)|,\ i = 1, \ldots, n).$$

Since MD(n) is a median and convergence results for the median are well known, see for example Serfling (1980 pp. 74-77), it is simple to prove convergence results for MAD(n). Serfling (1980, pp. 8-9) defines W_n to be *bounded in probability,* $W_n = O_P(1)$, if for every $\epsilon > 0$ there exist positive constants D_ϵ and N_ϵ such that

$$P(|W_n| > D_\epsilon) < \epsilon$$

for all $n \geq N_\epsilon$, and $W_n = O_P(n^{-\delta})$ if $n^\delta W_n = O_P(1)$. Typically $\text{MED}(n) = \text{MED}(Y) + O_P(n^{-1/2})$ and $\text{MAD}(n) = \text{MAD}(Y) + O_P(n^{-1/2})$. Equation (5) in the proof of the following proposition implies that if MED(n) converges to MED(Y) almost surely and MD(n) converges to MAD(Y) almost surely, then MAD(n) converges to MAD(Y) almost surely. Almost sure convergence of MAD(n) was also proven by Hall and Welsh (1985) while Falk (1997) showed that the joint distribution of MED(n) and MAD(n) is asymptotically normal. The following proposition gives a weaker result, but the proof is much simpler since the theory of empirical processes is avoided.

Proposition 1. If $\text{MED}(n) = \text{MED}(Y) + O_P(n^{-\delta})$ and $\text{MD}(n) = \text{MAD}(Y) + O_P(n^{-\delta})$, then $\text{MAD}(n) = \text{MAD}(Y) + O_P(n^{-\delta})$.

Proof. Let $W_i = |Y_i - \text{MED}(n)|$ and let $V_i = |Y_i - \text{MED}(Y)|$. Then

$$W_i = |Y_i - \text{MED}(Y) + \text{MED}(Y) - \text{MED}(n)| \leq V_i + |\text{MED}(Y) - \text{MED}(n)|,$$

and $\text{MAD}(n) = \text{MED}(W_1, \ldots, W_n) \leq \text{MED}(V_1, \ldots, V_n) + |\text{MED}(Y) - \text{MED}(n)|$. Similarly

$$V_i = |Y_i - \text{MED}(n) + \text{MED}(n) - \text{MED}(Y)| \leq W_i + |\text{MED}(n) - \text{MED}(Y)|$$

and thus $\text{MD}(n) = \text{MED}(V_1, \ldots, V_n) \leq \text{MED}(W_1, \ldots, W_n) + |\text{MED}(Y) - \text{MED}(n)|$. Combining the two inequalities shows that

$$\text{MD}(n) - |\text{MED}(Y) - \text{MED}(n)| \leq \text{MAD}(n) \leq \text{MD}(n) + |\text{MED}(Y) - \text{MED}(n)|,$$

or

$$|\text{MAD}(n) - \text{MD}(n)| \leq |\text{MED}(n) - \text{MED}(Y)|. \tag{5}$$

Adding and subtracting $\text{MAD}(Y)$ to the left hand side shows that

$$|\text{MAD}(n) - \text{MAD}(Y) - O_P(n^{-\delta})| = O_P(n^{-\delta}) \tag{6}$$

and the result follows. QED

5 Conclusions

The robust methods presented in this paper can be computed in closed form for a wide variety of distributions. They tend to be asymptotically normal with high outlier resistance. A promising two stage estimator is the cross checking estimator that uses an asymptotically efficient estimator if it is close to the robust estimator but uses the robust estimator otherwise. If the robust estimator is a high breakdown consistent estimator, then the cross checking estimator is asymptotically efficient and also has high breakdown. The bias of the cross checking estimator is greater than that of the robust estimator since the probability that the robust estimator is chosen when outliers are present is less than one. However, few two stage estimators will have performance that rivals the statistical properties and simplicity of the cross checking estimator. See He and Fung (1999).

Since the cross checking estimator is asymptotically efficient, the robust estimator should be highly outlier resistant. If the data are iid from a location scale family, the robust estimators of location and scale based on the sample median and median absolute deviation should be $\sqrt{n}$ consistent and should have very high resistance to outliers. An

M-estimator, for example, will have both lower efficiency and outlier resistance than the cross checking estimator.

Hence for location scale families and for the lognormal family, the cross checking estimator based on the methods in this paper is in some sense optimal. For the Weibull distribution, there may exist robust estimators that are more resistant than the method suggested in this paper. In this case the more resistant method should be used in the cross checking estimator.

The methods described in this paper can be used to detect outliers and as starting values for iterative methods such as maximum likelihood or M-estimators even if the data is left or right censored. Since they are simple to compute, software manufacturers could use the simple robust estimators as initial values and then print a warning that the parametric model may not hold if the simple and final estimators disagree greatly.

Robust methods are not often used by applied statisticians, perhaps because they are often difficult to compute and understand. Often applied statisticians will use maximum likelihood and then attempt to find outliers graphically. This procedure is very useful for detecting gross outliers but can fail for moderate outliers. The estimators presented in this paper can be used by statistical consultants to demonstrate the merits of using robust estimators. Then the clients may be persuaded to use more complex but also more efficient alternative robust estimators.

6 References

Brazauskas, V. and Serfling, R., Robust estimation of tail parameters for two-parameter Pareto and exponential models via generalized quantile statistics, Extremes 3, 231-249, (2000).

Brazauskas, V. and Serfling, R., Small sample performance of robust estimators of tail parameters for Pareto and exponential models, J. Statist. Computat. Simulat. 70, 1–19, (2001).

Chandra, N.K. and Chaudhuri, A., On the efficency of a testimator for the Weibull shape parameter, Commun. Statist. Theory Methods 19, 1247–1259, (1990).

Falk, M., Asymptotic independence of median and mad, Statist. Probab. Lett. 34, 341–345, (1997).

Hall, P. and Welsh, A.H., Limit theorems for the median deviation, Annals Instit. Statist. Mathematics, Part A 37, 27–36, (1985).

He, X. and Fung, W.K., Method of medians for lifetime data with Weibull models, Statistics Medicine 18, 1993–2009, (1999).

Kalbfleisch, J.D. and Prentice, R.L., The Statistical Analysis of Failure Time Data, John Wiley and Sons, New York, (1980).

Marazzi, A. and Ruffieux, C., The truncated mean of an asymmetric distribution, Computat. Statist. Data Analys. 32, 79-100, (1999).

McKane, S.W., Escobar, L.A. and Meeker, W.Q., Sample size and number of failure requirements for demonstration tests with log-location-scale distributions and failure censoring, Technom. 47, 182–190, (2005).

McDonald, J.B., Model selection: some generalized distributions, Commun. Statist. Theory Methods 16, 1049–1074, (1987).

Pewsey, A., Large-sample inference for the half-normal distribution, Commun. Statist. Theory Methods 31, 1045–1054, (2002).

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

Seki, T. and Yokoyama, S., Robust parameter-estimation using the bootstrap method for the 2-parameter Weibull distribution, IEEE Transactions Reliability 45, 34–41, (1996).

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

Serfling, R., Efficient and robust fitting of lognormal distributions, North Amer. Actuarial Journal 6, 95–109, (2002).

Smith, R.M., Some results on interval estimation for the two–parameter Weibull or extreme–value distribution, Communic. Statist. Theory Methods A6, 1311–1321, (1977).

Toma, A., Robust estimators for the parameters of multivariate lognormal distribution, Commun. Statist. Theory Methods 32, 1405–1417, (2003).