1. Introduction

The most popular technique in estimating a location parameter is the least squares estimator. Its popularity mainly reflects its advantages in the theoretical property from the parametric point of view that it is uniformly minimum variance unbiased estimator when the variable follows a normal distribution. However, the least squares estimator is sensitive to departures from normality and to the presence of outliers. Hence, we need to consider robust estimators.

An important class (see Hogg (1974) and Huber (1981)) that provides many choices of robust estimators for location parameter is the L-estimators, defined in terms of ordinary quantile. The benefits of using an estimator that is based on quantile include its easiness in computation and asymptotic efficiency shown in literature (see Hogg (1974), Jureckova and Sen (1996) and Chen (1996)). Let F^{-1} be the population quantile function. The class of ordinary L-estimators is considering the following quantile means

$$\int_0^1 \delta(\alpha) F^{-1}(\alpha) d\alpha \tag{1.1}$$

as estimand for some nonnegative function δ. This provides a rich class of quantities very popular and interesting in application and theoretical study for measuring center for the underlying distribution.

The trimmed mean, with δ an indicator function having value zero outside a quantile interval and trimming for its sample version refers to the removal of the extreme values of a sample, has a long history. Huber (1972) quoted an anonymous author from 1921, who explained that in certain provinces of France the mean yield of land was estimated by averaging the middle 18 yields from 20 consecutive years. The most popular version considers symmetrically trimming as

$$\mu_{med}(2\alpha) = \frac{1}{1-2\alpha} \int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} x f(x) dx. \tag{1.2}$$

We call it the median type trimmed mean. Basically the trimmed mean is a natural compromise between the nonrobust sample mean and median which, although it is insensitive to outlying observations, but is turned out that it went too far in discarding obseravtions. That is, the trimmed mean discrad only a certain number of the obseravtions and to use as estimator the mean of the remaining observations.

There are many criterions available for us to compare the robust estimators. Among them, breakdown point for an estimator is one interesting for statistician. The breakdown point represents the smallest percentage of contamination in the data that may cause the estimator to take on arbitrary large values. For a mathematical definition, see Hampel,et al.(1986). It is popularly recoganized that the breakdown point in statistical inferences is no more than 0.5 (see this point in Hampel et al. (1986)). Consider several examples. In the case of the least squares estimator we find that the breakdown point is zero. The first step in the robustification of the least squares estimator is the Huber's estimator (see this latter). Its breakdown point depends on the design (the distribution of the x's). Its breakdown point never greater than 25%. The most robust limiting case of the Huber's estimator, the ℓ_1-norm estimator (that minimizes the sum of absolute deviations) has breakdown point 25% for uniform x's, less than 24% for normal x's, and arbitrary close to zero for heavy tail and asymmetric designs.

It is known that the assumption of symmetric distribution is a serious concern in robust estimation. With this consideration of possible asymmetric distribution, symmetrically trimming is nolonger a right trimming choice. However, it is also known that an asymmetric trimming such as the following

$$\mu(\alpha_1, \alpha_2) = \frac{1}{\alpha_2 - \alpha_1} \int_{F^{-1}(\alpha_1)}^{F^{-1}(\alpha_2)} x f(x) dx \tag{1.3}$$

for some α_1, α_2 with $0 < \alpha_1 < \alpha_2 < 1$ does not fulfill desired equivariant conditions in general unless the symmetric case that $\alpha = \alpha_1 = 1 - \alpha_2$. What else choice we may consider for trimmed mean?

Our interest in this paper is to proposed a population type trimmed mean that the retained region with a fixed coverage probability for computing mean is the shortest one which will be called the mode type trimmed mean. To investigate this new population type trimmed mean, we will introduced a sample type version and study it. To study this sample type trimmed mean, we will investigate its breakdown point and show that it may exceed the popularly accepted value 0.5. Furthermore we will also investigate this sample type trimmed mean by comparing the mean squares errors with one traditional used trimmed mean.

2. Breakdown Point for Some Robust Location Estimators

First we give one example for iid random variable case. Let $y_1, ..., y_n$ be iid random variables.

Sample Mean and Sample Median

The sample mean $\bar{y}$ has breakdown point is $\frac{1}{n}$ which converges to zero and the sample median has breakdown point approximated 0.5.

Hodges-Lehmann Estimator

Consider the location model. The Hodges-Lehmann (HL) estimator (see Hodges and Lehmann (1963)) is defined as

$$\hat{\theta} = \text{med}_{1 \leq i < j \leq n} \frac{y_i + y_j}{2}.$$

Then the breakdown point of the Hodges-Lehmann estimator is $1 - (\frac{1}{2})^{1/2}$ which is approximately 0.293.

Suppose that we have a linear regression model

$$y_i = x_i' \beta + \epsilon_i, i = 1, ..., n$$

where x_i is a p-vector of independent variables with value 1 on the first element.

Least Squares Estimator

The least squares (LS) estimator is defined as

$$\hat{\beta}_{LS} = \arg\ \min_b \sum_{i=1}^{n} (y_i - x_i' b)^2$$

which has breakdown point $\frac{1}{n}$ that converges to zero as n goes to infinity.

Least Median of Squares Estimator

The least median of squares (LMS) estimator, proposed by Rousseeuw (1984), $\hat{\beta}_{LMS}$ is defined as

$$\hat{\beta}_{LMS} = \arg\ \min_b \text{med}_i (y_i - x_i' b)^2.$$

The breakdown point of the LMS estimator is $\frac{(n+1)/2}{n}$ which converges to 0.5 as $n \to \infty$.

Least Trimmed Squares Estimator

The least trimmed squares (LTS) estimator proposed by Rousseeuw (1983). Let b be any p vector in R^p. By letting $r_i = y_i - x_i' b, i = 1, ..., n$ and $(r^2)_{i:n}, i = 1, ..., n$ be the order statistics of $r_i^2, i = 1, ..., n$, the LTS estimators is defined as

$$\hat{\beta}_{LTS} = \min_b \sum_{i=1}^{h} (r^2)_{i:n}$$

where $h = [\frac{n}{2}] + 1$. Then the breakdown point of the LTS estimator is $\frac{[(n+1)/2]}{n}$ which converges to 0.5 as $n \to \infty$.

Least Winsorized Squares Estimator

For the location estimation problem, the least Winsorized squares (LWS) estimator (see Rousseeuw (1987)) is defined as

$$\hat{\beta}_{LWS} = \min_a \sum_{i=1}^{h} (r^2)_{i:n} + (n-h)(r^2)_{h:n}$$

where $h = [n/2] + 1$. Then the breakdown point of the LWS estimator is $\frac{[(n+1)/2]}{n}$ which converges to 0.5 as $n \to \infty$.

Least Absolute Values Regression Estimator

The least absolute values regression estimator is defined as

$$\hat{\beta}_L = \arg\ \min_b \sum_{i=1}^{n} |y_i - x_i' b|$$

which has breakdown point $\frac{1}{n} \to 0$ as $n \to \infty$.

Huber's M-Estimator

The Huber's M-estimator (M) is defined as

$$\hat{\beta}_M = \arg_b [\sum_{i=1}^{n} x_i \psi(y_i - x_i' b) = 0]$$

where

$$\psi(z) = \begin{cases} z, & \text{if } |z| < k \\ k \text{ sgn } (z), & \text{if } |z| \geq k \end{cases}$$

where k is positive constant, usually taking 1.5. The breakdown point of the Huber's M-estimator is $\leq \frac{1}{p}$.

3. Mode Type Trimmed Mean

Before defining a more general trimmed mean, we consider a set of desired equivariant conditions for location parameter. The following conditions expected for location parameter to fulfill may be seen in Staudte and Sheather (1990).

Definition 3.1. We say that μ, a real function of r.v. X , is a measure of location if it satisfies

(a). $\mu(X+b)=\mu(X)+b$ for $b \in R$.

(b). $\mu(ax)=a\mu(x)$ for $a>0$.

(c). $\mu(-X)=-\mu(X)$.

(d). If $X \geq 0$, then $\mu(X) \geq 0$.

We know that not every trimmed mean of (1.3) satisfies the above conditions of a measure of location. It is known that the symmetric 2α trimmed mean defined as $\mu_{med}(2\alpha)=\mu(\alpha,1-\alpha)$ does satisfies the desired conditions.

Definition 3.2. The shortest width 2α trimmed mean is defined as

$$\mu_{mod}(2\alpha)=\frac{1}{1-2\alpha}\int_{F^{-1}(\alpha^*)}^{F^{-1}(1-2\alpha+\alpha^*)} xf(x)dx \tag{3.1}$$

where $\alpha^*=\text{arginf}_{0<\alpha_1<2\alpha}(F^{-1}(1-2\alpha+\alpha_1)-F^{-1}(\alpha_1))$.

Theorem 3.3. The shortest width 2α trimmed mean is a measure of location.

Proof. Let's redenote $\mu_{mod}(2\alpha)=\mu_{mod}(X,2\alpha)$, $F^{-1}(X,\alpha)=F^{-1}(\alpha)$, $f(x)=f_X(x)$ and $\alpha^*=\alpha^*(X)$. We know that the quantile function F^{-1} satisfies $F^{-1}(X+b,\alpha)=F^{-1}(X,\alpha)+b$ for $b\in R$ and $F^{-1}(aX,\alpha)=aF^{-1}(X,\alpha)$ if $a>0$ and $aF^{-1}(X,1-\alpha)$ if $a\leq 0$.

(a). Let $b\in R$. It is easy to see that $\alpha^*(X+b)=\alpha^*(X)$. Then, by letting $Y=X+b$,

$$\mu_{mod}(X+b,2\alpha)=\frac{1}{1-2\alpha}\int_{F^{-1}(X+b,\alpha^*)}^{F^{-1}(X+b,1-2\alpha+\alpha^*)} yf_{X+b}(y)dy$$

$$=\frac{1}{1-2\alpha}\int_{F^{-1}(X,\alpha^*)+b}^{F^{-1}(X,1-2\alpha+\alpha^*)+b} yf_X(y-b)dy$$

$$=\frac{1}{1-2\alpha}\int_{F^{-1}(X,\alpha^*)}^{F^{-1}(X,1-2\alpha+\alpha^*)} (x+b)f_X(x)dx$$

$$=\frac{1}{1-2\alpha}\int_{F^{-1}(X,\alpha^*)}^{F^{-1}(X,1-2\alpha+\alpha^*)} xf_X(x)dx+b\frac{1}{1-2\alpha}\int_{F^{-1}(X,\alpha^*)}^{F^{-1}(X,1-2\alpha+\alpha^*)} f_X(x)dx$$

$$=\mu_{mod}(X,2\alpha)+b.$$

(b). Let $a > 0$. We also see that $\alpha^*(aX) = \arg_\alpha \inf_{0\leq\alpha<1-\gamma}\{a(F^{-1}(X,\gamma+\alpha) - F^{-1}(X,\alpha))\} = \alpha^*(X)$. We have

$$\begin{aligned}
\mu_{mod}(aX,2\alpha) &= \frac{1}{1-2\alpha}\int_{F^{-1}(aX,\alpha^*)}^{F^{-1}(aX,1-2\alpha+\alpha^*)} yf_{aX}(y)dy \\
&= \frac{1}{1-2\alpha}\int_{aF^{-1}(X,\alpha^*)}^{aF^{-1}(X,1-2\alpha+\alpha^*)} y\frac{1}{a}f_X(\frac{y}{a})dy \\
&= \frac{1}{1-2\alpha}\int_{F^{-1}(X,\alpha^*)}^{F^{-1}(X,1-2\alpha+\alpha^*)} ax\frac{1}{a}f_X(x)adx \\
&= a\frac{1}{1-2\alpha}\int_{F^{-1}(X,\alpha^*)}^{F^{-1}(X,1-2\alpha+\alpha^*)} xf_X(x)dx \\
&= a\mu_{mod}(X,2\alpha).
\end{aligned}$$

(c). Consider the transformation of multiplying X by negative value -1. We see that

$$\begin{aligned}
\alpha^*(-X) &= \text{arginf}_{0\leq\alpha_1<2\alpha}\{F^{-1}(-X,1-2\alpha+\alpha_1) - F^{-1}(-X,\alpha_1)\} \\
&= \text{arginf}_{0\leq\alpha_1<2\alpha}\{-F^{-1}(X,1-(1-2\alpha+\alpha_1)) + F^{-1}(X,1-\alpha_1)\} \\
&= \text{arginf}_{0\leq\alpha_1<2\alpha}\{F^{-1}(X,1-\alpha_1) - F^{-1}(X,2\alpha-\alpha_1)\} \\
&= \text{arginf}_{0\leq 2\alpha-\alpha_1<2\alpha}\{F^{-1}(X,1-2\alpha+(2\alpha-\alpha_1)) - F^{-1}(X,2\alpha-\alpha_1)\} \\
&= \text{arginf}_{0\leq\beta<2\alpha}\{F^{-1}(X,1-2\alpha+\beta) - F^{-1}(X,\beta)\}.
\end{aligned}$$

This implies that $1-(1-2\alpha+\alpha^*(-X)) = \alpha^*(X)$ and the we derive $\alpha^*(-X) = 2\alpha - \alpha^*(X)$.

$$\begin{aligned}
\mu_{mod}(-X,2\alpha) &= \frac{1}{1-2\alpha}\int_{F^{-1}(-X,\alpha^*(-X))}^{F^{-1}(-X,1-2\alpha+\alpha^*(-X))} yf_{-X}(y)dy \\
&= \frac{1}{1-2\alpha}\int_{-F^{-1}(X,1-\alpha^*(-X))}^{-F^{-1}(X,2\alpha-\alpha^*(-X))} yf_X(-y)dy \\
&= \frac{1}{1-2\alpha}\int_{-F^{-1}(X,1-2\alpha+\alpha^*(X))}^{-F^{-1}(X,\alpha^*(X))} yf_X(-y)dy \\
&= -\frac{1}{1-2\alpha}\int_{F^{-1}(X,\alpha^*(X))}^{F^{-1}(X,1-2\alpha+\alpha^*(X))} (-x)f_X(x)(-dx) \\
&= -\mu_{mod}(X,2\alpha).
\end{aligned}$$

(d). If $X \geq 0$, then $F^{-1}(X, \alpha^*) \geq 0$ and $F^{-1}(X, 1-2\alpha+\alpha^*) \geq 0$ which implies that $\mu_{mod}(X, 2\alpha) = \frac{1}{1-2\alpha}\int_{F^{-1}(X,\alpha^*)}^{F^{-1}(X,1-2\alpha+\alpha^*)} x f_X(x)dx \geq 0$. $\square$

Consider the example of exponential distribution with pdf $f(x) = \frac{1}{\theta}e^{-x/\theta}I(0 < x < \infty)$. Since $F^{-1}(\alpha) = -\theta ln(1-\alpha)$ for $0 < \alpha < 1$, we may derive the 2α median type trimmed mean as

$$\mu_{med} = \frac{\theta}{1-2\alpha}[(1-\alpha)(1-ln(1-\alpha)) + \alpha(ln\alpha - 1)]$$

and the 2α mode type trimmed mean as

$$\mu_{mod} = \frac{\theta}{1-2\alpha}[2\alpha ln(2\alpha) + 1 - 2\alpha].$$

Unless that the underlying distribution is symmetric, the population mean, median type and mode type trimmed means are generally varies. Consider the exponential distribution with pdf $f(x) = \frac{1}{\theta}e^{-x/\theta}I(x > 0)$ and we display the median type and mode type trimmed means for several values of γ.

Table 1. Mode type trimmed mean and median type trimmed mean of exponential distribution for some values of $\gamma = 1 - 2\alpha$ $(\theta = 1)$

γ	$\frac{\mu_{mod}}{\theta}$	$\frac{\mu_{med}}{\theta}$	γ	$\frac{\mu_{mod}}{\theta}$	$\frac{\mu_{med}}{\theta}$
0.1	0.0517	0.6948	0.6	0.3891	0.7610
0.2	0.1074	0.6998	0.7	0.4840	0.7908
0.3	0.1677	0.7085	0.8	0.5976	0.8307
0.4	0.2337	0.7212	0.9	0.7441	0.8877
0.5	0.3068	0.7383	0.95	0.8423	0.9289

Note that the population mean (untrimmed) is θ. The results revealed in Table 1 is that the population mode type and median type trimmed means are varied and the mode type is also smaller than the median type. The reason that this happened is that the mode type trimmed mean for the exponential distribution is the average of the region on the left hand side of the sample space that are set of relatively smaller values. We also see that these two trimmed means vary from the population mean.

The next we consider the chi-square distribution $\chi^2(5)$ as an example to compute the population type mode and median trimmed means.

Table 2. Mode type trimmed mean and median type trimmed mean of $\chi^2(5)$ for some values of $\gamma = 1 - 2\alpha$

γ	μ_{mod}	μ_{med}	γ	μ_{mod}	μ_{med}
0.1	3.0117	4.356	0.6	3.4703	4.4925
0.2	3.0471	4.3658	0.7	3.6730	4.5546
0.3	3.1077	4.3836	0.8	3.9408	4.6383
0.4	3.1960	4.4096	0.9	4.3140	4.7587
0.5	3.3147	4.4453	0.95	4.5759	4.8464

Note that the population mean of the $\chi^2(5)$ distribution is 5. The two types of trimmed mean are also varied in this asymmetric distribution. Again, the mode type trimmed mean is smaller than the median type trimmed mean and they both vary from the population mean.

An interesting problem is that when the mode type trimmed mean will be identical with the median type trimmed mean.

Theorem 3.4. Suppose that F has a symmetric continuous density f that is unimodal (meaning that $f(x)$ is strictly decreasing about its center of symmetry). Then the mode type 2α trimmed mean is exactly equal to the median type 2α trimmed mean of (1.2).
Proof.

$$
\begin{aligned}
&\frac{\partial}{\partial\alpha}(F^{-1}(\gamma+\alpha) - F^{-1}(\alpha)) = 0 \\
&\frac{1}{f(F^{-1}(\gamma+\alpha))} - \frac{1}{f(F^{-1}(\alpha))} = 0 \\
&F^{-1}(\gamma+\alpha) = F^{-1}(\alpha) \\
&\gamma + \alpha = 1 - \alpha \\
&\alpha^* = \frac{1-\gamma}{2}
\end{aligned}
$$

This leads the result of identity of 2α mode type trimmed mean and the median type trimmed mean of (1.2) by the fact that $\alpha^* = \alpha$ and $1 - \alpha^* = 1 - \alpha$. □

4. Nonparametric Estimation of Trimmed Mean

We consider a nonparametric estimation technique for estimating the unknown trimmed mean. Parametric methods of data analysis rely on distributional assumptions on the underlying data. Nonparametric methods, however, are fully data-driven and hence are particularly suited for the less understood random experiments of highly complexity. The estimator that we will introduce is a new trimmed mean.

Let $X_{(1)}, ..., X_{(n)}$ be the order statistics of a random sample of sample size n drawn from a distribution F. The median type 2α trimmed mean is

$$\hat{\mu}_{med}(2\alpha) = \frac{1}{n(1-2\alpha)} \sum_{i=1}^{n} X_i I(F_n^{-1}(\alpha) \le X_i \le F_n^{-1}(1-\alpha))$$

which has asymptotic breakdown point α. By letting $k = [n(1-2\alpha)]$, where $[n(1-2\alpha)]$ denotes the largest integer less than or equal to $n(1-2\alpha)$, we define $\ell^* = \text{argmin}_i\{h_i = X_{(k+i-1)} - X_{(i)}, i = 1, ..., n-k+1\}$. This means that $h_{\ell^*} = X_{(k+\ell^*-1)} - X_{(\ell^*)}$ is the shortest width of k order statistics interval $[X_{(i)}, X_{(k+i-1)}]$. The nonparametric 2α trimming interval is $(X_{(k_1)}, X_{(k_2)})$ and its corresponding mode type 2α trimmed mean is

$$\hat{\mu}_{mod}(2\alpha) = \frac{1}{n(1-2\alpha)} \sum_{i=1}^{n} X_i I(X_{(k_1)} \le X_i \le X_{(k_2)}).$$

We concern the property of Hampel's breakdown (Hampel (1971)). Roughly the breakdown point is the largest fraction of the data - no matter how they are chosen - that can be so drastically changed without greatly changing the estimate. And the asymptotic breakdown point is that fraction when the sample size goes to infinity.

Theorem 4.1. Consider that $\alpha < 0.5$. The asymptotic breakdown point of a mode type 2α trimmed mean is 2α which is twice as it of the median type 2α trimmed mean.

This means that when 2α is close to 0.5 the mode type trimmed mean has asymptotic breakdown point approaching to 0.5 where it for the traditional trimmed mean is only approaching to 0.25. However, we know that when 2α approaches to 1.0 the median type trimmed mean has asymptotic breakdown point approaching to 0.5. Then can the mode type trimmed mean be the one with desirable property of approaching to 1.0 in some mild situations? We design a situation for simulation of the asymptotic breakdown points for these two trimmed means.

Theorem 4.2. Suppose that we consider the data to be replaced by arbitrary values are also spread widely. Then the mode type trimmed mean has asymptotic breakdown point has value 1 as the upper bound.

The sample size is 100 with replication $m = 10000$ and this sample are drawn from the following distribution model,

$$X_i = \begin{cases} Z_i & \text{if outlier does not occurs} \\ Z_i + v_i & \text{if outlier does occurs} \end{cases}$$

where z_i are iid drawn from an ideal distribution and $v_i = 1000 + 10 * i$. If $X_i = Z_i + v_i$ then this x represents an extreme point. We compute the breakdown point for each replication and then averages this m replications. In the following table, we present the average breakdown points of mode type and median type trimmed means under the case that the ideal distribution is standard normal distribution.

Table 3. Breakdown points of the two trimmed means under standard normal distribution with extremes

γ	$\hat{\mu}_{mod}$	$\hat{\mu}_{med}$	γ	$\hat{\mu}_{mod}$	$\hat{\mu}_{med}$
0.1	0.910069	0.459972	0.6	0.409962	0.209985
0.2	0.810051	0.409962	0.7	0.310039	0.160005
0.3	0.709968	0.360039	0.8	0.209985	0.109989
0.4	0.609987	0.310039	0.9	0.109989	0.059998
0.5	0.509961	0.260009	0.95	0.049995	0.029999

It is interesting to see that the breakdown points for the mode type trimmed means are approximated twice as the values $1-\gamma$, however, those for the median type trimmed mean are only a half of $1-\gamma$. This indicates that the mode type trimmed mean has breakdown point twice as the value for the traditional trimmed mean. One more interesting fact is that it for this new trimmed mean may be larger than 0.5, the value that Hampel et al. (1986) have claimed that breakdown point may not be more than 0.5.

Now we consider that the ideal distribution is the exponential distribution $Exp(1)$. The following table displays the breakdown points of the two trimmed means.

Table 4. Breakdown points of the two trimmed means under exponential distribution with extremes

γ	$\hat{\mu}_{mod}$	$\hat{\mu}_{med}$	γ	$\hat{\mu}_{mod}$	$\hat{\mu}_{med}$
0.1	0.910068	0.459956	0.6	0.409941	0.209959
0.2	0.810048	0.409947	0.7	0.310014	0.159978
0.3	0.709962	0.360025	0.8	0.209954	0.109957
0.4	0.609974	0.310013	0.9	0.109964	0.059972
0.5	0.509944	0.259982	0.95	0.049965	0.029968

The results displayed in the above table have the same conclusions as we stated for Table 3.

One important class of robust estimators is the M-estimator, so called because they behave like, or actually are, maximum likelihood estimators for a distribution with longer tails than the normal distribution. Among the class of M-estimators, the most interesting one is the one proposed by Huber (see Huber (1981)). Huber's M-estimate is defined as any value of ℓ solving

$$\sum_{i=1}^{n} \phi(\frac{x_i - \ell}{d}) = 0$$

where $\phi(z) = \begin{cases} z & \text{if } |z| < k \\ k \text{ sgn}(z) & \text{if } |z| \geq k \end{cases}$. Hampel (1974) proposed taking

$$d = \text{ median} \left\{ \frac{|x_i - \text{ median } \{x_i\}|}{0.6745} \right\}.$$

The divisor 0.6745 is suggested because d is then approximated equal to the standard deviation if n is large and the distribution is normal. It is also suggested that k be taken equal to about 1.5. The sample median may be taken as the initial estimate of ℓ. However, for convenience of computation, we take sample mean as initial estimate of ℓ.

The Huber's M-estimate $\hat{\mu}_{huber}$ has the property that it is the mean of the set of numbers that results when observations located more than a distance kd away from $\hat{\mu}_{huber}$ are replaced by $\hat{\mu}_{huber} + kd\text{sgn}(x - \hat{\mu}_{huber})$. This suggests a simple iterative procedure for finding $\hat{\mu}_{huber}$.

Table 5. Breakdown points of M-estimator

Estimator	$N(0,1)$	$Exp(1)$	$\chi^2(3)$	$\chi^2(5)$
M-est	0.4566	0.3854	0.2745	0.2366

The above results shows that the breakdown points for the Huber's M-estimator may not be more than 0.5 where sometimes its breakdown points may be quite small.

With this new trimmed mean, we may also do a simulation to compare it with the traditional symmetric type trimmed mean for their efficiency by computing the mean square errors (MSE). Consider a random sample of sample size 1000 from the following contaminated normal distribution

$$(1 - \delta)N(0, 1) + \delta|N(0, \sigma^2)|$$

and we compute these estimator's average MSE's from a replication of 10000. The following table display the simulation results of MSE's for these two trimmed means.

Table 6. MSE's for trimmed means under the contaminated normal distribution

	$2\alpha = 0.3$ μ_{mod}	μ_{med}	$2\alpha = 0.2$ μ_{mod}	μ_{med}	$2\alpha = 0.1$ μ_{mod}	μ_{med}
$\delta = 0.1$						
$\sigma = 3$	0.0060	0.0136	0.0043	0.0148	0.0166	0.0181
$\sigma = 10$	0.0048	0.0247	0.0032	0.0312	0.0021	0.0724
$\sigma = 100$	0.0050	0.0317	0.0029	0.0440	0.0012	3.7036
$\delta = 0.2$						
$\sigma = 3$	0.0089	0.0570	0.0537	0.0677	0.0250	0.0848
$\sigma = 10$	0.0045	0.1503	0.0102	0.3081	0.1472	0.7913
$\sigma = 100$	0.0036	2.0361	0.0015	18.021	13.280	74.892
$\delta = 0.3$						
$\sigma = 3$	0.0276	0.1609	0.1337	0.1697	0.1707	0.2185
$\sigma = 10$	0.0379	0.7853	0.1013	1.4365	1.0014	2.5983
$\sigma = 100$	0.0019	51.785	7.2581	127.75	99.379	269.04

We know that in general the estimators with larger breakdown points are with lower efficiency. However, it is not the case in this simulation. The mode type trimmed mean has larger breakdown points but with higher efficiency where everything is compared with the median type trimmed mean.

5. Nonparametric Estimation of Winsorized Mean

Winsorize is a term introduced by Dixon (1960), who attributes the idea to Charles P. Winsor. Dixon was concerned particularly with the possibility that extreme values are poorly determined or unknown to the statistician. Winsorization refers to the modification of the extreme values of a sample. The sample is symmetrically Winsorized by setting the k smallest equal to the $(k+1)$th order statistic and setting the k largest order statistics equal to the $(n-k)$th order statistic for some specific $k < n/2$.

The population median type 2α Winsorized mean is

$$\mu^*_{med}(2\alpha) = \int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} xf(x)dx + \alpha[F^{-1}(\alpha) + F^{-1}(1-\alpha)]. \tag{5.1}$$

Definition 5.1. The shortest width 2α Winsorized mean, called the mode type Winsorized mean, is defined as

$$\mu^*_{mod}(2\alpha) = \int_{F^{-1}(\alpha^*)}^{F^{-1}(1-2\alpha+\alpha^*)} xf(x)dx + \alpha^* F^{-1}(\alpha^*) + (2\alpha-\alpha^*)F^{-1}(1-2\alpha+\alpha^*) \tag{5.2}$$

where $\alpha^* = \text{arginf}_{0<\alpha_1<2\alpha}(F^{-1}(1-2\alpha+\alpha_1) - F^{-1}(\alpha_1))$.

Theorem 5.2. The mode type 2α Winsorized mean is a measure of location.

Proof. Let's redenote $\mu^*_{mod}(2\alpha) = \mu^*_{mod}(X, 2\alpha), F^{-1}(X,\alpha) = F^{-1}(\alpha),\ f(x) = f_X(x)$ and $\alpha^* = \alpha^*(X)$. We know that the quantile function F^{-1} satisfies $F^{-1}(X+b,\alpha) = F^{-1}(X,\alpha) + b$ for $b \in R$ and $F^{-1}(aX,\alpha) = aF^{-1}(X,\alpha)$ if $a > 0$ and $aF^{-1}(X, 1-\alpha)$ if $a \leq 0$.

(a). Let $b \in R$. It is easy to see that $\alpha^*(X+b) = \alpha^*(X)$. Then, by letting $Y = X + b$,

$$\mu^*_{mod}(X+b, 2\alpha) = \int_{F^{-1}(X+b,\alpha^*)}^{F^{-1}(X+b,1-2\alpha+\alpha^*)} y f_{X+b}(y)dy + \alpha^* F^{-1}(X+b,\alpha^*)$$
$$+ (2\alpha - \alpha^*)F^{-1}(X+b, 1-2\alpha+\alpha^*)$$
$$= \int_{F^{-1}(X,\alpha^*)+b}^{F^{-1}(X,1-2\alpha+\alpha^*)+b} y f_X(y-b)dy + \alpha^*(F^{-1}(X,\alpha^*) + b$$
$$+ (2\alpha - \alpha^*)F^{-1}(X, 1-2\alpha+\alpha^*) + b$$
$$= \int_{F^{-1}(X,\alpha^*)}^{F^{-1}(X,1-2\alpha+\alpha^*)} (x+b) f_X(x)dx + \alpha^* F^{-1}(X,\alpha^*) + \alpha^* b$$
$$+ (2\alpha - \alpha^*)F^{-1}(X, 1-2\alpha+\alpha^*) + (2\alpha - \alpha^*)b$$
$$= \int_{F^{-1}(X,\alpha^*)}^{F^{-1}(X,1-2\alpha+\alpha^*)} x f_X(x)dx + b\int_{F^{-1}(X,\alpha^*)}^{F^{-1}(X,1-2\alpha+\alpha^*)} f_X(x)dx + \alpha^* F^{-1}(X,\alpha^*) + \alpha^* b$$
$$+ (2\alpha - \alpha^*)F^{-1}(X, 1-2\alpha+\alpha^*) + 2\alpha b - \alpha^* b$$
$$= \int_{F^{-1}(X,\alpha^*)}^{F^{-1}(X,1-2\alpha+\alpha^*)} x f_X(x)dx + b(1-2\alpha) + \alpha^* F^{-1}(X,\alpha^*)$$
$$+ \alpha^* b + (2\alpha - \alpha^*)F^{-1}(X, 1-2\alpha+\alpha^*) + 2\alpha b - \alpha^* b$$
$$= \int_{F^{-1}(X,\alpha^*)}^{F^{-1}(X,1-2\alpha+\alpha^*)} x f_X(x)dx + \alpha^* F^{-1}(X,\alpha^*) + (2\alpha - \alpha^*)F^{-1}(X, 1-2\alpha+\alpha^*) + b$$
$$= \mu^*_{mod}(X, 2\alpha) + b.$$

(b). Let $a > 0$. We also see that $\alpha^*(aX) = \arg_\alpha \inf_{0\leq\alpha<1-\gamma}\{a(F^{-1}(X,\gamma+\alpha) - F^{-1}(X,\alpha))\} = \alpha^*(X)$. We have

$$\mu^*_{mod}(aX,2\alpha) = \int_{F^{-1}(aX,\alpha^*)}^{F^{-1}(aX,1-2\alpha+\alpha^*)} y f_{aX}(y)dy + \alpha^* F^{-1}(aX,\alpha^*)$$
$$+ (2\alpha-\alpha^*)F^{-1}(aX,1-2\alpha+\alpha^*)$$
$$= \int_{aF^{-1}(X,\alpha^*)}^{aF^{-1}(X,1-2\alpha+\alpha^*)} y\frac{1}{a}f_X(\frac{y}{a})dy + \alpha^* aF^{-1}(X,\alpha^*)$$
$$+ (2\alpha-\alpha^*)aF^{-1}(X,1-2\alpha+\alpha^*)$$
$$= \int_{F^{-1}(X,\alpha^*)}^{F^{-1}(X,1-2\alpha+\alpha^*)} ax\frac{1}{a}f_X(x)adx + a(\alpha^* F^{-1}(X,\alpha^*)$$
$$+ (2\alpha-\alpha^*)F^{-1}(X,1-2\alpha+\alpha^*))$$
$$= a\int_{F^{-1}(X,\alpha^*)}^{F^{-1}(X,1-2\alpha+\alpha^*)} xf_X(x)dx + a(\alpha^* F^{-1}(X,\alpha^*)$$
$$+ (2\alpha-\alpha^*)F^{-1}(X,1-2\alpha+\alpha^*))$$
$$= a(\int_{F^{-1}(X,\alpha^*)}^{F^{-1}(X,1-2\alpha+\alpha^*)} xf_X(x)dx + \alpha^* F^{-1}(X,\alpha^*)$$
$$+ (2\alpha-\alpha^*)F^{-1}(X,1-2\alpha+\alpha^*))$$
$$= a\mu^*_{mod}(X,2\alpha).$$

(c). Consider the transformation of multiplying X by negative value -1. We see that $\alpha^*(-X) = 2\alpha - \alpha^*(X)$.

$$\mu^*_{mod}(-X,2\alpha) = \int_{F^{-1}(-X,\alpha^*(-X))}^{F^{-1}(-X,1-2\alpha+\alpha^*(-X))} yf_{-X}(y)dy + \alpha^*(-X)F^{-1}(-X,\alpha^*(-X))$$
$$+ (2\alpha-\alpha^*(-X))F^{-1}(-X,1-2\alpha+\alpha^*(-X))$$
$$= \int_{-F^{-1}(X,1-\alpha^*(-X))}^{-F^{-1}(X,2\alpha-\alpha^*(-X))} yf_X(-y)dy - \alpha^*(-X)F^{-1}(X,1-\alpha^*(-X))$$
$$- (2\alpha-\alpha^*(-X))F^{-1}(X,2\alpha-\alpha^*(-X))$$
$$= \int_{-F^{-1}(X,1-2\alpha+\alpha^*(X))}^{-F^{-1}(X,\alpha^*(X))} yf_X(-y)dy - (2\alpha-\alpha^*(X))F^{-1}(X,1-2\alpha+\alpha^*(X))$$
$$- \alpha^*(X)F^{-1}(X,\alpha^*(X))$$

$$= \int_{F^{-1}(X,1-2\alpha+\alpha^*(X)))}^{F^{-1}(X,\alpha^*(X))} (-x) f_X(x)(-dx) - (2\alpha - \alpha^*(X))F^{-1}(X, 1 - 2\alpha + \alpha^*(X))$$
$$- \alpha^*(X)F^{-1}(X, \alpha^*(X))$$
$$= -\int_{F^{-1}(X,\alpha^*(X))}^{F^{-1}(X,1-2\alpha+\alpha^*(X)))} x f_X(x)dx - \alpha^*(X)F^{-1}(X, \alpha^*(X))$$
$$- (2\alpha - \alpha^*(X))F^{-1}(X, 1 - 2\alpha + \alpha^*(X))$$
$$= -(\int_{F^{-1}(X,\alpha^*(X))}^{F^{-1}(X,1-2\alpha+\alpha^*(X)))} x f_X(x)dx + \alpha^*(X)F^{-1}(X, \alpha^*(X))$$
$$+ (2\alpha - \alpha^*(X))F^{-1}(X, 1 - 2\alpha + \alpha^*(X)))$$
$$= -\mu^*_{mod}(X, 2\alpha).$$

(d). If $X \geq 0$, then $F^{-1}(X, \alpha^*) \geq 0$ and $F^{-1}(X, 1 - 2\alpha + \alpha^*) \geq 0$ which implies that $\mu^*_{mod}(X, 2\alpha) = \int_{F^{-1}(X,\alpha^*)}^{F^{-1}(X,1-2\alpha+\alpha^*)} x f_X(x)dx + \alpha^* F^{-1}(X, \alpha^*) + (2\alpha - \alpha^*)F^{-1}(X, 1 - 2\alpha + \alpha^*)) \geq 0$. □

Now we consider a Monte Carlo simulation of the case normal with extremes where the design for drawing the observations is the one stated in Section 4.

Table 7. Breakdown points for two Winsorized means under normal distribution with extremes

γ	$\hat{\mu}^*_{mod}$	$\hat{\mu}^*_{med}$	γ	$\hat{\mu}^*_{mod}$	$\hat{\mu}^*_{med}$
0.1	0.910069	0.459972	0.6	0.409962	0.209985
0.2	0.810051	0.409962	0.7	0.310039	0.160005
0.3	0.709968	0.360039	0.8	0.209985	0.109989
0.4	0.609987	0.310039	0.9	0.109989	0.059998
0.5	0.509961	0.260009	0.95	0.049995	0.029999

It is verified that the mode type Winsorized mean has breakdown point approximated equal to the mode type trimmed mean. From the point of breakdown point, the mode type trimmed mean and the mode type Winsorized mean are equally interesting and important.

In the next we consider the simulation for exponential distribution plus extreme values. The results are stated in the next table.

Table 8. Breakdown points for two Winsorized means under exponential distribution with extreme values

γ	$\hat{\mu}^*_{mod}$	$\hat{\mu}^*_{med}$	γ	$\hat{\mu}^*_{mod}$	$\hat{\mu}^*_{med}$
0.1	0.909909	0.459954	0.6	0.409947	0.209961
0.2	0.809893	0.409942	0.7	0.310021	0.159980
0.3	0.709867	0.360023	0.8	0.209964	0.109963
0.4	0.609942	0.310021	0.9	0.109971	0.059979
0.5	0.509941	0.259985	0.95	0.049961	0.029965

This simulation provides the same indications as it shows in the exponential distribution case.

With this new Winsorized mean, we may also do a simulation to compare it with the median type Winsorized mean for their efficiency by computing the mean square errors (MSE). Consider a random sample of sample size 1000 from the following contaminated normal distribution

$$(1-\delta)N(0,1)+\delta|N(0,\sigma^2)|$$

and we compute these estimator's average MSE's from a replication of 10000. The following table display the simulation results of MSE's for these two Winsorized means.

Table 9. MSE's for Winsorized means under the contaminated normal distribution

	$2\alpha=0.3$ μ^*_{mod}	μ^*_{med}	$2\alpha=0.2$ μ^*_{mod}	μ_{med}	$2\alpha=0.1$ μ^*_{mod}	μ^*_{med}
$\delta=0.1$						
$\sigma=3$	0.0088	0.0147	0.0100	0.0184	0.0197	0.0212
$\sigma=10$	0.0134	0.0338	0.0184	0.0483	0.0379	0.2258
$\sigma=100$	0.0163	0.0467	0.0253	0.1028	0.0823	25.007
$\delta=0.2$						
$\sigma=3$	0.0291	0.0714	0.0645	0.0860	0.0667	0.1124
$\sigma=10$	0.0629	0.3302	0.1330	0.9095	0.8314	1.6227
$\sigma=100$	0.0893	31.186	0.2834	99.184	98.456	179.82
$\delta=0.3$						
$\sigma=3$	0.0753	0.2050	0.1710	0.2277	0.2295	0.2868
$\sigma=10$	0.2975	2.0790	0.9144	3.1213	2.9797	4.1722
$\sigma=100$	0.5636	223.51	112.47	342.84	341.96	464.18

In this study, the MSE's based on mode type Winsorized means are smaller than the corresponding median type Winsorized mean. This indicates that the mode type Winsorized mean not only have larger breakdown points but also more efficient in this case.

6. Conclusion

In this paper, we proposed the mode type trimmed mean. The most interesting result showing by this mean is that its breakdown point may be greater than 0.5 which was claimed by Hampel et al. (1986) that the breakdown point for any estimator may not be larger than 0.5. We also made a simulation showing that this new trimmed mean may be more efficient than the traditional trimmed mean. The results showing for mode type trimmed mean are re-studied for case of Winsorized means that also support to use the mode type robust estimators.

References

Chen, L. A. and Chiang, Y. C. (1996). Symmetric type quantile and trimmed means for location and linear regression model. *Journal of Nonparametric Statistics* **7**, 171-185.

Dixon, W. J. (1960). Simplified estimation from censored normal samples. *Annals of Mathematical Statistics*. 31, 385-391.

Hampel, F. R. (1971). A general qualitative definition of robustness. *Annals of Mathematcal Statistics*. 42, 1887-1896.

Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association*. 69, 383-393.

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

Hodges, J. L., Jr. and Lehmann, E. L. (1963). Estimates of location based on rank tests. *Annals of Mathematical Statistics*. 34, 598-611.

Hogg, R. V. (1974). Adaptive robust procedures. *Journal of the American Statistical Association*. 69, 909-927.

Huber, P. J. (1972). Robust statistics: A review. *Annals of Mathematical Statistics*. 43, 1041-1067.

Huber, P. J. (1981). *Robust Statistics*. Wiley: New York.

Jureckova, J. and Sen, P. K. (1996). *Robust Statistical Procedures: Asymptotics and Interrelations*. Wiley: New York.

Staudte, R. G. and Sheather, S. J. (1990). *Robust Estimation and Testing*. Wiley:New York.

Rousseeuw, P. J. (1983). Multivariate estimation with high breakdown point, paper presented at fourth Pannonian symposium on mathematical statistics and probability. Bad Tatzmannsdorf, Australia.

Rousseeuw, P. J. (1984). Least median of squares regression. *Journal of the American Statistical Association.* 79, 871-880.

Rousseeuw, P. J. (1987). *Robust Regression and Outlier Detection.* Wiley: New York.