The Annals of Statistics
1973, Vol. 1, No. 3, 472-477

THE ASYMPTOTIC DISTRIBUTION OF THE TRIMMED MEAN[1]

BY STEPHEN M. STIGLER

The University of Wisconsin, Madison

In this paper it is shown that in order for the trimmed mean to be asymptotically normal, it is necessary and sufficient that the sample be trimmed at sample percentiles such that the corresponding population percentiles are uniquely defined. (The sufficiency of this condition is well known.) In addition, the (non-normal) limiting distribution of the trimmed mean when this condition is not satisfied is derived, and it is shown that in some situations the use of the trimmed mean may lead to severely biased inferences. Some possible remedies are briefly discussed, including the use of "smoothly" trimmed means.

1. Introduction.

For many years, the trimmed mean has been an extremely popular estimator of location parameters (see Tukey and McLaughlin (1963), Bickel (1965), and Huber (1972) for accounts of its history and properties.) The usual reason given for using the trimmed mean (or the median, the ultimate in trimmed means) is robustness; in particular, the trimmed mean is less sensitive to extreme deviations and heavy-tailed distributions than is the ordinary sample mean. One purpose of this present paper is to point out that the trimmed mean lacks some of the robustness of the sample mean when departures from assumptions other than in the tails are considered.

It follows from the Theorem of Section 2 that a necessary and sufficient condition for the asymptotic normality of the trimmed mean is that the trimming be done at proportions corresponding to uniquely defined percentiles of the population distribution. (The sufficiency of this condition is well known—see Huber (1969), Shorack (1972), Stigler (1972).) If this condition does not hold, it will be shown that the limiting distribution is not normal, and use of the trimmed mean may lead to invalid tests or confidence intervals, even with large samples. This danger may exist when sampling from discrete populations or continuous populations with gaps, or when using grouped data.

In Section 2 we derive the asymptotic distribution of the trimmed mean for an arbitrary population distribution. Section 3 discusses the nature of the possible non-normal limiting distribution, the extent to which it may lead to biased inferences, and some alternatives to the ordinary trimmed mean which do not suffer from this disadvantage.

2. The asymptotic distribution of the trimmed mean.

Let $X_1, X_2, \cdots, X_n$ be independent and identically distributed, each with cumulative distribution function $F(x)$. We make no assumptions about F other than that it is proper ($\lim_{x\to\infty} F(x) -$

Received April 1972; revised August 1972.

[1] This research was supported in part by the Wisconsin Alumni Research Foundation, and in part by the United States Army under Grant No. DA-ARO-D-31-124-G917.

$F(-x) = 1$.) Let $X_{(1)} \leqq X_{(2)} \leqq \cdots \leqq X_{(n)}$ denote the order statistics of the sample. We consider the trimmed mean given by

$$S_n = \frac{1}{([\beta n] - [\alpha n])} \sum_{i=[\alpha n]+1}^{[\beta n]} X_{(i)}, \tag{1}$$

where $0 < \alpha < \beta < 1$ are any fixed numbers and $[\cdot]$ represents the greatest integer function.

Let a and b denote the largest αth and smallest βth percentiles of F; that is,

$$a = \sup\{x : F(x) \leqq \alpha\},$$
$$b = \inf\{x : F(x) \geqq \beta\}.$$

Let A and B be the lengths (possibly zero) of the intervals of the αth and βth percentiles:

$$A = a - \inf\{x : F(x) \geqq \alpha\},$$
$$B = \sup\{x : F(x) \leqq \beta\} - b.$$

Further, define

$$\begin{aligned} G(x) &= 0 && \text{for } x < a, \\ &= (F(x) - \alpha)/(\beta - \alpha) && \text{for } a \leqq x < b, \\ &= 1 && \text{for } x \geqq b, \end{aligned} \tag{2}$$

and set

$$\mu = \int_{-\infty}^{\infty} x\, dG(x), \tag{3}$$

$$\sigma^2 = \int_{-\infty}^{\infty} x^2\, dG(x) - \mu^2. \tag{4}$$

THEOREM. *As* $n \to \infty$, $\mathscr{L}(n^{\frac{1}{2}}(S_n - \mu)) \to \mathscr{L}(Z)$, *with* $Z = (\beta - \alpha)^{-1}(Y_1 + (b - \mu)Y_2 + (a - \mu)Y_3 + B\max(0, Y_2) - A\max(0, Y_3))$ *and* $E(Z) = [B(\beta(1-\beta))^{\frac{1}{2}} - A(\alpha(1-\alpha))^{\frac{1}{2}}]/[(\beta - \alpha)(2\pi)^{\frac{1}{2}}]$, *where* Y_1 *is* $N(0, (\beta - \alpha)\sigma^2)$ *and independent of* (Y_2, Y_3), *and* (Y_2, Y_3) *is* $N(\mathbf{0}, \mathbf{C})$,

$$\mathbf{C} = \begin{pmatrix} \beta(1-\beta) & -\alpha(1-\beta) \\ -\alpha(1-\beta) & \alpha(1-\alpha) \end{pmatrix}.$$

PROOF. Given $X_1, \cdots, X_n$, define random variables V_n and W_n as follows: If $F(a-) = \alpha$, let $V_n = (\# X_i < a)$; otherwise let V_n count the number of $X_i = a$ with probability $p_1 = (\alpha - F(a-))/(F(a) - F(a-))$ each (independently of the others) plus the number of $X_i < a$. If $F(b) = \beta$, let $W_n = (\# X_i \leqq b)$; otherwise let W_n count the number of $X_i = b$ with probability $p_2 = (\beta - F(b-))/(F(b) - F(b-))$ each (independently of the others) plus the number of $X_i < b$. (In the special case that $a = b$ and $F(a-) < \alpha < \beta < F(a)$, then conditional on $(\# X_i = a) = m$, perform m independent trinomial trials with probabilities $(p_1, p_2 - p_1, 1 - p_2)$, observing (R_1, R_2, R_3). Let $V_n = (\# X_i < a) + R_1$ and $W_n = V_n + R_2$.) Then V_n has a binomial (n, α) distribution, W_n has a binomial (n, β) distribution, and $\operatorname{Cov}(V_n, W_n) = n\alpha(1 - \beta)$.

In terms of V_n and W_n define

$$Y_n = \max\{[\alpha n], V_n\}, \qquad Z_n = \min\{[\beta n], W_n\}.$$

Then we can write

$$\begin{aligned}([\beta n] - [\alpha n])S_n &= \sum_{i=[\alpha n]+1}^{[\beta n]} X_{(i)}, \\ &= \sum_{i=[\alpha n]+1}^{Y_n} X_{(i)} + \sum_{i=Y_n+1}^{Z_n} X_{(i)} + \sum_{i=Z_n+1}^{[\beta n]} X_{(i)}, \\ &= T_n - A\{Y_n - [\alpha n]\} + B\{[\beta n] - Z_n\},\end{aligned}$$

where

$$\begin{aligned}T_n &= \sum_{i=[\alpha n]+1}^{[\beta n]} X^*_{(i)}, \\ X^*_{(i)} &= X_{(i)} + A \quad &&\text{if} \quad X_{(i)} < a, \\ &= X_{(i)} &&\text{if} \quad a \leqq X_{(i)} \leqq b, \\ &= X_{(i)} - B &&\text{if} \quad b < X_{(i)}.\end{aligned}$$

Note that if $F(a-) < \alpha$ or $F(b) > \beta$, then $A = 0$ or $B = 0$, respectively.

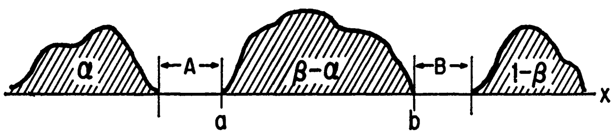

FIG. 1. The density of $F(x)$.

In the above, we understand $\sum_{i=c}^{d} e_i$ to be zero if $d < c$. In what follows it will be convenient to use the notation

$$\begin{aligned}\sum_{i=c}^{d} e_i &= \sum_{i=c+1}^{d} e_i \quad &&\text{if} \quad d > c, \\ &= 0 &&\text{if} \quad d = c, \\ &= -\sum_{i=d+1}^{c} e_i &&\text{if} \quad d < c.\end{aligned}$$

In particular, $\sum_{i=c}^{d} e = (d - c)e$.

Since $X^*_{(i)} = X_{(i)}$ for $V_n + 1 \leqq i \leqq W_n$, we can then write

$$\begin{aligned}n^{-\frac{1}{2}}(T_n - n(\beta - \alpha)\mu) = n^{-\frac{1}{2}}\Big(\sum_{i=V_n+1}^{W_n} X_{(i)} - n(\beta - \alpha)\mu\Big) \\ + n^{-\frac{1}{2}} \sum_{i=W_n}^{[\beta n]} (X^*_{(i)} - b) + n^{-\frac{1}{2}} \sum_{i=[\alpha n]}^{V_n} (X^*_{(i)} - a) \\ + n^{-\frac{1}{2}} b([\beta n] - W_n) + n^{-\frac{1}{2}} a(V_n - [\alpha n]).\end{aligned}$$

Now,

$$\begin{aligned}\Big|n^{-\frac{1}{2}} \sum_{i=W_n}^{[\beta n]} (X^*_{(i)} - b)\Big| &\leqq |n^{-\frac{1}{2}}([\beta n] - W_n)| \cdot \max\{|X^*_{([\beta n])} - b|, |X^*_{(W_n)} - b|\} \\ &\to 0 \quad \text{in probability},\end{aligned}$$

since $X^*_{([\beta n])} \to_p b$, $X^*_{(W_n)} \to_p b$, and $n^{-\frac{1}{2}}([\beta n] - W_n)$ is bounded in probability. Similarly,

$$n^{-\frac{1}{2}} \sum_{i=[\alpha n]}^{V_n} (X^*_{(i)} - a) \to_p 0.$$

Thus we have

$$n^{-\frac{1}{2}}\{([\beta n] - [\alpha n])S_n - n(\beta - \alpha)\mu\}$$
$$= n^{-\frac{1}{2}}(\sum_{i=V_n+1}^{W_n}(X_{(i)} - \mu)) + n^{-\frac{1}{2}}(b - \mu)(\beta n - W_n) + n^{-\frac{1}{2}}(a - \mu)(V_n - \alpha n)$$
$$+ B \max(0, n^{-\frac{1}{2}}(\beta n - W_n)) - A \max(0, n^{-\frac{1}{2}}(V_n - \alpha n)) + \varepsilon_n ,$$

where $\varepsilon_n \rightarrow_p 0$ as $n \rightarrow \infty$.

We now note that given W_n and V_n, $\sum_{i=V_n+1}^{W_n} X_{(i)}$ has the same probability distribution as the sum of $W_n - V_n$ independent random variables, each with distribution function $G(x)$ (given by (2)). It then follows easily from the central limit theorem (or appeal to the results of Wittenberg (1964)) that the joint asymptotic distribution of $n^{-\frac{1}{2}}(\sum_{i=V_n+1}^{W_n}(X_{(i)} - \mu))$, $n^{-\frac{1}{2}}(\beta n - W_n)$, and $n^{-\frac{1}{2}}(V_n - \alpha n)$ is $\mathscr{L}(Y_1, Y_2, Y_3)$, which proves the theorem. □

We remark that this proof can be easily adapted to the case where the X_i's are not identically distributed, and it can be used to derive the limiting distribution of other statistics of the form $n^{-1} \sum J(i/(n+1))X_{(i)}$, where J may have a jump discontinuity corresponding to a non-unique percentile of F (see also Stigler (1972)). The restriction to proper F is of course unnecessary—all that is needed is $\lim_{x\to-\infty} F(x) < \alpha < \beta < \lim_{x\to\infty} F(x)$. Also, $\alpha = 0$ could be allowed if $\int_{-\infty}^{0} x^2 \, dF(x) < \infty$, and similarly $\beta = 1$ if $\int_0^\infty x^2 \, dF(x) < \infty$.

3. The possible non-normality of the trimmed mean. It is apparent from the theorem of Section 2 that, while the trimmed mean is more "robust" than the sample mean in the sense that its behavior is not affected by the *tails* of the population distribution, it may actually be less robust in other senses. In particular, the trimmed mean is sensitive to gaps in the distribution near the trimming proportions. (These gaps may be thought of as weight in the tails of the "trimmed" distribution $G(x)$.) The question naturally arises as to the practical importance of this possible non-normality, and what may be done about it.

It could be argued that this behavior is of no importance since in order for non-normality to occur, the trimming must be done at an α (or β) corresponding *exactly* to a non-unique percentile—an unlikely state of affairs! However, this viewpoint overlooks the fact that we only employ asymptotic distributions to approximate finite sample size distributions, and that we may reasonably expect the actual distribution of the trimmed mean to be close to the possible non-normal limit for moderate sample sizes, providing only that the trimming is done somewhat near a gap. That this may be the case when sampling from a discrete population or using grouped data is obvious. Another situation in which this may occur is when outliers are present in a proportion close to the trimming proportion, exactly the type of situation for which the trimmed mean is often advocated.

How seriously can a gap or non-unique percentile affect the distribution of S_n? If we think of S_n as being used to estimate or test hypotheses about μ (given by (3)), then the fact that the mean of the limiting distribution of $n^{\frac{1}{2}}(S_n - \mu)$ is

$$E(Z) = [B(\beta(1-\beta))^{\frac{1}{2}} - A(\alpha(1-\alpha))^{\frac{1}{2}}]/[(\beta - \alpha)(2\pi)^{\frac{1}{2}}]$$

(rather than zero) suggests that inferences may be thrown off considerably in some situations. To see the extent to which this can happen, consider the case where $A = 0$. It can be easily seen that we then have

$$P((\beta - \alpha)Z \leqq z) = \int_{-\infty}^{z} \Phi(z; w)\phi(w)\, dw ,$$

where $\phi(w)$ is the $N(0, \sigma_1^2)$ density and $\Phi(z; w)$ is the $N(w\rho\sigma_2/\sigma_1, \sigma_2^2(1 - \rho^2))$ cumulative distribution function, and

$$\begin{aligned}
\sigma_1^2 &= \sigma^2 + (b - \mu)^2\beta(1 - \beta) - 2(a - \mu)(b - \mu)\alpha(1 - \beta) \\
&\quad + (a - \mu)^2\alpha(1 - \alpha) , \\
\sigma_2^2 &= \sigma_1^2 + [B^2 + 2B(b - \mu)]\beta(1 - \beta) - 2(a - \mu)B\alpha(1 - \beta) , \\
\sigma_1\sigma_2\rho &= \sigma_1^2 + B((b - \mu)\beta - (a - \mu)\alpha)(1 - \beta) ,
\end{aligned}$$

with σ^2 given by (4).

To see how the distribution of Z can vary with B, consider the special case where $\alpha = 0.1$, $\beta = 0.9$, and

$$\begin{aligned}
F(x) &= \Phi(x) && x \leqq b , \\
&= \Phi(b) && b \leqq x \leqq b + B , \\
&= \Phi(x - B) && x \geqq b + B ,
\end{aligned}$$

where $\Phi(\cdot)$ is the standard normal distribution function, $b = \Phi^{-1}(\beta)$, and $B \geqq 0$. Table 1 gives the cumulative distribution function of Z for certain values corresponding to fractional points of the distribution for $B = 0$ and a range of B.

TABLE 1
$P(Z \leqq z)$ as a function of B

		Z						
		−2.40	−1.69	−1.32	0.00	1.32	1.69	2.40
B	0.0	.010	.050	.100	.500	.900	.950	.990
	0.5	.007	.036	.073	.390	.790	.863	.947
	1.0	.007	.033	.067	.342	.689	.759	.855
	2.0	.006	.031	.062	.308	.592	.656	.722
	5.0	.006	.030	.059	.284	.514	.551	.595
	10.0	.006	.030	.058	.275	.486	.515	.547

It is interesting to note that as $B \to \infty$, the density of Z approaches that of a defective distribution with total probability slightly over .5, the limit being close to .5 times the standard normal density.

We see from Table 1 that, depending upon the size of B, confidence intervals for μ centered at S_n and tests of hypotheses about μ based on S_n may be severely biased if no allowance is made for the existence of gaps. What can be done about this? In the case of grouped data, one remedy is to "de-group" near the αth and βth sample percentiles. However, it should be recalled that "de-grouping" amounts to assuming the true population distribution is uniform over the interval in question, an assumption which may introduce a bias of its own!

Yet another solution—this one valid for any case—is to use a more smoothly trimmed mean, say $n^{-1}\sum J(i/(n+1))X_{(i)}$, where

$$\begin{aligned} J(u) &= (u-\alpha)(.5-\alpha)^{-1} && \alpha \leqq u \leqq .5\,, \\ &= (1-\alpha-u)(.5-\alpha)^{-1} && .5 \leqq u \leqq 1-\alpha\,, \\ &= 0 && \text{otherwise;} \end{aligned}$$

or

$$\begin{aligned} J(u) &= \left(u-\frac{\alpha}{2}\right)2\frac{h}{\alpha} && \frac{\alpha}{2} \leqq u \leqq \alpha\,, \\ &= h && \alpha \leqq u \leqq 1-\alpha\,, \\ &= \left(1-\frac{\alpha}{2}-u\right)2\frac{h}{\alpha} && 1-\alpha \leqq u \leqq 1-\frac{\alpha}{2}\,, \\ &= 0 && \text{otherwise} \end{aligned}$$

where $h = 2(2-3\alpha)^{-1}$. Such statistics are asymptotically normally distributed for *any* population distribution (or, if the X_i's are not identically distributed, any set of distributions which are uniformly bounded in probability) (see Stigler (1972)), but they retain the most attractive features of the trimmed mean; robustness to heavy-tailed distributions and outliers. The efficiencies of the first of these smoothly trimmed means relative to trimmed means and Winsorized means for various continuous populations have been studied by Crow and Siddiqui (1967), where they call it a "linearly weighted mean."

Regardless of which approach is taken, some alternative to the trimmed mean is desirable in circumstances where the population may contain gaps. Clearly the same could be said about any linear function of order statistics with a discontinuous weight function. In particular, it is obvious that the situation is even worse for the Winsorized mean. The slight increase in calculations necessary to use a smooth weight function would seem to be worthwhile in many situations.

REFERENCES

[1] Bickel, P. J. (1965). On some robust estimates of location. *Ann. Math. Statist.* **36** 847–858.

[2] Huber, P. J. (1969). *Théorie de l' Inférence Statistique Robuste*. Les Presses de l' Université de Montréal.

[3] Huber, P. J. (1972). Robust statistics: A review. *Ann. Math. Statist.* **43** 1041–1067.

[4] Shorack, G. R. (1972). Random means. Unpublished.

[5] Stigler, S. M. (1972). Linear functions of order statistics with smooth weight functions. University of Wisconsin Mathematics Research Center T.R. #1211.

[6] Wittenberg, H. (1964). Limiting distributions of random sums of independent random variables. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **3** 7–18.

[7] Tukey, J. W., and McLaughlin, D. H. (1963). Less vulnerable confidence and significance procedures for location based on a single sample (Trimming/Winsorization 1). *Sankhyā, Ser. A* **25** 331–352.

[8] Crow, E. L., and Siddiqui, M. M. (1967). Robust estimates of location. *J. Amer. Statist. Assoc.* **62** 353–389.

Department of Statistics
University of Wisconsin
Madison, Wisconsin 53706