Journal of Statistical Planning and Inference 139 (2009) 350 – 365

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Trimmed and Winsorized means based on a scaled deviation[☆]

Mingxin Wu[a,*], Yijun Zuo[b]

[a]Risk Manager, American Express, Mailbox 40-03-04, 2512 W Dunlap Ave, Phoenix, AZ 85021, USA
[b]Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824, USA

ARTICLE INFO

Article history:
Received 7 October 2005
Received in revised form
14 March 2008
Accepted 17 March 2008
Available online 17 May 2008

MSC:
primary 62G05
secondary 62H12
62G35

Keywords:
Trimmed mean
Winsorized mean
Influence function
Breakdown point
Robustness
Efficiency
Asymptotic representation
Limiting distribution

ABSTRACT

Trimmed (and Winsorized) means based on a scaled deviation are introduced and studied. The influence functions of the estimators are derived and their limiting distributions are established via asymptotic representations. As a main focus of the paper, the performance of the estimators with respect to various robustness and efficiency criteria is evaluated and compared with leading competitors including the ordinary Tukey trimmed (and Winsorized) means. Unlike the Tukey trimming which always trims a fixed fraction of sample points at each end of data, the trimming scheme here only trims points at one or both ends that have a scaled deviation beyond some threshold. The resulting trimmed (and Winsorized) means are much more robust than their predecessors. Indeed they can share the best breakdown point robustness of the sample median for any common trimming thresholds. Furthermore, for appropriate trimming thresholds they are highly efficient at light-tailed symmetric models and more efficient than their predecessors at heavy-tailed or contaminated symmetric models. Detailed comparisons with leading competitors on various robustness and efficiency aspects reveal that the scaled deviation trimmed (Winsorized) means behave very well overall and consequently represent very favorable alternatives to the ordinary trimmed (Winsorized) means.

1. Introduction

Tukey trimmed (and Winsorized) means are among the most popular estimators of location parameter; see, e.g., Stigler (1977). They overcome the extreme sensitivity of the mean while improving the efficiency of the median at light tailed distributions. The robustness and efficiency are two fundamentally desirable properties of any statistical procedure. They, however, do not work in tandem in general. The trimmed (and Winsorized) means somehow can keep a quite good balance between the two. Tukey trimming scheme is a symmetric one in the sense that it trims the same number of sample points at both ends of data and hence is quite efficient for symmetric distributions. It, however, becomes less efficient when there is even just a slight departure from symmetry, e.g., with one end containing outlying points. Metrical trimming, introduced in Bickel (1965), trims points based on their distance to the center—median and hence is more efficient at contaminated symmetric models. Like the ordinary trimming, it always trims a fixed fraction of sample points, no matter those points are "good" or "bad". This raises a concern as to whether there is a trimming scheme that only trims points that are "bad", which motivates us to consider in this paper the so-called scaled deviation trimmed and Winsorized means.

☆ Research partially supported by NSF Grants DMS-0234078 and DMS-0501174.
* Corresponding author. Tel.: +1 602 350 0547.
E-mail address: wumingxi@msu.edu (M. Wu).

0378-3758/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2008.03.039

The main idea behind the new trimming scheme is that sample points are trimmed based on the magnitude of their scaled (standardized) deviations to a center (say median). Only points with the scaled deviation beyond some fixed threshold are trimmed. This new trimming scheme can lead to the best possible breakdown point (see Section 5.1 for definition) robustness. The resulting estimators are also highly efficient at light-tailed symmetric models and much more efficient than the Tukey trimmed and Winsorized means at models with a slight departure from symmetry or with heavy tails. Hence they represent favorable alternatives to their predecessors.

For multidimensional setting, trimmed means are introduced based on data depth. With a projection depth induced ordering, Zuo (2006) discussed a trimming scheme that possesses many desirable properties. However, little attention has been paid to the one dimensional counterpart and the relationship with other types of trimmed means is not well clarified. This paper will shed light on the robustness and efficiency of univariate counterpart and give reader a clear picture on different types of trimmed means.

The rest of the paper is organized as follows. Section 2 defines the scaled deviation trimmed and Winsorized means and discusses some primary properties. Section 3 investigates the local robustness, the influence functions, of the estimators. The asymptotic normality of the estimators is established via their asymptotic representations in Section 4. The performance comparison of the estimators with other leading trimmed means with respect to various robustness and efficiency criteria is carried out in Section 5. Concluding remarks in Section 6 end the main body of the paper. Proofs of main results and auxiliary lemmas are reserved for the Appendix A.

2. Scaled deviation trimmed and Winsorized means

Let $\mu(F)$ and $\sigma(F)$ be some robust location and scale measures of a distribution F. For simplicity, we consider μ and σ being the median (Med) and the median absolute deviations (MAD) throughout the paper. Assume $\sigma(F)>0$, namely, F is not degenerate. For a given point x, we define the *scaled deviation* (generalized standardized deviation) of x to the center F by

$$D(x,F)=(x-\mu(F))/\sigma(F). \tag{2.1}$$

Now we trim points based on the absolute value of this scaled deviation and define the β *scaled deviations trimmed mean* at F as (cf. Zuo, 2003 for a multidimensional version)

$$T^{\beta}(F)=\frac{\int \mathbf{I}(|D(x,F)|\leqslant\beta)\mathrm{w}(D(x,F))x\,\mathrm{d}F(x)}{\int \mathbf{I}(|D(x,F)|\leqslant\beta)\mathrm{w}(D(x,F))\,\mathrm{d}F(x)}, \tag{2.2}$$

where $0<\beta\leqslant\infty$ and w is an even bounded weight function on $[-\infty,\infty]$ so that the denominator is positive. The heuristic idea behind this definition is that one trims points that are far ($\beta\sigma$) away from the center and then one *weights* (not just simply average) left points based on the robust scaled deviation with larger weights for points closer to the center. When w is a non-zero constant, T^{β} becomes the plain average of points after the trimming. To cover a broader class of the trimmed means, we consider general w in our treatment. Note that in the extreme case $\beta=\infty$ and $\mathrm{w}=c\neq 0$, T^{β} becomes the usual mean. A concern might be that T^{β} throws away useful information in the tails. A remedial measure is the *Winsorization*. For the completeness of our discussion, we consider here the β *scaled deviations Winsorized mean* at F, defined as

$$T_{w}^{\beta}(F)=\frac{\int (x\mathbf{I}(|D(x,F)|\leqslant\beta)+L(F)\mathbf{I}(x<L(F))+U(F)\mathbf{I}(x>U(F)))\mathrm{w}(D(x,F))\,\mathrm{d}F(x)}{\int \mathrm{w}(D(x,F))\,\mathrm{d}F(x)}, \tag{2.3}$$

where $L(F)=\mu(F)-\beta\sigma(F)$ and $U(F)=\mu(F)+\beta\sigma(F)$. In the extreme case $\beta=0$, T_{w}^{β} degenerates into the median. For a fixed β, we sometimes suppress β in T^{β} and T_{w}^{β} for convenience.

Since both μ and σ are affine equivariant, i.e., $\mu(F_{aX+b})=a\mu(F_X)+b$, $\sigma(F_{aX+b})=|a|\sigma(F_X)$ for any scalars a and b, where F_X is the distribution of X, it is readily seen that $|D(x,F)|$ is affine invaraint and T thus is *affine equivariant* as well. For $X\sim F$ symmetric about θ (i.e., $\pm(X-\theta)$ have the same distribution), it is seen that $T(F)=\theta$, i.e., T is *Fisher consistent*. Without loss of generality, we can assume $\theta=0$. Let F_n be the usual empirical version of F based on a random sample. It is readily seen that $T(F_n)$ is also affine equivariant. It is *unbiased* for θ if F is symmetric about θ and has an expectation. For $T_w(F)$ and $T_w(F_n)$, all these properties hold.

Two popular trimmed means in the literature are: the ordinary trimmed mean (Tukey, 1948) and the metrically trimmed mean (Bickel, 1965; Kim, 1992), defined, respectively, as

$$T_{o}^{\alpha}(F)=\frac{1}{1-\alpha}\int_{F^{-1}(\alpha/2)}^{F^{-1}(1-\alpha/2)}x\,\mathrm{d}F(x),\quad T_{m}^{\alpha}(F)=\frac{1}{1-\alpha}\int_{\mu(F)-v(F)}^{\mu(F)+v(F)}x\,\mathrm{d}F(x), \tag{2.4}$$

where $F^{-1}(r)$ is the rth quantile of F and $F(\mu(F)+v)-F(\mu(F)-v)=1-\alpha$. It is readily seen that these trimmed means are also affine equivariant and consequently Fisher consistent for symmetric F. The two trimming schemes are probability content based. The former, however, trims equally ($50\alpha\%$) of points at each tail. This is not always the case for the latter (though total points trimmed are also $100\alpha\%$). At the sample level, $T_{o}^{\alpha}(F_n)$ trims a fixed (equal) number of sample points at each tail while $T_{m}^{\alpha}(F_n)$ trims sample points at both tails or just one tail with the same total number of points trimmed as in the former case. For performance