The Annals of Mathematical Statistics
1971, Vol. 42, No. 5, 1540–1552

SOME FLEXIBLE ESTIMATES OF LOCATION[1]

BY LOUIS A. JAECKEL
University of California, Berkeley[2]

This paper considers two procedures for estimating the center of a symmetric distribution, which use the observations themselves to choose the form of the estimator. Both procedures begin with a family of possible estimators. We use the observations to estimate the asymptotic variance of each member of the family of estimators. We then choose the estimator in the family with smallest estimated asymptotic variance and use the value given by that estimator as the location estimate. These procedures are shown to be asymptotically as good as knowing beforehand which estimator in the family is best for the given distribution, and using that estimator.

1. Introduction and summary. In this paper we consider two procedures for estimating the center of a symmetric distribution, which use the observations themselves to choose the form of the estimator. Both procedures begin with a family of possible estimators. In the first case we take a family of trimmed means with different trimming proportions, and in the second case a family of linear combinations of a finite number of given L-estimators. (An L-estimator is a linear combination of order statistics with fixed coefficients. See Jaeckel (1971).) We use the observations to estimate the asymptotic variance of each member of the family of estimators. We then choose the estimator in the family with smallest estimated asymptotic variance and use the value given by that estimator as the location estimate. We shall show that in each case the estimator chosen in this way converges to the true best estimator in the family (that is, the one with smallest asymptotic variance) in such a way that the two estimates have the same limiting distribution and therefore the same asymptotic variance. Thus the procedure here is asymptotically as good as knowing beforehand which estimator in the family is best for the given distribution, and using that estimator. (Note that by "asymptotic variance" we mean the variance of the limiting distribution. We make no assertions about the actual variance of an estimator for a given sample size, or even about the limit of such variances.)

It may be inferred from the asymptotic variance formulas that if an estimator is close, in some sense, to the best estimator for a given F, it will have high asymptotic efficiency relative to the best estimator for that F. Thus, by providing a variety of estimators from which to choose, these procedures will have high relative efficiency for a wide class of distributions. Stein (1956), Hájek (1962), and van

Received October 30, 1970.

[1] This paper was prepared with the partial support of the United States Public Health Service Grant 10525–05 and the Office of Naval Research, Contract Nonr N00014–67–A–0114–0004. This work forms a part of the author's 1969 Ph.D. dissertation at the University of California, Berkeley.

[2] Now with Bell Telephone Laboratories, Murray Hill.

Eeden (1970) have constructed testing and estimation procedures which are asymptotically optimal for all densities, but these procedures are not useful in practice for moderate sample sizes. Since the procedures described in this paper involve estimating one or a small number of intermediate parameters, it is hoped that they will approach their asymptotic behavior rapidly enough so that they will be useful for moderate sample sizes. Some Monte Carlo results, described at the end of Section 2, suggest that this is the case. In addition, the procedures here are not difficult computationally.

2. The optimal trimmed mean. Let $X_{(1)} \leqq \cdots \leqq X_{(n)}$ be the order statistics of a random sample of size n with distribution $F(x-\theta)$, where F is symmetric, that is, $F(x)+F(-x) = 1$, and has a density f. We want to estimate the unknown parameter θ. Since we shall be dealing with translation-invariant statistics, we may assume that $\theta = 0$.

We define the *α-trimmed mean* as

$$m(\alpha) = \frac{1}{n-2[\alpha n]} \sum_{i=[\alpha n]+1}^{n-[\alpha n]} X_{(i)}.$$

Under very general regularity conditions, $n^{\frac{1}{2}}m(\alpha)$ is asymptotically normal with asymptotic variance

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

where $x_\alpha = F^{-1}(\alpha)$ and $x_{1-\alpha} = F^{-1}(1-\alpha)$. See Bickel (1965).

We construct the following estimator of $\sigma^2(\alpha)$ by replacing the terms in the formula above by their estimates. Let

$$s^2(\alpha) = \frac{1}{(1-2\alpha)^2}\left\{\frac{1}{n}\sum_{i=[\alpha n]+1}^{n-[\alpha n]} [X_{(i)} - m(\alpha)]^2 + \alpha[X_{([\alpha n]+1)} - m(\alpha)]^2 + \alpha[X_{(n-[\alpha n])} - m(\alpha)]^2\right\}.$$

We point out that $s^2(\alpha)$ is intended to estimate the variance of $m(\alpha)$, rather than to estimate a scale parameter. When many differently shaped distributions and different values of α are allowed, the two problems are quite distinct. Tukey and McLaughlin (1963) considered the problem of finding an estimate of the variance of the trimmed mean which will be valid for a wide class of distributions and proposed an estimate which is almost the same as ours. They also proposed choosing the trimming proportion to minimize the estimated variance, as we do below.

We now define the optimal trimmed mean. We assume a range of permissible values of α, say $\alpha_0 \leqq \alpha \leqq \alpha_1$, is fixed in advance. We compute $s^2(\alpha)$ for all α in this range such that αn is an integer. Let $\hat{\alpha}$ be the value of α which minimizes $s^2(\alpha)$. We assume this α is unique.

DEFINITION. We take $m(\hat{\alpha})$, the $\hat{\alpha}$-trimmed mean, as our location estimate, and call it the optimal trimmed mean.

In order to compute $m(\hat{\alpha})$, we must compute $m(\alpha)$ and $s^2(\alpha)$ for each of the approximately $(\alpha_1 - \alpha_0)n$ allowable values of α. This may be done systematically by considering successive values of α, beginning with the largest, since for that value the number of terms in the sums is the smallest. Then, for each α, $s^2(\alpha)$ and $m(\alpha)$ may be easily computed by making use of the results of the computations for the preceding α.

THEOREM 1. *Suppose* $0 < \alpha_0 < \alpha_1 < \frac{1}{2}$. *Suppose F satisfies the conditions of Lemma* 1 *below, with this same* α_0. *Suppose also that* $\sigma^2(\alpha)$ *has a unique minimum in the interval* $\alpha_0 \leqq \alpha \leqq \alpha_1$, *attained at* $\alpha = A$. *Then*

$$n^{\frac{1}{2}}m(\hat{\alpha}) - n^{\frac{1}{2}}m(A) \to_P 0.$$

It follows that both have the same limiting distribution and therefore the same asymptotic variance.

Thus, when the best α is unknown, but is known to lie between α_0 and α_1, the optimal trimmed mean is asymptotically as good as knowing and using the best α.

The optimal trimmed mean is closely related to Huber's Proposal 3, Huber (1964), page 97, which chooses a maximum likelihood type estimator to minimize the estimated asymptotic variance among the family of such estimators which correspond to the family of trimmed means. See Jaeckel (1971), Section 3. It follows from Theorem 1 above and Theorem 3 of Jaeckel (1971) that, subject to the conditions of those theorems, the optimal trimmed mean has an optimality property which Huber conjectured for his Proposal 3: The optimal trimmed mean is, simultaneously for each permissible ε and scale parameter σ, an asymptotic minimax solution to Huber's problem, Huber (1964), Section 5, with these parameters and with G the normal distribution.

The difference between the optimal trimmed mean and the procedures for rejecting outliers which have been proposed by several authors is that the optimal trimmed mean is not designed to be a method for recognizing and casting out spurious observations; instead, it is intended to make use of the data as well as possible, under the conditions of the theorem. The approach of Anscombe (1960) is similar to ours in this respect.

We make a few remarks about the conditions of the theorem before proceeding with the proof. The important case where the best value of α is zero is not considered. The conditions imposed on F seem more restrictive than they need be. The condition that $\sigma^2(\alpha)$ have a unique minimum could be removed if the procedure were modified slightly. A possible modification is described following Lemma 3 below. Under the conditions of the theorem, $\sigma^2(\alpha)$ is a continuous function of α, so the minimum of $\sigma^2(\alpha)$ for $\alpha_0 \leqq \alpha \leqq \alpha_1$ is attained.

We restate a definition and a lemma from Jaeckel (1971).

DEFINITION. Let $i^* = i/(n+1)$.

LEMMA 1. *Suppose F is symmetric, has a density f, and there are numbers* $\alpha_0 > 0$, $\varepsilon_0 > 0$, *and* $f_0 > 0$ *such that* $f(x) \geqq f_0$ *for all x such that* $\alpha_0 - \varepsilon_0 \leqq F(x) \leqq$

$1-(\alpha_0-\varepsilon_0)$. *Then* $X_{(i)}-F^{-1}(i^*)$ *is* $O(n^{-\frac{1}{2}})$ *in probability uniformly in* $i = [\alpha_0 n]+1, \cdots, n-[\alpha_0 n]$. *That is, for all* $\delta > 0$ *there exist* D *and* N *such that for all* $n \geqq N$:

$$P\left\{|X_{(i)}-F^{-1}(i^*)| \leqq \frac{D}{n^{\frac{1}{2}}f_0} \text{ for } i = [\alpha_0 n]+1, \cdots, n-[\alpha_0 n]\right\} \geqq 1-\delta.$$

LEMMA 2. *Under the conditions of Lemma* 1, $s^2(\alpha)$ *converges to* $\sigma^2(\alpha)$ *in probability uniformly in* α *such that* $\alpha_0 \leqq \alpha \leqq \alpha_1$; *that is, given* $\varepsilon > 0$ *and* $\delta > 0$, *there exists* N *such that for all* $n \geqq N$:

$$P\{|s^2(\alpha)-\sigma^2(\alpha)| < \varepsilon \text{ for all } \alpha_0 \leqq \alpha \leqq \alpha_1\} \geqq 1-\delta.$$

PROOF. By the symmetry of F,

$$\textstyle\sum_{i=[\alpha n]+1}^{n-[\alpha n]} F^{-1}(i^*) = 0,$$

so that

$$m(\alpha) = \frac{1}{n-2[\alpha n]}\sum X_{(i)} = \frac{1}{n-2[\alpha n]}\sum [X_{(i)}-F^{-1}(i^*)].$$

For a given $\delta > 0$, we apply Lemma 1. If we let

$$C_1 = \frac{D}{n^{\frac{1}{2}}f_0}$$

and

$$C_2 = 2C_1[2F^{-1}(1-\alpha_0+\varepsilon_0)+2C_1],$$

where D, f_0, and ε_0 are as given in Lemma 1, we have

$$|X_{(i)}-F^{-1}(i^*)| \leqq C_1 \text{ for } i = [\alpha_0 n]+1, \cdots, n-[\alpha_0 n]$$

with probability $\geqq 1-\delta$ for all sufficiently large n. Note that C_1 and C_2 depend on n, so that we can fix n later on in the proof.

Suppose the inequalities of Lemma 1 hold. Then, for $\alpha_0 \leqq \alpha \leqq \alpha_1$ and $i = [\alpha_0 n]+1, \cdots, n-[\alpha_0 n]$,

$$|m(\alpha)| \leqq \frac{1}{n-2[\alpha n]}\sum|X_{(i)}-F^{-1}(i^*)| \leqq C_1, |X_{(i)}-m(\alpha)-F^{-1}(i^*)| \leqq 2C_1,$$

and

$$\begin{aligned} &|[X_{(i)}-m(\alpha)]^2-F^{-1}(i^*)^2| \\ &\quad = |X_{(i)}-m(\alpha)-F^{-1}(i^*)| \cdot |X_{(i)}-m(\alpha)+F^{-1}(i^*)| \\ &\quad \leqq 2C_1 \cdot |2F^{-1}(i^*)+X_{(i)}-m(\alpha)-F^{-1}(i^*)| \\ &\quad \leqq 2C_1 \cdot [2|F^{-1}(i^*)|+2C_1] \leqq C_2. \end{aligned}$$

Let

$$t^2(\alpha) = \frac{1}{(1-2\alpha)^2}\left\{\frac{1}{n}\sum_{i=[\alpha n]+1}^{n-[\alpha n]} F^{-1}(i^*)^2+\alpha F^{-1}[([\alpha n]+1)^*]^2+\alpha F^{-1}[(n-[\alpha n])^*]^2\right\}.$$

We shall show that $s^2(\alpha)$ is near $t^2(\alpha)$ and $t^2(\alpha)$ is near $\sigma^2(\alpha)$.

$$\left|\frac{1}{n}\sum_{i=[\alpha n]+1}^{n-[\alpha n]}[X_{(i)}-m(\alpha)]^2-\frac{1}{n}\sum_{i=[\alpha n]+1}^{n-[\alpha n]}F^{-1}(i^*)^2\right|$$
$$\leqq \frac{1}{n}\sum\left|[X_{(i)}-m(\alpha)]^2-F^{-1}(i^*)^2\right|$$
$$\leqq C_2.$$
$$\alpha\left|[X_{([\alpha n]+1)}-m(\alpha)]^2-F^{-1}[([\alpha n]+1)^*]^2\right| \leqq \alpha C_2.$$
$$\alpha\left|[X_{(n-[\alpha n])}-m(\alpha)]^2-F^{-1}[(n-[\alpha n])^*]^2\right| \leqq \alpha C_2.$$

Therefore

$$\left|s^2(\alpha)-t^2(\alpha)\right| \leqq \frac{1}{(1-2\alpha)^2}\cdot 2C_2 \leqq \frac{2C_2}{(1-2\alpha_1)^2}.$$

Letting $t = F(x)$ in the formula for $\sigma^2(\alpha)$, we have

$$\int_{x_\alpha}^{x_{1-\alpha}} x^2 f(x)\,dx = \int_\alpha^{1-\alpha} F^{-1}(t)^2\,dt$$

and $2\alpha x_\alpha{}^2 = \alpha F^{-1}(\alpha)^2+\alpha F^{-1}(1-\alpha)^2$, so that

$$\sigma^2(\alpha) = \frac{1}{(1-2\alpha)^2}\left\{\int_\alpha^{1-\alpha} F^{-1}(t)^2\,dt+\alpha F^{-1}(\alpha)^2+\alpha F^{-1}(1-\alpha)^2\right\}.$$

The three terms in $t^2(\alpha)$ converge to the corresponding terms in $\sigma^2(\alpha)$ at the rate of $O(n^{-1})$, the first one because it is a sum which approximates the integral. Since the derivative of $F^{-1}(t)^2$ is bounded for the range of t under consideration, this convergence is uniform in $\alpha \geqq \alpha_0$. Since $\alpha \leqq \alpha_1$ implies $(1-2\alpha)^{-2} \leqq (1-2\alpha_1)^{-2}$, we conclude that $t^2(\alpha) \to \sigma^2(\alpha)$ uniformly in $\alpha_0 \leqq \alpha \leqq \alpha_1$.

For a given $\varepsilon > 0$, we can choose an N such that for all $n \geqq N$, $\left|s^2(\alpha)-\sigma^2(\alpha)\right| \leqq \left|s^2(\alpha)-t^2(\alpha)\right|+\left|t^2(\alpha)-\sigma^2(\alpha)\right| < \varepsilon$ for all $\alpha_0 \leqq \alpha \leqq \alpha_1$, since C_2 depends on n. Since these inequalities hold with probability $\geqq 1-\delta$, the lemma is proved.

LEMMA 3. *Under the conditions of Lemma* 1, $\hat{\alpha} \to_P A$ *as* $n \to \infty$.

PROOF. We must show that for all $\delta > 0$ and $E > 0$, $P\{|\hat{\alpha}-A| < E\} \geqq 1-\delta$ for all sufficiently large n. For a given $E > 0$, let $J = \{\alpha : \alpha \in [\alpha_0, \alpha_1]$ but $\notin (A-E, A+E)\}$. We assume J is not empty. Let $T = \inf\{\sigma^2(\alpha) : \alpha \in J\}$. Since $\sigma^2(\alpha)$ is continuous for $\alpha \in [\alpha_0, \alpha_1]$ and has a unique minimum at $\alpha = A$, we must have $T > \sigma^2(A)$. Choose D such that $0 < D < \frac{1}{2}[T-\sigma^2(A)]$.

Suppose $\left|s^2(\alpha)-\sigma^2(\alpha)\right| < D$ for all $\alpha \in [\alpha_0, \alpha_1]$. For a given $\delta > 0$, this inequality holds, by Lemma 2, with probability $\geqq 1-\delta$ for sufficiently large n. Then, for all $\alpha \in J$,

$$s^2(\alpha) > \sigma^2(\alpha)-D \geqq T-D,$$

and $s^2(A) < \sigma^2(A)+D$. Therefore, for $\alpha \in J$,

$$s^2(\alpha)-s^2(A) > T-2D-\sigma^2(A).$$

Since $D < \frac{1}{2}[T-\sigma^2(A)]$, $T-\sigma^2(A)-2D > 0$, so $s^2(\alpha)-s^2(A) > 0$, or $s^2(\alpha) > s^2(A)$ for all $\alpha \in J$. Therefore, the minimum value of $s^2(\alpha)$ for $\alpha \in [\alpha_0, \alpha_1]$ is not attained by any $\alpha \in J$. So the minimum must be attained at some $\alpha \in (A-E, A+E)$; that is, $\hat{\alpha} \in (A-E, A+E)$, or $|\hat{\alpha}-A| < E$.

Since this event occurs with probability $\geqq 1-\delta$ for sufficiently large n, the lemma is proved.

If $\sigma^2(\alpha)$ is not assumed to have a unique minimum in $\alpha \in [\alpha_0, \alpha_1]$, we could modify the procedure as follows. Choose a sequence $\{z_n\}$ such that $z_n > 0$ and $z_n \to 0$, but $n^{\frac{1}{2}} z_n \to \infty$. Let $A = \inf\{\alpha \in [\alpha_0, \alpha_1]: \sigma^2(\alpha) = \text{minimum}\}$. Let $M^2 = \min\{s^2(\alpha): \alpha \in [\alpha_0, \alpha_1]\}$. Let $\hat{\alpha}$ be the smallest $\alpha \in [\alpha_0, \alpha_1]$ such that $s^2(\alpha) \leqq M^2(1+z_n)$. Then, since $s^2(\alpha)-\sigma^2(\alpha)$ in Lemma 2 is $O(n^{-\frac{1}{2}})$ in probability, it can be shown by an argument similar to that in Lemma 3 that $\hat{\alpha} \to_P A$. The theorem will then follow, by the proof below.

PROOF OF THE THEOREM. Given $\varepsilon > 0$ and $\delta > 0$, we must show that for sufficiently large n, $P\{n^{\frac{1}{2}}|m(\hat{\alpha})-m(A)| < \varepsilon\} \geqq 1-\delta$. By Lemma 1, there exists C such that $P\{|X_{(i)}-F^{-1}(i^*)| \leqq n^{-\frac{1}{2}}C$ for $i = [\alpha_0 n]+1, \cdots, n-[\alpha_0 n]\} \geqq 1-\frac{1}{2}\delta$ for sufficiently large n. Let

$$E = \frac{(1-2\alpha_1)\varepsilon}{6C}.$$

By Lemma 3, $P\{|\hat{\alpha}-A| < E\} \geqq 1-\frac{1}{2}\delta$ for sufficiently large n. Therefore,

$$P\{|X_{(i)}-F^{-1}(i^*)| \leqq n^{-\frac{1}{2}}C \quad \text{for} \quad i = [\alpha_0 n]+1, \cdots, n-[\alpha_0 n] \quad \text{and} \quad |\hat{\alpha}-A| < E\} \geqq 1-\delta$$

for sufficiently large n.

Suppose these inequalities hold. We write

$$\begin{aligned} m(\hat{\alpha})-m(A) &= \frac{1}{n-2[\hat{\alpha}n]}\sum_{i=[\hat{\alpha}n]+1}^{n-[\hat{\alpha}n]} X_{(i)} - \frac{1}{n-2[An]}\sum_{i=[An]+1}^{n-[An]} X_{(i)} \\ &= \left(\frac{1}{n-2[\hat{\alpha}n]}-\frac{1}{n-2[An]}\right)\sum_{i=[An]+1}^{n-[An]} X_{(i)} + \frac{1}{n-2[\hat{\alpha}n]}\cdot V, \end{aligned}$$

where

$$\begin{aligned} V &= \textstyle\sum_{i=[\hat{\alpha}n]+1}^{[An]} X_{(i)} + \sum_{i=n-[An]+1}^{n-[\hat{\alpha}n]} X_{(i)} && \text{if } [\hat{\alpha}n] < [An] \\ &= 0 && \text{if } [\hat{\alpha}n] = [An] \\ &= -\textstyle\sum_{i=[An]+1}^{[\hat{\alpha}n]} X_{(i)} - \sum_{i=n-[\hat{\alpha}n]+1}^{n-[An]} X_{(i)} && \text{if } [\hat{\alpha}n] > [An]. \end{aligned}$$

Since $[\alpha n]/n = \alpha+O(n^{-1})$, and $\hat{\alpha} \leqq \alpha_1$,

$$\begin{aligned} n\cdot\left|\frac{1}{n-2[\hat{\alpha}n]}-\frac{1}{n-2[An]}\right| &= \left|\frac{1}{1-2\hat{\alpha}}-\frac{1}{1-2A}\right| + O(n^{-1}) \\ &= \left|\frac{2(\hat{\alpha}-A)}{(1-2\hat{\alpha})(1-2A)}\right| + O(n^{-1}) \\ &\leqq \frac{2E}{(1-2\alpha_1)(1-2A)} + O(n^{-1}). \end{aligned}$$

By the symmetry of F,

$$\begin{aligned}|\textstyle\sum_{i=[An]+1}^{n-[An]} X_{(i)}| &= |\sum [X_{(i)} - F^{-1}(i^*)]| \\ &\leqq \sum |X_{(i)} - F^{-1}(i^*)| \\ &\leqq (n - 2[An]) \cdot n^{-\frac{1}{2}} C \\ &= n^{\frac{1}{2}}[(1-2A)C + O(n^{-1})].\end{aligned}$$

Therefore,

$$\begin{aligned}&n^{\frac{1}{2}}\left|\left(\frac{1}{n-2[\hat{\alpha}n]} - \frac{1}{n-2[An]}\right)\sum X_{(i)}\right| \\ &\leqq \frac{2E}{(1-2\alpha_1)(1-2A)} \cdot (1-2A)C + O(n^{-1}) \\ &= \frac{\varepsilon}{3} + O(n^{-1}).\end{aligned}$$

Also,

$$\begin{aligned}|V| = |\textstyle\sum X_{(i)}| &= |\sum [X_{(i)} - F^{-1}(i^*)]| \\ &\leqq 2|[\hat{\alpha}n] - [An]| \cdot n^{-\frac{1}{2}} C \\ &= n^{\frac{1}{2}}[2C|\hat{\alpha} - A| + O(n^{-1})] \\ &\leqq n^{\frac{1}{2}}[2CE + O(n^{-1})],\end{aligned}$$

where the range of the summation index is as indicated in the definition of V. Therefore,

$$\begin{aligned}n^{\frac{1}{2}}\frac{1}{n-2[\hat{\alpha}n]}|V| &= \frac{n}{n-2[\hat{\alpha}n]} \cdot n^{-\frac{1}{2}}|V| \\ &\leqq \frac{1}{1-2\alpha_1} \cdot 2CE + O(n^{-1}) \\ &= \frac{\varepsilon}{3} + O(n^{-1}).\end{aligned}$$

Finally,

$$n^{\frac{1}{2}}|m(\hat{\alpha}) - m(A)| \leqq \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + O(n^{-1}) \leqq \varepsilon$$

for sufficiently large n, and the proof is complete.

A small amount of Monte Carlo work was performed, with sample size 20 and using the normal, logistic, and Cauchy distributions. The limits for α were $\alpha_0 = 0$ and $\alpha_1 = 0.25$, so that there were six possible choices for $\hat{\alpha}$. It was felt that at this sample size, $s^2(\alpha)$ would be too unstable for larger values of α. For the logistic and Cauchy distributions, the observed variance of the optimal trimmed mean compared favorably with that of the trimmed mean with fixed α which did best for that distribution. In the normal case the variance of the optimal trimmed mean was

10 per cent greater than that of the mean. The observed values of $\hat{\alpha}$ were, on the average, greatest for the Cauchy and smallest for the normal. Thus, at this sample size, the optimal trimmed mean is already able, to some degree, to adjust itself to the shape of the distribution. The observed values of $\hat{\alpha}$ varied wildly; Lemma 3 suggests that the convergence of $\hat{\alpha}$ to A is very slow. But despite the behavior of $\hat{\alpha}$, which should be thought of as an intermediate step in the calculations, the estimator itself performed stably. At larger sample sizes the optimal trimmed mean should perform even better, compared to other estimators.

3. Optimal linear combinations of L-estimators. An L-estimator, or linear combination of order statistics, is defined as follows:

Let $h(t)$ be defined on [0, 1] and such that

$$\int_0^1 h(t)\,dt = 1 \quad \text{and} \quad h(1-t) = h(t). \tag{1}$$

Let $X_{(1)} \leqq \cdots \leqq X_{(n)}$ be the order statistics of a random sample drawn from a symmetric F. Define L as

$$L = \frac{1}{n}\sum_{i=1}^{n} h(i^*)X_{(i)}.$$

Under some regularity conditions, $n^{\frac{1}{2}}L$ is asymptotically normal with asymptotic variance

$$\sigma^2(F) = \int_0^1 U^2(t)\,dt,$$

where

$$U(t) = \int_{\frac{1}{2}}^{t} \frac{h(u)}{f[F^{-1}(u)]}\,du.$$

See Chernoff, Gastwirth and Johns (1967), Huber (1968), and Jaeckel (1971).

Let $h_1(t), \cdots, h_r(t)$ be r functions satisfying (1). We assume they are linearly independent in the sense that $\sum c_k h_k(t) = 0$ a.e. with respect to Lebesgue measure implies $c_k = 0$, $k = 1, \cdots, r$.

Let H be the family of L-estimators defined by all functions of the form

$$h(t) = \sum_{k=1}^{r} c_k h_k(t),$$

subject to $\sum c_k = 1$. Each such h satisfies (1). We remark that h need not be nonnegative.

We shall show that the best $(c_1, \cdots, c_r)$ for a given F may be estimated from the observations by solving a set of r linear equations, and that the procedure which uses as the location estimate the value given by the estimated h is asymptotically as good as using the estimator in H which is best for F.

A simple example would be the family generated by three trimmed means with different trimming proportions. Since this family yields a variety of possible shapes for h, the procedure will be reasonably efficient for a wide class of distributions.

We begin by considering the asymptotic variance of members of H. We assume that F has a density f and is symmetric, and that the general asymptotic variance formula is valid over H. For each k, we have

$$\sigma_k^{\ 2} = \int_0^1 U_k^{\ 2}(t)\,dt,$$

where

$$U_k(t) = \int_{\frac{1}{2}}^{t} \frac{h_k(u)}{f[F^{-1}(u)]} du.$$

Thus, for $h = \sum c_k h_k$,

$$U(t) = \int_{\frac{1}{2}}^{t} \frac{\sum c_k h_k(u)}{f[F^{-1}(u)]} du = \sum c_k U_k(t),$$

and

$$\begin{aligned}\sigma^2 &= \int_0^1 [\sum c_k U_k(t)]^2\,dt \\ &= \sum_k \sum_l c_k c_l \int_0^1 U_k(t) U_l(t)\,dt = \sum_k \sum_l c_k c_l v_{kl},\end{aligned}$$

where $v_{kl} = \int U_k(t) U_l(t) dt$. If we write $V = ((v_{kl}))$ and $c = (c_1, \cdots, c_r)'$, we can express σ^2 as a quadratic form:

$$\sigma^2 = c'Vc.$$

Since, by its definition, $\sigma^2 \geqq 0$, V must be positive indefinite. If V is nonsingular, it is positive definite, and, as we shall see later, there is a unique c^0 which minimizes σ^2 subject to $\sum c_k = 1$. There is thus a unique member of H at which the minimum of σ^2 for F is attained. We show now that this is the case.

Lemma 4. *V is nonsingular.*

Proof. Suppose $Vc = 0$ for some vector c. Then

$$0 = c'Vc = \int [\sum c_k U_k(t)]^2\,dt.$$

Therefore $\sum c_k U_k(t) = 0$ a.e. Since by definition $U_k(t)$ is absolutely continuous and therefore continuous, $\sum c_k U_k(t) = 0$ for all $0 \leqq t \leqq 1$. Hence,

$$\frac{\sum c_k h_k(u)}{f[F^{-1}(u)]} = 0 \quad \text{a.e.}$$

Since

$$\int_{\{f[F^{-1}(u)] = \infty\}} du = \int_{\{f(x) = \infty\}} f(x)\,dx = 0,$$

$f[F^{-1}(u)] < \infty$ a.e. Therefore, $\sum c_k h_k(u) = 0$ a.e. in $0 \leqq u \leqq 1$. Since the h_k are linearly independent, $c = 0$, and V is nonsingular.

We now define an estimator s^2 of σ^2 for a given h. We assume $f(x) > 0$ whenever $h[F(x)] \neq 0$. For $t \geqq \frac{1}{2}$, we write

$$\begin{aligned}U(t) = \int_{\frac{1}{2}}^{t} \frac{h(u)}{f[F^{-1}(u)]} du &= \int_0^{F^{-1}(t)} \frac{h[F(x)]}{f(x)} \cdot f(x)\,dx \\ &= \tfrac{1}{2} \int_{F^{-1}(1-t)}^{F^{-1}(t)} h[F(x)]\,dx.\end{aligned}$$

We shall estimate $F(x)$ by the sample distribution function $F_n(x)$, which we define as:

$$
\begin{aligned}
F_n(x) &= \frac{i}{n} && \text{for} \quad X_{(i)} < x < X_{(i+1)}, \quad i = 1, \cdots, n-1, \\
&= 0 && \text{for} \quad x < X_{(1)}, \\
&= 1 && \text{for} \quad x > X_{(n)}, \\
&= i^* = \frac{i}{n+1} && \text{for} \quad x = X_{(i)}, \qquad i = 1, \cdots, n
\end{aligned}
$$

Let t take on the values i^*, for $i = [(n+1)/2]+1, \cdots, n$. We shall estimate $U(t)$ for these values of t. We estimate $F^{-1}(i^*)$ by $F_n^{-1}(i^*) = X_{(i)}$ and $F^{-1}(1-i^*)$ by $F_n^{-1}(1-i^*) = X_{(n+1-i)}$. Our estimate of $U(i^*)$ is therefore

$$
\begin{aligned}
\hat{U}(i^*) &= \tfrac{1}{2}\int_{X_{(n+1-i)}}^{X_{(i)}} h[F_n(x)]\,dx \\
&= \tfrac{1}{2}\sum_{j=n+1-i}^{i-1} h(j/n)[X_{(j+1)} - X_{(j)}].
\end{aligned}
$$

Since $\sigma^2 = \int_0^1 U^2(t)\,dt = 2\int_{\frac{1}{2}}^1 U^2(t)\,dt$, we estimate σ^2 by

$$
s^2 = \frac{2}{n}\sum_{i=[(n+1)/2]+1}^{n} [\hat{U}(i^*)]^2.
$$

If $h(t) = \sum c_k h_k(t)$, these expressions become

$$
\begin{aligned}
\hat{U}(i^*) &= \tfrac{1}{2}\sum_j \sum_k c_k h_k(j/n)[X_{(j+1)} - X_{(j)}] \\
&= \sum_k c_k \hat{U}_k(i^*),
\end{aligned}
$$

and

$$
\begin{aligned}
s^2 &= (2/n)\sum_i [\sum_k c_k \hat{U}_k(i^*)]^2 \\
&= \sum_k \sum_l c_k c_l \cdot (2/n)\sum_i \hat{U}_k(i^*)\hat{U}_l(i^*) \\
&= \sum_k \sum_l c_k c_l \hat{v}_{kl} = c'\hat{V}c,
\end{aligned}
$$

where $\hat{v}_{kl} = (2/n)\sum \hat{U}_k(i^*)\hat{U}_l(i^*)$ and $\hat{V} = ((\hat{v}_{kl}))$. We thus have a positive indefinite quadratic form, since $s^2 \geqq 0$. We shall show below that $\hat{V}$ is nonsingular with probability approaching one.

Assuming $\hat{V}$ is nonsingular, we shall now derive the unique $\hat{c}$ which minimizes s^2 subject to $\sum c_k = 1$. Let

$$
T = \sum_k \sum_l c_k c_l \hat{v}_{kl} - 2\lambda(\sum_k c_k - 1).
$$

Then, for $k = 1, \cdots, r$,

$$
\frac{\partial T}{\partial c_k} = 2c_k \hat{v}_{kk} + 2\sum_{l \neq k} c_l \hat{v}_{kl} - 2\lambda = 2\sum_l c_l \hat{v}_{kl} - 2\lambda.
$$

Thus, to minimize s^2, we must solve the system of equations

$$
\begin{aligned}
&\sum_l \hat{c}_l \hat{v}_{kl} = \lambda, \qquad k = 1, \cdots, r, \\
&\sum_k \hat{c}_k = 1.
\end{aligned} \tag{2}
$$

We do this by solving

$$\sum b_l \hat{v}_{kl} = 1, \qquad k = 1, \cdots, r \tag{3}$$

and setting $\hat{c}_k = b_k / \sum_l b_l$.

Let $b = (b_1, \cdots, b_r)'$ and $e = (1, \cdots, 1)'$. We write (3) as $\hat{V}b = e$. Then $b = \hat{V}^{-1}e$ is the unique solution to (3). Since $b \neq 0$, $\sum b_l = b'e = b'\hat{V}b \neq 0$; hence we can divide b_k by $\sum b_l$ to obtain $\hat{c}_k$. So we have at least one solution of (2). Since $\hat{V}$ is nonsingular, there can be no solution of (2) with $\lambda = 0$. But any solution of (2) with $\lambda \neq 0$ gives rise to a solution of (3), which must be the same as that given above. It follows that the solution $\hat{c} = (\hat{c}_1, \cdots, \hat{c}_r)$ of (2) is unique.

The same argument applies to the matrix V, which we showed was nonsingular. So there is a unique c^0 which minimizes σ^2 subject to $\sum c_k = 1$.

We now show that under some restrictions on F and the h_k the estimates defined above converge to the quantities they estimate.

LEMMA 5. *Suppose that the h_k are bounded; that for some $t_0 > 0$, $h_k(t) = 0$ for $t < t_0$ and $t > 1-t_0$; and that for some finite set of values of t, the h_k are uniformly continuous on the intervals between those values. Suppose that for some $f_0 > 0$, $f(x) \geqq f_0$ for all x such that $t_0 - \varepsilon_0 \leqq F(x) \leqq 1 - t_0 + \varepsilon_0$ for some $\varepsilon_0 > 0$. Then:*

(i) *For all k, $\hat{U}_k(i^*) - U_k(i^*)$ converges to zero in probability, uniformly in i.*

(ii) *For all k and l, $\hat{v}_{kl} \to_P v_{kl}$.*

(iii) *$P\{\hat{V}$ is nonsingular$\} \to 1$.*

(iv) *$\hat{c} \to_P c^0$.*

PROOF. By (i) we mean that for all $\varepsilon > 0$ and $\delta > 0$, $P\{|\hat{U}_k(i^*) - U_k(i^*)| < \varepsilon$ for all $i = [(n+1)/2]+1, \cdots, n\} \geqq 1-\delta$ for sufficiently large n, for all k. We write $U_k(i^*) = \frac{1}{2}\int h_k[F(x)]dx$ and $\hat{U}_k(i^*) = \frac{1}{2}\int h_k[F_n(x)]dx$. Comparing these integrals, we see that there are three sources of error in $\hat{U}_k(i^*)$: the estimates of the limits of integration $F^{-1}(1-i^*)$ and $F^{-1}(i^*)$ by $X_{(n+1-i)}$ and $X_{(i)}$; the estimates of $F(x)$ by $F_n(x)$ when both fall in the same continuity interval of h_k; and the estimates of the end points of these intervals by the appropriate order statistics.

By the assumptions of the lemma, we may apply Lemma 1, which implies, for $t_0 \leqq i^* \leqq 1-t_0$, $X_{(i)} - F^{-1}(i^*) \to 0$ in probability uniformly in i, in the sense stated above. Since $h_k(t) = 0$ for $t < t_0$ and $t > 1-t_0$ and is bounded everywhere, the first and third kinds of error converge to zero in probability. Since $\sup |F_n(x) - F(x)| \to 0$ in probability and h_k is uniformly continuous on each continuity interval, $\sup |h_k[F_n(x)] - h_k[F(x)]| \to 0$ in probability, where the supremum is over all x such that $F_n(x)$ and $F(x)$ are in the same continuity interval. Therefore, since these intervals have finite length, the second kind of error converges to zero in probability. We conclude that (i) holds.

Due to the assumptions of the lemma, the $U_k(i^*)$ are uniformly bounded. Therefore, by (i), for all k and l, $\hat{U}_k(i^*)\hat{U}_l(i^*) - U_k(i^*)U_l(i^*) \to 0$ in probability uniformly in i. Therefore,

$$\frac{2}{n}\sum_i \hat{U}_k(i^*)\hat{U}_l(i^*) - \frac{2}{n}\sum_i U_k(i^*)U_l(i^*) \to_P 0.$$

Since the first sum is $\hat{v}_{kl}$ and the second converges to $\int U_k(t)U_l(t)dt = v_{kl}$, we have (ii):

$$\hat{v}_{kl} \to_P v_{kl}.$$

Since a determinant is a continuous function of its entries,

$$|\hat{V}| \to_P |V| \neq 0,$$

and we have (iii).

Since the entries in the inverse of a nonsingular matrix are continuous functions of the entries in the matrix, we have

$$\hat{V}^{-1} \to_P V^{-1}.$$

If we write

$$\hat{c} = \frac{b}{e'b} = \frac{\hat{V}^{-1}e}{e'\hat{V}^{-1}e}$$

and

$$c^0 = \frac{V^{-1}e}{e'V^{-1}e},$$

we see that (iv) holds.

We conclude by showing that the estimator

$$L(\hat{c}) = \frac{1}{n}\sum_i \sum_k \hat{c}_k h_k(i^*)X_{(i)}$$

using the estimated coefficients is asymptotically as good as

$$L(c^0) = \frac{1}{n}\sum_i \sum_k c_k{}^0 h_k(i^*)X_{(i)},$$

the best estimator in the family for F.

THEOREM 2. *Under the assumptions of Lemma* 5,

$$n^{\frac{1}{2}}[L(\hat{c}) - L(c^0)] \to_P 0.$$

The two estimators therefore have the same limiting distribution and the same asymptotic variance.

PROOF.

$$\begin{aligned} n^{\frac{1}{2}}&[L(\hat{c}) - L(c^0)] \\ &= n^{-\frac{1}{2}}\{\textstyle\sum_k \hat{c}_k \sum_i h_k(i^*)X_{(i)} - \sum_k c_k{}^0 \sum_i h_k(i^*)X_{(i)}\} \\ &= \textstyle\sum_k (\hat{c}_k - c_k{}^0)\cdot n^{-\frac{1}{2}}\sum_i h_k(i^*)X_{(i)}. \end{aligned}$$

Since for $k = 1, \cdots, r$, $\hat{c}_k - c_k{}^0 \to 0$ in probability and $n^{-\frac{1}{2}}\sum h_k(i^*)X_{(i)}$ is bounded in probability, the theorem follows.

Acknowledgment. I would like to thank Professor E. L. Lehmann for his generous guidance and encouragement throughout the course of this research.

REFERENCES

ANSCOMBE, F. J. (1960). Rejection of outliers. *Technometrics.* **2** 123–147.

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

CHERNOFF, H., GASTWIRTH, J. L., and JOHNS, M. V. (1967). Asymptotic distributions of linear combinations of functions of order statistics with applications to estimation. *Ann. Math. Statist.* **38** 52–72.

HÁJEK, J. (1962). Asymptotically most powerful rank order tests. *Ann. Math. Statist.* **33** 1124–1147.

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

HUBER, P. J. (1968). Robust estimation. *Mathematical Centre Tracts* **27** Mathematisch Centrum Amsterdam 3–25.

JAECKEL, L. A. (1969). Robust estimates of location. Unpublished Ph.D. dissertation, Univ. of California, Berkeley.

JAECKEL, L. A. (1971). Robust estimates of location: Symmetry and asymmetric contamination. *Ann. Math. Statist.* **42** 1020–1034.

STEIN, C. (1956). Efficient nonparametric testing and estimation. *Proc. Third Berkeley Symp. Math. Statist. Prob.* **1** 187–195.

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 Series A* **25** 331–352.

VAN EEDEN, C. (1970). Efficiency-robust estimation of location. *Ann. Math. Statist.* **41** 172–181.