DESCRIPTIVE STATISTICS FOR NONPARAMETRIC MODELS III. DISPERSION

By P. J. Bickel and E. L. Lehmann*

ABSTRACT. Measures of dispersion are defined as functionals satisfying certain equivariance and order conditions. Attention is restricted to symmetric distributions. Different measures are compared in terms of asymptotic relative efficiency, i.e., the inverse ratio of their standardized variances. The efficiency of a trimmed to the untrimmed standard deviation turns out not to have a positive lower bound even over the family of Tukey models. Positive lower bounds for the efficiency (over the family of all symmetric distributions for which the measures are defined) exist if the trimmed standard deviations are replaced by pth power deviations. However, these latter measures are no longer robust, although for $p < 2$ they are more robust than the standard deviation. The results of the paper suggest that a positive bound to the efficiency may be incompatible with robustness but that trimmed standard deviations and pth power deviations for $p = 1$ or 1.5 are quite satisfactory in practice.

Key words and phrases. Dispersion, estimation, standardized asymptotic variance, asymptotic efficiency, Tukey model, standard deviation, trimmed standard deviation, pth power deviation, substitution of an estimate of location, scaling a positive distribution.

*This paper was prepared with support of U.S. Office of Naval Research Contract No. N00014-69-A-0200-1033 (NR042-036) and National Science Foundation Grant NSF-GP-38485.

University of California
Department of Statistics
Statistical Laboratory
Berkeley, California 94720

1. Measures of dispersion

In analogy with the definition of a measure of location, we shall define a measure of dispersion to be a functional (defined over a sufficiently large family of distributions) which satisfies certain invariance conditions and which in addition has the property of assigning a larger value to G than to F if G is more dispersed than F. In the present paper we shall consider the problem for symmetric distributions* and assume that X is a random variable whose distribution F is symmetric about μ. It then seems natural to interpret dispersion in terms of the distance of X from μ, that is, in terms of the magnitude of $|X-\mu|$, and to consider Y as more dispersed about ν than X about μ if

(1.1) $|Y-\nu|$ is stochastically larger than $|X-\mu|$.

(This is essentially the "peakedness"-ordering introduced by Z. W. Birnbaum (1948).)

Note that

(a) any symmetric random variable is more dispersed than a constant;

(b) aX is more dispersed than X if $a > 1$.

If F and G are symmetric about 0 with densities f and g, a simple sufficient condition for (1.1) with $\mu = \nu = 0$

---

*We expect to take up the asymmetric case in a subsequent paper.

(1.2) $g(x)/f(x)$ is increasing for $x > 0$.

If F and G are symmetric about zero, and G is more dispersed than F, and if

(1.3) $$H_\theta(x) = \theta G(x) + (1-\theta)F(x),$$

then H_θ is more dispersed than F for any $0 < \theta < 1$. As an illustration, note that a standard normal distribution contaminated with another normal distribution with zero mean and variance > 1 (Tukey model) is more dispersed than the uncontaminated standard normal distribution.

An important class of examples is provided by the following result, which is a generalization of a lemma of Birnbaum (1948).

Theorem 1. Let X_i, Y_i $(i=1,2)$ be independent with distributions F_i, G_i $(i=1,2)$ which are symmetric about zero, and suppose that

(i) Y_i is more dispersed than X_i for $i=1,2$

and

(ii) F_1 and G_2 have unimodal densities and possibly some probability mass at zero.

Then $Y_1 + Y_2$ is more dispersed than $X_1 + X_2$.

Proof. Consider the probability

$$P(|X_1 + X_2| < c) = 2\int_0^\infty [F_1(x+c) - F_1(x-c)]\,dF_2(x).$$

Letter on file

The unimodality of F_1 implies that the integrand on the right-hand side is a decreasing function of x. From the fact that (F_2, G_2) satisfies (1.1), it then follows that this last integral is decreased when F_2 is replaced by G_2. Thus,

$$P(|X_1 + X_2| \leq c) \geq 2 \int_0^\infty [F_1(x+c) - F_1(x-c)]\, dG_2(x)$$

$$= 2 \int_0^\infty [G_2(x+c) - G_2(x-c)]\, dF_1(x).$$

Repeating the argument (this time using the unimodality of G_2), we arrive at the desired result. Birnbaum has shown that Theorem 1 no longer holds when assumption (ii) is dropped.

Consider now a functional $\tau(F)$ [also denoted by $\tau(X)$ when X is a random variable with distribution F] defined over a sufficiently large class of distributions which is closed under changes of location and scale. We shall require τ to be nonnegative and to satisfy

$$\tau(aX) = |a|\tau(X) \quad \text{for} \quad a > 0 \tag{1.4}$$

and

$$\tau(X + b) = \tau(X) \quad \text{for all} \quad b. \tag{1.5}$$

It follows from (1.5) and the symmetry of F that

$$\tau(-X) = \tau(X) \tag{1.6}$$

so that (1.4) holds for all $a \neq 0$.

From (1.4) and (1.5) it is easily seen that

$$\tau(c) = 0 \quad \text{for any constant} \quad c. \tag{1.7}$$

For by (1.4), we have $\tau(0) = \tau(2\times 0) = 2\tau(0)$ and hence $\tau(0) = 0$, and by (1.5), $\tau(c) = \tau(0)$. The converse, that $\tau(X) = 0$ requires X to be a constant with probability 1, will in general not hold. An example is provided by the trimmed standard deviation defined in Section 3 below.

A nonnegative functional τ satisfying (1.4) and (1.5) will be called a measure of dispersion if it satisfies in addition

(1.9) $\tau(F) \leqq \tau(G)$ whenever G is more dispersed than F.

Note that if $\tau(F)$ is a measure of dispersion, so is $k\tau(F)$ for any $k > 0$.

A large and important class of dispersion measures is provided by the functionals

$$\tau(F) = \left[\int_0^1 [F_*^{-1}(t)]^\gamma \, d\Lambda(t)\right]^{1/\gamma} \tag{1.10}$$

where F is assumed to be symmetric about μ, F_* denotes the distribution of $|X - \mu|$, Λ is any probability distribution on $(0,1)$ and γ any positive number.

That (1.10) satisfies (1.4) and (1.5) is easily checked; that it satisfies (1.9) follows from the fact that $F_*^{-1}(t) \leqq G_*^{-1}(t)$ for all t when G_* is stochastically larger than F_*.

A special case of (1.10) is the standard deviation (SD) of F defined as

$$SD(F) = \left[\int (x - \mu)^2 dF(x)\right]^{1/2} . \tag{1.11}$$

This is easily seen to be given by (1.10) with $\gamma = 2$ and Λ the uniform distribution on $(0,1)$. The following three important classes of measures, all special cases of (1.10) provide the alternatives to the standard deviation with which we shall be concerned.

(i) A generalization $\tau^2(F; p)$ of the standard deviation is the pth power deviation obtained by replacing γ by p in (1.10) and letting Λ be the uniform distribution on $(0,1)$.

(ii) The doubly trimmed standard deviation $\tau(F; \alpha, \beta)$ is given by (1.10) with $\gamma = 2$ and Λ the uniform distribution on $(\alpha, 1 - \beta)$. The most important example of this is the case $\alpha = 0$.

(iii) The αth quantile is obtained from (1.10) by letting Λ assign probability 1 to the point α. The resulting measure is independent of γ.

The standard deviation is of course a member of both (i) and (ii). The αth quantile is the limit of the doubly trimmed standard deviation as $\beta \to \alpha$.

2. Estimation

A most important aspect in comparing two measures of scale $\tau_i(F)$ and $\tau_j(F)$ is the accuracy with which they can be estimated. Unfortunately, it is no longer possible to compare these accuracies directly in terms of the asymptotic variances of the estimators. This is clearly seen by considering the case $\tau_2 = c\tau_1$ where c is any positive constant. If τ_1 is a possible measure of scale and its estimator is $\hat{\sigma}_1$, one would be equally happy to use τ_2

and estimate it by $\delta_2 = c\delta_1$; of course the asymptotic variance does not remain the same but gets multiplied by c^2. For this reason a natural measure of accuracy of an estimator of $\tau(F)$ with asymptotic variance $v^2(F)$ is not $v^2(F)$ itself but the scale invariant standardized asymptotic variance (already proposed by Daniell (1920))

$$v^2(F)/\tau^2(F). \tag{2.1}$$

The asymptotic efficiency $e_{2,1}$ of δ_2 (estimating τ_2) to δ_1 (estimating τ_1) will then be defined as

$$e_{2,1}(F) = \frac{v_1^2(F)}{\tau_1^2(F)} \Big/ \frac{v_2^2(F)}{\tau_2^2(F)} . \tag{2.2}$$

If $\sqrt{n_i}\,(T_i - \tau_i)$ is asymptotically normally distributed for $i=1,2$ as the number n_i of observations tends to infinity, the usual argument shows that the asymptotic efficiency (2.2) is the limiting ratio of the numbers of observations required by the two estimators to achieve the same standardized variance.

That the above definitions are reasonable can be seen from another point of view. The logarithm of δ is an estimator of the location parameter $\log \tau(F)$. Suppose that the distribution of $\sqrt{n}\,(\delta - \tau)$ tends to the normal distribution with zero mean and variance v^2. Then the distribution of $\sqrt{n}\,(\log \delta - \log \tau)$ tends to the normal distribution with mean zero and variance $v^2/\tau^2(F)$; that is, $v^2/\tau^2(F)$ is the asymptotic variance of the location estimate $\log \delta$.

When studying the estimators of functionals such as those defined in Section 1, it is convenient first to consider F restricted to distributions which are symmetric about a known point μ. On the other hand, in order even to define the estimators of τ we wish to study, it is necessary to extend τ to asymmetric distributions. In all the examples to be considered here, there is a natural extension of τ to asymmetric distributions. Given this extension, we define as estimator of $\tau(F)$ the functional τ evaluated at the empirical distribution function $\hat{F}$ of $X_1,\cdots,X_n$. In what follows, we shall assume without loss of generality that the known value μ of the center of symmetry of F is $\mu = 0$.

The point of view in the present paper will be that the standard measure of scale is the standard deviation $\tau(F; 0, 0)$ which (since $\mu = 0$) is estimated by

$$\delta_1 = \sqrt{\sum_{i=1}^{n} X_i^2/n} \,. \tag{2.3}$$

This estimator is well known to be very unsatisfactory because of its extreme sensitivitt to outlying observations. We shall therefore look at the other functionals under consideration as competitors of τ_1 and hence shall be interested principally in comparing their behavior with that of δ_1. Unfortunately, we are only able to make these comparisons asymptotically. For $\Sigma X_i^2/n$, it is of course obvious from the central limit theorem that $\sqrt{n}[\Sigma X_i^2/n - \sigma^2]$ is asymptotically normal with zero mean and variance $\mathrm{Var}(X^2)$ provided

the later variance is finite. Here $\sigma^2 = E(X^2)$ denotes the variance of X.

It follows that $\sqrt{n}(\sigma_1 - \sigma)$ is asymptotically normal with mean zero and variance $\mathrm{Var}(X^2)/4\sigma^2$. Thus the standardized variance is

$$\frac{v_1^2(F)}{\tau_1^2(F)} = \frac{\mathrm{Var}(X^2)}{4\sigma^4} = \frac{1}{4}\left[\frac{E(X^4)}{\sigma^4} - 1\right]. \tag{2.4}$$

In the next sections we shall obtain the corresponding expansion for the estimators of some other functionals, and then study the efficiencies (2.2).

3. The doubly trimmed standard deviation

To replace the standard deviation as a measure of scale, we shall seek among the measures discussed in Section 1 one which would be more robust but which still can be estimated fairly efficiently. Encouraged by the results of BLII, we shall begin this search by studying the Trimmed standard deviations.

As before, let the random variables X_i be independently distributed according to a distribution F, which is symmetric with respect to the origin. Let $Y = X^2$ and denote the distribution of Y by G. We can then write $\tau^2(F; \alpha, \beta)$ (with $\alpha, \beta \leq 1/2$) as

$$\tau^2(F; \alpha, \beta) = \frac{1}{1-\alpha-\beta}\int_\alpha^{1-\beta} G^{-1}(t)\,dt = \int_{u_\alpha}^{u_{1-\beta}} y\,dG(y) \tag{3.1}$$

where u_α is the αth percentile of G.

Consider now the estimator,

$$\tau^2(\hat{F}, \alpha, \beta) \overset{\text{def.}}{=} \hat{\tau}^2(\alpha, \beta). \tag{3.2}$$

This is obtained by trimming off the $100\ \alpha\%$ observations with the smallest absolute values and the $100\ \beta\%$ observations with the largest absolute values and then computing the standard deviation of the remaining observations. In terms of the Y's, $\hat{\tau}^2$ is a doubly trimmed mean. Its expectation is given by (3.1) and its asymptotic variance is (see for example BL II (1975))

$$(3.3) \qquad v^2_{\alpha,\beta} = \frac{1}{(1-\alpha-\beta)^2} \int_{u_\alpha}^{u_{1-\beta}} [y - c(\alpha,\beta)]^2 dG(y) + \alpha[u_\alpha - c(\alpha,\beta)]^2 + \beta[u_{1-\beta} - c(\alpha,\beta)]^2$$

where

$$(3.4) \qquad C(\alpha,\beta) = \int_{u_\alpha}^{u_{1-\beta}} y dG(y) + \alpha u_\alpha + \beta u_{1-\beta}$$

and

$$(3.5) \qquad u_t = G^{-1}(t) = \left[F^{-1}\ \frac{t+1}{2}\right]^2 .$$

(This formula holds if u_α, $u_{1-\beta}$ are uniquely defined and G is continuous at u_α, $u_{1-\beta}$ (see Stigler (1973)).

To get an idea of the behavior of such procedures with respect to the untrimmed standard deviation we consider some representative cases namely the (singly) trimmed standard deviations with $\beta = .1, .2, \ \alpha = 0$. The following tables computed by Winston Chow and W. Carmichael show the efficiencies (2.2) of these estimators with respect to the standard deviation for the following three classes of distributions.

(a) t-distributions with 5, 6, 7, 8, 9, 10, 25, 50 degrees of freedom.

(b) Symmetric beta distributions with densities,

$$f(x) = [B(r,r)]^{-1}\left(\frac{1}{4} - t^2\right)^{r-1}, \quad |t| \leq \frac{1}{2} \tag{3.6}$$

for $r = \frac{1}{2}, 2, 4, 6, 8, 10, 20, 30$.

(c) Tukey normal gross error distributions, i.e.,

$$F(x) = (1-\varepsilon)\Phi(x) + \varepsilon\Phi\left(\frac{x}{\lambda}\right), \tag{3.7}$$

for $\varepsilon = .025, .075, .10, .20, .40, .50$ and $\lambda = 2, 4, 6$.

These models were selected as representing a range of long and short tailed distributions and for ease of computation.

The figures suggest that, on the whole, for heavy-tailed distributions both trimmed SD's are better than the untrimmed SD and that for the ranges considered $\beta = .1$ is preferable to $\beta = .2$. For light-tailed distributions such as the Beta-distribution, the untrimmed SD does best, $\beta = .1$ does better than $\beta = .2$, and $\beta = .1$ performs reasonably well.

f	5	6	7	8	9	10	25	50	∞
β = .1	2.35	1.56	1.29	1.16	1.08	1.03	.85	.81	.78
β = .2	2.11	1.36	1.11	.99	.92	.87	.69	.66	.63

Table 3.1: Asymptotic efficiency of α-trimmed with respect to untrimmed SD: t-distribution with f degrees of freedom

r	.5	1.0	2.0	4.0	6.0	8.0	10.0	0.0	30.0	∞
β = .1	1.30	.86	.72	.70	.71	.72	.73	.75	.76	.78
β = .2	.83	.57	.52	.54	.55	.57	.57	.60	.61	.63

Table 3.2: Asymptotic efficiency of α-trimmed with respect to untrimmed SD: Beta distribution with density (3.6)

	β = .1			β = .2		
ε \ τ	2	4	6	2	4	6
0	.78	.78	.78	.63	.63	.63
.025	.98	3.97	10.04	.80	3.27	8.36
.075	1.19	4.01	6.48	.99	3.57	6.03
.10	1.23	3.48	4.79	1.03	3.27	4.93
.20	1.21	1.46	.76	1.05	2.04	2.26
.40	.96	.63	.50	.87	.68	.38
.50	.87	.63	.55	.79	.52	.36

Table 3.3: Asymptotic efficiency of α-trimmed with respect to untrimmed SD: Tukey model (3.7)

A surprising feature of Table 3.3 are the extremely high efficiency values for small $\varepsilon > 0$ and large τ. These seem to arise in cases where there is enough trimming to insure with very high probability that only a small proportion of the gross errors is retained in the trimmed sample. The standardized variance of the untrimmed SD then rises very sharply with τ while that of the trimmed SD is affected only little as τ gets large. The curious

"dip" which occurs in the neighborhood of $\varepsilon = \beta$ is predicted by the asymptotic theory of the next section.

4. Nonexistence of a lower bound

The numerical results of the preceding section are encouraging and raise the hope that, as in the case of the trimmed means as measures of location, the efficiencies of the trimmed to the untrimmed SD have a positive lower bound. Unfortunately, this turns out not to be the case even if attention is restricted to unimodal distributions. In fact, even within the class of Tukey models (3.7), the efficiencies can take on arbitrarily small values.

Theorem 2. Let $e(\beta; \varepsilon, \lambda)$ denote the asymptotic efficiency of the trimmed standard deviation $\hat{\tau}(0, \beta)$ relative to the untrimmed standard deviation, in the Tukey model (3.7). Then

$$\lim_{\lambda \to \infty} e(\beta; \beta, \lambda) = 0. \tag{4.1}$$

The proof of this result is based on the following two lemmas.

Lemma 1. For the Tukey model (3.7), the standardized asymptotic variance of the SD

$$\frac{1}{4}\left[\frac{E(X^4)}{E^2(X^2)} - 1\right] \tag{4.2}$$

is bounded above uniformly in λ for any fixed ε.

Proof. For the model (3.7), the standardized variance (3.9) becomes

$$\frac{1}{4}\left(\frac{3[(1 - \varepsilon) + \varepsilon\lambda^4]}{[(1 - \varepsilon) + \varepsilon\lambda^2]^2} - 1\right) .$$

Since this is continuous in λ we need only consider its behavior as $\lambda \to \infty$ when it is clearly bounded.

Lemma 2. If $\varepsilon = \beta$ in (3.7) and if $u_{1-\beta} = u_{1-\beta}^{(\lambda)}$ is given by (3.5) then as a function of λ,

$$u_{1-\beta}^{(\lambda)} \sim 2 \log \lambda \quad \text{as} \quad \lambda \to \infty .$$

Proof. Let

$$x(\lambda) = \sqrt{u_{1-\beta}^{(\lambda)}} .$$

Then $x(\lambda)$ satisfies,

$$(4.3) \qquad (1-\beta)\Phi(x) + \beta\Phi\left(\frac{x}{\lambda}\right) = 1 - \frac{\beta}{2} .$$

First note that as $\lambda \to \infty$ we must have,

$$(4.4) \qquad x(\lambda) \to \infty$$

and

$$(4.5) \qquad \frac{x(\lambda)}{\lambda} \to 0.$$

For if (4.4) does not hold there exists a sequence λ_n such that,

$$x(\lambda_n) \to c < \infty.$$

Then,

$$(1-\beta)\Phi(x(\lambda_n)) + \beta\Phi\left(\frac{x(\lambda_n)}{\lambda_n}\right) \to (1-\beta)\Phi(c) + \frac{\beta}{2} < 1 - \alpha$$

and this contradicts (4.3). Similarly if (4.4) holds but (4.5) is violated we could find a sequence $\{\lambda_n'\}$ such that,

$$\lim_n \ (1-\beta)\Phi(x(\lambda_n')) \ + \beta\Phi\left(\frac{x(\lambda_n')}{\lambda_n'}\right) > 1 - \beta,$$

and we would again have a contradiction.

Now, by (3.12) and (3.13),

$$1 - \Phi[x(\lambda)] = \frac{\varphi[x(\lambda)]}{x(\lambda)} \ [1 + o(1)] \tag{4.6}$$

and

$$\Phi\left(\frac{x(\lambda)}{x}\right) - \frac{1}{2} = \frac{\varphi(o)x(\lambda)}{\lambda} \ [1 + o(1)] \ . \tag{4.7}$$

Substituting (4.6) and (4.7) in (4.3), cancelling, and then taking logs we obtain,

$$\frac{-x^2(\lambda)}{2} - \log x(\lambda) = \log x(\lambda) - \log \lambda + \log \frac{\beta}{1-\beta} + o(1). \tag{4.8}$$

By (4.4)

$$\log x(\lambda) = o(x^2(\lambda))$$

and we conclude that

$$x^2(\lambda)(1 + o(1)) = 2 \log \lambda \tag{4.8}$$

which was to be proved. □

Proof of Theorem 2. In view of Lemma 1 we only need to show that,

$$v_{0,\beta}^2 \to \infty \tag{4.9}$$

while

$$\tau^2(F;\ 0,\ \beta) = 0(1) \quad \text{as} \quad \lambda \to \infty. \tag{4.10}$$

To show (4.9) we calculate,

$$\begin{aligned}(1-\beta)\tau^2(F;\ 0,\ \beta) &= (1-\beta)\int_{-x(\lambda)}^{x(\lambda)} x^2\varphi(x)\,dx + \frac{\beta}{\lambda}\int_{-x(\lambda)}^{x(\lambda)} x^2\varphi(\tfrac{x}{\lambda})\,dx \\ &\leq (1-\beta) + 2\beta\lambda^2\int_0^{x(\lambda)/\lambda} y^2\varphi(y)\,dy \\ &\leq (1-\beta) + \frac{2\beta\varphi(0)}{3}\,\frac{x^3(\lambda)}{\lambda}\end{aligned}$$

and the result follows. This argument and Lemma 2 show that

$$C(0,\ \beta) = (1-\beta)\tau(F;\ 0,\ \beta) + \beta u_{1-\beta} \sim \beta u_{1-\beta} \tag{4.11}$$

where $C(\alpha,\ \beta)$ is given by (3.4).

Finally, from (4.11)

$$v_{0,\beta}^2 \geq \beta(u_{1-\beta} - C(0,\ \beta))^2 \sim \beta(1-\beta)^2 u_{1-\beta}^2$$

which tends to ∞, and (4.9) and the theorem follow. □

Remark 1. The same arguments show that the result of Theorem 2 continues to hold if $\tau^2(\hat{F};\ 0,\ \beta)$ is replaced by $\tau^2(\hat{F};\ \alpha,\ \beta)$ for any $\alpha < 1 - \beta$, and if as before $\varepsilon = \beta$.

Remark 2. In addition to predicting that the efficiency of the β-trimmed SD to the SD tends to 0 as $\tau \to \infty$ for $\varepsilon = \beta$ the asymptotic theory also gives a positive limit for $\varepsilon \neq \beta$. Roughly speaking, for $\varepsilon > \beta$ the behavior is governed by the contaminant while for $\varepsilon < \beta$ it is governed by the main portion. Clearly Table 3.3 reflects this only very crudely. For the values of τ

we have given, the efficiency at $\varepsilon = \beta$ remains quite large. However, further numerical calculations for extremely large values of τ (of the order of 50 and above) confirm the asymptotic theory although convergence is slow. It is seen from the tables that small efficiency values do appear for smaller values of τ and $\varepsilon > \beta$. These presumably reflect small positive limits which are approached more rapidly than the asymptotic value for $\varepsilon = \beta$.

Remark 3. An interesting limiting case of the doubly trimmed standard deviations arises as both α and β tend to $1/2$. The limiting functional is naturally taken to be,

$$\tau(F; \tfrac{1}{2}, \tfrac{1}{2}) = \sqrt{G^{-1}(\tfrac{1}{2})}, \tag{4.12}$$

the median of the distribution of $|X - \mu|$ which, since X is symmetric about μ, coincides with

$$F^{-1}(\tfrac{3}{4}) - F^{-1}(\tfrac{1}{4}),$$

the interquartile range of F. The estimator $\tau(\hat{F}; \frac{1}{2}, \frac{1}{2})$ has asymptotic standardized variance,

$$\frac{1}{16x^2 f^2(x)} \tag{4.13}$$

where $F(x) = 3/4$.

Again for the Tukey model with $\varepsilon = .5$ the asymptotic efficiency of this estimator relative to the SD tends to zero as $\lambda \to \infty$. Numerical values of the efficiency similar to those shown in Table 3.3

for $\alpha = 0$, $\beta = .1$ and $.2$ are given in Table 4.1 computed by Winston Chow.

ε \ τ	2	4	6
0	.37	.37	.37
.1	.62	2.05	3.16
.2	.65	1.43	1.78
.3	.61	1.01	1.11
.4	.56	.73	.70
.5	.51	.52	.43

Table 4.1. Asymptotic efficiency of $\tau(\hat{F}; \frac{1}{2}, \frac{1}{2})$ with respect to the SD for the Tukey model (3.7).

They are clearly much less satisfactory than the corresponding values for $\alpha = .1$. In particular, the low values at the uncontaminated normal make this measure unsuitable.

The result of Theorem 2 is rather disappointing, particularly since it is in such contrast to the general (asymmetric) location case. One may ask whether a positive lower bound can be obtained if the trimmed SD is replaced by a measure of the form (1.10) with $\gamma = 2$ but more general Λ. An essentially negative answer (if attention is restricted to robust measures) is provided by the following generalization of Theorem 2.

Theorem 3. Suppose that $\tau(F)$ is given by (1.10) where the weight function Λ has a density Λ' which vanishes outside $[\alpha, 1-\beta]$ with $0 < \alpha < 1 - \beta < 1$ and is bounded away from 0 inside. Then the relative efficiency of $\tau^2(\hat{F})$ relative to the standard deviation tends to 0 as $\lambda \to \infty$ in the Tukey model (3.7) with $\varepsilon = \beta$.

For the proof we require the following lemma.

Lemma 3. If τ_1^2 and τ_2^2 are two functionals of the form (1.10) with $\gamma = 2$ and weight functions Λ_1, Λ_2, and if the Λ_i have densities Λ_i' satisfying

$$0 \leq a \leq \Lambda_2'(t)/\Lambda_1'(t) \leq A < \infty , \tag{4.14}$$

then the efficiency of the estimator of τ_2 to that of τ_1 satisfies

$$e_{2,1}(F) \geq a^2/A^2 . \tag{4.15}$$

Proof. Condition (4.14) clearly implies that $\tau_2(F) \geq a\tau_1(F)$ for all F. On the other hand, it follows from Theorem 3 of BL II that $v_2^2(F) \leq A^2 v_1^2(F)$, and the conclusion follows.

The theorem now follows on taking τ_1 and τ_2 as $\tau(\hat{F}; \alpha, \beta)$ and the SD respectively and applying Remark 1 following Theorem 2.

5. The pth power deviations

Although we have not found a measure of dispersion which is robust in the sense of BL (I,II) and whose efficiency relative to the SD has a positive lower bound, it is possible to find measures which are more robust (in the sense of BL II) than the SD and which

achieve such a lower bound. In fact, restrict attention to distributions with finite pth moment and consider the pth absolute power deviation

$$\tau_{(p)}^2 = (E|X - \mu|)^{1/p} \tag{5.1}$$

for $1 \leq p \leq 2$, and the associated estimator

$$\delta_{(p)} = \left(\frac{1}{n}\Sigma|X_i - \mu|^p\right)^{1/p}, \tag{5.2}$$

where from now on we shall assume wlog that $\mu = 0$.

Below we shall obtain lower bounds for the efficiency of (5.2) relative to the SD for the family $\mathcal{J}_0$ of all symmetric distributions. This bound can be improved if we restrict attention to the family $\mathcal{J}_1$ of symmetric unimodal distributions and improved still further for the family $\mathcal{J}_2$ of scale mixtures of normal distributions with a common mean.

Theorem 4. Let $e(F; p, q)$ denote the efficiency of the pth relative to the qth absolute power deviation where $1 \leq p < q$. Then

$$\inf_{\mathcal{J}_i} e(F; p, q) = \begin{cases} p^2/q^2 & \text{for } i = 0 \\ \dfrac{p^2}{q^2}\,\dfrac{2p+1}{2q+1}\left(\dfrac{q+1}{p+1}\right)^2 & \text{for } i = 1 \\ \dfrac{p^2}{q^2}\,\dfrac{\Gamma(q+\frac{1}{2})}{\Gamma^2(\frac{q+1}{2})}\,\dfrac{\Gamma^2(\frac{p+1}{2})}{\Gamma(p+\frac{1}{2})} & \text{for } i = 2 \end{cases} \tag{5.3}$$

Before proving these results, let us note that for $p = 1$, $q = 2$ we obtain for the efficiency of the absolute deviation to the standard deviation the values

$$\frac{1}{4} = .250 \qquad \text{for} \quad i = 0$$

$$\frac{27}{80} = .337 \qquad \text{for} \quad i = 1$$

$$\frac{1.5}{\pi} = .477 \qquad \text{for} \quad i = 2$$

These bounds are rather low and can be improved (at the cost of some robustness) by taking a larger value of p. For example, if $p = 1.5$, the values are

$$.563 \quad \text{for} \quad i = 0$$

$$.648 \quad \text{for} \quad i = 1$$

$$.851 \quad \text{for} \quad i = 2$$

The proof of all three parts of the theorem hinge on the following lemma, which is equivalent to the case $i = 0$.

Lemma 4. Let V be any nonnegative random variable and

$$\mu_\alpha = E(V^\alpha) < \infty. \tag{5.4}$$

Then, if $1 \leq \alpha \leq \beta$,

$$\frac{\mu_{2\alpha}}{\mu_\alpha^2} \leq \frac{\mu_{2\beta}}{\mu_\beta^2} \tag{5.5}$$

with equality if and only if V is a positive constant with probability 1.

Proof. To see this, note that $\log \mu_\alpha$ is a convex function of α (see for example Loève (1955), p. 156)). From this it follows easily that

$$\frac{\log \mu_{2\beta} - \log \mu_{2\alpha}}{2(\beta - \alpha)} \geq \frac{\log \mu_\beta - \log \mu_\alpha}{\beta - \alpha}$$

and hence that $\mu_{2\beta}/\mu_\beta^2 \geq \mu_{2\alpha}/\mu_\alpha^2$ as was to be proved.

Proof of part (i) of Theorem 4. By the central limit theorem we find easily that if the required moments exist, we have

$$e(F;\ p,\ q) = \frac{p^2}{q^2}\left\{\left(\frac{E|X|^{2q}}{E^2|X|^q} - 1\right)\Big/\left(\frac{E|X|^{2p}}{E^2|X|^p} - 1\right)\right\} \tag{5.6}$$

The result then follows immediately from the lemma.

Proof of parts (ii) and (iii) of Theorem 4. Note first that both of these families are of the following type:

$$\mathcal{F} = \{F:\ X \sim F,\ X = ZW,\ Z \text{ is fixed and } W \text{ is independent of } Z \text{ and varies freely}\}. \tag{5.7}$$

In case (ii) this is achieved by taking Z to be the uniform distribution on $(-1, 1)$ (see Feller (1966), p. 155). In case (iii) we just take Z to be a standard normal variable. To complete the proof we need the following lemma.

Lemma 5. Let $\mathcal{F}$ be as in (5.7) with Z symmetric about 0. Let

$$\nu_\alpha = E|Z|^\alpha . \tag{5.8}$$

Then for $1 \leq p < q$,

(5.9) $$\inf_{\mathcal{J}} e(F;\ p,\ q) = \frac{p^2}{q^2}\,\frac{\nu_{2q}}{\nu_q^2}\,\frac{\nu_p^2}{\nu_{2p}}\ .$$

Proof. Let

$$w_\alpha = \frac{E|W|^{2\alpha}}{E^2|W|^{\alpha}}\ .$$

Then from (5.6) and (5.7)

$$e(F;\ p,\ q) = \frac{p^2}{q^2}\,\frac{(\nu_{2q}/\nu_q^2)\ w_q - 1}{(\nu_{2p}/\nu_p^2)\ w_p - 1} = \frac{p^2}{q^2}\,\frac{(\nu_{2q}/\nu_q^2)\ \rho - (1/w_p)}{(\nu_{2p}/\nu_p^2) - (1/w_p)}$$

where $\rho = w_q/w_p$ and where $X = ZW$ has distribution F. Keeping w_p fixed, this is minimized by taking for ρ its minimum value which is 1 by Lemma 4. Since by Lemma 4, $\nu_{2q}/\nu_q^2 \geq \nu_{2p}/\nu_p^2$, we now obtain a lower bound for the efficiency by letting $w_p \to \infty$. To show that this lower bound is sharp we need to exhibit a sequence of distributions for W such that $\rho \to 1$ and $w_p \to \infty$. Take,

(5.10) $$W = 1 \quad \text{with probability} \quad \pi$$
$$= a \quad \text{with probability} \quad 1 - \pi.$$

Then,

$$w_p = \frac{\pi + (1-\pi)\,a^{2p}}{(\pi + (1-\pi)\,a^{p})^2}\ .$$

As $a \to \infty$,

$$w_p \to \frac{1}{1-\pi} \quad \text{and} \quad \rho \to 1\ .$$

Now letting $\pi \to 1$ we can extract the requisite sequence of distributions. The lemma follows. □

The theorem now follows from the standard formulas for the absolute moments of uniform and standard normal variables. □

Remark. Note that for case (iii) for the trimmed standard deviations, an extremal sequence is provided by an appropriate sequence of Tukey models. In fact, the distributions defined by (5.10) are of this type.

As we have noted, the lower bound to the efficiency of the mean deviation is rather low even for J_2. However, as we found for trimmed standard deviations, in reasonable situations the bound is very conservative. Here are some numerical results for the pth power deviation for p = 1, 1.5 and a selection of the distributions given in Tables 3.1-3.3.

f	5	10	25	50	∞
p = 1	2.35	1.12	.94	.91	.88
p = 1.5	1.88	1.12	1.01	.99	.97

Table 5.1. Asymptotic efficiency of pth power deviations with respect to standard deviation.
t—distribution with f degrees of freedom.

r	.5	1.0	2.0	4	6	8	10	20	30	∞
p = 1	.53	.60	.68	.75	.78	.80	.81	.84	.85	.88
p = 1.5	.76	.80	.85	.89	.92	.93	.94	.96	.97	.97

Table 5.2. Asymptotic efficiency of pth power deviation with respect to SD.
Beta distribution with density (3.6)

	p = 1			p = 1.5		
ε \ τ	2	4	6	2	4	6
0	.88	.88	.88	.97	.97	.97
.025	1.06	3.08	5.18	1.10	1.99	2.33
.075	1.22	2.53	2.66	1.18	1.61	1.51
.10	1.25	2.21	2.14	1.19	1.48	1.35
.20	1.24	1.50	1.31	1.17	1.20	1.09
.40	1.09	1.04	.92	1.08	1.02	.96
.50	1.03	.95	.85	1.05	.99	.94

Table 5.3. Asymptotic efficiency of pth power deviation with respect to standard deviation. Tukey model (3.7).

Qualitatively the behavior of these measures closely parallels that of the corresponding Tables 3.1-3.3. For reasonable distributions the mean deviation particularly seems to do even better than the trimmed deviations. Of course, for sufficiently heavy tailed distributions, for example t-distributions with sufficiently low degrees of freedom, it can break down badly.

6. Measuring the scale of positive random variables

The concepts and results developed so far, also apply to a somewhat different problem. Consider a random variable X with distribution F, which is known to be positive. Then one may be interested in scaling this distribution by defining a suitable measure σ of its distance from zero. Of such a measure we shall require (in analogy with the earlier axioms for τ)

(6.1) $\sigma(aX) = a\sigma(X)$ for $a > 0$

and

(6.2) $\sigma(Y) \geq \sigma(X)$ if Y is stochastically larger than X.

There is a simple correspondence between such measures and the earlier measures τ, defined by

(6.3) $\tau(X) = \sigma(|X - \mu|)$

where $\mu = \mu_F$ as before denotes the center of symmetry of X. Given a measure σ defined over positive random variables and satisfying (6.1) and (6.2), let X be a symmetric variable whose center of symmetry is denoted by μ. Then (6.3) defines a measure satisfying (1.4), (1.5) and (1.9). Conversely, let τ be a measure of dispersion defined over symmetric random variables, and let Z be any positive random variable. Extend Z to negative values so that it is symmetric about 0 and denote the resulting random variable by X. Then $\mu = 0$ and $|X-\mu| = |X| = Z$. The measure $\sigma(Z)$ defined through (6.3) satisfies (6.1) and (6.2).

Examples of scale measures σ are provided by the median of a positive random variable X, by the first moment $E(X)$, by the square root of the second moment $\sqrt{E(X^2)}$, or by trimmed versions of these latter measures. Using (6.3), it is a trivial matter to adapt the results of Sections 3 and 4 to comparisons of the estimator of $\sqrt{E(X^2)}$ with its trimmed versions. This is, however, not quite appropriate since the scale measures of greatest

interest for positive random variables are those corresponding to $E(X)$ and its trimmed versions.

We believe that the results of numerical and theoretical comparisons of the sample mean and its trimmed competitors qualitatively will be quite similar to those obtained for the sample SD and its trimmed competitors, but we have not carried out this program. In addition to this, Theorem 4 of Section 5 reveals that the sample mean, which is the estimator of $E(X)$, has for $p > 1$ efficiency bounded from below with respect to estimators of $[E(X^p)]^{1/p}$ for the families

$\mathcal{I}_1$ = {All distributions on the positive axis with finite second moment}

$\mathcal{I}_2$ = {All members of $\mathcal{I}_1$ with monotone nonincreasing densities}

$\mathcal{I}_3$ = {All members of $\mathcal{I}_2$ which are scale mixtures of half normal distributions}.

7. Unknown center of symmetry

In estimating dispersion we have so far restricted attention to symmetric distributions and have assumed the center of symmetry to be known. If the center μ of symmetry is unknown, it is tempting to estimate μ by a suitable estimate of location and to substitute this for μ in the estimator of τ. The question naturally arises what effect this has on the asymptotic distribution and hence on the efficiency of the estimator. For the

measures of dispersion considered in this paper it turns out that under suitable regularity conditions the asymptotic distribution is unchanged by this substitution. We have not investigated the small sample behavior of these procedures.

The theorem below gives a simple sufficient condition for substitution of an estimate of location to work in this sense. We subsequently check that this condition is satisfied by trimmed standard deviations and pth power deviations among others.

To formalize the process of substitution we appeal to the discussion of Section 6 in which we indicated that there is a 1-1 correspondence between measures of dispersion τ for symmetric distributions and measures of scale σ for positive random variables via (6.3). Let us start then with such a σ defined for nonnegative variables. For reasons similar to those given at the beginning of Section 2, we need to consider extensions of σ to larger families of distributions, namely to the family of all distributions F for which $\sigma(|X|)$ is defined. To avoid proliferation of notation we also call the extension σ and define it by,

$$\sigma(X) = \sigma(|X|).$$

For example, if $\sigma^2(X) = \int_0^\infty x^2 dF$ for nonnegative variables, the extension of σ^2 is the second moment. If $\sigma(X)$ is the median of the distribution on $(0, \infty)$, the extension is the median of the distribution of $|X|$.

Let $\mu(F)$ be a measure of location satisfying (1.1) and (1.2) of BL II. Suppose also that $X_1,\cdots,X_n$ are independently distributed each with distribution F and let $\hat{F}$ be the empirical c.d.f. Define,

$$\hat{\mu} = \mu(\hat{F})$$

as the usual estimator of $\mu(F)$ and let

$$F_\mu(x) = F(x + \mu)$$

denote the c.d.f. of $X - \mu$. Note that the empirical c.d.f. of $X_1 - \hat{\mu},\cdots,X_n - \hat{\mu}$ is then $\hat{F}_{\hat{\mu}}$.

If F is symmetric about μ the measure of dispersion corresponding to σ by (6.3) is,

$$\tau(F) = \sigma(F_\mu).$$

The estimator of τ we have used for known μ is $\sigma(\hat{F}_\mu)$. Failing this knowledge we use $\hat{\tau}$ defined by,

$$\hat{\tau} = \sigma(\hat{F}_{\hat{\mu}}).$$

For example if

$$\mu(F) = \int_{-\infty}^{\infty} x dF(x)$$

$$\sigma^2(F) = \int_0^{\infty} x^2 dG(x)$$

where G is the c.d.f. of $|X|$ then,

$$\sigma^2(\hat{F}_\mu) = \frac{1}{n}\Sigma(X_i - \mu)^2$$

and $\hat{\tau}$ is the sample SD.

If

$$\mu(F) = F^{-1}(\tfrac{1}{2})$$

$$\sigma(F) = G^{-1}(\tfrac{1}{2})$$

then $\hat{\tau}$ is the median of the absolute deviations from the median of the observations.

Theorem 5. Suppose that the underlying distribution F is symmetric about μ and suppose without loss of generality that $\mu = 0$. Suppose further that,

(7.1) $\sigma(F_\mu)$ is differentiable in μ at $\mu = 0$.

(7.2) $\lim_{M\to\infty} \limsup_n P[\sqrt{n}|\hat{\mu}| \geq M] = 0$.

(7.3) $\lim_{\delta\downarrow 0} \limsup_n P[\sup\{\sqrt{n}|\sigma(\hat{F}_\mu) - \sigma(F_\mu) - \sigma(\hat{F}) + \sigma(F)| : |\mu| \leq \delta\} \geq \varepsilon] = 0$ for all $\varepsilon > 0$.

Then,

(7.4) $$\sqrt{n}(\hat{\tau} - \sigma(\hat{F})) \xrightarrow{P} 0.$$

If we assume that $\tau(F)$ is positive and

(7.5) $$\sigma(\hat{F}) \xrightarrow{P} \tau(F),$$

then we can arrive at conclusion (7.4) even if throughout (7.3) we replace σ by σ^p where $p > 0$.

Note: $\hat{\mu}$ will satisfy (7.2) provided that $\sqrt{n}\,\hat{\mu}$ has a limiting normal distribution.

By Slutsky's theorem (7.4) implies that $\sqrt{n}(\hat{\tau} - \sigma(F))$ and $\sqrt{n}(\sigma(\hat{F}) - \sigma(F))$ have the same limiting distribution. But since F is symmetric about 0,

$$\sigma(F) = \tau(F)$$

and

$$\sigma(\hat{F}) = \tau(\hat{F}),$$

the estimate we would use if the center of symmetry of F were known. Thus, (7.4) implies that $\sqrt{n}(\hat{\tau} - \tau(F))$ and $\sqrt{n}(\tau(\hat{F}) - \tau(F))$ have the same limiting distribution. That is, if (7.4) holds substitution works.

Proof. From (7.2) and (7.3),

$$\sqrt{n}[\hat{\sigma} - \sigma(F_{\hat{\mu}}) - \sigma(\hat{F}) + \sigma(F)] \xrightarrow{P} 0. \tag{7.6}$$

Since $\hat{\mu} \xrightarrow{P} 0$ by (7.2 we can apply (7.1) to conclude that

$$\sqrt{n}\left[(\sigma(F_{\hat{\mu}}) - \sigma(F)) - \hat{\mu}\left\{\frac{\partial\sigma}{\partial\mu}(F_{\mu})\Big|_{\mu=0}\right\}\right] \xrightarrow{P} 0. \tag{7.7}$$

If F is symmetric about 0, it follows from (1.4) and (1.6) that $\sigma(F_{\mu})$ is an even function of μ, and hence that $\frac{\partial\sigma}{\partial\mu}(F_{\mu})\Big|_{\mu=0} = 0$. Substituting in (7.7) and (7.6) completes the proof of (7.4).

If now σ is replaced by σ^p in (7.3) we can imitate the proof of (7.4) exactly to conclude that,

$$\sqrt{n}(\sigma^p(\hat{F}_{\hat{\mu}}) - \sigma^p(\hat{F})) \xrightarrow{P} 0. \tag{7.8}$$

But,

$$\sqrt{n}(\hat{\sigma} - \sigma(\hat{F})) = \rho\tilde{\sigma}^{p-1}\sqrt{n}(\sigma^p(\hat{F}_{\hat{\mu}}) - \sigma^p(\hat{F}))$$

where $\rho = \frac{1}{p}$ and $\tilde{\sigma}$ lies between $\sigma(\hat{F}_{\hat{\mu}})$ and $\sigma(\hat{F})$. By (7.5) and (7.8)

$$\tilde{\sigma} \xrightarrow{P} \sigma(F) > 0$$

and (7.4) follows. □

Examples

(i) The pth power deviations

Define $\tau(H) = (\int_{-\infty}^{\infty}|x|^p dH(x))^{1/p}$ for all H such that $\int|x|^p dH(x) < \infty$. Then $\tau(F_\mu)$ is the pth power deviation. We shall check that (7.3) holds for τ^p under the condition

$$\int|x|^{2p} dF(x) < \infty . \tag{7.9}$$

For $\mu < \nu$, we find

$$E(\sqrt{n}\,(\tau^p(\hat{F}_\nu) - \tau^p(\hat{F}_\mu) - \tau^p(F_\nu) + \tau^p(F_\mu)))^2$$

$$= \mathrm{Var}(|X_1 - \nu|^p - |X_1 - \mu|^p) .$$

If $p > 1$ this variance is bounded by,

$$E(|X_1 - \nu|^p - |X_1 - \mu|^p)^2 \leq p^2(\nu-\mu)^2 \sup\{E|X_1 - \lambda|^{2p-2}:$$

$$\lambda \in [\mu, \nu]\} \leq C(\nu - \mu)^2 \quad \text{if} \quad |\nu|, |\mu| \leq M,$$

where C is a constant depending on M. If $p = 1$,

$$\mathrm{Var}(|X_1 - \nu| - |X_1 - \mu|) \leq E[(\nu - \mu)(2I_{[X_1<\mu]} - 1)$$

$$+ 2(\nu - X_1) I_{[\mu<X_1<\nu]}]^2 \leq C(\nu - \mu)^2 \quad \text{if} \quad |\nu|, |\mu| \leq M.$$

In any event,

$$E(\sqrt{n}(\tau^p(\hat{F}_\nu) - \tau^p(\hat{F}_\mu) - \tau^p(F_\nu) + \tau^p(F_\mu)))^2 \leq C(\nu - \mu)^2$$

and an argument using Theorem 12.3, p. 95 in [3] completes the proof. Since (7.9) implies (7.5) as well as the asymptotic normality of $\tau(\hat{F})$ we see that any measure $\hat{\mu}$ satisfying (7.2), for instance the mean, can be substituted without affecting the asymptotic behavior of the estimators.

(7.i) (L) estimators

Suppose τ is given by

$$\tau(H) = [\int_0^1 |H^{-1}(t)|^\gamma \, d\tilde{\Lambda}(t)]^{1/\gamma} \tag{7.10}$$

where $\tilde{\Lambda}$ is symmetric about $\frac{1}{2}$.

If F is symmetric about μ, and if we define $\Lambda(t) = 2\tilde{\Lambda}(\frac{1+t}{2}) - 1$, then $\tau(F_\mu)$ is the measure defined in (1.10).

Suppose that

(7.11) Λ places no mass outside an interval $(\frac{\alpha}{2}, 1-\frac{\alpha}{2})$, $0 < \alpha < 1/2$

(7.12) F is differentiable with derivative f which is positive and continuous.

We shall now show that (7.3) holds for τ^γ under these assumptions, so that we may safely substitute an estimate of location in the singly and doubly trimmed standard deviations.

Here is a sketch proof.

Let

$$G(x, \mu) = P[|X_1 - \mu|^\gamma \leq x],$$

and let $\hat{G}(x, \mu)$ be the empirical c.d.f. of $|X_1-\mu|^\gamma, \ldots, |X_n-\mu|^\gamma$. Define

$$G(x) = G(x, 0), \qquad \hat{G}(x) = \hat{G}(x, 0).$$

Finally let $G^{-1}(t, \mu)$, $\hat{G}^{-1}(t, \mu)$ be the corresponding inverses (in x).

Note that

$$\tau(F_\mu) = 2\int_0^{1-\alpha} G^{-1}(t,\mu)\,d\tilde{\Lambda}\left(\frac{1+t}{2}\right) = \int_0^{1-\alpha} G^{-1}(t,\mu)\,d\Lambda(t),$$

and similar assertions hold for $\hat{F}$, $\hat{F}_\mu$, etc.

Then,

$$(7.13)\qquad \left|\tau^\gamma(\hat{F}_\mu) - \tau^\gamma(\hat{F}) - \tau^\gamma(F_\mu) + \tau^\gamma(F)\right|$$
$$= \left|\int_0^{1-\alpha}\left(\hat{G}^{-1}(t,\mu) - \hat{G}^{-1}(t) - G^{-1}(t,\mu) + G^{-1}(t)\right)d\Lambda(t)\right|$$
$$\leq \sup\{\left|\hat{G}^{-1}(t,\mu) - \hat{G}^{-1}(t) - G^{-1}(t,\mu) + G^{-1}(t)\right| :$$
$$0 \leq t \leq 1-\alpha\}.$$

Therefore, we need only establish that

$$(7.14)\qquad \lim_{\delta \downarrow 0} \limsup_n P[\sup\{\sqrt{n}\left|\hat{G}^{-1}(t,\mu) - \hat{G}^{-1}(t) - G^{-1}(t,\mu) + G^{-1}(t)\right| :$$
$$0 \leq t \leq 1-\alpha,\ |\mu| \leq \delta\} \geq \varepsilon\ ] = 0.$$

By a standard but tedious argument (see [1] and [8]), we can show that (7.14) follows if, for every M

$$(7.15)\qquad \sup\{\left|\hat{G}^{-1}(t,\mu) - G^{-1}(t,\mu)\right| : 0 \leq t \leq 1-\alpha,\ |\mu| \leq M\} \xrightarrow{P} 0$$

and

$$(7.16)\qquad j(t,\mu) = \frac{\partial G^{-1}}{\partial t}(t,\mu) \text{ is continuous on}$$
$$\{(t,\mu) : 0 \leq t \leq 1 - \frac{\alpha}{2},\ |\mu| \leq M\}$$

and if in addition

$$(7.17)\qquad \lim_{\delta \downarrow 0} \limsup_n P\left[\sup\left\{\sqrt{n}\left|\frac{\hat{G}(x,\mu) - G(x,\mu)}{j(G(x,\mu),\mu)} - \frac{\hat{G}(x) - G(x)}{j(G(x),0)}\right| : |x| \leq G^{-1}(1-\alpha),\ |\mu| \leq \delta\right\} \geq \varepsilon\right] = 0.$$

Now, (7.15) follows from the fact that for every M and M'

$$\sup\{|\hat{G}(x, \mu) - G(x, \mu)| : x \leq M', |\mu| \leq M\} \xrightarrow{P} 0. \tag{7.13}$$

This assertion is a consequence of the Glivenko-Cantelli theorem and the definition of $\hat{G}$, G.

Since

$$j(t, \mu) = \frac{1}{g(G^{-1}(t,\mu), \mu)} = [f(G^{-1}(t,\mu) + \mu) + f(-G^{-1}(t,\mu) + \mu)]^{-1},$$

assertion (7.16) is immediate from assumption. Finally, (7.17) follows from (7.16), the tightness of the empirical process

$$\lim_{\delta \downarrow 0} \limsup_n P[\sup\{\sqrt{n}|\hat{F}(x+\mu) - F(x+\mu) - \hat{F}(x) + F(x)| :$$
$$|\mu| \leq \delta, |x| \leq M\} \geq \varepsilon] = 0,$$

and the boundedness (in probability) of $\sup_x \sqrt{n}|\hat{G}(x) - G(x)|$. Therefore, (7.13) holds for τ defined by (7.10). As before (7.1) and (7.5) are usually obvious under our assumptions and we conclude that substitution of location estimates satisfying (7.2) such as the median is legitimate.

REFERENCES

[1] Bickel, P. J. (1973), "On some analogues to linear combinations of order statistics and linear model," Ann. Statist. 1, 597-616.

[2] Bickel, P. J. and Lehmann, E. L. (1975), "Descriptive statistics for nonparametric models,"

(a) I. Introduction, Ann. Statist.

(b) II. Location, Ann. Statist.

[3] Billingsley, P. (1968), Weak Convergence of Probability Measures. Wiley. New York.

[4] Birnbaum, Z. W. (1948), "On random variables with comparable peakedness," Ann. Math. Statist. 37, 1593-1601.

[5] Daniell, P. J. (1920), "Observations weighted according to order," Amer. J. Math. 42, 222-236.

[6] Feller, William (1966). An Introduction to Probability Theory and Its Applications II. John Wiley, New York.

[7] Loève, Michel (1955). Probability Theory. van Nostrand, New York.

[8] Pyke, R. and Shorack, G. (1968), "Weak convergence of a two-sample empirical process and a new approach to Chernoff-Savage theorems," Ann. Math. Statist. 39, 755-771.

[9] Stigler, Stephen M. (1973), "The asymptotic distribution of the trimmed mean," Ann. Statist. 1, 474-477.