The Annals of Mathematical Statistics
1971, Vol. 42, No. 3, 1020-1034

ROBUST ESTIMATES OF LOCATION: SYMMETRY AND ASYMMETRIC CONTAMINATION[1]

BY LOUIS A. JAECKEL

University of California, Berkeley[2]

1. Introduction and summary.

The problem of finding location estimators which are "robust" against deviations from normality has received increasing attention in the last several years. See, for example, Tukey (1960), Huber (1968), and papers cited therein. In the theoretical work done on the estimation of a location parameter, the underlying distribution is usually assumed to be symmetric, and the estimand is taken to be the center of symmetry, a natural quantity to estimate in this situation. Since the finite sample size properties of many proposed estimators are difficult to study analytically, most research has focussed on their more easily ascertainable asymptotic properties, which, it is hoped, will provide useful approximations to the finite sample size case. Most of the estimators commonly studied are, under suitable regularity conditions, asymptotically normal about the center of symmetry, with asymptotic variance depending on the underlying distribution. We thus have a simple criterion, the asymptotic variance, for comparing the performance of different estimators for a given underlying distribution, and of a given estimator for different underlying distributions. Huber (1964) has formulated and solved some minimax problems, in which the estimators are judged by their asymptotic variance.

In Section 2 we define and state the asymptotic variances which have been found for the three most commonly studied types of location estimators. In Section 3 we demonstrate some relationships among the three types of estimators, and in Section 4 we show that Huber's minimax result applies to all three types. Then, in Section 5 we consider an aspect of the more general estimation problem in which the distributions are not assumed symmetric. A model of asymmetric contamination of a symmetric distribution is formulated, in which the amount of asymmetry tends to zero as the sample size increases. The estimators here are thought of as estimating the center of the symmetric component of the distribution. The maximum likelihood type estimators are shown to be asymptotically normal under this model, but with a bias that tends to zero as the sample size increases. The estimators may be judged by their asymptotic mean squared error, a concept which is made meaningful by the model. We conclude in Section 6 with a minimax result analogous to Huber's, for which we allow both symmetric and asymmetric contamination of a

Received May 25, 1970; revised December 10, 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, New Jersey.

given distribution and judge the estimators by their asymptotic mean squared error.

2. Definitions. Let $X_1, X_2, \cdots, X_n$ be independent identically distributed random variables with distribution $F(x-\theta)$, where F is symmetric, that is, $F(x)+F(-x) = 1$. We assume F has a density f. We want to estimate the unknown parameter θ and judge the quality of an estimator by its asymptotic variance. Each of the three types of estimators defined below is, under general regularity conditions, asymptotically normal with mean θ and asymptotic variance as given below. For more specific statements of regularity conditions under which the variance formulas are valid, see the works referred to below. Since we shall be dealing with translation-invariant statistics, we shall henceforth assume that $\theta = 0$. Note that the estimators defined here are actually sequences of estimators indexed by n, the sample size, but we shall simply think of them as single estimators, without an index, whenever no confusion will arise thereby. The notation used below is essentially that of Huber (1968).

(i) Maximum likelihood type estimators, which, following Huber, we shall call M-estimators.

Let $\psi(x)$ be such that $\psi(-x) = -\psi(x)$. Define M as a solution of the equation

$$\textstyle\sum_{i=1}^{n} \psi(X_i - M) = 0.$$

If $\psi(x)$ is monotonic, M is essentially uniquely determined. The asymptotic variance of $n^{\frac{1}{2}}M$ is

$$\sigma_M{}^2(F) = \frac{\int \psi^2(x) f(x)\,dx}{(\int \psi'(x) f(x)\,dx)^2}\,.$$

See Huber (1964).

(ii) Linear combinations of order statistics, which we shall call L-estimators.

Let $X_{(1)} \leqq \cdots \leqq X_{(n)}$ be the order statistics derived from the sample. Let $h(t)$ be such that $h(t) = h(1-t)$ and $\int_0^1 h(t)\,dt = 1$. We shall adopt the following notational convention: Let $i^* = i/(n+1)$. Define L as

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

The asymptotic variance of $n^{\frac{1}{2}}L$ is

$$\sigma_L{}^2(F) = \textstyle\int_0^1 U^2(t)\,dt,$$

where

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

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

(iii) Estimators derived from rank tests, which we shall call R-estimators.

Let $J(t)$ be such that $J(1-t) = -J(t)$. For any given r, form the $2n$ numbers $X_1 - r, \cdots, X_n - r; -X_1 + r, \cdots, -X_n + r$. Order these $2n$ numbers and let $V_i = 1$ if the ith smallest is of the form $X_j - r$, and $V_i = 0$ otherwise. Form the sum

$$W(r) = \sum_{i=1}^{2n} J\left(\frac{i}{2n+1}\right) V_i.$$

Define R as a solution of the equation $W(R) = 0$. If J is monotonic, R is essentially uniquely determined. The asymptotic variance of $n^{\frac{1}{2}}R$ is

$$\sigma_R{}^2(F) = \frac{\int J^2(t)\, dt}{\left(\int \frac{d}{dx}\{J[F(x)]\} f(x)\, dx\right)^2}.$$

See Hodges and Lehmann (1963). Although the estimator here is derived from a two-sample rank test, it is asymptotically equivalent to the estimator originally derived by Hodges and Lehmann from a one-sample rank test.

3. Relationships among the estimators. For any given symmetric F satisfying appropriate regularity conditions, there is a three-way correspondence among the three types of estimators which preserves the asymptotic variance under F. For a given odd ψ defining an M-estimator, we define functions defining an L-estimator and an R-estimator as follows.

Let $t = F(x)$, so that $x = F^{-1}(t)$. We assume for simplicity that F is strictly increasing. Let $h(t) = \psi'[F^{-1}(t)] = \psi'(x)$, where $\psi'(x) = (d/dx)\psi(x)$. Let $J(t) = \psi[F^{-1}(t)] = \psi(x)$.

Since ψ may be multiplied by a constant without affecting the estimator, we may assume that $\int \psi'(x) f(x)\, dx = 1$. This condition implies that $\int h(t)\, dt = \int h[F(x)] f(x)\, dx = \int \psi'(x) f(x)\, dx = 1$. Since $\psi(-x) = -\psi(x)$, the symmetry conditions on h and J are clearly satisfied.

THEOREM 1. *Assuming the asymptotic variance formulas hold, we have, for the estimators defined above,* $\sigma_M{}^2(F) = \sigma_L{}^2(F) = \sigma_R{}^2(F)$.

PROOF. $\sigma_M{}^2(F) = \int \psi^2(x) f(x)\, dx$, since the denominator in the formula is 1. For the L-estimator we have

$$U(t) = \int_{\frac{1}{2}}^{t} \frac{h(u)}{f[F^{-1}(u)]}\, du = \int_0^{F^{-1}(t)} \frac{h[F(x)]}{f(x)} f(x)\, dx$$
$$= \int_0^{F^{-1}(t)} \psi'(x)\, dx = \psi[F^{-1}(t)],$$

since $\psi(0) = 0$. Therefore,

$$\sigma_L{}^2(F) = \int U^2(t)\, dt = \int \psi^2[F^{-1}(t)]\, dt = \int \psi^2(x) f(x)\, dx = \sigma_M{}^2(F).$$

Since the denominator for the variance of the R-estimator is the square of

$$\int \frac{d}{dx}\{J[F(x)]\} f(x)\, dx = \int \frac{d}{dx} \psi(x) f(x)\, dx = 1,$$

we have

$$\sigma_R{}^2(F) = \int J^2(t)\,dt = \int \psi^2(x)f(x)\,dx = \sigma_M{}^2(F).$$

For the special case of Huber's estimator with parameter k, Huber (1964),

$$\begin{aligned} \psi(x) &= x & &\text{for} \quad |x| < k \\ &= k \operatorname{sign}(x) & &\text{for} \quad |x| \geqq k, \end{aligned}$$

we have

$$\begin{aligned} h(t) &= \text{constant for } F(-k) < t < F(k), \\ &= 0 \qquad\qquad \text{otherwise.} \end{aligned}$$

This is the trimmed mean with trimming proportion $\alpha = F(-k)$. The equality of variances in this case was recognized by Bickel (1965). The corresponding R-estimator is defined by

$$\begin{aligned} J(t) &= F^{-1}(t) & &\text{for} \quad F(-k) < t < F(k), \\ &= k & &\text{for} \quad t \geqq F(k), \\ &= -k & &\text{for} \quad t \leqq F(-k). \end{aligned}$$

If F is the contaminated normal distribution considered by Huber, this $J(t)$ defines a sort of truncated Van der Waerden test. See Gastwirth (1966) page 946.

It follows from the theorem that if any one of the three estimators is asymptotically optimal for F among translation-invariant estimators, then all three are. Correspondences of this type have been given in the asymptotically optimal case by Gastwirth (1966) and Huber (1968).

We shall show that under some more restrictive conditions the relationship between the M-estimator and the corresponding L-estimator is even closer than that indicated above. But first we shall define some terms and prove a lemma.

DEFINITION.

(i) The sequence of random variables $\{Z_n\}$ is bounded in probability if

$$\forall\, \delta > 0\ \exists\, B, N \text{ such that } \forall\, n \geqq N\colon \quad P\{|Z_n| \leqq B\} \geqq 1-\delta.$$

(ii) The sequence of random variables $\{Z_{nk}\}$ is bounded in probability uniformly in k if

$$\forall\, \delta > 0\ \exists\, B, N \text{ such that } \forall\, n \geqq N\colon \quad P\{|Z_{nk}| \leqq B, \forall\, k\} \geqq 1-\delta.$$

(iii) The sequence $\{Z_n\}(\{Z_{nk}\})$ is $O(n^a)$ in probability (uniformly in k) if the sequence $\{Z_n/n^a\}(\{Z_{nk}/n^a\})$ is bounded in probability (uniformly in k).

DEFINITION. The sample distribution function is

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

LEMMA 1. *Suppose F has a density $f(x)$, 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 $n \geqq N$:*

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

PROOF. The statistic $K_n = n^{\frac{1}{2}} \sup |F_n(x) - F(x)|$ has a limiting distribution which was found by Kolmogorov. See Hájek and Šidák (1967). For a given δ, we can therefore choose a D so that for sufficiently large n, $P\{K_n \leqq D\} \geqq 1-\delta$.

Suppose $K_n \leqq D$; that is, $|F_n(x) - F(x)| \leqq n^{-\frac{1}{2}} D$ for all x. Then $|i^* - F(X_{(i)})| \leqq n^{-\frac{1}{2}} D$ for $i = 1, \cdots, n$. Since $([\alpha_0 n]+1)^* \to \alpha_0$ and $(n-[\alpha_0 n])^* \to 1-\alpha_0$, and $n^{-\frac{1}{2}} D < \frac{1}{2}\varepsilon_0$ for sufficiently large n, we have, for large n, $\alpha_0 - \varepsilon_0 < F(X_{(i)}) < 1-(\alpha_0-\varepsilon_0)$ for $i = [\alpha_0 n]+1, \cdots, n-[\alpha_0 n]$. Since, for $\alpha_0 - \varepsilon_0 < t < 1-(\alpha_0-\varepsilon_0)$ we have

$$\frac{d}{dt} F^{-1}(t) = \frac{1}{f[F^{-1}(t)]} \leqq \frac{1}{f_0},$$

we can apply the mean value theorem to $F^{-1}(t)$, obtaining

$$|F^{-1}(i^*) - F^{-1}[F(X_{(i)})]| \leqq \frac{1}{f_0} |i^* - F(X_{(i)})| \leqq \frac{D}{n^{\frac{1}{2}} f_0}$$

for $i = [\alpha_0 n]+1, \cdots, n-[\alpha_0 n]$. The lemma follows.

The effect of this lemma is to convert a uniform vertical bound into a uniform horizontal bound. The lemma clearly applies to any unimodal density.

THEOREM 2. *Suppose F has a bounded density f and satisfies the conditions of Lemma 1 for some α_0. Suppose $\psi(x)$ and $h(t)$ are related by $h(t) = \psi'[F^{-1}(t)] = \psi'(x)$, the asymptotic variance formulas apply, and $h(t) = \psi'(x) = 0$ for $t < \alpha_0$ and $t > 1-\alpha_0$. Finally, suppose that ψ is continuous, and that, at all but a finite number of points ψ' is defined and bounded and has a bounded derivative. Then*

$$n^{\frac{1}{2}}(M-L) \to_P 0.$$

In fact, $M-L$ is $O(n^{-1})$ in probability.

PROOF. If we write $X_{(i)} = F^{-1}(i^*)+e_i$, we have, by Lemma 1, e_i is $O(n^{-\frac{1}{2}})$ in probability uniformly in i such that $\alpha_0 \leqq i^* \leqq 1-\alpha_0$. Since $F^{-1}(1-i^*) = -F^{-1}(i^*)$ and $h(1-i^*) = h(i^*)$, $\sum h(i^*)F^{-1}(i^*) = 0$. L is therefore defined by

$$\begin{aligned} nL &= \textstyle\sum_i h(i^*)X_{(i)} = \sum_i h(i^*)[F^{-1}(i^*)+e_i] \\ &= \textstyle\sum_i h(i^*)e_i \quad = \sum_i \psi'[F^{-1}(i^*)]e_i. \end{aligned}$$

M is defined by $\sum \psi(X_{(i)}-M) = 0$. We can expand each term of this sum as follows:

$$\begin{aligned} (1) \qquad \psi(X_{(i)}-M) &= \psi[F^{-1}(i^*)+(e_i-M)] \\ &= \psi[F^{-1}(i^*)]+(e_i-M)\psi'[F^{-1}(i^*)]+r_i. \end{aligned}$$

Let $r = \sum r_i$. These remainder terms will be dealt with later. Summing over i in (1) we get

$$\begin{aligned} 0 = \textstyle\sum_i \psi(X_{(i)}-M) = \sum_i \psi[F^{-1}(i^*)] \\ +\textstyle\sum_i e_i\psi'[F^{-1}(i^*)]-M\sum_i \psi'[F^{-1}(i^*)]+r. \end{aligned}$$

Since $\psi(-x) = -\psi(x)$, $\sum\psi[F^{-1}(i^*)] = 0$. Since $\int h(t)\,dt = 1$, $\sum\psi'[F^{-1}(i^*)] = \sum h(i^*) = n+O(1)$; the excess here may be absorbed into the remainder term r. Therefore,

$$0 = 0+nL-nM+r,$$

or $M-L = r/n$.

We must now examine the remainder. First suppose $\alpha_0 \leqq i^* \leqq 1-\alpha_0$. If no "bad point" of ψ' lies between $X_{(i)}-M$ and $F^{-1}(i^*)$, then by (1), $|r_i| \leqq \frac{1}{2}(e_i-M)^2 \cdot \sup \psi''$, which is $O(n^{-1})$ in probability uniformly in i. The contribution of these terms to r is therefore $O(1)$ in probability. If a "bad point" of ψ' lies between $X_{(i)}-M$ and $F^{-1}(i^*)$, then r_i is $O(n^{-\frac{1}{2}})$ in probability uniformly in i, since ψ is continuous and ψ' is bounded. Since f is bounded, the number of such i is $O(n^{\frac{1}{2}})$ in probability, so their contribution to r is $O(1)$ in probability. Now suppose $i^* < \alpha_0$. Since $\psi(x)$ is constant for $x < F^{-1}(\alpha_0)$, (1) implies $r_i = 0$ if $X_{(i)}-M < F^{-1}(\alpha_0)$. If $X_{(i)}-M \geqq F^{-1}(\alpha_0)$, we have, letting j be the smallest i such that $i^* \geqq \alpha_0$, $0 \leqq X_{(i)}-M-F^{-1}(\alpha_0) \leqq X_{(j)}-M-F^{-1}(\alpha_0) = F^{-1}(j^*)-F^{-1}(\alpha_0)+e_j-M$, which is $O(n^{-\frac{1}{2}})$ in probability. Thus the intrusion of $X_{(i)}-M$ into the non-constant part of the domain of ψ is $O(n^{-\frac{1}{2}})$ in probability uniformly in i Hence, by (1), r_i is $O(n^{-\frac{1}{2}})$ in probability uniformly in i. Since the number of such i is $O(n^{\frac{1}{2}})$ in probability, their contribution to r is $O(1)$ in probability. A similar argument holds for $i^* > 1-\alpha_0$. Therefore, r is bounded in probability, and the theorem follows.

The theorem sheds light on a point which has caused some confusion. Huber's estimator, which we defined earlier, appears at first glance to resemble a Winsorized mean rather than a trimmed mean. See Huber (1964). But we now see that Huber's estimator is in fact very closely related to the trimmed mean; indeed, an iterative procedure to compute Huber's estimator should use the more easily computed trimmed mean as a first approximation. Thus we can assert that symmetric trimming of outliers is in no sense "throwing away" the information contained in those observations. It would seem that trimming does what Winsorizing was intended to accomplish: it takes into account the existence of the outlying observations in each tail, but not their values.

We conjecture that a result analogous to Theorem 2 holds with respect to the R-estimators.

4. Huber's minimax result. The asymptotic minimax problem stated below was solved by Huber (1964) in terms of M-estimators. In view of our correspondence among the three types of estimators, we may ask whether the corresponding L-estimators and R-estimators are also solutions to this problem. We shall answer this question in the affirmative. We restrict ourselves to symmetric distributions.

Let C be the set of distributions $F = (1-\varepsilon)G+\varepsilon H$, where ε is fixed, G is a fixed symmetric, strongly unimodal distribution, and H is a variable symmetric distribution. We want an estimator which minimizes the supremum over C of the asymptotic variance. Huber solved this problem by showing the existence of a saddle point: there is an F_0 in C and a ψ_0 defining an M-estimator such that the estimator is asymptotically optimal for F_0, and for all F in C, $\sigma_M^2(F) \leqq \sigma_M^2(F_0)$. F_0 is defined by

$$\begin{aligned} f_0(x) &= (1-\varepsilon)g(-x_0)\exp(k(x+x_0)) &&\text{for } x \leqq -x_0, \\ &= (1-\varepsilon)g(x) &&\text{for } -x_0 < x < x_0, \\ &= (1-\varepsilon)g(x_0)\exp(-k(x-x_0)) &&\text{for } x \geqq x_0, \end{aligned}$$

for some x_0 and k depending on G and ε. $\psi_0(x) = -f_0'(x)/f_0(x)$ is monotonic, and for $x < -x_0$ and $x > x_0$ is constant. See Huber, page 81.

The L-estimator and R-estimator corresponding to this M-estimator are defined by

$$h_0(t) = \psi_0'[F_0^{-1}(t)] = \psi_0'(x)$$

and

$$J_0(t) = \psi_0[F_0^{-1}(t)] = \psi_0(x).$$

It follows that $h_0(t) \geqq 0$ and $J_0(t)$ is monotonic, and for $t < F_0(-x_0)$ and $t > F_0(x_0)$, $h_0(t) = 0$ and $J_0(t)$ is constant.

THEOREM 3. *Assuming the asymptotic variance formulas apply, the estimators defined above are also solutions to Huber's minimax problem. We assume all distributions under consideration have densities.*

By Theorem 1, the estimators defined by h_0 and J_0 are asymptotically optimal for F_0. To show that these estimators have the required minimax property, we must show that for all F in C, $\sigma_L^2(F) \leqq \sigma_L^2(F_0)$ and $\sigma_R^2(F) \leqq \sigma_R^2(F_0)$. We need the following inequality:

LEMMA 2. *For all F in C, and for all t such that* $\frac{1}{2} \leqq t \leqq F_0(x_0)$:

$$f[F^{-1}(t)] \geqq f_0[F_0^{-1}(t)].$$

PROOF. Choose F in C. Then $F(0) = F_0(0) = \frac{1}{2}$.
For $0 \leqq x \leqq x_0$,

$$f_0(x) = (1-\varepsilon)g(x) \leqq f(x).$$

Therefore, for $0 \leqq x \leqq x_0$, $F_0(x) \leqq F(x)$. Letting $t = F(x)$, we have, for $\frac{1}{2} \leqq t \leqq F(x_0)$,

$$F_0[F^{-1}(t)] \leqq F[F^{-1}(t)] = t.$$

Therefore,

$$F^{-1}(t) \leqq F_0^{-1}(t).$$

Since $F_0(x_0) \leqq F(x_0)$, this inequality holds for $\frac{1}{2} \leqq t \leqq F_0(x_0)$, and we have, for these t,

$$0 \leqq F^{-1}(t) \leqq F_0^{-1}(t) \leqq x_0.$$

Since, for $0 \leqq x \leqq x_0$, $f(x) \geqq f_0(x)$ and $f_0(x)$ is monotone decreasing,

$$f[F^{-1}(t)] \geqq f_0[F^{-1}(t)] \geqq f_0[F_0^{-1}(t)],$$

and the lemma is proved.

PROOF OF THE THEOREM. We consider the L-estimator first. Since $h_0(u) \geqq 0$, Lemma 2 implies, for $\frac{1}{2} \leqq t \leqq F_0(x_0)$,

$$\frac{h_0(u)}{f[F^{-1}(u)]} \leqq \frac{h_0(u)}{f_0[F_0^{-1}(u)]},$$

and therefore

$$U(t) = \int_{\frac{1}{2}}^{t} \frac{h_0(u)}{f[F^{-1}(u)]}\, du \leqq U_0(t) = \int_{\frac{1}{2}}^{t} \frac{h_0(u)}{f_0[F_0^{-1}(u)]}\, du$$

for $\frac{1}{2} \leqq t \leqq F_0(x_0)$. Since $h_0(u) = 0$ for $u > F_0(x_0)$, $U(t) \leqq U_0(t)$ for all $t \geqq \frac{1}{2}$. By symmetry, $U(1-t) = -U(t)$ and $U_0(1-t) = -U_0(t)$, so $U^2(t) \leqq U_0^2(t)$ for all t, and therefore $\sigma_L^2(F) \leqq \sigma_L^2(F_0)$.

We now consider the R-estimator. Since the numerator in the asymptotic variance formula does not depend on F, we need consider only the integral in the denominator. We write

$$J_0'(t) = \frac{d}{dt} J_0(t),$$

so that

$$\frac{d}{dx} J_0[F(x)] = J_0'[F(x)]f(x).$$

Substituting $t = F(x)$ in the integral, we get

$$I = \int \frac{d}{dx} J_0[F(x)] \cdot f(x)\, dx = \int J_0'[F(x)]f(x) \cdot f(x)\, dx$$
$$= \int_0^1 J_0'(t) \cdot f[F^{-1}(t)]\, dt.$$

Similarly, with $t = F_0(x)$,

$$I_0 = \int \frac{d}{dx} J_0[F_0(x)] \cdot f_0(x)\, dx = \int J_0'[F_0(x)] f_0(x) \cdot f_0(x)\, dx$$

$$= \int_0^1 J_0'(t) \cdot f_0[F_0^{-1}(t)]\, dt.$$

Since the integrands are symmetric with respect to $t = \frac{1}{2}$, and since for $t > F_0(x_0)$, $J_0(t)$ is constant, so that $J_0'(t) = 0$, the domain of integration we must consider is $\frac{1}{2} \leqq t \leqq F_0(x_0)$. Since $J_0(t)$ is monotonic, $J_0'(t) \geqq 0$. Therefore, by Lemma 2, $0 \leqq I_0 \leqq I$, so that $I_0^2 \leqq I^2$, and $\sigma_R^2(F) \leqq \sigma_R^2(F_0)$.

5. A model for asymmetric contamination. When we restrict our attention to symmetric distributions, we have a natural estimand, the center of symmetry, and since the estimators we have considered are symmetrically distributed about the center of symmetry, we have a natural criterion for judging the estimators, their asymptotic variance. But these conditions do not exist in the case of asymmetric distributions. In order to simplify the problem so that it can be related to the framework to which we are already accustomed, we shall consider distributions $F = (1-\varepsilon)G + \varepsilon H$ consisting of a symmetric part, G, with a small amount of asymmetric contamination, H, added to it, and we shall consider the problem of estimating the center of the symmetric component of the distribution. See Huber (1964), page 82. Since the estimators will in general be biased now, a natural criterion for judging them is their mean squared error, a measure which takes into account both their inherent variability and their distance from the estimand.

As in the symmetric case, our goal is to obtain asymptotic results which hopefully will furnish useful approximations for finite sample sizes. Therefore, we wish to consider estimators in terms of what we shall call their asymptotic mean squared error. But for a fixed amount of asymmetric contamination, the estimates generally do not converge to the center of G as n increases. See Huber (1964), page 83. Thus, in order to give meaning to the concept of asymptotic mean squared error, both the bias and the variability of the estimator must be made to approach zero at the same rate. We are thus led to the following model. Let the sequence of distributions $\{F_n\}$ be defined by

$$F_n = (1 - cn^{-\frac{1}{2}})G + cn^{-\frac{1}{2}}H,$$

where G and H are fixed distributions, G is symmetric, n is the sample size, and c is a constant. So the underlying distribution is $F_n(x - \theta)$ and the problem is to estimate the parameter θ. As before, we shall assume that $\theta = 0$. (This F_n is not to be confused with the F_n defined in Section 3.)

Let N be an estimator of θ under the model. We shall see that under certain conditions, $n^{\frac{1}{2}}N$ is asymptotically normal (b, σ^2) for some constants b and σ^2. When this is the case, we shall define b to be the *asymptotic bias* and $\sigma^2 + b^2$ to be the *asymptotic mean squared error*, or AMSE, of N under the model. For finite

n, N is thus approximately normal $(n^{-\frac{1}{2}}b, n^{-1}\sigma^2)$, and $(\sigma^2+b^2)/n$ will serve as an approximate mean squared error.

The statistical interpretation of the model is as follows. The amount of asymmetric contamination is large enough to affect the performance of the estimator, but is too small to be measured accurately at the given sample size. We may think of the model as taking a given contaminated distribution and embedding it in a sequence of distributions, each of which has an amount of contamination which will affect the behavior of the estimator in approximately the same way.

We shall now prove the consistency and asymptotic normality of arbitrary M-estimators under the model. The method of proof is to compare the performance of a given estimator under $\{F_n\}$ with its performance under G.

For each n we simultaneously construct a random sample of size n from G and from F_n. Let $X_1, \cdots, X_n$ be i.i.d. random variables with distribution G. Let B_n be a binomial variable corresponding to n trials with success probability $n^{-\frac{1}{2}}c$. Let $Y_1, \cdots, Y_{B_n}$ be i.i.d. random variables with distribution H. For $i = B_n+1, \cdots, n$ let $Y_i = X_i$. Then $Y_1, \cdots, Y_n$ is essentially a random sample of size n from F_n. The ordering of the Y_i should be randomized, but this will not be necessary.

Let ψ be an odd, monotonic, bounded function defining an M-estimator. We define the statistics M and N as the solutions to the following equations:

$$\sum\psi(X_i-M) = 0$$

and

$$\sum\psi(Y_i-N) = 0.$$

M and N are thus the location estimates defined by ψ under G and $\{F_n\}$ respectively.

LEMMA 3. *Suppose G has a bounded density and the asymptotic variance formula is valid for ψ under G. Suppose that ψ is continuous, and that ψ' and ψ'' are continuous and bounded at all but a finite number of points. Suppose $E_G\psi' > 0$ and $E_G\psi'(x-\varepsilon)$ is continuous in ε. Let $A = \int \psi(x)H(dx)$ and let $b = cA/E_G\psi'$. Then*

$$n^{\frac{1}{2}}(N-M) \rightarrow_P b.$$

PROOF. We show first that N is $O(n^{-\frac{1}{2}})$ in probability; that is, for all $q > 0$, there exists δ such that $P\{|N| \leqq n^{-\frac{1}{2}}\delta\} \geqq 1-q$ for sufficiently large n. Since ψ is monotonic, $|N| \leqq n^{-\frac{1}{2}}\delta$ is equivalent to $\sum\psi(Y_i-n^{-\frac{1}{2}}\delta) \leqq 0 \leqq \sum\psi(Y_i+n^{-\frac{1}{2}}\delta)$. We consider the first inequality.

$$n^{-\frac{1}{2}}\sum_{i=1}^{n}\psi(Y_i-n^{-\frac{1}{2}}\delta) = n^{-\frac{1}{2}}\sum_{i=1}^{B_n}\psi(Y_i-n^{-\frac{1}{2}}\delta)$$
$$-n^{-\frac{1}{2}}\sum_{i=1}^{B_n}\psi(X_i-n^{-\frac{1}{2}}\delta)+n^{-\frac{1}{2}}\sum_{i=1}^{n}\psi(X_i-n^{-\frac{1}{2}}\delta).$$

Since ψ is bounded and B_n is $O(n^{\frac{1}{2}})$ in probability, the first two sums on the right are bounded in probability for each fixed δ, with bounds not depending on δ. The third sum is a sum of i.i.d. bounded random variables, and may therefore be expressed as

$$n^{\frac{1}{2}}E_G\psi(x-n^{-\frac{1}{2}}\delta)+S_n(\delta),$$

where $S_n(\delta)$ is bounded in probability for each fixed δ, with bounds not depending on δ. This follows from the Tchebycheff inequality, since the boundedness of ψ imposes an upper bound on the variance of $S_n(\delta)$.

Because of our assumptions on ψ, we may differentiate $E_G\psi(x-\varepsilon)$:

$$\frac{d}{d\varepsilon}\int \psi(x-\varepsilon)g(x)\,dx = -E_G\psi'(x-\varepsilon).$$

Therefore, since $E_G\psi = 0$,

$$E_G\psi(x-n^{-\frac{1}{2}}\delta) = -n^{-\frac{1}{2}}\delta E_G\psi'(x-\varepsilon_n)$$

for some $0 \leqq \varepsilon_n \leqq n^{-\frac{1}{2}}\delta$. Hence, for a given q, there is a D such that for sufficiently large n,

$$P\{|n^{-\frac{1}{2}}\textstyle\sum \psi(Y_i-n^{-\frac{1}{2}}\delta)+\delta E_G\psi'(x-\varepsilon_n)| < D\} \geqq 1-q/2,$$

for each fixed δ. Now let $\delta = 2D/E_G\psi'$. Since $E_G\psi'(x-\varepsilon_n) \to E_G\psi'$, we have

$$P\{n^{-\frac{1}{2}}\textstyle\sum \psi(Y_i-n^{-\frac{1}{2}}\delta) < 0\} \geqq 1-q/2$$

for sufficiently large n. A similar result holds for $\sum \psi(Y_i+n^{-\frac{1}{2}}\delta)$, so N is $O(n^{-\frac{1}{2}})$ in probability.

Since $X_i-M = (X_i-N)+(N-M)$, we can write

$$\psi(X_i-M) = \psi(X_i-N)+(N-M)\psi'(X_i-N)+R_i, \tag{2}$$

where R_i is $O(N-M)^2$ uniformly in i if no bad points of ψ' lie between X_i-M and X_i-N, and R_i is $O(N-M)$ uniformly in i otherwise. Since M and N are $O(n^{-\frac{1}{2}})$ in probability and g is bounded, the number of R_i of the latter kind is $O(n^{\frac{1}{2}})$ in probability. It follows that $R = \sum R_i$ is bounded in probability.

By the definition of N,

$$\begin{aligned}\textstyle\sum_{i=1}^{n} \psi(X_i-N) &= \textstyle\sum_{i=1}^{B_n} \psi(X_i-N)+\sum_{i=B_n+1}^{n} \psi(Y_i-N)\\ &= \textstyle\sum_{i=1}^{B_n} \psi(X_i-N)-\sum_{i=1}^{B_n} \psi(Y_i-N).\end{aligned}$$

Now

$$\frac{1}{B_n}\sum_{i=1}^{B_n} \psi(X_i-N) \to_P \int \psi(x)G(dx) = 0$$

and

$$\frac{1}{B_n}\sum_{i=1}^{B_n} \psi(Y_i-N) \to_P \int \psi(x)H(dx) = A$$

and

$$n^{-\frac{1}{2}}B_n \to_P c,$$

so

$$n^{-\frac{1}{2}}\sum_{i=1}^{n} \psi(X_i-N) = n^{-\frac{1}{2}}B_n\cdot\frac{1}{B_n}\sum_{i=1}^{n} \psi(X_i-N) \to_P -cA.$$

Summing over i in (2), we get

$$0 = \sum \psi(X_i - M) = \sum \psi(X_i - N) + (N-M)\sum \psi'(X_i - N) + R.$$

Since

$$1/n \sum \psi'(X_i - N) \to_P E_G\psi' > 0,$$

we have

$$n^{\frac{1}{2}}(N-M) = \frac{-n^{-\frac{1}{2}}\sum \psi(X_i - N) - n^{-\frac{1}{2}}R}{n^{-1}\sum \psi'(X_i - N)} \to_P \frac{cA}{E_G\psi'} = b.$$

As a consequence of Lemma 3, we have

THEOREM 4. *Under the conditions stated in Lemma* 3, $n^{\frac{1}{2}}N$ *is asymptotically normal with mean* b *and variance* $\sigma_M{}^2(G)$. *The asymptotic bias of* N *is therefore* b, *an*. *the* AMSE *of* N *is*

$$\sigma_M{}^2(G) + b^2 = \frac{\int \psi^2(x)g(x)\,dx + c^2A^2}{\{\int \psi'(x)g(x)\,dx\}^2}.$$

PROOF. Since $n^{\frac{1}{2}}M$ is asymptotically normal $[0, \sigma_M{}^2(G)]$ and $(n^{\frac{1}{2}}N - b) - n^{\frac{1}{2}}M \to 0$ in probability, $n^{\frac{1}{2}}N - b$ is asymptotically normal $[0, \sigma_M{}^2(G)]$. The theorem follows.

We may draw some elementary conclusions from Theorem 4. Suppose g is unimodal, and we are considering the family of estimators with parameter k defined by

$$\begin{aligned}\psi(x) &= x &&\text{for } |x| < k\\ &= k\,\mathrm{sign}\,(x) &&\text{for } |x| \geqq k.\end{aligned}$$

Suppose the amount of contamination is $n^{-\frac{1}{2}}c$, and H puts all of its mass to the right of k, so that the bias is as great as possible. The limiting case of this family of estimators as $k \to 0$ is the sample median. Since $\sup g(x) = g(0)$,

$$1/k \int \psi'(x)g(x)\,dx = 1/k \int_{-k}^{k} g(x)\,dx \leqq 2g(0),$$

and, assuming $g(x)$ is continuous at $x = 0$,

$$1/k \int \psi'(x)g(x)\,dx \to 2g(0)$$

as $k \to 0$. So we have

$$b = \frac{ck}{E_G\psi'} = \frac{c}{k^{-1}E_G\psi'} \geqq \frac{c}{2g(0)}$$

for all k, and $b \to c/2g(0)$ as $k \to 0$. Thus the asymptotic bias for this family of estimators is minimized by the sample median. If we consider the AMSE as a function of k, we see that $b^2(k) = \mathrm{AMSE}(k) - \sigma^2(k)$ is an increasing function of k, since

$$1/k \int_{-k}^{k} g(x)\,dx$$

decreases as k increases. Therefore, if $\sigma^2(k)$, the asymptotic variance under G, is minimized at $k = k_0$, AMSE (k) must attain its minimum at some $k \leqq k_0$. Thus, some value of k smaller than k_0 will perform better when asymmetric contamination is present, but at the cost of some increase in variance if there is no asymmetric contamination.

In general, when we allow some asymmetric contamination in a model containing symmetric distributions and we judge estimators by their AMSE, we can expect some trade-off between the variance due to the symmetric distributions and the bias due to the contamination. The theorem in the next section illustrates this point.

We state without proof the asymptotic bias and AMSE for the α-trimmed mean, assuming H puts all of its mass to the right of $x_{1-\alpha}$, where $x_\alpha = G^{-1}(\alpha)$ and $x_{1-\alpha} = G^{-1}(1-\alpha)$.

$$b(\alpha) = \frac{cx_{1-\alpha}}{1-2\alpha}$$

and

$$\begin{aligned}\text{AMSE}\,(\alpha) &= \sigma^2(\alpha)+b^2(\alpha)\\ &= \frac{1}{(1-2\alpha)^2}\,\{\textstyle\int_{x_\alpha}^{x_{1-\alpha}} x^2 g(x)\,dx+(2\alpha+c^2){x_\alpha}^2\}.\end{aligned}$$

The proof is similar to that of Theorem 4.

6. A minimax result. We shall now obtain an asymptotic minimax result analogous to that derived by Huber (1964), page 80.

For a given symmetric, strongly unimodal distribution G satisfying the conditions of Huber's theorem and a given $0 < \varepsilon_0 < 1$, we consider the class C' of sequences of distributions $\{F_n\}$ of the form

$$F_n = (1-n^{-\frac{1}{2}}c)[(1-\varepsilon)G+\varepsilon H_1]+n^{-\frac{1}{2}}cH_2$$

where H_1 is an arbitrary symmetric distribution, H_2 is a completely arbitrary distribution, and ε and c satisfy

$$\frac{\varepsilon+c^2}{1+c^2} = \varepsilon_0. \tag{3}$$

We may rewrite (3) as $c^2 = (\varepsilon_0-\varepsilon)/(1-\varepsilon_0)$, from which we see that the larger ε is, the smaller c must be, and when $\varepsilon = \varepsilon_0$, its maximum allowable value, c must be zero. (3) thus represents a kind of trade-off between bias due to H_2 and increased variance due to the symmetric H_1. We shall write AMSE (ψ, F) for the AMSE of the estimator defined by ψ under $\{F_n\}$.

THEOREM 5. *AMSE (ψ, F) has a saddlepoint: There is a sequence $\{F_{0,n}\}$ in C' and a ψ_0 such that*

$$\sup_{\{F_n\}} \text{AMSE}\,(\psi_0, F) = \text{AMSE}\,(\psi_0, F_0) = \inf_\psi \text{AMSE}\,(\psi, F_0),$$

where $\{F_n\}$ ranges over C' and ψ ranges over all functions to which the conclusion of Theorem 4 applies. Since $c = 0$ and $\varepsilon = \varepsilon_0$ for the sequence $\{F_{0,n}\}$ we may write $F_{0,n} = F_0 = (1-\varepsilon_0)G+\varepsilon_0 H_0$. F_0 and ψ_0 are the same as those shown by Huber, *page* 81, *to be the saddlepoint for the set of symmetric distributions defined in his theorem by G and ε_0. See Section* 4 *of the present paper. We assume Theorem* 4 *applies to ψ_0 for all $\{F_n\}$ in C'.*

PROOF. As in Huber's theorem, F_0 is symmetric and is a member of C', and ψ_0 is odd, monotonic and bounded, and is asymptotically optimal for F_0. Therefore, AMSE $(\psi_0, F_0) \leqq$ AMSE (ψ, F_0) for all ψ; in fact, this inequality holds for all translation-invariant estimators.

It remains to show that AMSE $(\psi_0, F) \leqq$ AMSE (ψ_0, F_0) for all $\{F_n\}$ in C'. (Compare Huber, page 81.) By Theorem 4,

$$\text{AMSE}\,(\psi_0, F) = \frac{\int \psi_0{}^2(x)[(1-\varepsilon)G+\varepsilon H_1](dx)+c^2\{\int \psi_0(x)H_2(dx)\}^2}{\{\int \psi_0{}'(x)[(1-\varepsilon)G+\varepsilon H_1](dx)\}^2}\,.$$

Since $|\psi_0(x)| \leqq k$, where k is as defined in Huber's theorem,

$$\int \psi_0{}^2(x)[(1-\varepsilon)G+\varepsilon H_1](dx) \leqq (1-\varepsilon)E_G\psi_0{}^2+\varepsilon k^2$$

and

$$\{\int \psi_0(x)H_2(dx)\}^2 \leqq k^2.$$

Since ψ_0 is monotonic, $\psi_0{}' \geqq 0$, so

$$\int \psi_0{}'(x)[(1-\varepsilon)G+\varepsilon H_1](dx) \geqq (1-\varepsilon)E_G\psi_0{}'.$$

Therefore

$$\text{AMSE}\,(\psi_0, F) \leqq \frac{(1-\varepsilon)E_G\psi_0{}^2+(\varepsilon+c^2)k^2}{(1-\varepsilon)^2(E_G\psi_0{}')^2}\,.$$

Since $\varepsilon_0 = (\varepsilon+c^2)/(1+c^2)$, $1-\varepsilon_0 = (1-\varepsilon)/(1+c^2)$. Dividing numerator and denominator by $(1+c^2)^2$, we obtain

$$\text{AMSE}\,(\psi_0, F) \leqq \frac{1}{1+c^2}\cdot\frac{(1-\varepsilon_0)E_G\psi_0{}^2+\varepsilon_0 k^2}{(1-\varepsilon_0)^2(E_G\psi_0{}')^2}\,.$$

The second fraction on the right is Huber's asymptotic variance of ψ_0 under F_0; therefore,

$$\text{AMSE}\,(\psi_0, F) \leqq \frac{1}{1+c^2}\,\text{AMSE}\,(\psi_0, F_0) \leqq \text{AMSE}\,(\psi_0, F_0),$$

and the proof is complete.

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

REFERENCES

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

[2] 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.

[3] GASTWIRTH, J. L. (1966). On robust procedures. *J. Amer. Statist. Assoc.* **61** 929–948.

[4] HÁJEK, J. and Z. ŠIDÁK (1967). *Theory of Rank Tests.* Academic Press, New York.

[5] HODGES, J. L., Jr. and LEHMANN, E. L. (1963). Estimates of location based on rank tests. *Ann. Math. Statist.* **34** 598–611.

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

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

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

[9] TUKEY, J. W. (1960). A survey of sampling from contaminated distributions. *Contributions to Probability and Statistics* (I. Olkin, ed.). Stanford Univ. Press.