A Small Sample Bias Correction and Implications for Inference

Brenton R Clarke[1*] and Christopher J Milne[2]

[1]Mathematics and Statistics, School of Engineering
and Information Technology,
Murdoch University, Murdoch, WA 6150, Australia

[2]Senior Business Analyst, Commercial Section, Verve Energy,
Level 11 Australia Place, Perth, WA 6000, Australia

* Corresponding author: Brenton R Clarke, email: B.Clarke@murdoch.edu.au

Abstract

The most popular and perhaps universal estimator of location and scale in robust estimation, where one accepts that ideally we have a normal population, but wish to guard against possible small departures from such, is Huber's Proposal-2 M-estimator. We outline the first order small sample bias correction for the scale estimator, which has been verified both through theory and simulation. While there may be other ways of reducing small sample bias, say as in jackknifing or bootstrapping, these can be computationally intensive, and would not be routinely used with this iteratively derived estimator. It is suggested that bias reduced estimates of scale are most useful when forming confidence intervals for location and or scale based on the asymptotic distribution. In this paper we expand on the results of an earlier work by the authors to include Hampel's three part re-descending psi function (with a three part re-descender for scale).

Keywords and phrases: M-estimators, location and scale estimation

1 Introduction

In a relatively recent article by Clarke and Milne (2004) the authors outlined the steps to calculating the small sample bias of Huber's Proposal 2 scale estimator. Somewhat surprisingly, while much effort has been invested in the small sample bias of variance estimators, see Cabrera & Fernholz (1999) and De Rossi & Gatto (2001), there has not been a great deal of comment on the small sample bias of the scale estimate, despite the fact that this appears in asymptotic confidence intervals, for instance in the description of confidence intervals for the location parameter. To repeat the story, we remind ourselves of the Huber's Proposal-2 M-estimator first given in Huber (1964), where estimators $\hat{\mu}$ and $\hat{\sigma}$ are solutions of equations

$$\boldsymbol{K_n}(\hat{\mu}, \hat{\sigma}) = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\psi}\left(\frac{X_i - \hat{\mu}}{\hat{\sigma}}\right) = 0. \tag{1.1}$$

Here $X_1, X_2, ..., X_n$ represent independent identically distributed random variables having the normal distribution with mean μ and standard deviation σ and $\boldsymbol{\psi} = (\psi_1, \psi_2)$ is a vector function defined by

$$\psi_1(x) = \min(k, \max(x, -k))$$

$$\psi_2(x) = \min(k^2 \ - \ \beta, x^2 \ - \beta).$$

Hence for example $\boldsymbol{K_n} = (K_{n1}, K_{n2})$ is a two component vector function. The term k appearing in the formula for $\boldsymbol{\psi}$ is a tuning constant and β satisfies $\int \psi_2(x) d\Phi(x) = 0$, where Φ denotes the standard normal cumulative distribution function. For example, the choice of $k = +\infty$ yields the maximum likelihood equations for a normal parametric family defined by $\boldsymbol{\psi}(x) \ = \ (x, -1 + x^2)$. Robust choices of k vary: popular choices being values such as k from the set of $\{1, 1.285, 1.5, 1.645, 1.96\}$. For example, choosing a value of $k = 1.96$ has the interpretation that asymptotically 5% of the data is winsorized leading to an asymptotic variance of a location estimator of 1.0116 when data are generated from a standard normal distribution. See Table 1 of Clarke and Milne (2004) for corresponding variances of location and scale for different values of k. In the deliberations in this paper we assume at the very least that the underlying distribution is symmetric (as does Huber). In actual calculations of bias we revert to the assumption that the data are normal. This assumption is usually challenged in typical robustness studies, but the bias calculations in small symmetric deviations from normal appear to be robust in the sense that they vary continuously with small departures from the normal distribution when $k < +\infty$.

If $\hat{\boldsymbol{\theta}} = (\hat{\mu}, \hat{\sigma})$ is a solution to equations (1.1) for suitably smooth $\boldsymbol{\psi}$, we may assume the bias determined through $\boldsymbol{b_\psi(\theta)} = \mathrm{E}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\theta}$, where E represents expectation with respect to the underlying population, has the following expansion:

$$\boldsymbol{b_\psi(\theta)} = \frac{\boldsymbol{B_1(\theta)}}{n} \ + \ \frac{\boldsymbol{B_2(\theta)}}{n^2} \ + \ \boldsymbol{o}(\frac{1}{n^2}) \tag{1.2}$$

For instance, here $\boldsymbol{\theta} = (\mu, \sigma)$. From symmetry considerations it follows that $E(\hat{\mu}) = \mu$, however there is a non-zero bias in the estimation of scale. To illustrate the bias calculations, it is well known that when estimating variance via the maximum likelihood estimator(MLE) $\hat{\sigma}^2 = (1/n)\sum_{i=1}^n (X_i - \bar{X})^2$, and this statistic has a bias of $-\sigma^2/n$. Consequently a bias-corrected estimator of σ^2 is $s_n^2 = (1/(n-1))\sum_{i=1}^n (X_i - \bar{X})^2$. Here $\bar{X}$ is the sample average corresponding to the maximum likelihood estimator of location. What is perhaps not so well known is that for the normal parametric family the maximum likelihood estimator for scale $\hat{\sigma} = \{\hat{\sigma}^2\}^{\frac{1}{2}}$ has a first order bias of $-\frac{3}{4}\sigma/n$. That is,

$$\mathrm{E}(\hat{\sigma} - \sigma) \approx \ -\frac{3}{4}\frac{\sigma}{n},$$

ignoring second order bias involving $1/n^2$. See Clarke and Milne (2004) for an easy derivation. A bias reduced estimator of scale in the case of maximum likelihood estimation from a normal population is thus $(n/(n - \frac{3}{4}))\hat{\sigma}$.

Clarke and Milne (2004) also establish the following bias calculation for the more general M-Estimator given as a solution of equation (1.1).

$$\begin{aligned} \mathrm{E}_{\mu,\sigma}(\hat{\sigma} - \sigma) \ = \ & \frac{\sigma}{n} \frac{1}{\mathrm{E}(Z\psi_2')} \left(-\frac{\mathrm{E}(\psi_2'\psi_1)}{\mathrm{E}(\psi_1')} - \frac{\mathrm{E}(Z\psi_2'\psi_2)}{\mathrm{E}(Z\psi_2')} + \right. \\ & \left. \frac{\mathrm{E}(\psi_1^2)\mathrm{E}(\psi_2'')}{2\mathrm{E}(\psi_1')^2} \ + \ \frac{\mathrm{E}(\psi_2^2)\left(2\mathrm{E}(Z\psi_2') + \mathrm{E}(Z^2\psi_2'')\right)}{2\mathrm{E}(Z\psi_2')^2} \right) \\ & + O(n^{-2}) \end{aligned} \tag{1.3}$$

Here Z is the standardized variable and E represents expectation with respect to the standardized distribution, in this case for example $\mathrm{E}(Z\psi_2') = \int x\psi_2'(x)d\Phi(x)$. Also $\mathrm{E}_{\mu,\sigma}$ is the expectation with respect to the unstandardized distribution.

Also (1.3) must be interpreted at least heuristically for ψ functions which do not have continuous derivatives, as in the case of Huber's Proposal-2 with a finite tuning constant k. Such functions are at least continuous and piecewise continuously differentiable. For example calculations we refer to Huber (1964, p. 78), Huber (1970, p. 462) and Hampel et al. (1986, p. 103). For instance $\mathrm{E}(\psi_2'')$ and $\mathrm{E}(Z^2\psi_2'')$ need to be interpreted this way in formula (1.3).

If $\hat{\sigma}$ is the scale solution to equations (1.1) it follows that a bias corrected estimator of scale is $\hat{\sigma}^* = (n/(n+b))\hat{\sigma}$, where b denotes the bias parameter for example calculated in Table 1. Clarke and Milne (2004) in their Table 2 verify the calculations of bias through simulation. They also explain that using the asymptotic distribution of the M-estimator leads to a 95% confidence interval for location of

$$(\hat{\mu} \; - \; 1.96\frac{\hat{\sigma}}{\sqrt{n}}\lambda, \;\; \hat{\mu} \; + \; 1.96\frac{\hat{\sigma}}{\sqrt{n}}\lambda), \;\; \text{where } \lambda^2 \; = \; \frac{\mathrm{E}(\psi_1^2)}{\mathrm{E}(\psi_1')^2}$$

This confidence interval involves the estimate of σ. The use of the bias-corrected scale estimator $\hat{\sigma}^*$ leads to confidence levels closer to nominal values. It is recognized that this simulation implements the asymptotic confidence interval, rather than an interval such as that obtained using a t-distribution. As the limits of the interval using a t-distribution are always wider than those obtained from the normal distribution, the actual levels are less than the nominal 95%. The simulations in Clarke and Milne (2004) generally show an improvement in coverage if one uses the adjusted scale estimate in the corresponding confidence interval for scale.

2 M-Estimators from Redescending Psi Functions

Hampel (1974) introduced the Hampel three part re-descender for location arguing that observations beyond a rejection point should not be given any weight in the estimating equations. The estimator is also found in Andrews et al. (1972). In essence the three part re-descender for location is governed by the equation depending on three tuning constants so that

$$\psi_{1;a,b,c}(x) \; = \begin{cases} x & \text{for } 0 \le |x| \le a \\ a \; sign(x) & \text{for } a \le |x| \le b \\ a\frac{c-|x|}{c-b} sign(x) & \text{for } b \le |x| \le c \\ 0 & \text{for } c \le |x| \end{cases}$$

The estimator for scale suggested by Hampel did not make use of the psi function as in equations (1.1) but was rather an alternative consistent estimator scale based on MAD $= med_i|X_i - med_jX_j|$. However it is possible to construct an M-estimator of scale, along the lines of a three part re-descender in the following form.

$$\psi_{2;a,b,c}(x) \; = \begin{cases} x^2 - 1 - p & \text{for } 0 \le |x| \le a \\ a^2 - 1 - p & \text{for } a \le |x| \le b \\ (a^2 - 1 - p)\frac{c-|x|}{c-b} & \text{for } b \le |x| \le c \\ 0 & \text{for } c \le |x| \; , \end{cases}$$

a	b	c	First Order Bias Parameter For Scale	Asymptotic Variance of Location	Asymptotic Variance of Scale
1.285	1.96	2.575	-1.3529	1.2182	1.1493
1.31	2.039	2.575	-1.2938	1.2013	1.1000
1.31	2.039	4	-0.8435	1.0966	0.8747
1.31	2.575	3.5	-0.8039	1.0802	0.8381
1.5	2.5	3.5	-0.7857	1.0637	0.7513
1.645	∞	∞	-0.6357	1.0262	0.6402
1.645	2	3.3	-0.9138	1.0942	0.7841
1.645	2.24	3.3	-0.8567	1.0751	0.7461
1.645	2.4	4	-0.7368	1.0466	0.6874
1.96	∞	∞	-0.6397	1.0116	0.5710
1.96	2.4	3.3	-0.8183	1.0500	0.6542
1.96	2.575	4	-0.7144	1.0259	0.6050
∞	∞	∞	-0.7500	1.0000	0.5000

Table 1: *First order bias term, with associated asymptotic variances of location and scale, for joint three part re-descender at the standard normal distribution.*

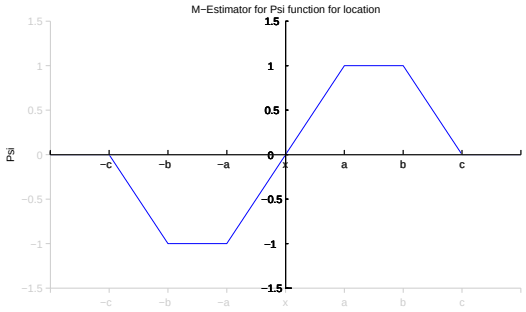

Figure 1: Plot of Hampel ψ_1 re-descender for location

where p is defined implicitly from the equation $\int \psi_{2;a,b,c}(x)d\Phi(x) = 0$. Arguments such as in Clarke (1986) which deal with piecewise differentiable continuous functions can be used to establish that there exists a Fréchet differentiable M-functional which leads to a consistent asymptotically normal root of equations (1.1). In Table 1 we detail the subsequent bias and asymptotic variances of estimates at some potential tuning constants.

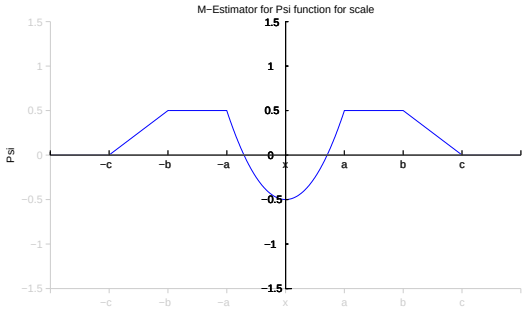

Figure 2: Plot of Hampel ψ_2 re-descender for scale

References

Andrews, D.F., Bickel, P.J., Hampel, F.R., Huber, P.J., Rogers, W.H. and Tukey, J.W. (1972) *Robust Estimates of Location: Survey and Advances.* Princeton University Press, Princeton, N.J.

Cabrera, J. and Fernholz, L.T. (1999) "Target estimation for bias and mean square error reduction," *Annals of Statistics*, **27**, 1080-1104.

Clarke, B.R. (1986) "Nonsmooth analysis and Fréchet differentiability of M-functionals," *Probab. Th. Re. Fields*, **73**, 197-209.

Clarke, B.R. and Milne, C.J. (2004) "Small sample bias correction for Huber's Proposal-2 scale M-Estimator," *Aust. N.Z. J. Stat.*, **46**, 649-656.

De Rossi, F. and Gatto, R. (2001) "Higher order expansions for robust tests," *Biometrika*, **88** 1153-1168.

Hampel,F.R. (1974) "The influence curve and its role in robust estimation," *J. Am. Statist. Assoc.*, **69**, 383-393.

Hampel,F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A. (1986) *Robust Statistics: The Approach Based on Influence Functions.* Wiley, New York.

Huber, P.J. (1964) "Robust estimation of a location parameter," *Ann. Math. Statist.*, **35**, 73-101.

Huber, P.J. (1970) "Studentizing robust estimators," In *Non-parametric Techniques in Statistical Inference.* Madan lal Puri(Ed.), Cambridge University Press, Cambridge, England. 453-463.