5-1990

On Choosing An Optimally Trimmed Mean

Kishan Mehrotra
Syracuse University, mehrtra@syr.edu

Paul Jackson
Syracuse University

Anton Schick
SUNY Binghamton

SU-CIS-90-04

On Choosing an Optimally Trimmed Mean

Kishan Mehrotra and Paul Jackson
Syracuse University

Anton Schick
SUNY Binghamton

May 1990

School of Computer and Information Science
Syracuse University
Suite 4-116
Center for Science and Technology
Syracuse, New York 13244-4100

ON CHOOSING AN OPTIMALLY TRIMMED MEAN

Kishan Mehrotra and Paul Jackson
Syracuse University
Syracuse, NY 13210

Anton Schick
SUNY Binghamton
Binghamton, NY 13901

Key Words and Phrases: Trimmed mean, bootstrap, optimal trimming.

ABSTRACT

In this paper we revisit the problem of choosing an optimally trimmed mean. This problem was originally addressed by Jaeckel (1971). We propose alternatives to Jaeckel's estimator and its modifications discussed in Andrews et al. (1972). Jaeckel's procedure chooses the optimal trimming by minimizing an estimate of the asymptotic variance of the trimmed mean. We use the bootstrap procedure to choose the optimal trimming. A simulation study shows that our procedure compares favorably with Jaeckel's procedure. We also discuss modification of our procedure to the two sample setting.

1. INTRODUCTION

A common problem in nonparametric estimation is as follows. To estimate a characteristic θ, the statistician must choose an estimate from a class of reasonable estimates $\{\hat{\theta}(\alpha) : \alpha \in A\}$ indexed by a parameter α. The parameter α, of course, is selected according to some optimality criterion. If the estimates are unbiased or nearly unbiased, it is reasonable to use that estimate which minimizes the variance. However, the variance of the estimate $\hat{\theta}(\alpha)$ generally depends on unknown characteristics of the population and is unknown to the statistician. Thus the optimal parameter α_* is unknown. To overcome this difficulty, the statistician can estimate the unknown variances of the estimates and select a parameter value, say $\hat{\alpha}$, which minimizes this estimate of the variance. The resulting estimate of θ is then $\hat{\theta}(\hat{\alpha})$.

Here are two examples of this type of problem.

Example 1: Estimating the center of symmetry. Let F be a distribution function which is symmetric about zero. Let $(X_1, \ldots, X_n)$ be a random sample from the distribution function $G = F(\cdot - \theta)$. A class of estimates of θ, the center of symmetry of G, is the class of trimmed means. The parameter α is the amount of trimming. Thus one is looking for the trimmed mean with smallest possible variance.

Example 2: Estimating the difference of location in a two sample problem. Consider two independent random samples $(X_1, \ldots, X_m)$ and $(Y_1, \ldots, Y_n)$ drawn from distribution functions F and $F(\cdot - \theta)$, respectively. Here we want to estimate the difference in location θ. One possible class of estimates is the difference of the α-trimmed means of the two samples. Another class of estimates are the trimmed means based on the pairwise differences $\{(Y_j - X_i) : i = 1, \ldots, m,\ j = 1, \ldots, n\}$. For both classes the parameter α is the trimming portion.

The first example has been studied by Jaeckel (1971). Jaeckel recommends to estimate the asymptotic variance of the trimmed means and use the trimmed mean which corresponds to the trimming which minimizes the estimate of the asymptotic variance. He proposes an estimate of the asymptotic variance of the trimmed means and verifies that the resulting randomly trimmed mean has the same asymptotic distribution as the trimmed mean smallest asymptotic variance. Modification of his method are discussed in the monograph by Andrews et al. (1972). This monograph reports also the results of extensive simulations.

Jaeckel's estimate of the asymptotic variance is a sample analogue of the expression of the asymptotic variance of a trimmed mean for symmetric error distributions. In other problems, however, explicit expressions for the (asymptotic) variances may not be available, and (or) it will be difficult to obtain an estimate of the variance. An approach which avoids these difficulties is as follows. Estimate the variances using the bootstrap method and then select the trimming portion which minimizes the bootstrap variance estimator. This method does not require an explicit expression for the variances and works in other situations as well. For this reason the bootstrap approach has become a major source in adaptive statistical procedures.

In this paper we shall study in the context of the above two examples the performance of randomly trimmed means whose amount of trimming is chosen to minimize a bootstrap estimator of the variances of the trimmed means. In connection with example 1, we shall treat various versions of this method and compare them with Jaeckel's estimate and its modifications presented in Andrews et al. (1972). In connection with example 2, we shall compare the randomly trimmed means for the two cases where in both cases the amount of trimming is chosen by the bootstrap method.

Our paper is organized as follows. In section 2 we describe our proposed estimates in the one sample case (see Example 1) and contrast it to Jaeckel's estimate. We report the numerical results of a simulation study and compare them with those reported in Andrews et al. (1972).

In Section 3 we discuss possible generalizations of our method to the two sample case (see Example 2). We discuss three classes of estimates. The first two classes of estimates consist of the differences of the trimmed means from the two samples and the third class of the trimmed means of the pairwise differences. Again we use the bootstrap method to select the amount of trimming. We then report the results of a simulation and compare the two resulting types of estimates.

2. THE ONE SAMPLE CASE

We only consider the problem of estimating the center of symmetry. We shall consider four different estimates all of which are versions of randomly trimmed means. The amount of trimming is chosen to minimize the bootstrap variance.

Let $\boldsymbol{X} = (X_1, \ldots, X_n)$ denote a random sample from a distribution function F which is symmetric about zero. Let $X_{(1)}, \ldots, X_{(n)}$ denote the order statistics. The α-trimmed mean based on the sample $\boldsymbol{X}$ is defined by

$$T_\alpha(\boldsymbol{X}) = \frac{X_{(1+\alpha n)} + \ldots + X_{(n-\alpha n)}}{n(1-2\alpha) \vee 2},$$

where $\alpha \in \Lambda = \{\frac{i}{n} : i = 0, \ldots, [\frac{n-1}{2}]\}$ represents the amount of trimming. We are looking for the trimmed mean with smallest possible variance.

Note that this question makes sense even when the symmetry assumption is dropped. If the symmetry assumption is violated the trimmed mean will estimate some quantity depending on the amount of left and right trimming. In this case we actually select the quantity we are estimating as well.

To select the "best" α in Λ, Jaeckel (1971) proposed to use that trimmed mean whose asymptotic variance is minimum. He estimates the asymptotic variance

$$\sigma^2(\alpha) = \frac{1}{(1-2\alpha)^2}\left(\int_{\xi_\alpha}^{\xi_{1-\alpha}} x^2\,dF(x) + 2\alpha\xi_\alpha^2\right)$$

where $\xi = F_0^{-1}(\alpha)$ and $\xi_{1-\alpha} = F_0^{-1}(1-\alpha)$, by its sample analogue

$$s^2(\alpha) = \frac{1}{(1-2\alpha)^2}\left(\frac{1}{n}\sum_{i=[\alpha n]+1}^{n-[\alpha n]} \tilde{X}_{i,\alpha}^2 + \alpha\tilde{X}_{[\alpha n]+1,\alpha}^2 + \alpha\tilde{X}_{n-[\alpha n],\alpha}^2\right)$$

where $\tilde{X}_{i,\alpha} = X_{(i)} - T_\alpha(\boldsymbol{X})$, $i = 1,\ldots,n$. He then proposed the estimate $T_{\hat{\alpha}}(\boldsymbol{X})$, where $\hat{\alpha}$ minimizes the estimated variance s^2 over a compact subset $[\alpha_0, \alpha_1]$ of $[0, .5]$. Under the assumption that $0 < \alpha_0 < \alpha_1 < .5$, he then showed that this estimate has an asymptotic variance that equals $\min_{\alpha_0\le\alpha\le\alpha_1}\sigma^2(\alpha)$, which is the smallest possible asymptotic variance for any of the estimates $T_\alpha(X), \alpha_0 \le \alpha \le \alpha_1$.

In our opinion the asymptotic variance may not be a true representative of the actual variance of a trimmed mean, especially in small samples. Moreover, Jaeckel's estimator of the asymptotic variance of the trimmed mean may be poor for values of α close to $1/2$. For this reason, Jaeckel recommends to confine α to the interval $[0, 0.25]$. To avoid these difficulties we use the bootstrap procedure to estimate the variances of the trimmed means. Let us now describe our four estimates.

Estimate 1: We draw N independent samples

$$\boldsymbol{Y}_1 = (Y_{1,1},\ldots,Y_{1,n}),\ldots,\boldsymbol{Y}_N = (Y_{N,1},\ldots,Y_{N,n})$$

of size n from $\boldsymbol{X} = (X_1,\ldots,X_n)$. Each sample is taken with replacement.

For each $\alpha \in \Lambda$, we compute the trimmed means $T_{\alpha,i} = T_\alpha(\boldsymbol{Y}_i)$ based on the bootstrap samples $\boldsymbol{Y}_i$, and calculate their sample variances

$$\hat{V}(\alpha) = \frac{1}{(N-1)}\sum_{i=1}^{N}(T_{\alpha,i} - \bar{T}_\alpha)^2,$$

where $\bar{T}_\alpha = \frac{1}{N}\sum_{i=1}^{N} T_{\alpha,i}$ is their sample average. Then we find α_* which minimizes the sample variance, i.e., α_* satisfies

$$\hat{V}(\alpha_*) \le \hat{V}(\alpha), \quad \alpha \in \Lambda.$$

The estimate is $T_{\alpha_*}(\boldsymbol{X})$.

The above estimate makes use of the symmetry assumption only in choosing a (symmetric) trimmed mean. To make additional use of the symmetry assumption we will augment the sample by $(2S - X_1, \ldots, 2S - X_n)$ and then calculate the above estimate based on the augmented sample $\boldsymbol{X}_S = \{X_1, \ldots, X_n, 2S - X_1, \ldots, 2S - X_n)$. Here S denotes an estimate of 0, the center of symmetry. Choices of S are discussed below.

Estimate 2 ($S = 0$): Augment the sample $\boldsymbol{X}$ with the negative values of the sample to obtain $\boldsymbol{X}_0 = (X_1, \ldots, X_n, -X_1, \ldots, -X_n)$. Now calculate Estimate 1 for the augmented sample $\boldsymbol{X}_0$.

Estimate 3 (Center at median): Let $S = M$, the median of the sample $\boldsymbol{X}$. Augment the sample $\boldsymbol{X}$ by the values $2M - X_1, \ldots, 2M - X_n$ and denote the result by $\boldsymbol{X}_M = (X_1, \ldots, X_n, 2M - X_1, \ldots, 2M - X_n)$. Now calculate Estimate 1 based on the augmented sample $\boldsymbol{X}_M$.

Estimate 4 (Double bootstrap): Let $S = \hat{T} = T_{\alpha_*}(\boldsymbol{X})$. Augment the sample $\boldsymbol{X}$ by the values $2\hat{T} - X_1, \ldots, 2\hat{T} - X_n$ and denote the result by $\boldsymbol{X}_{\hat{T}} = (X_1, \ldots, X_n, 2\hat{T} - X_1, \ldots, 2\hat{T} - X_n)$. Now calculate Estimate 1 based on the augmented sample $X_{\hat{T}}$.

We have performed a simulation study to compare our estimates with the estimate of Jaeckel and its modifications proposed by Bickel and Jaeckel and denoted as SJA, BIC, JBT and JLJ (See Andrews et al. page 2B3-2C for their description). Our simulations are for the choice $n = 20$ and $N = 200$. Recall n denotes the sample size and N the size of the bootstrap samples. In the table below we report the sample variance of our estimates based on 5000 iterations multiplied by $n = 20$. We have carried out our simulations for five symmetric distributions. These distributions are the standard normal distribution, the Cauchy distribution, the double-exponential distribution, the logistic distribution, and a bimodal distribution with density

$$f(x) = \frac{1}{2\sqrt{2\pi}} \left(\exp(-\frac{(x-1)^2}{2}) + \exp(-\frac{(x+1)^2}{2}) \right), \quad x \in R.$$

Where available we have also reported the results of Andrews et al (1972).

Table 1

	Normal	Cauchy	D-Exp.	Logistic	Bimodal
JAE	1.105	3.500	1.480		
BIC	1.088	16.600	1.450		
SJA	1.110	3.700	1.390		
JBT	1.110	3.300	1.450		
JLJ	1.167	2.800	1.530		
Estimate 1	1.067	3.108	1.528	3.338	2.303
Estimate 1	1.045	2.996	1.493	3.001	2.122
Estimate 2	0.905	2.465	1.088	2.563	1.888
Estimate 3	1.208	2.861	1.364	3.416	2.667
Estimate 4	1.070	3.034	1.450	3.136	2.190
Variance	1.000	2.584	1.304	3.004	2.000

Remarks: The values given in the table are the variances multiplied by n. The numbers in the last row are the variances corresponding to the "best" trimmed mean for the distribution. These values are based on 10,000 samples.

It can be seen that the smallest variances are observed for estimate 2. But in this case an unfair use of the knowledge of location parameter is used in augmenting the sample.

The ordinary bootstrap and the double bootstrap have better performances than the Jaeckel's estimator in most of the cases. In case of double exponential distribution only the double bootstrap has better performance than Jaeckel's estimate. Use of sample median to argument the sample helps in the case of Cauchy and double exponential distributions. Our variance calculation, based on 10,000 repetitions shows that in these two distributions the minimum variances are actually seen near the median for samples of size 20. This is also an expected result.

In conclusion, it is observed that the bootstrap method gives very satisfactory results for a variety of distributions. The simulation results are given only for the symmetric distributions. However, the same ideas apply to the case of nonsymmetric distributions as well. In the later it would be desirable to obtain the optimum trimming from the left and right end independently, by the bootstrap procedure.

3. THE TWO SAMPLE CASE

Let $\boldsymbol{X} = (X_1, \ldots, X_m)$ and $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ denote two independent random samples from distribution functions F and G, respectively. For simplicity we assume that $m = n$. Suppose that $G = F(\cdot - \theta)$ for some $\theta \in R$. We consider three classes of estimates of θ, the differences of the trimmed means from the two samples with equal trimming portions

$$\Delta_\alpha(\boldsymbol{X}, \boldsymbol{Y}) = T_\alpha(\boldsymbol{Y}) - T_\alpha(\boldsymbol{X}),$$

the differences of trimmed means from the samples

$$\Delta_{\alpha_1,\alpha_2}(\boldsymbol{X}, \boldsymbol{Y}) = T_{\alpha_1}(\boldsymbol{Y}) - T_{\alpha_2}(\boldsymbol{X}),$$

and the trimmed means

$$T_\alpha(\boldsymbol{Z})$$

based on the n^2 pairwise differences $\boldsymbol{Z} = \boldsymbol{Z}(\boldsymbol{X}, \boldsymbol{Y}) = \{Y_j - X_i : i = 1, \ldots, n,\ j = 1, \ldots, n\}$ of the observations from the two samples. For all three estimates we select the trimming using the bootstrap. In all cases we draw N independent samples

$$(\boldsymbol{X}_1, \boldsymbol{Y}_1), \ldots, (\boldsymbol{X}_N, \boldsymbol{Y}_N),$$

where $\boldsymbol{X}_i = (X_{i,1}, \ldots, X_{i,n})$ and $\boldsymbol{Y}_i = (Y_{i,1}, \ldots, Y_{i,n})$ are independent samples from $\boldsymbol{X}$ and $\boldsymbol{Y}$, respectively. In the first case, we compute for each $\alpha \in \Lambda_1 = \{\frac{i}{n} : i = 0, \ldots, [\frac{n-1}{2}]\}$ the estimates $\Delta_{\alpha,i} = \Delta_\alpha(\boldsymbol{X}_i, \boldsymbol{Y}_i)$ and calculate the sample variance

$$W_1(\alpha) = \frac{1}{(N-1)} \sum_{i=1}^{N} (\Delta_{\alpha,i} - \bar{\Delta}_\alpha)^2.$$

Our estimator is then $\Delta_{\alpha_1}(\boldsymbol{X}, \boldsymbol{Y})$ where α_1 is chosen to minimize the sample variance. In the second case we find the minimum variance $\boldsymbol{X}$ trimmed mean and similarly the minimum variance $\boldsymbol{Y}$ trimmed mean, based on 200 bootstrap samples, and estimate θ by the difference of these two trimmed means. This estimate is only appropriate if the underlying error distribution is symmetric about some value. If the error distribution is not symmetric this estimate may be heavily biased. In the third case, we form the samples

$\boldsymbol{Z}_i = \boldsymbol{Z}(\boldsymbol{X}_i, \boldsymbol{Y}_i)$ and then compute for each $\alpha \in \Lambda_2 = \{\frac{i}{n^2} : i = 0, \ldots, [\frac{n^2-1}{2}]\}$ the estimates $T_{\alpha,i} = T_\alpha(\boldsymbol{Z}_i)$ and the sample variance

$$W_3(\alpha) = \frac{1}{(N-1)} \sum_{i=1}^{N} (T_{\alpha,i} - \bar{T}_\alpha)^2.$$

Our estimator is then $T_{\alpha_3}(\boldsymbol{Z})$ where α_3 is chosen to minimize the sample variance. These estimators are called Estimate 1, Estimate 2, and Estimate 3, respectively. We have performed a simulation study to see how these estimators perform. The results are summarized in Table 2.

Table 2

	Estimate 1		Estimate 2		Estimate 3	
Distribution	mean	variance	mean	variance	mean	variance
Normal	1.005	0.207863	1.002	0.210760	1.006	0.218678
Cauchy	0.991	1.528480	0.957	1.404565	1.016	1.269137
Logistic	0.992	0.650573	0.970	0.658473	0.997	0.655253
D. Exp.	0.994	0.345908	1.007	0.345837	0.997	0.340930
Bimodal	0.999	0.418877	1.000	0.431048	1.001	0.400520

We observe that the variances of all estimates are almost equal in four out of five distributions and differ only in the case of the Cauchy distribution. In the case of the Cauchy distribution Estimate 3 clearly outperforms the other the estimates. The first and the second methods of estimation are faster than the third method by a significant factor. But difference in calculation time will only show up for larger values of n and m. Recall also that Estimate 2 requires the error distribution to be symmetric about some value.

BIBLIOGRAPHY

Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H. and Tuckey, J. W. (1972). *Robust Estimates of Location.* Princeton University Press, Princeton, NJ.

Jaeckel, L. B. (1971) "Some flexible estimates of location." *Ann. Math. Statist.*, **42**, 1540–1552.