The Annals of Statistics
1983, Vol. 11, No. 1, 317–325

A MINIMAX CRITERION FOR CHOOSING WEIGHT FUNCTIONS FOR L-ESTIMATES OF LOCATION

By David M. Mason

University of Delaware

Let $X_1, \cdots, X_n$ be independent with common distribution F symmetric about μ. Let $T_n = n^{-1} \sum_{i=1}^n J(i/(n+1))X_{in}$ be an L-estimate of μ based on a weight function J and the order statistics $X_{1n} \le \cdots \le X_{nn}$ of $X_1, \cdots, X_n$. Under very general regularity conditions $n^{1/2}T_n$ has asymptotic variance $\sigma^2(J, F)$. A weight function J_0 is found that minimizes the maximum of $\sigma^2(J, F)/s^2(F)$, whenever $s(F)$ is a measure of scale of a general type, as F ranges over a subclass of the symmetric distributions determined by $s(F)$ and J ranges over a class of weight functions also determined by $s(F)$. The sample mean and the trimmed mean arise as the solutions for particular choices of scale measures.

1. Introduction. Let $X_1, X_2, \cdots, X_n$ be independent random variables with a common distribution function F, where F is assumed to be a right continuous distribution function symmetric about μ. μ is usually called the location parameter of F. We will say that F is symmetric about μ if $F(x - \mu) = 1 - F(-(x-\mu)-)$ for all real x. Let $X_{1n} \le \cdots \le X_{nn}$ be the order statistics of $X_1, \cdots, X_n$ and let J be a real valued weight function defined on (0, 1). Any estimate of μ of the form

$$T_n = n^{-1} \textstyle\sum_{i=1}^n J(i/(n+1))X_{in}$$

will be called an L-estimate of μ based on the weight function J.

Some of the minimum requirements for T_n to be a consistent estimate of μ are that J be symmetric about ½, that is, $J(u) = J(1-u)$ for all $\mu \in (0, 1)$, and $\int_0^1 J(u)\,du = 1$. For convenience we will let $\mathscr{J}$ denote the class of all measurable real valued functions defined on (0, 1) that satisfy these two conditions.

Under certain regularity conditions on J depending on the tail behavior of F, $T_n \to \mu$ a.s. See for instance Wellner (1977), van Zwet (1980) or Mason (1982). Other regularity conditions on J and F in combination with symmetry of J and F imply that

$$n^{1/2}(T_n - \mu) \to_d N(0, \sigma^2(J, F)),$$

where

$$0 < \sigma^2(J, F) = \int_0^1 \int_0^1 J(u)J(v)(u \wedge v - uv)\, dF^{-1}(u)\, dF^{-1}(v) < \infty,$$

with

$$F^{-1}(u) = \inf\{x : F(x) \ge u\} \quad \text{for} \quad u \in (0, 1].$$

Refer to Shorack (1972), Stigler (1974) and Mason (1981).

Let f be a density function that is symmetric about zero and consider the family of distributions $\mathscr{F}_f = \{F_{\mu,s} : F_{\mu,s}$ has density $s^{-1}f(s^{-1}(x-\mu)), \mu \in (-\infty, \infty), s \in (0, \infty)\}$. Under certain regularity conditions there exists a weight function $J_f \in \mathscr{J}$ such that J_f minimizes

(1) $$\sigma^2(J, F_{0,1})$$

Received July 1981; revised May 1982.
AMS 1970 *subject classifications.* Primary 62G35; secondary 62G05.
Key words and phrases. L-estimates for location, order statistics, minimax, weight function.

among all $J \in \mathscr{J}$, and in addition the J_f that minimizes expression (1) is asymptotically efficient in the sense that

$$\sigma^2(J_f, F_{0,1}) = I_f^{-1},$$

where I_f is the Fisher information number

$$I_f = \int_{-\infty}^{\infty} \left\{\frac{f'(x)}{f(x)}\right\}^2 f(x)\, dx.$$

Since $\sigma^2(J, F_{\mu,s})s^{-2} = \sigma^2(J, F_{0,1})$, an alternate way of writing expression (1) is

$$\sup\{\sigma^2(J, F_{\mu,s})s^{-2}; \quad F_{\mu,s} \in \mathscr{F}_f\}. \tag{2}$$

Additional regularity conditions imply that when T_n has the optimal weight function J_f, for each $F_{\mu,s} \in \mathscr{F}_f$

$$s^{-1}n^{1/2}(T_n - \mu) \to_d N(0, I_f^{-1}).$$

See Chernoff, Gastwirth and Johns (1967) and Huber (1977, pages 22–23) for more details.

The problem of choosing a weight function that minimizes expressions like

$$\sup\{\sigma^2(J, F_s)s^{-2} : F_s \in \mathscr{F}\}$$

among all $J \in \mathscr{J}$ (here F_s means that F_s has scale parameter s and $\mathscr{F}$ is some class of symmetric distributions) has also been considered by Gastwirth and Rubin (1969) in a more general context where point masses are allowed at particular quantiles in location estimates. They require that $\mathscr{F}$ be a class of distribution functions such that each distribution function in the class has a density which is symmetric about zero and satisfies a certain uniform tail condition. See pages 27 and 28 of Gastwirth and Rubin (1969).

We will be concerned with an extension of the foregoing ideas. Let $s(F)$ denote a measure of scale. We will find that weight function J_0 that minimizes the maximum of $\sigma^2(J, F)/s^2(F)$, as F ranges over a subclass of the symmetric distributions determined by $s(F)$ and J ranges over a subclass of $\mathscr{J}$ also determined by $s(F)$. For some choices of scale measures $s(F)$, finiteness of $s(F)$ is saying something about the tail behavior of F. J_0 then becomes the minimax choice, by our criterion, of a weight function for an L-estimate of location for the class of all symmetric distributions which behave in a specified manner in the tails determined by $s(F)$. Often it will turn out that for particular choices of $s(F)$ known limit theorems for L-estimates will imply that the T_n with weight function J_0 is consistent and asymptotically normal with variance $\sigma^2(J, F)$.

Our approach differs from the minimax approach for the choice of a ψ function for an M-estimate of location considered by Huber (1964) in the following manner. Huber's approach is in a sense semi-parametric, in that a ψ function is found that minimizes the maximum asymptotic variance of the M-estimate as F ranges over a particular topologically "small" neighborhood of a specified symmetric distribution. Whereas our approach is in a sense semi-nonparametric, in that we restrict our distributions to lie in a certain subclass of the symmetric distributions that possess a specified tail behavior. For some more recent results related to Huber's approach see Collins (1977) and Rousseuw (1981).

2. The minimax choice of a weight function for a particular class of scale measures. $\mathscr{S}$ will denote the class of distributions symmetric about zero. Let $\mathscr{H}$ be the class of nonnegative measurable functions defined on (0, 1) which are symmetric about ½. We will consider the class of scale measures defined via a function $h \in \mathscr{H}$ as follows: For $h \in \mathscr{H}$ and $F \in \mathscr{S}$, let

$$s(h, F) \equiv \int_0^1 h(u)\, dF^{-1}(u),$$

whenever $s(h, F)$ is finite. It is easy to verify that when $s(h, F) < \infty$, $s(h, F_{\mu,\tau}) = s(h, F)\tau$ for all $\mu \in (-\infty, \infty)$ and $\tau \in (0, \infty)$, where $F_{\mu,\tau}(x) = F((x - \mu)/\tau)$. This class includes some common measures of scale. See the examples below.

Given any $h \in \mathscr{H}$, let us define the following subclasses of $\mathscr{S}$ and $\mathscr{J}$: Let $\mathscr{S}_h$ be the subclass of $\mathscr{S}$ such that for each $F \in \mathscr{S}_h$ (i) F^{-1} and h have no common discontinuity points; and (ii) $0 < s(h, F) < \infty$.

Whenever h is continuous except perhaps at a finite number of jump discontinuities in (0, 1), let $\mathscr{J}_h$ be the subclass of $\mathscr{J}$ such that for each $J \in \mathscr{J}_h$, J is continuous on (0, 1) except perhaps where h is discontinuous, in which case J has only jump discontinuities. We will define the asymptotic risk of using an L-estimate of location based on a weight function J for a particular distribution F in $\mathscr{S}_h$ to be

$$R(J, s(h, F), F) = \sigma^2(J, F)/s^2(h, F). \tag{3}$$

It is easy to see that R is both location and scale invariant, so it was with no loss of generality that we assumed from the beginning that each F is symmetric about zero.

At this point it is convenient to introduce the following subclass of $\mathscr{S}$ consisting of symmetric three point distributions. The nice properties of this class will be essential to the proof of our main result.

For any $x_\nu \in (0, \infty)$ and $\nu \in (0, ½)$, let G_{ν,x_ν} be a symmetric distribution defined as follows:

$$G_{\nu,x_\nu}(x) = \begin{cases} 1 & x_\nu \le x \\ 1-\nu & 0 \le x < x_\nu \\ \nu & -x_\nu \le x < 0 \\ 0 & x < -x_\nu. \end{cases}$$

Let $\mathscr{G}$ be the class of all such distributions. Each $G_{\nu,x_\nu} \in \mathscr{G}$ has an inverse distribution g_{ν,x_ν} defined as follows:

$$g_{\nu,x_\nu}(u) = \begin{cases} x_\nu & 1-\nu < u \le 1 \\ 0 & \nu < u \le 1-\nu \\ -x_\nu & 0 < u \le \nu. \end{cases}$$

Let $J \in \mathscr{J}$, $h \in \mathscr{H}$ and $G_{\nu,x_\nu} \in \mathscr{G}$. Trivial calculations show that whenever J is continuous at ν, $\sigma^2(J, G_{\nu,x_\nu}) = 2\nu x_\nu^2 J^2(\nu)$, and whenever h is continuous at ν, $s(h, G_{\nu,x_\nu}) = 2x_\nu h(\nu)$.

THEOREM 1. *Let $h \in \mathscr{H}$ be such that h is continuous on* (0, 1) *except perhaps at a finite number of jump discontinuities, and*

$$0 < 2\int_0^{1/2} h(u)u^{-1/2}\,du \equiv \{C(h)\}^{-1} < \infty.$$

Let $J_h(u) = C(h)h(u)u^{-1/2}$ for $0 < u \le ½$ and $= C(h)h(u)(1-u)^{-1/2}$ for $½ < u < 1$. Then J_h minimizes

$$\sup\{R(J, s(h, F), F) : F \in \mathscr{S}_h\}$$

among all $J \in \mathscr{J}_h$, and

$$\sup\{R(J_h, s(h, F), F) : F \in \mathscr{S}_h\} = 2^{-1}C^2(h).$$

REMARK 1. The requirement that h and F^{-1} and hence J and F^{-1} have no common discontinuity points is natural in the sense that this is one of the minimum assumptions needed for the asymptotic normality of L-estimates. See Mason (1981), Shorack (1972) and Stigler (1974).

We will postpone the proof until Section 3 and first give some examples.

EXAMPLE 1. Choose $0 < p < \infty$ and set $h_p(u) = u^{1/p}$ for $0 < u \le ½$ and $h_p(u) = (1-u)^{1/p}$ for $½ < u < 1$. For $F \in \mathscr{S}$,

$$s(h_p, F) = \int_0^{1/2} u^{1/p}\, dF^{-1}(u) + \int_{1/2}^1 (1-u)^{1/p}\, dF^{-1}(u).$$

By integration by parts, $s(h_p, F)$ also equals

$$\int_0^{1/2} | F^{-1}(u)|\, u^{1/p-1}\, du + \int_{1/2}^1 F^{-1}(u)(1-u)^{1/p-1}\, du.$$

When $p = 1$, $s(h_1, F)$ becomes the absolute first moment of F. In the context of a scale measure $s(h_1, F)$ is often called the mean absolute deviation from the median. When $0 < p < 1$ or $1 < p < \infty$, finiteness of $s(h_p, F)$ is very closely related to finiteness of the pth absolute moment of F. Refer to the Appendix of Mason (1982) for a complete discussion of this point.

Finiteness of $s(h_p, F)$ is also saying something about the tail behavior of F. It is equivalent to saying that

$$| x |^{-1/p}(1 - F(x) + F(-x)) \to 0 \quad \text{as} \quad x \to \infty$$

at such a rate as to make

$$\int_{-\infty}^0 \{F(x)\}^{1/p-1}\, | x |\, dF(x) + \int_0^\infty \{1 - F(x)\}^{1/p-1} x\, dF(x) < \infty.$$

In this case $J_{h_p}(u) = C(p)u^{1/p-1/2}$ for $0 < u \le ½$ and $= C(p)(1-u)^{1/p-1/2}$ for $½ < u < 1$, where $C(p) = (2 + p)p^{-1}2^{1/p-3/2}$. $\mathscr{S}_{h_p}$ is the class of all $F \in \mathscr{S}$ such that $0 < s(h_p, F) < \infty$. Observe that when $p = 2$, $J_{h_2} \equiv 1$, so that the L-estimate T_n based on J_{h_2} is just the sample mean. This is reasonable since $s(h_2, F) < \infty$ is almost equivalent to F having a finite second moment. See the remark immediately following this example.

It is interesting to note that if T_n is the L-estimate with weight function J_{h_p} and the underlying distribution F satisfies $0 < s(h_p, F) < \infty$, that the weight function and the scale condition exactly interlock to imply asymptotic normality of T_n (actually for a slightly trimmed version of T_n for the case when $0 < p < 2$). Refer to Theorem 1 of Mason (1981).

REMARK 2. Suppose instead of the scale measure $s(h_2, F)$, the standard deviation $s(F) = (\int_0^1 \{F^{-1}(u)\}^2\, du)^{1/2}$ is used. Then the weight function $J_{h_2} \equiv 1$ also minimizes

$$\sup\{R(J, s(F), F) : F \in \mathscr{S}^*\}$$

among all $J \in \mathscr{J}_{h_2} = \{J \in \mathscr{J}: J \text{ is continuous}\}$, where $\mathscr{S}^* = \{F: 0 < s(F) < \infty\}$. To see this, choose any $F \in \mathscr{S}^*$. Then since

$$\sigma^2(J_{h_2}, F) = \int_0^1 \int_0^1 (u \wedge v - uv)\, dF^{-1}(u)\, dF^{-1}(v) = s^2(F),$$

$R(J_{h_2}, s(F), F) = 1$ for all $F \in \mathscr{S}^*$. Let $J \in \mathscr{J}_{h_2}$ be such that $J \ne 1$ for some point in (0, 1). Since J is continuous and $\int_0^1 J(u)\, du = 1$, there must exist a point $\nu \in (0, ½)$ such that $J(\nu) > 1$. Now choose any $x_\nu \in (0, \infty)$ and $G_{x_{\nu,\nu}} \in \mathscr{G}$, we see that

$$\sigma^2(J, F) = 2\nu x_\nu^2 J^2(\nu) > 2\nu x_\nu^2 = \sigma^2(J_{h_2}, F).$$

Hence $J_{h_2} \equiv 1$ is the minimax choice of J with respect to the standard deviation. It can be shown that $\mathscr{S}_{h_2} \subset \mathscr{S}^*$, but $\mathscr{S}^* \not\subset \mathscr{S}_{h_2}$. Refer to Hoeffding (1973) or Mason (1982).

EXAMPLE 2. Choose $\alpha \in (0, ½)$ and set $h(u) = 1$ if $\alpha \le u \le 1 - \alpha$ and zero otherwise.

Whenever F^{-1} is continuous at α and $1 - \alpha$

$$s(h, F) = F^{-1}(1-\alpha) - F^{-1}(\alpha).$$

Observe that $s(h, F)$ is a symmetric interquantile range. In this case $J_h(u) = C(h)u^{-1/2}$ for $\alpha \le u \le ½$, $= C(h)(1-u)^{-1/2}$ for $½ < u \le 1-\alpha$, and equal to zero elsewhere, where $C(h) = \{4(2^{-1/2} - \alpha^{1/2})\}^{-1}$. $\mathscr{S}_h$ is the class of all $F \in \mathscr{S}$ such that $0 < s(h, F) < \infty$ and F^{-1} is continuous at α and $1-\alpha$.

L-estimates based on J_h are not quite trimmed means. Increasingly more weight is placed on the outer quantiles up to the αth and the $(1-\alpha)$th quantiles than on the inner quantiles. The symmetrically trimmed α − mean is obtained by using the scale measure $s(h, F)$ with $h(u) = u^{1/2}$ for $\alpha \le u \le ½$, $h(u) = (1-u)^{1/2}$ for $½ < u \le 1-\alpha$, and zero elsewhere.

Example 3. Choose $\alpha \in (0, ½)$ and set $h(u) = u$ for $\alpha \le u \le ½$, $h(u) = 1-u$ for $½ < u \le 1-\alpha$, and zero elsewhere. Whenever $F \in \mathscr{S}$ and F^{-1} is continuous at α and $1-\alpha$, we see by integration by parts that

$$s(h, F) = \alpha\{F^{-1}(1-\alpha) - F^{-1}(\alpha)\} + \int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} |x| \, dF(x).$$

So $s(h, F)$ is an α-Winsorized mean absolute deviation from the median. In this case $J_h(u) = C(h)u^{1/2}$ for $\alpha \le u \le ½$, $= C(h)(1-u)^{1/2}$ for $½ < u \le 1-\alpha$, and zero elsewhere, where $C(h) = ¾(2^{-3/2} - \alpha^{3/2})^{-1}$. $\mathscr{S}_h$ is the same class as that given in Example 2.

Example 4. Choose $0 < \alpha < ½$, and set $h(u) = (1-2\alpha)^{-1}(u-\alpha)$ for $\alpha \le u \le ½ = (1-2\alpha)^{-1}(1-u-\alpha)$ for $½ < u \le 1-\alpha$ and equal to zero elsewhere. By combining Examples 2 and 3 we see that

$$s(h, F) = (1-2\alpha)^{-1} \int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} |x| \, dF(x).$$

$s(h, F)$ is an α − trimmed mean absolute deviation from the median. $J_h(u) = D(h)(u-\alpha)u^{-1/2}$ for $\alpha \le u \le ½$, $= D(h)(1-\alpha-u)(1-u)^{-1/2}$ for $½ < u \le 1-\alpha$, and equal to zero elsewhere, where $D(h) = \{(⅔ - 4\alpha)2^{-1/2} + \frac{8}{3}\alpha^{3/2}\}^{-1}$ with $D(h) = C(h)(1-2\alpha)$.

Remark 3. If it is assumed that F has a finite absolute pth moment for some positive p, then each of the L-estimates based on the weight functions in Examples 2, 3 and 4 are asymptotically normal. Refer to Theorem 5 of Stigler (1974).

3. Proof of Theorem 1. First observe that $J_h \in \mathscr{J}_h$. Let $\mathscr{G}_h$ be the subclass of $\mathscr{G}$ consisting of all those $G_{\nu,x_\nu} \in \mathscr{G}$ such that G_{ν,x_ν}^{-1} and h have no discontinuity points in common and $s(h, G_{\nu,x_\nu}) > 0$. Note that $\mathscr{G}_h \subset \mathscr{S}_h$. The proof of Theorem 1 will consist of the following steps:

Step 1. We will show that

$$R(J_h, s(h, G_{\nu,x_\nu}), G_{\nu,x}) = 2^{-1}C^2(h) \tag{4}$$

for all $G_{\nu,x_\nu} \in \mathscr{G}_h$.

Step 2. For any $F \in \mathscr{S}_h$,

$$R(J_h, s(h, F), F) \le 2^{-1}C^2(h). \tag{5}$$

Step 3. For any $J \in \mathscr{J}_h$ such that J and J_h disagree at a point $\nu \in (0, 1)$ where both J

and J_h are continuous there exists an $F \in \mathscr{S}_h$ such that

$$R(J, s(h, F), F) > 2^{-1}C^2(h).$$

By the definition of $\mathscr{J}_h$, for any $J \in \mathscr{J}_h$ such that $J \neq J_h$, J and J_h can only disagree where they both have jump discontinuities or both are continuous. Steps 1, 2, and 3 imply that

$$\sup\{R(J, s(h, F), F) : F \in \mathscr{S}_h\} > \sup\{R(J_h, s(h, F), F) : F \in \mathscr{S}_h\} = 2^{-1}C^2(h),$$

whenever J and J_h disagree at a point where they are both continuous. If they only disagree at points where they both have jump discontinuities, continuity of F^{-1} at those points implies that

$$\sup\{R(J, s(h, F), F) : F \in \mathscr{S}_h\} = \sup\{R(J_h, s(h, F), F) : F \in \mathscr{S}_h\} = 2^{-1}C^2(h).$$

These remarks show that J_h is indeed minimax.

PROOF OF STEP 1. Choose any $G_{\nu,x_\nu} \in \mathscr{G}_h$, then $\sigma^2(J_h, G_{\nu,x_\nu}) = 2J_h^2(\nu)\nu x_\nu^2 = 2C^2(h)h^2(\nu)x_\nu^2$ and $s^2(h, G_{\nu,x_\nu}) = 4x_\nu^2 h^2(\nu) > 0$. □

PROOF OF STEP 2. Let $\mathscr{I}_h = \{F^{-1} : F \in \mathscr{S}_h\}$ and for any $F^{-1} \in \mathscr{I}_h$ set $\tau(J_h, F^{-1}) = \sigma(J_h, F)$. It is easy to see that $\mathscr{I}_h$ is a convex class of functions. We claim that $\tau(J_h, \cdot)$ in a convex functional defined on $\mathscr{I}_h$.

Let U be a uniform (0, 1) random variable. Application of Fubini's theorem gives

$$\tau^2(J_h, F^{-1}) = \operatorname{Var}\left(\int_0^1 \{I(U \le u) - u\} J_h(u)\, dF^{-1}(u)\right);$$

where $I(x \le y) = 1$ or 0 according to whether $x \le y$ or $x > y$.

Let F_1^{-1} and $F_2^{-1} \in \mathscr{I}_h$ and choose $0 \le \alpha \le 1$. For $i = 1, 2$ let

$$Z_i = \int_0^1 \{I(U \le u) - u\} J_h(u)\, dF_i^{-1}(u).$$

Now

$$\begin{aligned}\tau(J_h, \alpha F_1^{-1} + (1-\alpha)F_2^{-1}) &= [\operatorname{Var}\{\alpha Z_1 + (1-\alpha)Z_2\}]^{1/2} \\ &\le \alpha(\operatorname{Var} Z_1)^{1/2} + (1-\alpha)(\operatorname{Var} Z_2)^{1/2} \\ &= \alpha\tau(J_h, F_1^{-1}) + (1-\alpha)\tau(J_h, F_2^{-1}),\end{aligned}$$

proving the claim.

To show (5) for all $F \in \mathscr{S}_h$ is equivalent to showing

$$\tau(J_h, F^{-1}) \Big/ \int_0^1 h(u)\, dF^{-1}(u) \le 2^{-1/2}C(h) \tag{6}$$

for all $F^{-1} \in \mathscr{I}_h$.

We will begin by first assuming that h is continuous and strictly positive. Choose any $F^{-1} \in \mathscr{I}_h$. Now choose any $\varepsilon \in (0, ½)$ such that $\int_\varepsilon^{1-\varepsilon} h(u)\, dF^{-1}(u) > 0$ and ε is a continuity point of F^{-1}. Choose a sequence of partitions of the interval $[\varepsilon, ½]$ $\varepsilon = \nu_{0n} < \nu_{1n} \cdots < \nu_{nn} < ½$ such that each ν_{in} for $i = 0, \cdots, n$ is a continuity point of F^{-1}, $\max_{0 \le i \le n-1}(\nu_{i+1,n} - \nu_{in}) \to 0$ and $½ - \nu_{n,n} \to 0$. For each $0 \le i \le n-1$ and $n \ge 1$ set

$$p_{in} = 2\int_{\nu_{in}}^{\nu_{i+1,n}} h(u)\, dF^{-1}(u) \Big/ \int_\varepsilon^{1-\varepsilon} h(u)\, dF^{-1}(u),$$

and

$$p_{nn} = \int_{\nu_{nn}}^{\nu_{n+1,n}} h(u)\, dF^{-1}(u) \bigg/ \int_{\varepsilon}^{1-\varepsilon} h(u)\, dF^{-1}(u),$$

where $\nu_{n+1,n} = 1 - \nu_{nn}$.

Observe that $\sum_{i=0}^{n} p_{in} = 1$. For $u \in [\varepsilon, 1-\varepsilon]$ set

$$H_n(u) = \sum_{i=0}^{n} p_{in} g_{\nu_{in}, x_{\nu_{in}}}(u)$$

where $x_{\nu_{in}} = \{2h(\nu_{in})\}^{-1}$ for $0 \le i \le n$.

Also, for $u \in [\varepsilon, 1-\varepsilon]$ set

$$H(u) = F^{-1}(u) \bigg/ \int_{\varepsilon}^{1-\varepsilon} h(\nu)\, dF^{-1}(\nu).$$

CLAIM. Whenever $u \in [\varepsilon, 1-\varepsilon]$ and u is a continuity point of F^{-1}, then $H_n(u) \to H(u)$.

PROOF. First assume that $u \in [\varepsilon, \frac{1}{2})$ is a continuity point of F^{-1}. Now for n large,

$$\begin{aligned} H_n(u) &= \textstyle\sum_{\nu_{in} \ge u} - p_{in}\, x_{\nu_{in}} \\ &= -\textstyle\sum_{\nu_{nn} > \nu_{in} \ge u} \displaystyle\int_{\nu_{in}}^{\nu_{i+1,n}} h(s)\, dF^{-1}(s) \bigg/ \left\{ h(\nu_{in}) \int_{\varepsilon}^{1-\varepsilon} h(\nu)\, dF^{-1}(\nu) \right\} \\ &\quad - \int_{\nu_{nn}}^{\nu_{n+1,n}} h(s)\, dF^{-1}(s) \bigg/ \left\{ 2h(\nu_{nn}) \int_{\varepsilon}^{1-\varepsilon} h(\nu)\, dF^{-1}(\nu) \right\}. \end{aligned}$$

Since h is uniformly continuous and bounded away from zero on $[\varepsilon, 1-\varepsilon]$ and F^{-1} is continuous at u, a standard argument shows that

$$\begin{aligned} &-\textstyle\sum_{\nu_{nn} > \nu_{in} \ge u} \displaystyle\int_{\nu_{in}}^{\nu_{i+1,n}} h(s)\, dF^{-1}(s) \{h(\nu_{in})\}^{-1} - \int_{\nu_{nn}}^{\nu_{n+1,n}} h(s)\, dF^{-1}(s) \{2h(\nu_{nn})\}^{-1} \\ &\qquad = \textstyle\sum_{\nu_{nn} > \nu_{nn} \ge u} \{F^{-1}(\nu_{in}) - F^{-1}(\nu_{i+1,n})\} - \{F^{-1}(\nu_{n+1,n}) - F^{-1}(\nu_{nn})\}/2 + o(1) \\ &\qquad = F^{-1}(u) - F^{-1}(\tfrac{1}{2}) + \{F^{-1}(\tfrac{1}{2}) - F^{-1}(\tfrac{1}{2}+)\}/2 + o(1) = F^{-1}(u) + o(1). \end{aligned}$$

Hence $H_n(u) \to H(u)$. Now let $u \in (\frac{1}{2}, 1-\varepsilon]$ be a continuity point of F^{-1}. For n large,

$$\begin{aligned} H_n(u) &= \textstyle\sum_{1-\nu_{nn} < 1-\nu_{in} < u} \displaystyle\int_{\nu_{in}}^{\nu_{i+1,n}} h(s)\, dF^{-1}(s) \bigg/ \left\{ h(\nu_{in}) \int_{\varepsilon}^{1-\varepsilon} h(\nu)\, dF^{-1}(\nu) \right\} \\ &\quad + \int_{\nu_{nn}}^{\nu_{n+1,n}} h(s)\, dF^{-1}(s) \bigg/ \left\{ 2h(\nu_{nn}) \int_{\varepsilon}^{1-\varepsilon} h(\nu)\, dF^{-1}(\nu) \right\}. \end{aligned}$$

The same argument as above gives

$$\begin{aligned} &\textstyle\sum_{1-\nu_{nn} < 1-\nu_{in} < u} \displaystyle\int_{\nu_{in}}^{\nu_{i+1,n}} h(s)\, dF^{-1}(s) \{h(\nu_{in})\}^{-1} + \int_{\nu_{nn}}^{\nu_{n+1,n}} h(s)\, dF^{-1}(s) \{2h(\nu_{nn})\}^{-1} \\ &\qquad = \textstyle\sum_{1-\nu_{nn} < 1-\nu_{in} < u} \{F^{-1}(\nu_{i+1,n}) - F^{-1}(\nu_{i,n})\} + \{F^{-1}(\nu_{n+1,n}) - F^{-1}(\nu_{nn})\}/2 + o(1) \\ &\qquad = F^{-1}(\tfrac{1}{2}) - F^{-1}(1-u) + \{F^{-1}(\tfrac{1}{2}+) - F^{-1}(\tfrac{1}{2})\}/2 + o(1), \\ &\qquad = F^{-1}(u) + o(1), \end{aligned}$$

since F is symmetric about zero. Hence $H_n(u) \to H(u)$. If $u = ½$ is a continuity of F^{-1} then $H_n(½) = 0$, but in this case $F^{-1}(½) = 0$, so $H_n(½) = H(½) = 0$. An argument very much like that given for Proposition 8.15 on page 165 of Breiman (1968) shows that

$$\left(\int_\varepsilon^{1-\varepsilon}\int_\varepsilon^{1-\varepsilon} J_h(u)J_h(v)(u \wedge v - uv)\, dH_n(u)\, dH_n(v)\right)^{1/2} \tag{7}$$

$$\to \left(\int_\varepsilon^{1-\varepsilon}\int_\varepsilon^{1-\varepsilon} J_h(u)J_h(v)(u \wedge v - uv)\, dH(u)\, dH(v)\right)^{1/2} \tag{8}$$

$$= \left(\int_\varepsilon^{1-\varepsilon}\int_\varepsilon^{1-\varepsilon} J_h(u)J_h(v)(u \wedge v - uv)\, dF^{-1}(u)\, dF^{-1}(v)\right)^{1/2} \Big/ \int_\varepsilon^{1-\varepsilon} h(s)\, dF^{-1}(s). \tag{9}$$

Expression (7) is equal to $\tau(J_h, H_n)$, which by convexity of $\tau(J_h, \cdot)$ is less than or equal to

$$\textstyle\sum_{i=0}^n p_{in}\tau(J_h, g_{\nu_{in},x_{\nu_{in}}}). \tag{10}$$

Observe that for each $0 \le i \le n$

$$\tau(J_h, g_{\nu_{in},x_{\nu_{in}}}) = 2^{-1/2}C(h).$$

Thus, expression (7) and hence expression (9) is $\le 2^{-1/2}C(h)$. Since the limit of expression (9) as $\varepsilon \downarrow 0$ through any sequence of continuity points of F^{-1} is equal to

$$\tau(J_h, F) \Big/ \int_0^1 h(u)\, dF^{-1}(u),$$

we immediately have that (6) is true for all $F^{-1} \in \mathscr{I}_h$ whenever h is continuous and strictly positive.

To show that (6) is true for all $F^{-1} \in \mathscr{I}_h$, for any h satisfying the conditions of the theorem, one only has to slightly modify the above argument to the interior of the region in (0, 1) where h is strictly positive and continuous. This completes the proof of Step 2. □

Proof of Step 3. Let $J \in \mathscr{J}_h$ be such that J and J_h differ at a point $\nu \in (0, 1)$ where they are both continuous. Since $\int_0^1 J(u) = 1$ and J is symmetric about ½, we can assume that $\nu \in (0, ½)$ and $J(\nu) > J_h(\nu)$, for if not, we can always find such a continuity point. We must consider two cases.

Case 1. $J_h(\nu) > 0$.
Choose any $x_\nu \in (0, \infty)$ and the $G_{\nu,x_\nu} \in \mathscr{G}_h$ that corresponds to x_ν. Since $J(\nu) > J_h(\nu) > 0$,

$$R(J, s(h, G_{\nu,x_\nu}), G_{\nu,x_\nu}) = 2^{-1}\nu J^2(\nu)h^{-2}(\nu) > 2^{-1}C^2(h) = R(J_h, s(h, G_{\nu,x_\nu}), G_{\nu,x_\nu}).$$

Case 2. $J_h(\nu) = 0$.
Choose a point $u \in (0, ½)$ such that $J_h(u) > 0$ and h is continuous at u, and choose two points x_ν and $x_u \in (0, \infty)$. Let g_{ν,x_ν} and g_{ν,x_u} be the inverses of G_{u,x_ν} and G_{u,x_u} respectively. Define F to be that distribution which has the inverse $2^{-1} g_{\nu,x_\nu} + 2^{-1} g_{u,x_u}$.

It is easy to check that $F \in \mathscr{S}_h$. Now since $h(\nu) = 0$,

$$s(h, F) = \int_0^1 h(t)\, dF^{-1}(t)$$

$$= 2^{-1} \int_0^1 h(t)\, dg_{u,x_u}(t) + 2^{-1} \int_0^1 h(s)\, dg_{\nu,x_\nu}(s) = x_u h(u).$$

Also,

$$\sigma^2(J, F) = \int_0^1 \int_0^1 J(s)J(t)(s \wedge t - st)\, dF^{-1}(s)\, dF^{-1}(t)$$

$$\geq 4^{-1} \int_0^1 \int_0^1 J(s)J(t)(s \wedge t - st)\, dg_{\nu,x_\nu}(s)\, dg_{\nu,x_\nu}(t)$$

$$= 2^{-1}\nu J^2(\nu)x_\nu^2,$$

and since $J_h(\nu) = 0$, $\sigma^2(J_h, F) = 2^{-1}uJ_h^2(u)x_u^2$. Hence

$$R(J_h, s(h, F), F) = 2^{-1}C^2(h),$$

but

$$R(J, s(h, F), F) \geq \nu J^2(\nu)x_\nu^2/(2x_u^2 h^2(u)).$$

We are free to choose x_u and x_ν so that the right side of the last inequality is as large as desired. This completes the proof of Step 3 and subsequently the proof of the theorem. □

Acknowledgment. The author would like to thank the referee for a careful reading of the manuscript and for useful suggestions that improved the presentation. Also comments by R. Petty were helpful.

REFERENCES

BREIMAN, L. (1968). *Probability*. Addison-Wesley, Reading, Mass.

CHERNOFF, H., GASTWIRTH, J. L. and JOHNS, M. V., JR., (1967). Asymptotic distribution of linear combinations of order statistics. *Ann. Math. Statist.* **38** 57–72.

COLLINS, J. R. (1977). Upper bounds on asymptotic variances of *M*-estimators of location. *Ann. Statist.* **5** 646–657.

GASTWIRTH, J. L. and RUBIN, H. (1969). On robust linear estimators. *Ann. Math. Statist.* **40** 24–39.

HOEFFDING, W. (1973). On the centering of a simple linear rank statistic. *Ann. Statist.* **1** 54–66.

HUBER, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35** 73–102.

HUBER, P. J. (1977). *Robust Statistical Procedures*. S.I.A.M., Philadelphia, Penn.

MASON, D. M. (1981). Asymptotic normality of linear combinations of order statistics with a smooth score function. *Ann. Statist.* **9** 899–904.

MASON, D. M. (1982). Some characterizations of strong laws for linear functions of order statistics. *Ann. Probability* **10** 1051–1057.

ROUSSEEUW, P. J. (1981). A new infinitesimal approach to robust estimation. *Z. Wahrsch. verw. Gebiete* **56** 127–132.

SHORACK, G. R. (1972). Functions of order statistics. *Ann. Math. Statist.* **43** 412–427.

STIGLER, S. M. (1974). Linear functions of order statistics with smooth weight functions. *Ann. Statist.* **2** 676–693.

VAN ZWET, W. R. (1980). A strong law for linear functions of order statistics. *Ann. Probability* **8** 986–990.

WELLNER, J. A. (1977). A Glivenko-Cantelli theorem and strong laws of large numbers for functions of order statistics. *Ann. Statist.* **5** 473–480.

DEPARTMENT OF MATHEMATICAL SCIENCES
501 EWING HALL
UNIVERSITY OF DELAWARE
NEWARK, DELAWARE 19711