

DESCRIPTIVE STATISTICS FOR NONPARAMETRIC MODELS II. LOCATION

BY P. J. BICKEL¹ AND E. L. LEHMANN²

University of California, Berkeley

Measures of location (without assumption of symmetry) are defined as functionals satisfying certain equivariance and order conditions. Three classes of such measures are discussed whose estimators are respectively linear functions of order statistics, R -estimators and M -estimators. It is argued that such measures can be compared in terms of the (asymptotic) efficiencies of their estimators. Of the three classes considered, it is found that trimmed expectations (and certain other weighted quantiles) are the only ones which are both robust and whose estimators have guaranteed high efficiency relative to the mean \bar{X} for all underlying distributions.

1. Conditions for a location parameter. Since the publication of two fundamental papers by Tukey (1960 and 1962), much work has been done on the problem of robust estimation culminating in the Princeton report "Robust Estimation of Location" and Huber's Wald lecture (1972) "Robust Statistics: A review". Most of this work has been concerned with the estimation of location parameters, and in keeping with the program suggested by Tukey (1962, Section 13), the effort to date has been concentrated nearly entirely on the case in which the error distributions are assumed to be symmetric, or at least nearly so. One of the chief advantages of this restriction, as Tukey pointed out in recommending it, lies in the fact that it is then clear what is being estimated: for a symmetric distribution, the only natural location parameter is its center of symmetry.

It is interesting to note that this point was actually made as early as 1920 in a paper by Daniell, recently brought to the attention of present-day statisticians by Stephen Stigler. In this paper, in which Daniell develops a surprisingly modern and comprehensive theory of linear functions of order statistics for estimating location and scale parameters, the author writes: "...however, we shall only consider cases in which the theoretical distribution is symmetrical, and this for logical reasons. It is useless to compare the relative merits of the various kinds of average, for example, the mean and the median, unless they all tend to coincide when n increases indefinitely."

Received April 1973; revised September 1974.

¹ This research was partially supported by the Office of Naval Research, Contract NONR N00014-69-A-O200-1038.

² Work done while the author was Research Professor for the Adolph C. and Mary Sprague Miller Institute for Basic Research in Science, University of California, Berkeley.

AMS 1970 subject classifications. Primary 62G99; Secondary 62G05, 62G20, 62G35.

Key words and phrases. Measures of location, robustness, efficiency, linear combinations of order statistics, trimmed means, estimators derived from rank tests, Huber's maximum likelihood estimators, heavy-tailed distributions, Tukey model.

As was indicated in BL I³ and will be discussed in more detail in Section 5, we in fact disagree with this last statement and believe that meaningful comparisons of this kind can, and should, be made. However, we must first lay the groundwork by defining the parameters we wish to estimate.

By a measure of location (or *location parameter*) we shall mean a functional $\mu(F)$ defined over a suitably large class of distributions (the precise definition of the class will depend on the context), which satisfies the three conditions below (these were already indicated briefly in BL I). In stating these conditions and elsewhere, we shall sometimes find it convenient to write $\mu(X)$ for $\mu(F)$ where X is a random variable with distribution F .

(i) If the parameter is to indicate location, it should take on larger values for random variables which typically are larger. Formally we shall require that $\mu(X) \leq \mu(Y)$ whenever Y is stochastically larger than X .

(ii) More specific behaviour is required of μ under change of location or scale, namely

$$(1.1) \quad \mu(aX + b) = a\mu(X) + b \quad \text{if } a > 0.$$

(iii) Finally, it seems natural to require a measure of location to change sign under reflection w.r.t. the origin, i.e., to satisfy

$$(1.2) \quad \mu(-X) = -\mu(X).$$

While condition (i) seems to us very intuitive, some referees have objected to it on the grounds that location is sometimes considered as referring only to the central part of a distribution. The following example shows that no reasonable version of (i) is possible if that position is adopted since truncation of two stochastically ordered distributions on a common point may reverse the ordering.

Let $G(t) = t$ for $0 < t < 1$ and let

$$\begin{aligned} F(t) &= t && \text{for } 0 < t < t_0 (< 1) \\ &= t_0 + \frac{B - t_0}{A - t_0} (t - t_0) && \text{for } t_0 < t \leq t_1 = t_0 + \frac{(A - t_0)(1 - t_0)}{B - t_0} \\ &= 1 && \text{for } t > t_1, \end{aligned}$$

where $t_0 < A < B < 1$. It is then easily checked that $G(t) < F(t)$ for $t_0 < t < 1$, but that $F^*(t) < G^*(t)$ for $0 < t < A$, where F^*, G^* denote the conditional distributions given that the random variable is $< A$.

Related to this remark is the fact that certain important classes of location measures that have been discussed in the literature do not satisfy condition (i). This is true in particular for the measures corresponding to Huber's M -estimators when scale is estimated simultaneously with location.

THEOREM 1. *Conditions (i)—(iii) imply the following four additional desirable requirements.*

³ BL I refers to part I of the present paper.

1. If F is symmetric w.r.t. θ , then $\mu(F) = \theta$;
2. In particular, if $X = c$ with probability 1, $\mu(X) = c$.
3. If $a \leq X \leq b$ with probability 1, then $a \leq \mu(X) \leq b$;
4. If X is stochastically positive, then $\mu(X) \geq 0$.

PROOF. 1. By (i), it is enough to prove 1. for $\theta = 0$. But if X and $-X$ have the same distribution, it follows from (iii) that $\mu(X) = -\mu(X) = 0$, as was to be proved.

3. If $a \leq X \leq b$ with probability 1, then X is stochastically larger than the constant variable a and smaller than b . The assertion now follows from (i) and 2.

4. By definition, X is stochastically positive if there exists a random variable U , symmetric about 0 and such that X is stochastically larger than U . By the first part of the theorem, $\mu(U) = 0$ and the result now follows from (i).

The following three examples show that conditions (i)—(iii) are independent.

- (a) The functional $\mu(F) = F^{-1}(a)$ with $0 < a < 1$, $a \neq \frac{1}{2}$, satisfies (i) and (ii) but not (iii);
- (b) the functional $\mu(F) = 2E(X) - F^{-1}(\frac{1}{2})$ satisfies (ii) and (iii) but not (i);
- (c) finally, $\mu(X) = [E(X)]^3$ satisfies (i) and (iii) but not (ii).

Let us now mention a few standard measures of location, which satisfy all three conditions:

- (a) The expectation of F

$$(1.3) \quad \mu_1(F) = \int x dF(x) ;$$

- (b) the median of F

$$(1.4) \quad \mu_2(F) = F^{-1}(\frac{1}{2})$$

and the average of the 1st and 3rd quartile

$$(1.5) \quad \mu_3(F) = \frac{1}{2}[F^{-1}(\frac{1}{4}) + F^{-1}(\frac{3}{4})] .$$

An important class of location measures are the symmetrically trimmed expectations

$$(1.6) \quad E_\alpha(X) = \frac{1}{1 - 2\alpha} \int_{u_\alpha}^{u_{1-\alpha}} x dF(x)$$

where $F(u_\alpha) = \alpha$, $F(u_{1-\alpha}) = 1 - \alpha$. The mean and median are the limiting cases corresponding respectively to $\alpha = 0$ and $\alpha = \frac{1}{2}$.

(c) As a last example, we mention the pseudo-median $\mu_4(F)$ defined by Høylund (1965) as the median of the distribution of $\frac{1}{2}(X_1 + X_2)$ where X_1, X_2 are independently distributed according to F . If F is continuous, $\mu_4(F)$ is the solution of the equation

$$(1.7) \quad \int F(2\theta - x) dF(x) = \frac{1}{2} .$$

In the next section, we shall consider two classes of location measures, which contain the above examples as special cases.

Two additional properties, which seem attractive although not as compelling as (i)—(iii) are

- (iv) If $\mu(F) \leq \mu(G)$ and $H = \alpha F + (1 - \alpha)G$, then $\mu(F) \leq \mu(H) \leq \mu(G)$.
- (v) If $\mu(X) \leq \mu(Y)$ and h is a nondecreasing function, then $\mu[h(X)] \leq \mu[h(Y)]$.

If $G(x) \leq F(x)$ for all x , then both (iv) and (v) are consequences of (i). However, in general both of these conditions are very restrictive. Of the examples (1.3) to (1.7) it is not difficult to see that μ_2 satisfies both (iv) and (v), that μ_1 satisfies (iv) but not (v) and that none of the other measures satisfy either (iv) or (v). For most location parameters neither (iv) or (v) will hold; for instance, the median is the only location parameter of the type defined by (3.2) that satisfies (v). We shall not consider these conditions any further here.

There is one last condition, with which we shall be concerned. We shall be interested in measures which are robust in the sense that small changes in F will not result in large changes in μ . Mathematically, this means that μ should be continuous with respect to a suitable metric such as that of Lévy, and then a fortiori also that of Kolmogorov. For a detailed discussion of this and related concepts, see Hampel (1968, 1971); the idea is fore-shadowed in Bahadur and Savage (1956). It is well-known and not difficult to see, (and will follow from the results of Section 4) that of the examples above, $\mu_1(F)$ is not robust, but $\mu_2(F)$ to $\mu_4(F)$ are under mild restrictions on F , and so is $F_\alpha(X)$ for any $0 < \alpha < \frac{1}{2}$.

2. Some examples. Let us now examine somewhat more closely the location parameters which were given as examples in the preceding section. Consider first the three parameters μ_1, μ_2 and μ_4 , i.e. the expectation, median and pseudo median. What can be said about their relative positions? Clearly, $\mu_2(F)$ can be either larger or smaller than $\mu_4(F)$ since $\mu_2(F) < \mu_4(F)$ implies that $\mu_4(-F) < \mu_2(-F)$. Furthermore, regardless of their relative positions, $\mu_1(F)$ can be to the left of both, to the right of both, or between them. This follows from the fact that the position of $\mu_1(F)$ can be moved from $-\infty$ to $+\infty$ by moving an arbitrarily small probability mass, which would result in only very small changes of $\mu_2(F)$ and $\mu_4(F)$.

Nevertheless, $\mu_4(F)$ can be bounded between certain quantiles. In fact, if F is continuous

$$(2.1) \quad F^{-1}\left(1 - \frac{1}{2^{\frac{1}{4}}}\right) \leq \mu_4(F) \leq F^{-1}\left(\frac{1}{2^{\frac{1}{4}}}\right).$$

To see this let $\mu_4(F) = \theta$ and let $p = P(X < \theta)$, $q = P(X > \theta)$. Then

$$P\left[\frac{X + X'}{2} < \theta\right] \geq p^2, \quad P\left(\frac{X + X'}{2} > \theta\right) \geq q^2$$

and the result follows from the definition of $\mu_4(F)$.

To get an idea of the values of the parameters μ_1, μ_2 and μ_4 in particular situations, consider as a first example the case that F is a χ^2 -distribution with ν degrees

TABLE 2.1
Location parameters for χ_ν^2

$\mu_1(F) = \nu$	1	2	3	4	5
$\mu_4(F)$.69	1.68	2.67	3.67	4.67
$\mu_2(F)$.45	1.39	2.37	3.36	4.35

TABLE 2.2
Differences of location parameters for χ_ν^2

ν	1	2	3	4	5	∞
$\mu_1(F) - \mu_4(F)$.31	.32	.33	.33	.33	$\frac{1}{3}$
$\mu_2(F) - \mu_4(F)$.24	.29	.30	.31	.32	$\frac{1}{3}$

of freedom. Then the table above gives the values of $\mu_1(F) = \nu$, $\mu_4(F)$ and $\mu_2(F)$ for $\nu = 1, 2, 3, 4$ and 5 . In spite of the rather different shapes of the distributions (which however are all skewed to the right), the differences $\mu_1(F) - \mu_4(F)$ and $\mu_2(F) - \mu_4(F)$ are remarkably constant, as is shown in Table 2.2 above. The limiting value as $\nu \rightarrow \infty$, shown in the last column comes from the well-known Cornish-Fisher expansion for the solution of the equation $P(\chi_\nu^2 \leq x_{\nu,\epsilon}^2) = \epsilon$:

$$x_{\nu,\epsilon}^2 = \nu \left[1 + 3w_\epsilon \left(\frac{2}{9\nu} \right)^{\frac{1}{2}} - \frac{2}{3\nu} + \frac{2w_\epsilon^2}{3\nu} + o\left(\frac{1}{\nu} \right) \right]$$

where $w_\epsilon = \Phi^{-1}(\epsilon)$ with Φ denoting the standard normal distribution. Putting $\epsilon = \frac{1}{2}$, we see that

$$\mu_2(F) = x_{\nu,\frac{1}{2}}^2 = \nu - \frac{2}{3} + o(1).$$

Now $\mu_4(F)$ is the median of $(X + X')/2$ where X and X' are independently distributed as $\chi_{2\nu}^2$. The median of $X + X'$ thus satisfies

$$x_{2\nu,\frac{1}{2}}^2 = 2\nu - \frac{2}{3} + o(1)$$

and

$$\mu_4(F) = \nu - \frac{1}{3} + o(1).$$

Comparison with $\mu_2(F)$ and $\mu_1(F) = \nu$ gives the desired result.

It is interesting to note that in the above examples and those computed below, μ_4 is always between μ_1 and μ_2 . This is not too surprising (although all orders are possible) since $\mu_1(X) = \mu_1[\frac{1}{2}(X_1 + X_2)]$ and $\mu_4(X) = \mu_2[\frac{1}{2}(X_1 + X_2)]$. By the central limit theorem one might expect $\frac{1}{2}(X_1 + X_2)$ to be closer to symmetry than is X , and μ_2 to be closer to μ_1 for $\frac{1}{2}(X_1 + X_2)$ than for X .

As a second example, consider the family of distributions

$$(2.2) \quad F_k = \gamma N(1, 1) + (1 - \gamma)N(k, k^2)$$

where $N(a, b^2)$ denotes a normal distribution with mean a and standard deviation b . The following table shows the values of μ_1 , μ_2 and μ_4 for $\gamma = \frac{3}{4}$ and $k = 2, 3, 5$ and 10 . The last column gives the limiting value as $k \rightarrow \infty$. To see how

TABLE 2.3
Location parameter for normal mixtures

k	2	3	5	10	∞
μ_1	1.25	1.50	2	3.25	∞
μ_2	1.14	1.19	1.26	1.27	1.29
μ_4	1.18	1.27	1.35	1.45	1.53

this is obtained for μ_2 and μ_4 , consider for example μ_2 and let u_k be the solution of the equation

$$(2.3) \quad \gamma\Phi(u - 1) + (1 - \gamma)\Phi\left(\frac{u - k}{k}\right) = \frac{1}{2}$$

with $\gamma = \frac{3}{4}$. As $u \rightarrow \infty$, the first term on the left hand side tends to $\frac{3}{4}$ which is too large. Thus u_k must remain bounded as $k \rightarrow \infty$. This means that the second term tends to $\frac{1}{4}\Phi(-1) = .0397$ and u_∞ is the solution of the equation

$$\frac{3}{4}\Phi(u - 1) = .4603 .$$

The argument for μ_4 is quite analogous.

Consider next the family (2.2) for a value of γ less than $\frac{1}{2}$. This time, as $k \rightarrow \infty$ any finite value of u is too small so that $u_k \rightarrow \infty$. The first term then tends to γ , and u_k must satisfy the equation

$$\Phi\left(\frac{u_k}{k} - 1\right) \rightarrow \frac{\frac{1}{2} - \gamma}{1 - \gamma} .$$

Proceeding in the same way for $\mu_4(F)$, we find that as $k \rightarrow \infty$, we have for example for $\gamma = \frac{1}{4}$

$$\frac{\mu_1(F_k)}{k} \rightarrow .75 , \quad \frac{\mu_2(F_k)}{k} \rightarrow .069 , \quad \frac{\mu_4(F_k)}{k} \rightarrow .705 .$$

As a last example, consider the Gumbel distribution with density

$$f(x) = e^{-x}e^{-e^{-x}} , \quad -\infty < x < \infty .$$

Here the values of μ_1 , μ_4 and μ_2 are respectively

$$\mu_1 = .577 , \quad \mu_4 = .464 , \quad \mu_2 = .366 .$$

Of those, μ_1 and μ_2 can be found for example in Johnson and Kotz (1970). To calculate μ_4 we note that,

$$\begin{aligned} P\left[\frac{X_1 + X_2}{2} \leq y\right] &= \int_{-\infty}^{\infty} e^{-e^{-(2y-x)}-x-e^{-x}} dx \\ &= \int_0^{\infty} \exp\left\{-\frac{e^{-2y}}{w} + w\right\} dw \\ &= e^{-2y} \int_0^{\infty} e^{-e^{-2y}z} e^{-1/z} dz \\ &= 2e^{-y}K_1(2e^{-y}) \end{aligned}$$

TABLE 2.4
Trimmed expectations for exponential distributions

α	0	.01	.07	.03	.05	.1	.5
$E_\alpha(X)$	1	.963	.939	.919	.888	.831	.693

where K_1 is the Bessel function of the second kind with imaginary argument (cf. Churchill (1958)). The value of μ_4 is now obtainable from the British Association Tables 6 (1958).

So far, we have compared the parameters μ_1 , μ_2 and μ_4 for a number of different distributions. Let us next illustrate the trimmed expectations $E_\alpha(X)$ defined by (1.6). As a first illustration, let F be the exponential distribution with density e^{-x} , $x > 0$. Then $E_\alpha(X)$ is shown above for a number of values of α . The values corresponding to $\alpha = 0$ and $\alpha = .5$ are of course the expectation and median respectively. The pseudo median in this case is .84.

As a second illustration, consider the mixture $F(x) = \frac{3}{4}\Phi(x) + \frac{1}{4}\Phi(x - 1)$. Some values of $E_\alpha(X)$ in this case are

TABLE 2.5
Trimmed expectations for a normal mixture

α	0	.05	.1	.15	.5
$E_\alpha(X)$.25	.248	.239	.238	.23

The pseudo median is .243.

3. Three classes of location parameters. Among the estimators of the center of symmetry of a symmetric distribution F , three classes have been found to be of particular interest (see for example Jaeckel (1971), Andrews et al. (1972), and Huber (1972)): linear combinations of order statistics, estimators derived from rank tests, and the maximum likelihood type estimators introduced by Huber. We shall now consider the quantities estimated by these estimators when F is no longer assumed to be symmetric, under restrictions which assure that they are measures of location in the sense of Section 1.

(A) If $\mu(X)$ satisfies conditions (i) and (ii) of Section 1 but not necessarily condition (iii), then it is easily seen that

$$(3.1) \quad \frac{1}{2}[\mu(X) - \mu(-X)]$$

satisfies all three conditions, i.e. is a location parameter. Furthermore, if $\{\mu_i(X)\}$ is a countable collection of functionals which satisfy (i)—(iii) then so does $\sum \alpha_i \mu_i(X)$ for any nonnegative α 's which add up to 1. This latter remark clearly extends also to noncountable collections.

As an application of these remarks consider the functional $\mu(X) = F^{-1}(a)$ for some fixed a between 0 and 1. This clearly satisfies conditions (i) and (ii). Since $\mu(-X) = -F^{-1}(1 - a)$, it follows that $\frac{1}{2}[F^{-1}(a) + F^{-1}(1 - a)]$ is a location parameter for any $0 < a < 1$. Forming convex combinations of these parameters,

we then see that the weighted quantile average

$$(3.2) \quad \mu(F) = \int_0^1 F^{-1}(t) dK(t)$$

is a location parameter, where K is any distribution function on $(0, 1)$ which is symmetric with respect to $\frac{1}{2}$.

Important special cases of (3.2) are the symmetrically trimmed expectations (1.6), as is seen by rewriting (1.6) as

$$(3.3) \quad E_\alpha(X) = \frac{1}{1 - 2\alpha} \int_\alpha^{1-\alpha} F^{-1}(t) dt,$$

and their limiting cases, the expectation and the median.

(B) To obtain the second class of location parameters, consider the probability

$$(3.4) \quad P(|X - \theta| \leq x) = F(x + \theta) - F(-x + \theta).$$

It seems natural to define as center of F the number θ for which the probabilities (3.4) are, in some average sense, as large as possible. Let L be an increasing convex function on $[0, 1]$ which is bounded and such that $L(0) = 0$. Define $\mu(F)$ as the number θ that maximizes

$$(3.5) \quad \int_0^\infty \{L[F(x + \theta) - F(-x + \theta)] - L[F(x) - F(-x)]\} dx.$$

Here the subtraction of $L[F(x) - F(-x)]$ under the integral sign is intended to aid convergence. It may be shown that if L' is bounded in absolute value, the functional (3.5) is always defined. (See Knüsel (1969) for a closely related remark.)

It is sometimes more convenient to extend L so that it is an even function on $[-1, 1]$, and it is then equivalent to maximize

$$(3.6) \quad \int_{-\infty}^\infty \{L[F(x + \theta) - F(-x + \theta)] - L[F(x) - F(-x)]\} dx \\ = \int_{-\infty}^\infty \{L[F(x) - F(2\theta - x)] - L[F(x) - F(-x)]\} dx.$$

Suppose that F has a density f , L is continuously differentiable, that $M = L'$, and that (3.6) can be differentiated under the integral sign. It can then be shown exactly as was done by Knüsel (1969) for a related but somewhat different class of functionals that the equation

$$\int_{-\infty}^\infty L'[F(x) - F(2\theta - x)]f(2\theta - x) dx = 0,$$

or, equivalently

$$(3.7) \quad \int_{-\infty}^\infty L'[F(x) - F(2\theta - x)]f(x) dx = 0$$

has at least one solution. If L' is strictly increasing, this solution is unique and hence is the desired maximizing value. (These measures are the quantities estimated by estimators derived from rank tests; see for example Andrews et al. (1972)).

In particular if $L(y) = y^2$, it is seen that (3.7) reduces to

$$(3.8) \quad \int F(2\theta - x)f(x) dx = \frac{1}{2}$$

so that θ is the pseudo median $\mu_1(F)$ of F . As another example suppose that in (3.5) we take $L(y) = y$. Then differentiation shows θ to be the median $\mu_2(F)$.

To see that the measure $\mu(F)$ uniquely defined through (3.7) is a location parameter for F continuous, rewrite (3.7) as

$$(3.9) \quad \int L'[t - F(2\theta - F^{-1}(t))] dt = 0 .$$

To check condition (i), suppose that $G(x) \leq F(x)$ for all x . Then for any fixed value of θ , we have $F[2\theta - F^{-1}(t)] \geq G[2\theta - G^{-1}(t)]$, and since L' is an increasing function, it follows that

$$\int L'[t - F(2\theta - F^{-1}(t))] dt \leq \int L'[t - G(2\theta - G^{-1}(t))] dt .$$

In this last inequality, let $\theta = \mu(G)$ so that the right hand side is zero. Since the left hand side is a decreasing function of θ , the value $\mu(F)$ of θ which will make the left hand side equal to zero will be less than or equal to $\mu(G)$, as was to be proved. That conditions (ii) and (iii) are also satisfied, is easily seen by replacing $F(x)$ by $F((x - b)/a)$ in (3.7).

(C) The third class is obtained from the quantities estimated by Huber's M -estimators. These are the quantities $\theta = \mu(F)$ which minimize

$$(3.10) \quad \int \rho(x - \theta) dF(x)$$

where we shall assume ρ to be positive, even, convex and twice differentiable with derivative $\rho' = \psi$. (It is possible to enlarge the domain of definition of $\theta(F)$ by minimizing $\int_{-\infty}^{\infty} [\rho(x - \theta) - \rho(x)] dF(x)$ instead, as was done by Knüsel (1969).) Essentially we shall now show that only a one-parameter subclass of (C) satisfies the conditions of Section 1.

THEOREM 2. *Suppose that $\mu(F)$ is defined as minimizing (3.10) on a set \mathcal{F} which is convex, contains all point masses, is closed under changes of scale, and contains a distribution F° symmetric about 0 such that*

$$\int \psi'(x) dF^\circ(x + t) < \infty \quad \text{for all } t \quad \text{and}$$

$$\frac{d}{dt} \int \psi(x - t) dF^\circ(x) = - \int \psi'(x - t) dF^\circ(x) .$$

For any given distribution F , denote by F_σ the distribution defined by $F_\sigma(x) = F(x/\sigma)$. Suppose that

$$(3.11) \quad \mu(F_\sigma) = \sigma \mu(F) \quad \text{for all } F \in \mathcal{F}, \quad \sigma > 0 .$$

(This is part of the assumption (1.1)). Then

$$(3.12) \quad \psi(x) = c|x|^\alpha \operatorname{sgn} x \quad \text{for some } \alpha > 0, \quad c > 0 .$$

PROOF. The measure $\mu(F)$ is the solution of

$$(3.13) \quad \int \psi(x - \theta) dF(x) = 0$$

Using our assumptions, we calculate

$$\lim_{\epsilon \rightarrow 0} \frac{\mu[(1 - \epsilon)F^\circ + \epsilon\delta_x] - \mu(F^\circ)}{\epsilon} = \frac{\psi(x)}{\int_{-\infty}^{\infty} \psi'(x) dF^\circ(x)}$$

and similarly

$$\lim_{\varepsilon \rightarrow 0} \frac{\mu([(1 - \varepsilon)F^\circ + \varepsilon\delta_x]_\sigma) - \mu(F^\circ)}{\varepsilon} = \frac{\phi(\sigma x)}{\int_{-\infty}^{\infty} \phi'(\sigma y) dF^\circ(y)}.$$

From (3.11)

$$\frac{\phi(\sigma x)}{\int_{-\infty}^{\infty} \phi'(\sigma y) dF^\circ(y)} = \frac{\sigma\phi(x)}{\int_{-\infty}^{\infty} \phi'(y) dF^\circ(y)}.$$

Differentiating both sides with respect to x for $x > 0$ leads to

$$\frac{\phi'(\sigma x)}{\int \phi'(\sigma y) dF^\circ(y)} = \frac{\phi'(x)}{\int \phi'(y) dF^\circ(y)}$$

and hence

$$\frac{\phi'(\sigma x)}{\phi'(x)} = \frac{\phi(\sigma x)}{\sigma\phi(x)} \quad \text{for } x > 0.$$

Put $x = 1$ to get

$$\frac{\phi'(\sigma)}{\phi(\sigma)} = \frac{\phi'(1)}{\phi(1)} \frac{1}{\sigma} \quad \text{or} \quad \log \phi = \frac{\phi'(1)}{\phi(1)} \log \sigma + c, \quad \sigma > 0$$

and this completes the proof.

REMARKS.

(1) The conclusion of this theorem was stated by Takeuchi (1967) without proof or regularity conditions.

(2) One of the referees has pointed out that the theorem also holds under a weaker set of conditions not including differentiability.

4. Robustness. Following Hampel (1968, 1971) and earlier writers, we shall define a functional μ to be robust at a point F_0 if μ is continuous at F_0 with respect to Lévy (or equivalently Prohorov) distance. However, this concept does not completely correspond to our intuitive notion of robustness since even the median is not continuous at all F . An easy example is provided by distributions for which $F(x) = \frac{1}{2}$ on a nondegenerate interval. On the other hand, the median is robust at every point of the set of all distributions having a unique median.

(A) Consider now the functionals μ defined by (3.2) and suppose that K is continuous. Then a necessary and sufficient condition for μ to be continuous at all points of \mathcal{F} is that K assigns probability zero to the intervals $(0, \alpha)$ and $(1 - \alpha, 1)$ for some $\alpha > 0$.

The necessity of this condition (which does not require the continuity of K) was pointed out by Huber (1972). Sufficiency follows from the fact that convergence in law of F_n implies bounded convergence, except on a countable set, for F_n^{-1} on the intervals $(\alpha, 1 - \alpha)$.

(B) For measures μ defined through (3.7), suppose that L is bounded and L' is nondecreasing. Huber (1972) has shown that μ is continuous at any distribution F for which (3.6) can be differentiated under the integral sign and such

that the resulting equation (3.7) has a unique solution. An alternative argument employs the following representation which may have some independent interest.

Define a distribution Q on the unit square by the cdf.

$$(4.1) \quad Q(u, v) = \frac{1}{2L(1)} \int_0^u [L'(v-t) - L'(-t)] dt.$$

Then $\mu(F)$ is the median of the distribution of $\frac{1}{2}[F^{-1}(U) + F^{-1}(V)]$ where (U, V) has distribution Q . Since Q is continuous, convergence in law of the cdf F_n to F implies convergence in law of $\frac{1}{2}[F_n^{-1}(U) + F_n^{-1}(V)]$ to $\frac{1}{2}[F^{-1}(U) + F^{-1}(V)]$. It follows that μ is continuous at any F for which it is uniquely defined.

(C) Of the measures defined by (3.12) and (3.13), it is easy to see that none are robust. However, it is clear that in a suitable sense if $\alpha < \beta$, μ_α is more robust than μ_β .

It seems reasonable that if ν is to be considered more robust than μ , then for any sequence F_n tending in law to F for which $\mu(F_n) \rightarrow \mu(F)$, we should also have $\nu(F_n) \rightarrow \nu(F)$. Unfortunately this definition seems to lead to noncomparability for the nonrobust functionals we have considered. All we need do is to consider sequences F_n converging in law to F such that $\mu(F_n) \rightarrow \mu(F)$ "fortuitously". For example let F_n assign mass $1 - [1/n]^{1/\alpha} - [1/n]^{1+1/\alpha}$ to 0 and mass $[1/n]^{1/\alpha}$ to n and $[1/n]^{1+1/\alpha}$ to $-n^2$. Then F_n tends to point mass at 0 and the expectations under F_n converge to that under F . However, $\mu_\alpha(F_n)$ converges to 1. To remedy this difficulty we formulate the following notion of comparative robustness, which gives us some needed control over the tails of the F_n . For any cdf F and any constant $M > 0$ let us define,

$$\begin{aligned} \bar{F}(x, M) &= 0, & x < 0, \\ &= F(M), & 0 \leq x \leq M \\ &= F(x), & x \geq M \\ \underline{F}(x, M) &= F(x), & x \leq -M, \\ &= F(-M), & -M \leq x < 0, \\ &= 1, & x \geq 0. \end{aligned}$$

Given a functional T we define a mode of convergence and hence a topology on its set of definition \mathcal{F} (which as usual is assumed to be large enough) as follows.

$$F_n \rightarrow_\tau F \Leftrightarrow$$

- (i) $F_n \rightarrow F$ in law ;
- (ii) $T(F_n) \rightarrow T(F)$;
- (iii) $\lim_{M \uparrow \infty} \limsup_n \{T(\bar{F}_n(\cdot, M)) + |T(F_n(\cdot, M))|\} = 0$.

(Note that by our assumptions if T is a measure of location, $T(\bar{F}) \geq 0$, $T(\underline{F}) \leq 0$).

DEFINITION. We shall call T_2 more robust than T_1 if and only if T_2 is continuous with respect to the topology induced by T_1 .

This permits us to make comparisons between different measures of type (A) and (C) which are not robust in the absolute sense considered previously. For convenience in what follows we suppose all functionals to be defined on the space of distributions with compact support.

THEOREM 3. Let T_1 and T_2 be two (L) measures defined by (3.2), say,

$$(4.2) \quad T_i(F) = \int_0^1 F^{-1}(t) dK_i(t) \quad i = 1, 2,$$

and suppose that K_1, K_2 are continuous probability measures. Then T_2 is more robust than T_1 if there exists $0 < \varepsilon \leq \frac{1}{2}$ such that

$$(4.3) \quad \begin{aligned} K_1(t) &\geq K_2(t) && \text{for all } 0 \leq t \leq \varepsilon, && \text{and} \\ K_1(t) &\leq K_2(t) && \text{for all } 1 - \varepsilon \leq t \leq 1. \end{aligned}$$

PROOF. Note that

$$(4.4) \quad \begin{aligned} T_i[\bar{F}(\cdot, M)] &= \int_{F(M)}^1 F^{-1}(t) dK_i(t), && \text{and} \\ T_i[\underline{F}(\cdot, M)] &= \int_0^{F^{-1}(M)} F^{-1}(t) dK_i(t). \end{aligned}$$

Therefore, for functionals defined by (3.2), (with T in place of μ), if (i) holds it follows that (ii) and (iii) are equivalent to

$$(4.5) \quad \lim_{\varepsilon \rightarrow \infty} \limsup_n \{ |\int_{\varepsilon}^1 F_n^{-1}(t) dK(t)| + |\int_0^{\varepsilon} F_n^{-1}(t) dK(t)| \} = 0.$$

Now suppose that $F_n \rightarrow_{T_1} F$. To show that $T_2(F_n) \rightarrow T_2(F)$ note that

$$(4.6) \quad \begin{aligned} |T_2(F_n) - T_2(F)| &\leq [|\int_{\varepsilon}^{1-\varepsilon} |F_n^{-1}(t) - F^{-1}(t)| dK_2(t) | \\ &\quad + |\int_0^{\varepsilon} F_n^{-1}(t) dK_2(t)| + |\int_{\varepsilon}^1 F^{-1}(t) dK_2(t)| \\ &\quad + |\int_{\varepsilon}^1 F_n^{-1}(t) dK_2(t)| + |\int_{\varepsilon}^1 F^{-1}(t) dK_2(t)| \end{aligned}$$

and that the first term on the right hand side of (4.6) converges to 0 for each fixed ε by dominated convergence.

Therefore, we need only establish that

$$(4.7) \quad \lim_{\varepsilon \rightarrow 0} \limsup_n \{ |\int_{\varepsilon}^1 F_n^{-1}(t) dK_2(t)| + |\int_0^{\varepsilon} F_n^{-1}(t) dK_2(t)| \} = 0.$$

To prove (4.7) begin by supposing that $0 < F(0-) \leq F(0) < 1$. Then, for ε sufficiently small, $F_n^{-1}(t) \geq 0$ for all $t \geq 1 - \varepsilon$ and ≤ 0 for all $t \leq \varepsilon$ and all n . Hence,

$$\begin{aligned} |\int_0^{\varepsilon} F_n^{-1}(t) dK_2(t)| &= - \int_0^{\varepsilon} F_n^{-1}(t) dK_2(t) \leq - \int_0^{\varepsilon} F_n^{-1}(t) dK_1(t) \\ &= |\int_0^{\varepsilon} F_n^{-1}(t) dK_1(t)| \end{aligned}$$

by our hypothesis on the K_i . Similar inequalities hold for $\int_{1-\varepsilon}^1$ and (4.7) follows from the hypothesis $F_n \rightarrow_{T_1} F$ in view of (4.5).

The two cases when $F(0) = 1$ or $F(0-) = 0$ are argued similarly. \square

Consider next the estimators of type (C) satisfying (3.12).

THEOREM 4. Suppose that $0 < \alpha_1 < \alpha_2$ and T_α corresponds to ϕ_α . Then T_{α_1} is more robust than T_{α_2} .

This result is an easy consequence of the following lemma.

LEMMA. If $\alpha > 0$, $F_n \rightarrow_{T_\alpha} F \Leftrightarrow$
 (4.8 a) $F_n \rightarrow F$ in law

and

(4.8 b) $\lim_{M \uparrow \infty} \limsup_n \int_{|x| \geq M} |x|^\alpha dF_n(x) = 0.$

PROOF. Suppose that $F_n \rightarrow F$ in law and that (4.8 a) and (4.8 b) hold. Then

$$\int \psi_\alpha(x - \theta) d\bar{F}_n(x, M) \rightarrow \int \psi_\alpha(x - \theta) d\bar{F}(x, M)$$

for all θ, M and hence

(4.9) $T_\alpha(\bar{F}_n(\cdot, M)) \rightarrow T_\alpha(\bar{F}(\cdot, M))$

Convergence similarly holds for $T_\alpha(F_n(\cdot, M))$ and $T_\alpha(F(\cdot, M))$. Since

$$\lim_{M \uparrow \infty} \{|T_\alpha(\bar{F}(\cdot, M))| + |T_\alpha(F(\cdot, M))|\} = 0$$

and $T_\alpha(\bar{F}_n(\cdot, M))$ is decreasing in M for fixed n it can easily be argued that (4.9) implies $\lim_M \limsup_n |T_\alpha(\bar{F}_n(\cdot, M))| = 0$. Similarly,

$$\lim_M \limsup_n |T_\alpha(F_n(\cdot, M))| = 0$$

and the “only if” part of the lemma follows. Conversely, suppose that $F_n \rightarrow_{T_\alpha} F$. Then, for M sufficiently large and all n , by the definition of T_α , for any $\varepsilon > 0$, the solutions $T_\alpha(\bar{F}_n(\cdot, M))$ and $T_\alpha(F_n(\cdot, M))$, of

$$\int_M^\infty (x - \theta)^\alpha dF_n(x) = \theta^\alpha F(M)$$

and

$$\int_{-\infty}^{-M} (\theta - x)^\alpha dF_n(x) = |\theta|^\alpha (1 - F(-M))$$

lie in the interval $(-\varepsilon, \varepsilon)$.

Then

$$\int_M^\infty |x|^\alpha dF_n(x) \leq 2^\alpha \varepsilon^\alpha$$

and

$$\int_{-\infty}^{-M} |x|^\alpha dF_n(x) \leq 2^\alpha \varepsilon^\alpha.$$

The lemma follows.

5. Estimation. The measures defined by (3.2) and (3.7) all satisfy conditions (i)—(iii) of Section 1, and are robust under mild conditions. There is therefore little to choose among them on this basis. We have, however, so far neglected an important aspect of these measures: how to estimate them.

The situation here is rather different from that in the symmetric location case. There are many reasonable estimators of the center of symmetry of a symmetric distribution, and much work has been devoted to comparing the accuracies of different such estimators. Without the assumption of symmetry, there is typically only one natural estimator of a location parameter $\mu(F)$, namely μ evaluated at the empirical distribution function.

The problem of choosing among different estimators of the same location parameter therefore does not arise. Instead, we are faced with the problem of comparing location parameters in terms of the ease with which they can be

estimated. It may at first seem strange to compare estimators of different parameters. However, such comparisons appear justified, philosophically because primarily we are not comparing the estimators but just these parameters, and mathematically because a shift in origin leaves the problem of estimating location essentially unchanged. This latter point becomes clearer through a comparison with the scale problem. Here the corresponding fact is not true since the formal expression for the accuracy with which a scale parameter can be estimated depends on the unit of measurement. (See also the related discussion in BL I.)

NOTE. We stated above that typically there is only one natural estimator of a location parameter $\mu(F)$, namely $\mu(\hat{F})$ where \hat{F} is the empirical cdf. However, unless additional restrictions are imposed on the estimators, the phenomenon of superefficiency can occur. Consider, for example, the problem of estimating the median of a completely unknown continuous strictly increasing distribution function. If T_n is an estimator based on n observations from F , one might conjecture that, asymptotically, the sample median is uniformly best among all estimators which are consistent estimators of the population median, and which satisfy the equivariance condition

$$T_n(aX_1 + b, \dots, aX_n + b) = aT_n(X_1, \dots, X_n) + b.$$

Consider however the estimator T_n^* defined by

$$T_n^*(X_1, \dots, X_n) = \hat{\mu}_1 \quad \text{if } \inf_{\mu, \sigma} \sup_x \left| \hat{F}_n(x) - \Phi\left(\frac{x - \mu}{\sigma}\right) \right| \leq n^{-1} \\ = \hat{\mu}_2 \quad \text{otherwise}$$

where $\hat{\mu}_1$ and $\hat{\mu}_2$ denote the sample mean and sample median respectively. Then it is easy to see that T_n^* is equivariant, consistent for μ_2 and superefficient when F is normal. In view of the results of Le Cam, Huber, Hájek (see Hájek (1970)), one might conjecture that such phenomena can be excluded by requiring continuity (in a suitable sense) of the asymptotic variance of the estimator.

We shall now consider from the above point of view the three classes of estimators obtained from the classes (A), (B) and (C) of location measures described in Section 3 by replacing the true distribution function F in $\mu(F)$ by the empirical distribution function \hat{F}_n .

5 A. Linear functions of order statistics. The estimators corresponding to the measures (3.2) are linear functions of order statistics (see for example Andrews et al. (1972)). Let us compare the estimators of two such measures

$$(5.1) \quad \mu_i(F) = \int_0^1 F^{-1}(t) dK_i(t) \quad i = 1, 2$$

where the K_i are two distributions on $(0, 1)$ with densities $K_i'(t)$.

THEOREM 5. Suppose that the densities K_i' satisfy

$$(5.2) \quad 0 \leq K_2'(t)/K_1'(t) \leq A \quad \text{where } 1 < A.$$

Then if T_i is the estimator of $\mu_i(F)$ obtained by replacing F by the empirical cdf, the asymptotic variances of T_i satisfy

$$(5.3) \quad \sigma_2^2(F) \leq A^2 \sigma_1^2(F) \quad \text{for all } F,$$

under the regularity conditions permitting the representation (5.5) below. Under these conditions $1/A^2$ is then a lower bound for the efficiency of T_2 relative to T_1 , i.e.

$$(5.4) \quad e_{T_2, T_1}(F) \geq \frac{1}{A^2} \quad \text{for all } F,$$

and the lower bound is sharp in the sense that if A is the sharp bound on K_2'/K_1' there exists a sequence F_n for which the efficiency tends to $1/A^2$.

PROOF. The asymptotic variance of $n^2 T_i$ (cf. Huber (1972)) is

$$(5.5) \quad \sigma_i^2(F) = \text{Var} [U_i(T)]$$

where T is uniform over $(0, 1)$ and

$$(5.6) \quad U_i(t) = \int_0^t \frac{K_i'(t)}{f[F^{-1}(t)]} dt.$$

Let T_1, T_2 be independent, uniform over $(0, 1)$. Then

$$\begin{aligned} 2\sigma_i^2(F) &= \text{Var} [U_i(T_2) - U_i(T_1)] = E[U_i(T_2) - U_i(T_1)]^2 \\ &= E \left(\int_{T_1}^{T_2} \frac{K_i'(s)}{f(F^{-1}(s))} ds \right)^2. \end{aligned}$$

Since $K_2'(t) \leq AK_1'(t)$ it follows that

$$\sigma_2^2(F) \leq A^2 \sigma_1^2(F).$$

Let us next show that this upper bound for σ_2^2 is sharp. For this purpose suppose that

$$K_2'(t) = (1 - \epsilon)AK_1'(t) \quad \text{for } a \leq t \leq b,$$

and let us write

$$2\sigma_i^2(F) = E(\int_{F^{-1}(T_1)}^{F^{-1}(T_2)} K_i'[F(w)] dw)^2 = 2 \int \int_{0 < t_1 < t_2 < 1} (\int_{F^{-1}(t_1)}^{F^{-1}(t_2)} K_i'[F(w)] dw)^2 dt_1 dt_2.$$

Then

$$\sigma_2^2(F) \geq \int \int_{0 < t_1 < t_2 < 1} A^2(1 - \epsilon)^2 (\int_{\max(F^{-1}(t_1), F^{-1}(a))}^{\min(F^{-1}(t_2), F^{-1}(b))} K_1'[F(w)] dw)^2 dt_1 dt_2.$$

Suppose we choose F so that $F^{-1}(a) = u$, $F^{-1}(b) = v$ and we put all mass to the left of u and the right of v very close to u and v respectively. Then $t_1 < a$ implies that $F^{-1}(T_1)$ is close to $F^{-1}(a)$ with high probability and $t_2 > b$ implies that $F^{-1}(T_2)$ is close to $F^{-1}(b)$ with high probability so that for a suitable sequence F_n

$$\int \int_{0 < t_1 < t_2 < 1} (\int_{\max}^{\min} K_1'[F(w)] dw)^2 dt_1 dt_2 = \int \int (\int_{F^{-1}(t_1)}^{F^{-1}(t_2)} K_1'[F(w)] dw)^2 dt_1 dt_2$$

and this completes the proof.

In the above, we have not used the symmetry of K . If K is symmetric about $\frac{1}{2}$, it is easily seen that we can choose the sequence F_n to be symmetric, so that A^2 is the sharp lower bound even if F is restricted to be symmetric. Thus, provided (5.2) holds, the lower bound is the same when F is permitted to be asymmetric as under the assumption of symmetry.

As an example, consider the weight functions K_i with densities

$$(5.7) \quad K_1'(t) \equiv 1, \quad K_2'(t) = 6t(1 - t) \quad \text{for } 0 \leq t \leq 1.$$

The estimator corresponding to K_2 was proposed by Bickel (1973), and it follows from Theorem 5 that its asymptotic efficiency relative to \bar{X} is always $\geq (\frac{4}{6})^2 = \frac{4}{9}$.

As an important class of estimators whose comparison with \bar{X} is covered by Theorem 5, let us now consider the trimmed means \bar{X}_α . These estimate the measures $E_\alpha(X)$ defined by (1.6) from which they are obtained by replacing F by \hat{F}_n . For this comparison, (5.2) holds with $A = 1 - 2\alpha$. The efficiency $e_\alpha(F)$ of \bar{X}_α to \bar{X} thus satisfies

$$(5.8) \quad e_\alpha(F) \geq (1 - 2\alpha)^2 \quad \text{for all } F.$$

For symmetric F , this was shown earlier by Bickel (1965).

If more is known about F , it may of course be possible to sharpen the bound (5.8). Typically, one would expect $e_\alpha(F)$ to increase as the tail of the distribution becomes heavier. An ordering of distributions which reflect this property was introduced by Barlow and Proschan (1966) (see also Doksum (1969)), and Lawrence (1966), who suggest that if G has heavier tails than F one would expect that

$$(5.9) \quad G^{-1}(u)/F^{-1}(u) \quad \text{is increasing.}$$

THEOREM 6. *If F and G are symmetric distributions satisfying (5.9), then*

$$(5.10) \quad e_\alpha(F) \leq e_\alpha(G) \quad \text{for any } 0 < \alpha \leq \frac{1}{2}.$$

The proof is based on the following lemmas which will be useful also for proving some later results.

LEMMA 1. *Let F, G be two distributions defined on $(0, \infty)$ such that the cdf defined by*

$$(5.11) \quad F^*(t) = \int_0^t a(x) dF(x) / \int_0^\infty a(x) dF(x)$$

is stochastically smaller than the cdf G^ obtained when F is replaced by G , for all nonnegative functions a which are integrable with respect to F and G . Then*

$$(5.12) \quad \frac{\int b(x) dF(x)}{\int a(x) dF(x)} \leq \frac{\int b(x) dG(x)}{\int a(x) dG(x)}$$

for all nonnegative functions a, b for which $b(x)/a(x)$ is nondecreasing.

PROOF. The left hand side can be written as

$$\int r(x) dF^*(x)$$

where $r(x) = b(x)/a(x)$, and the result follows from a well known property of stochastically ordered distributions.

LEMMA 2. *The conclusion of Lemma 1 will hold if $b(x)/a(x)$ is nondecreasing and if the functions a , F and G satisfy one of the following three conditions:*

(a) $dG(x)/dF(x)$ is nondecreasing;⁴

(b) *The function a is nondecreasing and F is obtained from G by truncation at some point u : i.e.,*

$$\begin{aligned} F(x) &= G(x) && \text{for } x < u \\ F(x) &= 1 && \text{for } x \geq u . \end{aligned}$$

(c) *F and G are obtained from a distribution H by truncation at points $u < v$, respectively.*

PROOF. (a) Since $dG^*(x)/dF^*(x)$ is proportional to $dG(x)/dF(x)$, it is nondecreasing, and hence G^* is stochastically larger than F^* .

(b) For any $t < u$, since $F(x) = G(x)$ for $x < u$,

$$\frac{\int_0^t a(x) dG(x)}{\int_0^\infty a(x) dG(x)} = \frac{\int_0^t a(x) dF(x)}{\int_0^\infty a(x) dG(x)} .$$

Since $a(x)$ is nondecreasing, $\int_0^\infty a(x) dG(x) \geq \int_0^\infty a(x) dF(x)$, and hence $G^*(t) \leq F^*(t)$ for $t < u$.

For any $t \geq u$, we have

$$\frac{\int_0^t a(x) dG(x)}{\int_0^\infty a(x) dG(x)} \leq 1 = \frac{\int_0^t a(x) dF(x)}{\int_0^\infty a(x) dF(x)}$$

and this completes the proof.

(c) This follows from (b) since F can be obtained from G through truncation.

PROOF OF THEOREM 6. Without loss of generality suppose that F and G are symmetric about 0, and let

$$(5.13) \quad \tilde{F}(x) = 2F(x) - 1, \quad \tilde{G}(x) = 2G(x) - 1, \quad 0 < x < \infty .$$

If in (5.5) and (5.6) we substitute $dK_1 = dt$ and

$$\begin{aligned} dK_2 &= dt/(1 - 2\alpha), \quad \alpha \leq t \leq 1 - \alpha \\ &= 0 \quad \text{otherwise,} \end{aligned}$$

we obtain for the efficiency $e_\alpha(F)$

$$(5.14) \quad e_\alpha(F) = (1 - 2\alpha)^2 \int_0^1 [\tilde{F}^{-1}(t)]^2 d\Lambda_2(t) / \int_0^1 [\tilde{F}^{-1}(t)]^2 d\Lambda_1(t) ,$$

where Λ_2 and Λ_1 are respectively the uniform distribution on $(0, 1)$ and the uniform distribution truncated at $1 - 2\alpha$. The result now follows from Lemma 2(b) with

$$\begin{aligned} \Lambda_1 &\text{ in place of } F; & \Lambda_2 &\text{ in place of } G; \\ b(t) &= [\tilde{G}^{-1}(t)]^2, & a(t) &= [\tilde{F}^{-1}(t)]^2 . \end{aligned}$$

⁴ This part of the lemma is given in Goldstein (1973).

To apply the lemma, we need only check that

$$\frac{b(t)}{a(t)} = \left[\frac{\tilde{G}^{-1}(t)}{\tilde{F}^{-1}(t)} \right]^2 \text{ is nondecreasing for } t > 0,$$

which is assured by the assumptions of the theorem; and that $a(t) = [\tilde{F}^{-1}(t)]^2$ is nondecreasing for $t > 0$, which is obvious. Finally, Λ_1 is obtained from Λ_2 through truncation, and this completes the proof for $\alpha < \frac{1}{2}$. For the case $\alpha = \frac{1}{2}$, the result follows by letting $\alpha \rightarrow \frac{1}{2}$ in the inequality (5.10).

A number of interesting examples of this theorem are obtained by considering the families $\mathcal{G}(F)$ of distribution G obtained as scale mixtures from a given distribution F , which is symmetric about 0,

$$(5.15) \quad G(x) = \int_0^\infty F\left(\frac{x}{\sigma}\right) d\nu(\sigma)$$

where ν is an arbitrary probability distribution on $(0, \infty)$. It will be proved in a later paper of this series that F and any $G \in \mathcal{G}(F)$ satisfy (5.9), and hence that

$$(5.16) \quad e_\alpha(G) \geq e_\alpha(F) \quad \text{for all } G \in \mathcal{G}(F),$$

provided the one-parameter family of distributions $F(\sigma x)$, $0 < \sigma < \infty$, have densities with monotone likelihood ratio. This condition holds, for example, when F is normal, a mixture of two normals with the same mean, or when it is double exponential. For the first of these cases, the family $\mathcal{G}(F)$ has been shown in Andrews et al. (1972) to be a very wide class of heavy-tailed distributions. It was recently pointed out by Stigler (1973) that this family was in fact proposed as representing distributions with tails heavier than the normal by Newcomb (1882 and particularly 1886).

The following table shows the lower bound $e_\alpha(F)$ for $\alpha = .05, .1, .15, .25, .5$ and the following four distributions:

- (1) The standard normal distribution $N(0, 1)$;
- (2a) and (2b) The contaminated normal distribution (Tukey model)
 $(1 - \varepsilon)N(0, 1) + \varepsilon N(0, \sigma^2) \quad \text{for } \varepsilon = .05 \text{ and } .25 \text{ and } \sigma = 3;$
- (3) The double exponential distribution.

These efficiencies are obtained from the corresponding asymptotic variances in Dachs (1972). The last column, for the sake of comparison, gives the universal lower bound $(1 - 2\alpha)^2$.

TABLE 5.1
Efficiency lower bound $e_\alpha(F)$

α	(1)	(2a)	(2b)	(3)	$(1 - 2\alpha)^2$
.05	.971	1.186	1.402	1.212	.81
.1	.943	1.197	1.622	1.342	.64
.15	.909	1.197	1.786	1.449	.49
.25	.833	1.085	1.667	1.626	.25
.5	.637	.833	1.327	2	0

From Theorem 6 it also follows easily that if F is uniform, inequality (5.10) holds whenever G has a symmetric unimodal density. This result was proved differently by Bickel (1965).

We mention finally that the bound (5.8) and Theorem 6 have obvious extensions to the comparison of two trimmed means, say \bar{X}_α and \bar{X}_β . If $\alpha > \beta$ and $e_{\alpha,\beta}(F)$ denotes the efficiency of \bar{X}_α w.r.t. \bar{X}_β we have from Theorem 5,

$$(5.17) \quad e_{\alpha,\beta}(F) \geq \left(\frac{1 - 2\alpha}{1 - 2\beta} \right)^2$$

and from Lemma 2 b that

$$(5.18) \quad e_{\alpha,\beta}(F) \leq e_{\alpha,\beta}(G)$$

whenever F and G satisfy (5.9).

Another interesting class of examples is provided by t -distributions with low degrees of freedom. Takeuchi (1973) in his review of Andrews et al. (1972) expresses the view that in the choice of estimator "at least four cases should be distinguished": (i) from the normal to the t -distribution with four to five degrees of freedom; (ii) to the t with two degrees of freedom and grossly contaminated normals; (iii) to the Cauchy; and (iv) to beyond the Cauchy. For the Cauchy and t with 2 degrees of freedom the efficiency $e_\alpha(F)$ are of course infinite. The following table gives the values of $e_\alpha(F)$ for $\alpha = .05$ when F is the t -distribution with $\nu = 3, 4$ or 5 degrees of freedom:

TABLE 5.2
 $e_\alpha(F)$ for t -distribution with ν d.f.

ν	3	4	5	∞
$e_\alpha(F)$	1.70	1.32	1.20	.97

It follows from our earlier results and the work of Van Zwet (1964) that if F and G are t -distributions with $\nu > \nu'$ degrees of freedom respectively, then (5.16) will hold. Takeuchi's remarks and Tables 5.1 and 5.2 suggest that a 5% trimmed mean seems to be a highly satisfying competitor to the mean.

5 B. R -estimators. Consider next the estimators obtained from (3.9) by replacing the true distribution function F in $\mu(F)$ by the empirical distribution function \hat{F}_n . Such an estimator may not be defined since the equation (3.9) may not have a solution. However, we can always define it as the median of $\frac{1}{2}[F_n^{-1}(U) + F_n^{-1}(V)]$ where (U, V) has the cdf (4.1). The resulting estimators form the class known as R -estimators, which can be derived from rank tests and are described, for example, in Section 4.3 of Huber's survey paper (1972). The functions L' of (3.7) and J of Huber's paper are related through the equation $L'(u) = J[\frac{1}{2}(u + 1)]$. In particular, if $L(y) = |y|$, the estimator reduces to $\hat{\mu}_2$, for $L(y) = y^2$, it reduces to $\hat{\mu}_4 = \text{med}_{i \leq j} [(X_i + X_j)/2]$, while for $L(y) = \Phi^{-1}[\frac{1}{2}(y + 1)]$ it reduces to the estimator based on Normal Scores.

The attraction of these latter two estimators for estimating the center of symmetry of a symmetric distribution, stems from the fact that their asymptotic efficiency relative to \bar{X} is respectively $\geq .864$ and ≥ 1 for all symmetric F . As was pointed out by Høyland (1965), the first of these inequalities breaks down completely when F is not restricted to be symmetric. In that case, there exist distributions for which the asymptotic efficiency of $\hat{\mu}_4$ relative to \bar{X} is zero.

We shall now prove

THEOREM 7. *For any R-estimator there exist distributions F such that the asymptotic efficiency of the R-estimator relative to \bar{X} is zero.*

The discrepancy of this result with those in the symmetric case is perhaps not too surprising since the signed-rank tests from which these estimators derive are tests of symmetry with respect to a given point.

PROOF. Recall (for example from Huber (1972)) that the variance of the R-estimator corresponding to L is, under suitable regularity conditions,

$$(5.19) \quad \sigma^2(\int U'(x)f(x) dx)^2/\text{Var} [U(X)]$$

where X is distributed according to F and where

$$(5.20) \quad U(x) = \int^x L''[F(y) - F(2\theta - y)]f(2\theta - y) dy,$$

θ being the solution of (3.9). We shall determine F in such a way that

$$(5.21) \quad \int L'[F(x) - F(-x)]f(x) dx = 0$$

and

$$(5.22) \quad \int U'(x)f(x) dx = 0,$$

while $\text{Var} [U(X)] > 0$.

If (5.21) holds, (5.22) becomes

$$(5.23) \quad \int L''[F(x) - F(-x)]f(x)f(-x) dx = 0,$$

and this is satisfied whenever $f(x)f(-x) = 0$ for all x . For example, we can take (as Høyland did)

$$(5.24) \quad \begin{aligned} f(x) &= p && \text{if } -1 \leq x \leq 0, \\ &= q && \text{if } 1 \leq x \leq 2; \quad p + q = 1, \end{aligned}$$

and (5.23) will be satisfied.

To show that there exist p for which (5.21) holds, note that (5.21) states that the solution θ of (3.7) is zero. Now for $p = 1$, this solution must lie between -1 and 0 , and for $p = 0$ it must lie between 1 and 2 (by Theorem 1, part 2). The existence of the required p then follows by continuity, and this completes the proof.

When F is not restricted to be symmetric, the asymptotic efficiency of, for example, $\hat{\mu}_2$ or $\hat{\mu}_4$ relative to \bar{X} , can take on any value between 0 and ∞ . It may be argued that this result is not very meaningful since distributions such as (5.24)

are not likely to obtain in practice. What can be said if F is restricted to some more reasonable class, say unimodal distribution? We shall now show that for $\hat{\mu}_2$ the lower bound is then positive; and this is also true for $\hat{\mu}_4$ if F is restricted to be strongly unimodal.

The asymptotic efficiency of $\hat{\mu}_2$ relative to $\hat{\mu}_1 = \bar{X}$ is

$$(5.25) \quad e_{2,1} = 4\sigma^2 f^2[F^{-1}(\frac{1}{2})].$$

Without loss of generality suppose that $F^{-1}(\frac{1}{2}) = 0$ and that the mode of F is ≤ 0 so that $f(x)$ is nonincreasing for all $x \geq 0$. Then

$$e_{2,1} \geq 4f^2(0) \int_0^\infty (x - \mu)^2 f(x) dx,$$

where μ is the expectation of F . Let g be the probability density defined by

$$\begin{aligned} g(x) &= 2f(x) && \text{for } x \geq 0 \\ &= 0 && \text{for } x < 0. \end{aligned}$$

Denoting the mean of g by ν , we have

$$(5.26) \quad e_{2,1} \geq \frac{1}{2}g^2(0) \int_0^\infty (x - \nu)^2 g(x) dx.$$

We shall now prove that the right hand side of (5.26) is minimized, among all densities on $(0, \infty)$ which are nonincreasing, when g is the density of a uniform distribution whose left hand end point is zero. This is a consequence of the following lemma.

LEMMA 3. *Subject to $\int_0^\infty xg(x)dx = \nu$, if $g(x) = 0$ for $x < 0$ and is nonincreasing for $x > 0$, the integral $\int_0^\infty x^2g(x) dx$ (and hence the variance of g) is minimized when $g(x) = 1/2\nu$ for $0 < x < 2\nu$.*

The proof of this lemma is completely analogous to the proof of the second part of Theorem 8 below, and we shall therefore omit it here.

To prove the desired result note that for fixed ν , the uniform distribution also minimizes $g(0)$. The right hand side of (5.26) is of course scale invariant, and hence is minimized for the uniform distribution over any interval $(0, a)$. The associated minimum value of (5.26) is $\frac{1}{24}$.

The lower bound $\frac{1}{24}$ is presumably not sharp. To see what values can be obtained, consider the density

$$(5.27) \quad \begin{aligned} f(x) &= \frac{1}{2}a && \text{if } -a \leq x < 0 \\ &= 1 && \text{if } 0 \leq x < \frac{1}{2}. \end{aligned}$$

Formula (5.25) is not applicable because of the discontinuity of f at zero. However, f can be approximated by a sequence of densities, continuous in $(-a, \frac{1}{2})$ and with $f(0) = 1$. Letting $a \rightarrow 0$, it is then seen that $e_{2,1} \rightarrow \frac{5}{8}$, so that the sharp lower bound lies between $\frac{2}{4^8} = .057$ and $\frac{5}{4^8} = .104$.

Much higher values for the lower bound are obtained if F is restricted to the subfamilies (5.15) considered in Section 5 A. That (5.16) holds for $\hat{\mu}_2$ when F and G satisfy (5.9) and hence for the four examples considered in Table 5.1 is

in fact the case $\alpha = \frac{1}{2}$ of Theorem 6. The values of the efficiency $e_{2,1}$ of $\hat{\mu}_2$ relative to \bar{X} for these four distributions F (and hence the lower bounds for the associated (F)) are shown in the last row of Table 5.1.

Consider next $\hat{\mu}_4$. The asymptotic efficiency $e_{4,1}$ of this estimator relative to \bar{X} is given by (see again Andrews et al. (1972))

$$(5.28) \quad e_{4,1} = \sigma^2(\int f(-x)f(x) dx)^2 / \int [\frac{1}{2} - F(-x)]^2 f(x) dx$$

provided $\mu_4(F) = 0$. Now $\int f(-x)f(x) dx$ is the density of $X + X'$ at 0, where X, X' are independent identically distributed with density f . If the density of $X + X'$ is unimodal, since the denominator of (5.28) is bounded above by $\frac{1}{4}$, we can apply the preceding result to get the crude lower bound .021. This bound holds whenever the density of $X + X'$ is unimodal and hence by a well-known result of Ibragimov (1956) whenever f is strongly unimodal.

Although crude, the bound is of the right order of magnitude. To see this, consider the following example. Let

$$(5.29) \quad \begin{aligned} f(x) &= q/a & \text{if } -a < x < 0 \\ &= p/b & \text{if } 0 < x < b, \quad p + q = 1. \end{aligned}$$

It is easily computed that the pseudo median of F is zero when $q \rightarrow \frac{1}{2}2^{\frac{1}{2}}$ as $a \rightarrow 0$, and that then $e_{4,1} \rightarrow .071$.

We shall finally show that (5.16) holds when e_α is replaced by $e_{4,1}$, the efficiency of $\hat{\mu}_4$ relative of \bar{X} when F is normal. If F is a distribution symmetric with respect to 0 and if $F * F$ denotes the convolution of F with itself, the efficiency $e_{4,1}$ can be written as

$$(5.30) \quad e_{4,1}(F) = 6[(F * F)'(0)]^2 \int x^2 d(F * F)(x).$$

Consider now the class $\mathcal{S}(\Phi)$ defined by (5.15) with F as the standard normal distribution Φ . Then when $F \in \mathcal{S}(\Phi)$, also $F * F \in \mathcal{S}(\Phi)$ and hence

$$(5.31) \quad \inf_{F \in \mathcal{S}(\Phi)} [e_{4,1}(F)] \geq \inf_{F \in \mathcal{S}(\Phi)} 6[f(0)]^2 \int x^2 dF(x).$$

However, except for the constant the expression being minimized on the right hand side of (5.31) is just the minimum efficiency of $e_{2,1}(F)$ over $\mathcal{S}(\Phi)$, and this minimum was shown earlier to be attained at $F = \Phi$. It thus follows that

$$\inf_{F \in \mathcal{S}(\Phi)} [e_{4,1}(F)] = 6[\varphi(0)]^2 = \frac{3}{\pi} = .955.$$

5C. M -estimators. Consider finally the class of “(M)” location parameters characterized by Theorem 2. Let $\mu_\alpha(F)$ minimize $\int_{-\infty}^{\infty} |t - \mu|^{a+1} dF(x)$ where $0 < \alpha \leq 1$. (Note that μ_1 here is the mean). Let T_α be the estimator obtained by replacing F by the empirical cdf in μ_α . The asymptotic variance of $n^{1/2}(T_\alpha - \mu_\alpha(F))$ is then under some regularity conditions given by (cf. Huber (1964))

$$(5.32) \quad \sigma_{\alpha^2}(F) = \frac{\int_{-\infty}^{\infty} \phi_\alpha^2(t - \mu_\alpha(F)) dF(t)}{(\int_{-\infty}^{\infty} \phi_\alpha'(t - \mu_\alpha(F)) dF(t))^2} \quad \text{for } \alpha > 0,$$

where ϕ_α is given in (3.12).

The following result is the analogue of the lower bound for the efficiency of trimmed means.

THEOREM 8. If $0 < \alpha \leq 1$,

$$(5.33) \quad \sigma_\alpha^2(F) \leq \sigma_1^2(F)/\alpha^2;$$

that is the efficiency of T_α to \bar{X} is at least α^2 whatever be the underlying distribution for which the representation (5.32) of the asymptotic variance is valid.

If F is known to be unimodal and symmetric about some point and (5.32) is valid,

$$(5.34) \quad \sigma_\alpha^2(F) \leq \frac{3}{2\alpha + 1} \sigma_1^2(F);$$

that is, the efficiency of T_α to \bar{X} for symmetric unimodal distributions is at least $(2\alpha + 1)/3$. Both bounds are sharp. They are achieved for (5.33) when F assigns mass $\frac{1}{2}$ to each of ± 1 and for (5.34) when F is the uniform distribution on an interval.

PROOF. To prove (5.33) we note that if we fix $\sigma_2^2(F) = 1$ since $2\alpha - 1 \leq \alpha \leq 1$

$$\begin{aligned} \int_{-\infty}^{\infty} \phi_\alpha^2(t - \mu_\alpha(F)) dF(t) &\leq [\int_{-\infty}^{\infty} |t - \mu_\alpha(F)|^\alpha dF(t)]^2 \leq [\int_{-\infty}^{\infty} |t - \mu_1(F)|^\alpha dF(t)]^2 \\ &\leq [\int_{-\infty}^{\infty} (t - \mu_1(F))^2 dF(t)]^\alpha = 1. \end{aligned}$$

Similarly,

$$\begin{aligned} \int_{-\infty}^{\infty} \phi_\alpha'(t - \mu_\alpha(F)) dF(t) &= \alpha \int_{-\infty}^{\infty} \frac{1}{|t - \mu_\alpha(F)|^{1-\alpha}} dF(t) \\ &\geq \alpha / \int_{-\infty}^{\infty} |t - \mu_\alpha(F)|^{1-\alpha} dF(t) \\ &\geq \alpha / [\int_{-\infty}^{\infty} (t - \mu_1(F))^2 dF(t)]^{(1-\alpha)/2} = \alpha, \end{aligned}$$

and (5.33) follows. That the bound is sharp is clear. A proof of (5.34) is given by Loynes (1970) who also proved (5.33) under the restriction of symmetry.

We can generalize Loynes' result somewhat along the lines of Theorem 6. Suppose that F is symmetric about 0 and that $\mathcal{G}(F)$ is defined as in the remarks following that theorem. Then for $G \in \mathcal{G}(F)$ we have $\mu_\alpha(G) = 0$ and if G is given by (5.15)

$$\sigma_\alpha^2(G) = \frac{\int_0^\infty \phi_\alpha^2(t) d\nu(t) \int_{-\infty}^{\infty} |t|^{2\alpha} dF(t)}{[\int_0^\infty \phi_\alpha'(t) d\nu(t)]^2 [\int_{-\infty}^{\infty} |t|^{\alpha-1} dF(t)]^2}.$$

Hence,

$$\sigma_\alpha^2(G) \leq \frac{\sigma_1^2(G)}{\alpha^2} \cdot \frac{\int_{-\infty}^{\infty} |t|^{2\alpha} dF(t)}{[\int_{-\infty}^{\infty} |t|^{\alpha-1} dF(t)]^2 \int_{-\infty}^{\infty} t^2 dF(t)}.$$

More generally, if $0 < \alpha \leq \beta \leq 1$, by Lemma 1 of Loynes

$$\sigma_\alpha^2(G) \leq \sigma_\beta^2(G) \cdot \frac{\beta^2 \int_{-\infty}^{\infty} |t|^{2\alpha} dF(t) [\int_{-\infty}^{\infty} |t|^{\beta-1} dF(t)]^2}{\alpha^2 [\int_{-\infty}^{\infty} |t|^{\alpha-1} dF(t)]^2 \int_{-\infty}^{\infty} |t|^{2\beta} dF(t)}$$

with equality if and only if there exists σ such that $G(x) = F(x/\sigma)$ for all x .

If F is the uniform distribution on $(-1, 1)$, $\mathcal{G}(F)$ is the family of symmetric

unimodal distributions and we obtain Loynes' Theorem 2:

$$\sigma_{\alpha}^2(G) \leq \frac{2\alpha + 1}{2\beta + 1} \sigma_{\beta}^2(G)$$

with equality if and only if G is uniform.

If F is a normal distribution with mean zero, we find

$$\begin{aligned} \sigma_{\alpha}^2(G) &\leq \frac{\Gamma(\alpha + \frac{1}{2})}{\Gamma(\beta + \frac{1}{2})} \left[\frac{\beta\Gamma(\beta/2)}{\alpha\Gamma(\alpha/2)} \right]^2 \sigma_{\beta}^2(G) \\ &= \frac{\Gamma(\alpha + \frac{1}{2})}{\Gamma(\beta + \frac{1}{2})} \left[\frac{\Gamma(\beta/2 + 1)}{\Gamma(\alpha/2 + 1)} \right]^2 \sigma_{\beta}^2(G) \end{aligned}$$

with equality if and only if G is normal. Passage to the limit as $\alpha \rightarrow 0$ is easily seen to be justified and the results then agree with those for the efficiency of the median with respect to the mean stated after Theorem 6.

6. Conclusion. We set out in this paper to find measures of location in the asymmetric case, which would be robust and at the same time could be estimated efficiently. If by efficient we mean a guaranteed high efficiency relative to the mean, we have shown that measures estimated by R -estimators are robust but do not satisfy this efficiency criterion, while the situation is just the reverse for the rather narrow class of location measures estimated by M -estimators. However, within the class of weighted quantiles, we have found measures satisfying both criteria. In particular, this is the case for the trimmed means, which appear to constitute a simple and satisfactory solution to our problem.

REFERENCES

- ANDREWS, D. F., BICKEL, P. J., HAMPPEL, F. R., HUBER, P. J., ROGERS, W. H. and TUKEY, J. W. (1972). *Robust estimates of location*. Princeton Univ. Press.
- BAHADUR, RAGHU RAJ and SAVAGE, LEONARD J. (1956). The nonexistence of certain statistical procedures in nonparametric problems. *Ann. Math. Statist.* **27** 1115-1122.
- BARLOW, RICHARD E. and PROSCHAN, FRANK (1966). Tolerance and confidence limits for classes of distributions based on failure rate. *Ann. Math. Statist.* **37** 1593-1601.
- BICKEL, P. J. (1965). On some robust estimates of location. *Ann. Math. Statist.* **43** 847-858.
- BICKEL, P. J. (1973). On some analogues of linear combinations of order statistics in the linear model. *Ann. Statist.* **1** 597-616.
- BICKEL, P. J. and LEHMANN, E. L. (1975). Descriptive statistics for nonparametric models. I. Introduction. *Ann. Statist.* **3** 1038-1044.
- British Association for the Advancement of Science (1958). *Mathematical Tables* **6** Cambridge Univ. Press.
- CHURCHILL, RUEL V. (1958). *Operational mathematics*. 2nd ed. McGraw-Hill, New York.
- DACHS, J. N. W. (1972). Asymptotic expansions for M -estimators. Unpublished Ph.D. thesis. Univ. of California, Berkeley.
- DANIELL, P. J. (1920). Observations weighted according to order. *Amer. J. Math.* **42** 222-236.
- DOKSUM, KJELL (1969). Starshaped transformations and the power of rank tests. *Ann. Math. Statist.* **40** 1167-1176.
- HAMPPEL, F. R. (1968). Contributions to the theory of robust estimation. Ph.D. dissertation, Univ. of California, Berkeley.
- HAMPPEL, F. R. (1971). A general qualitative definition of robustness. *Ann. Math. Statist.* **42** 1887-1896.

- HØYLAND, ARNLJOT (1965). Robustness of the Hodges–Lehmann estimator for shift. *Ann. Math. Statist.* **36** 174–197.
- HUBER, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35** 73–101.
- HUBER, P. J. (1972). Robust statistics: A review. *Ann. Math. Statist.* **43** 1041–1067.
- HUBER, P. J., *Robust Statistics*. A book in preparation.
- IBRAGIMOV, J. A. (1956). On the composition of unimodal distributions (Russian). *Teor. Veroyatnost. i Primenen.* **1** 283–288.
- JAECKEL, L. A. (1971). Robust estimates of location: symmetry and asymmetric contamination. *Ann. Math. Statist.* **42** 1020–1034.
- JOHNSON, NORMAN L. and KOTZ, SAMMUEL (1970). *Continuous univariate distributions I*. Houghton and Mifflin, Boston.
- GOLDSTEIN, MATTHEW (1973). An application of ordered expectations. *Amer. Statist.* **27** (1) 30.
- KNÜSEL, L. F. (1969). Uber minimum-distance Schätzungen. Thesis, ETH, Zürich.
- LAWRENCE, MICHAEL (1966). Inequalities and tolerance limits for s -ordered distributions. Unpublished report.
- LOYNES, R. M. (1970). On the asymptotic relative efficiencies of certain location parameter estimates. *J. Roy. Statist. Soc. Ser. B.* 134–136.
- NEWCOMB, S. (1882). Discussion and results of observations on transits of Mercury from 1677 to 1881. *Astronomical papers* **1** 363–487.
- NEWCOMB, S. (1886). A generalized theory of the combination of observations so as to obtain the best result. *Amer. J. Math.* **8** 343–366.
- STIGLER, STEPHEN M. (1972). Simon Newcomb, Percy Daniell, and the history of robust estimation 1885–1920. Unpublished report of the Univ. of Wisconsin, Dept. of Statistics.
- TAKEUCHI, K. (1967). Robust estimation and robust parameter. Unpublished.
- TAKEUCHI, K. (1973). Review of robust estimates of location. Andrews et al. *Science* **179** 469.
- TUKEY, J. W. (1960). A survey of sampling from contaminated distributions. *Contributions to Probability and Statistics*. (I. Olkin, ed.) Stanford Univ. Press.
- TUKEY, J. W. (1962). The future of data analysis. *Ann. Math. Statist.* **33** 1–67.
- VAN ZWET, W. R. (1964). *Convex Transformations of Random Variables*. Math. Centrum, Amsterdam.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720