The Annals of Statistics
1988, Vol. 16, No. 2, 619–628

THE LENGTH OF THE SHORTH[1]

BY R. GRÜBEL

Imperial College of Science and Technology, University of London

Let $\hat{H}_n(\alpha)$ $(0 < \alpha < 1)$ denote the length of the shortest α-fraction of the ordered sample $X_{1:n}, X_{2:n}, \ldots, X_{n:n}$, i.e.,

$$\hat{H}_n(\alpha) = \min\{X_{k+j:n} - X_{k:n}: 1 \le k \le k+j \le n;\ (j+1)/n \ge \alpha\}.$$

Such quantities arise in the context of robust scale estimation. Using the concept of compact derivatives of statistical functionals, the asymptotic behaviour of $\hat{H}_n(\alpha)$ as $n \to \infty$ is investigated.

1. Introduction. One of the proposals of Andrews, Bickel, Hampel, Huber, Rogers and Tukey (1972) for the robust estimation of location is the shorth, the mean of those data points which constitute the shortest half or, more generally, the shortest α-fraction $(0 < \alpha < 1)$ of the sample. It turned out (Andrews, Bickel, Hampel, Huber, Rogers and Tukey [(1972), page 50] and Shorack and Wellner [(1986), page 767)]) that its asymptotic rate is only $n^{-1/3}$ and also that the limiting distribution is not normal.

In the present paper a similar procedure is investigated in the light of scale estimation: We consider the length of the shortest interval which contains at least a fraction α of the sample. Interestingly, the asymptotic rate is $n^{-1/2}$ now and the limiting distribution is normal. The resulting scale estimators are outlier resistant and have the desirable invariance and equivariance properties. Differences of symmetrically located order statistics are often used as robust scale estimators. Shortest α-fractions yield estimators with the same asymptotic behaviour, but with breakdown point twice as high; see Section 4.1 for details.

The length of the shortest α-fraction may be regarded as a functional of the empirical distribution function; an asymptotic normality result for the latter exists. We obtain our result by decomposing this functional into two factors which are sufficiently smooth near the respective limits to permit local replacement by linear operators (Propositions 8 and 9). To make this rigorous we use the concept of compact differentiation of statistical functionals introduced by Reeds (1976). Indeed, one motivation for this paper is to give another example of the simplicity and usefulness of this method; see also Fernholz (1983), Reid (1981), Esty, Gillette, Hamilton and Taylor (1985) and Gill (1986).

Section 2 introduces the indispensable notation and states the result which is then proved in Section 3. In the final section we compare our estimators with established procedures and comment on assumptions and extensions.

2. Notation and the result. Let $X_1, X_2, \ldots$ be independent and identically distributed random variables with distribution function F and density f. Further

Received February 1987; revised October 1987.

[1] Research supported by the Nuffield foundation.

AMS 1980 *subject classifications*. Primary 62G05; secondary 62E20, 62F35, 60F17.

Key words and phrases. Robust scale estimation, asymptotic normality, statistical functionals, compact derivative, empirical concentration function, breakdown point.

let $0 < \alpha_0 < \alpha_1 < 1$ be given. We assume that f satisfies the following conditions:

$$f > 0,\ f \text{ increasing on } (-\infty, 0) \text{ and symmetric about } 0,$$

$$(1)\qquad f' \text{ exists and } f' \geq c_0 \text{ on } (a, b), \text{ where } c_0 > 0 \text{ and } -\infty < a < b < 0,$$

$$2F(a) < 1 - \alpha_1, 1 - \alpha_0 < 2F(b).$$

Symmetry about 0 is of course assumed for notational convenience alone. Since all procedures considered here are shift invariant, 0 can be replaced by an arbitrary real number without changing the results; see also Section 4.3. Under these conditions F and the corresponding concentration function G,

$$G(\lambda) = \sup\{F(x + \lambda) - F(x): x \in \mathbb{R}\},$$

have unique inverses,

$$F^{-1}(y) = \inf\{x: F(x) \geq y\}, \qquad 0 < y < 1,$$
$$G^{-1}(\alpha) = \inf\{\lambda > 0: G(\lambda) \geq \alpha\}, \qquad 0 < \alpha < 1,$$

respectively. Let $H = G^{-1}$ and let $\hat{F}_n$, $\hat{G}_n$ and $\hat{H}_n$ denote the corresponding empirical quantities,

$$\hat{F}_n(x) = n^{-1}\#\{i: 1 \leq i \leq n,\ X_i \leq x\},$$
$$\hat{G}_n(\lambda) = \sup\{\hat{F}_n(x + \lambda) - \hat{F}_n(x): x \in \mathbb{R}\},$$
$$\hat{H}_n(\alpha) = \inf\{\lambda > 0: \hat{G}_n(\lambda) \geq \alpha\}.$$

Thus $\hat{F}_n$ is the empirical distribution function and $\hat{G}_n$ the empirical concentration function with inverse $\hat{H}_n$. Obviously $\hat{H}_n(\alpha)$ is the shortest α-part of the sample $X_1, \ldots, X_n$, i.e.,

$$\hat{H}_n(\alpha) = \inf\{X_{i+j:n} - X_{i:n}: 1 \leq i \leq i + j \leq n, (j+1)/n \geq \alpha\},$$

where $X_{1:n}, \ldots, X_{n:n}$ is the ordered sample. For any closed interval I in the usual two-point compactification of the real line, let $D_r(I)$ be the set of all functions $f\colon I \to \mathbb{R}$ which are continuous from the right and possess left limits. Similarly $D_l(I)$ denotes the set of all left continuous functions with right limits. This implies that, for example, for $I = [-\infty, \infty]$, $f(x)$ tends to a finite limit as $x \to \infty$ or $x \to -\infty$ for all $f \in D_l(I), D_r(I)$. Endow these spaces with the Skorohod topology; see Billingsley [(1968), Chapter 3] for the case $D_r([0,1])$. In the results given here, weak convergence $\to_{\mathscr{D}}$ refers to this construction. The following is a slight reformulation of a by now classical result.

THEOREM 1. [See, e.g., Billingsley (1968), Theorem 16.4.]

$$n^{1/2}(\hat{F}_n - F) \to_{\mathscr{D}} Z^F \quad \text{on } D_r([-\infty, \infty]),$$

where Z^F is a Gaussian process with mean $EZ_x^F = 0$ and covariance

$$\operatorname{cov}(Z_x^F, Z_y^F) = F(x \wedge y) - F(x)F(y).$$

We are interested in a similar result for $\hat{H}_n$, regarded as a random element with values in $D_l([\alpha_0, \alpha_1])$. Choose $0 < \lambda_0 < \lambda_1$ such that $2F(a) < 2F(-\lambda_1/2) < 1 - \alpha_1$, $1 - \alpha_0 < 2F(-\lambda_0/2) < 2F(b)$. The following is an intermediate result which might be of interest in its own right—it implies asymptotic normality for the largest number of order statistics that can be fitted into an interval of given length.

THEOREM 2.

$$n^{1/2}\big(\hat{G}_n - G\big) \to_{\mathscr{D}} Z^G \quad on\ D_l([\lambda_0, \lambda_1]),$$

where Z^G is a Gaussian process with mean $EZ_\lambda^G = 0$ and covariance

$$\operatorname{cov}\big(Z_\lambda^G, Z_\mu^G\big) = 3F(\lambda/2) + 3F(\mu/2) - 4F(\lambda/2)F(\mu/2) + F((\lambda \wedge \mu)/2) - F((\lambda \vee \mu)/2) - 2.$$

The following is our main result.

THEOREM 3.

$$n^{1/2}\big(\hat{H}_n - H\big) \to_{\mathscr{D}} Z^H \quad on\ D_l([\alpha_0, \alpha_1]),$$

where Z^H is a Gaussian process with mean $EZ_\alpha^H = 0$ and covariance

$$\operatorname{cov}\big(Z_\alpha^H, Z_\beta^H\big) = \frac{\alpha + \beta - 2\alpha\beta + \alpha \wedge \beta - \alpha \vee \beta}{2f\big(F^{-1}((1+\alpha)/2)\big)f\big(F^{-1}((1+\beta)/2)\big)}.$$

3. Proofs.

3.1. We start with a few explanatory comments on the method used; see Reeds (1976), Fernholz (1983) or Gill (1986) for a detailed account.

We plan to "reduce" Theorem 3 to Theorem 1 via Theorem 2. In the more familiar real-valued setting, the following could be used. Given random variables $\xi_0, \xi_1, \ldots$, a constant $a \in \mathbb{R}$ and a measurable function $g: \mathbb{R} \to \mathbb{R}$ which is differentiable at a, we know that $n^{1/2}(\xi_n - a) \to_{\mathscr{D}} \xi_0$ implies $n^{1/2}(g(\xi_n) - g(a)) \to_{\mathscr{D}} g'(a)\xi_0$. An analogue of this in a more abstract setting requires a suitable differentiation concept first.

Let B_1, B_2 be Banach spaces, $a \in B_1$ and $\Phi: B_1 \to B_2$ a function.

DEFINITION 4. The function Φ is *compact* (or *Hadamard*) *differentiable* at a if there exists a linear and continuous function $\Phi_a': B_1 \to B_2$ such that

$$\lim_{h \to 0} \sup_{x \in K} \left\| \frac{1}{h}[\Phi(a + hx) - \Phi(a)] - \Phi_a'(x) \right\| = 0$$

for all compact $K \subset B_1$.

For Proposition 5, let $X, X_1, X_2, \ldots$ be B_1-valued random variates, i.e., measurable with respect to the σ-field generated by the norm topology. Let τ be a second topology on B_2 with Borel σ-field $\mathscr{H}$ and the properties that every τ-open set is $\|\cdot\|$-open and that for fixed $a \in B_2$, $\alpha \in \mathbb{R}$, the functions

$$x \to x + a, \qquad x \to \alpha x$$

are $(\mathscr{H}, \mathscr{H})$-measurable. We further assume that all $\Phi \circ X_n$, $n \in \mathbb{N}$, are $\mathscr{H}$-measurable; on B_2, $\to_{\mathscr{D}}$ refers to τ.

PROPOSITION 5. *Assume that Φ is compact differentiable at a and that B_1 is separable. Then $n^{1/2}(X_n - a) \to_{\mathscr{D}} X$ implies $n^{1/2}[\Phi(X_n) - \Phi(a)] \to_{\mathscr{D}} \Phi_a'(X)$.*

PROOF. On a suitable probability space there exist random variates X_n' and X' with the same distribution as X_n and X, respectively, such that $n^{1/2}(X_n' - a) \to X'$ on a set A of probability 1 [this is the Skorohod–Dudley construction; see, e.g., Pollard (1984), Theorem 4.3.13]. For each $\omega \in A$ the set $\{n^{1/2}(X_n'(\omega) - a): n \in \mathbb{N}\} \cup \{X'(\omega)\}$ is compact, therefore the compact differentiability of Φ at a yields $n^{1/2}(\Phi(X_n'(\omega)) - \Phi(a)) \to \Phi_a'(X'(\omega))$ in $\|\cdot\|$ and hence also in τ. This implies $n^{1/2}(\Phi(X_n') - \Phi(a)) \to_{\mathscr{D}} \Phi_a'(X')$ and the assertion follows from the equality of the respective distributions. □

3.2. On $C(I) = D_r(I) \cap D_l(I)$, the Skorohod topology coincides with the topology generated by the supremum norm,

$$\|f\|_\infty = \sup\{|f(x)|: x \in I\},$$

which makes $C(I)$ a separable Banach space. We avoid the measurability or separability problems associated with $D_r(I)$ by replacing $\hat{F}_n$, which takes its values in $D_r(I) - C(I)$, by a sufficiently close continuous function: Let $\tilde{F}_n \in C([-\infty, \infty])$ be such that

$$\tilde{F}_n(X_i) = \hat{F}_n(X_i), \qquad 1 \le i \le n.$$

It may further be assumed that $\tilde{F}_n$ is a distribution function and *strictly* increasing on $\{\tilde{F}_n < 1\}$. Put

$$\tilde{G}_n(\lambda) = \sup\{\tilde{F}_n(x + \lambda) - \tilde{F}_n(x): x \in \mathbb{R}\}, \qquad \tilde{H}_n(\alpha) = \inf\{\lambda > 0: \tilde{G}_n(\lambda) \ge \alpha\}.$$

Theorem 1 holds with $\tilde{F}_n$ instead of $\hat{F}_n$; also $D_r([-\infty; \infty])$ can then be replaced by $C([-\infty, \infty])$.

LEMMA 6. *$n\|\tilde{G}_n - \hat{G}_n\|_\infty \le 2$ on a set of probability 1.*

PROOF. The set of those points in the underlying probability space for which no two random variables X_i coincide has probability 1, and on this set $\hat{F}_n$ and $\tilde{F}_n$ differ by at most $2/n$ in the mass they give to intervals. □

On D-spaces the $\|\cdot\|_\infty$-topology is finer than the Skorohod topology. As a first consequence of this lemma, it follows that $n^{1/2}(\tilde{G}_n - \hat{G}_n) \to_P 0$ with respect to the latter. To prove Theorem 2, it is therefore enough to show $n^{1/2}(\tilde{G}_n - G) \to_{\mathscr{D}} Z^G$ in $(C([\lambda_0, \lambda_1]), \|\cdot\|_\infty)$.

3.3. We need a technical lemma first. It shows that intervals of a given length λ with almost maximal probability must lie close to the optimal interval $[-\frac{1}{2}\lambda, \frac{1}{2}\lambda]$.

LEMMA 7. *Let $\varepsilon > 0$ be such that $a + \varepsilon < -\frac{1}{2}\lambda_1$, $-\frac{1}{2}\lambda_0 < b - \varepsilon$ and put*

$$u(\lambda, x) = F(\tfrac{1}{2}\lambda + x) - F(-\tfrac{1}{2}\lambda + x).$$

Then for all $\lambda \in [\lambda_0, \lambda_1]$ and all $x \in \mathbb{R}$, $u(\lambda, x) \geq u(\lambda, 0) - \frac{1}{2}c_0\varepsilon^2$ implies

$$|x| \leq \left[c_0^{-1}(u(\lambda,0) - u(\lambda, x))\right]^{1/2}.$$

PROOF. Because of $u(\lambda, x) = F(\frac{1}{2}\lambda + x) + F(\frac{1}{2}\lambda - x) - 1 = u(\lambda, -x)$, it suffices to consider $x > 0$. Also

$$(\partial/\partial x)u(\lambda, x) = f(\tfrac{1}{2}\lambda + x) - f(-\tfrac{1}{2}\lambda + x) = f(\tfrac{1}{2}\lambda + x) - f(\tfrac{1}{2}\lambda - x);$$

thus $x \to u(\lambda, x)$ is decreasing on $(0, \infty)$.

For all $x \in [0, \varepsilon]$, $\lambda \in [\lambda_0, \lambda_1]$,

$$\begin{aligned} &u(\lambda,0) - u(\lambda,x) \\ (2)\qquad &= \int_0^x \left(f(-\tfrac{1}{2}\lambda + y) - f(-\tfrac{1}{2}\lambda - y)\right) dy \geq c_0 \int_0^x 2y\,dy = c_0 x^2. \end{aligned}$$

Assume now that $u(\lambda, x) \geq u(\lambda, 0) - \frac{1}{2}c_0\varepsilon^2$. Since u is decreasing in x, the inequality $x \geq \varepsilon$ would imply

$$u(\lambda, \varepsilon) \geq u(\lambda, x) \geq u(\lambda, 0) - \tfrac{1}{2}c_0\varepsilon^2,$$

but (2) yields

$$u(\lambda, \varepsilon) \leq u(\lambda, 0) - c_0\varepsilon^2.$$

So we must have $x < \varepsilon$, and (2) applies and yields the assertion. □

The modulus of continuity of a function $g \in C(I)$ is given by

$$\omega(g, \lambda) = \sup\{|g(x) - g(y)|: x, y \in I, |x - y| \leq \lambda\}.$$

Define Φ: $C([-\infty, \infty]) \to C([\lambda_0, \lambda_1])$ by

$$(\Phi(g))(\lambda) = \omega(g, \lambda), \qquad \lambda_0 \leq \lambda \leq \lambda_1.$$

Obviously $\Phi(F)$ is the restriction of G to $[\lambda_0, \lambda_1]$ and, in the same sense, Φ applied to $\tilde{F}_n$ yields $\tilde{G}_n$. It is easy to show that Φ is continuous.

PROPOSITION 8. *The function Φ is compact differentiable at F and*

$$\Phi_F'(g)(\lambda) = g(\tfrac{1}{2}\lambda) - g(-\tfrac{1}{2}\lambda) \quad \textit{for all } g \in C([-\infty, \infty]),\ \lambda \in [\lambda_0, \lambda_1].$$

PROOF. Given $t \neq 0$, $\lambda \in [\lambda_0, \lambda_1]$ and $g \in C([-\infty, \infty])$, there exist $x(t, g, \lambda)$ and $y(t, g, \lambda)$ such that

$$|x(t, g, \lambda) - y(t, g, \lambda)| \leq \lambda,$$
$$\Phi(F + tg) = (F + tg)(x(t, g, \lambda)) - (F + tg)(y(t, g, \lambda));$$

for fixed g, λ and t small enough we must have $x(t,g,\lambda) > y(t,g,\lambda)$. By definition of T,

$$(3)\qquad \begin{aligned}&(F+tg)(\tfrac{1}{2}\lambda) - (F+tg)(-\tfrac{1}{2}\lambda)\\ &\quad\le (F+tg)(x(t,g,\lambda)) - (F+tg)(y(t,g,\lambda)),\end{aligned}$$

which implies

$$\begin{aligned}F(x(t,g,\lambda)) - F(x(t,g,\lambda)-\lambda) &\ge F(x(t,g,\lambda)) - F(y(t,g,\lambda))\\ &\ge F(\tfrac{1}{2}\lambda) - F(-\tfrac{1}{2}\lambda) - 4t\|g\|_\infty,\end{aligned}$$

i.e., in the notation of Lemma 7,

$$u(\lambda, x(t,g,\lambda) - \tfrac{1}{2}\lambda) \ge u(\lambda, 0) - 4t\|g\|_\infty.$$

Now let $K \subset C([-\infty,\infty])$ be compact. Since it is bounded, Lemma 7 gives

$$(4)\qquad \limsup_{t\to 0}\ \sup_{g\in K}\ \sup_{\lambda_0\le\lambda\le\lambda_1} |x(t,g,\lambda) - \tfrac{1}{2}\lambda| = 0.$$

The same arguments yield

$$(5)\qquad \limsup_{t\to 0}\ \sup_{g\in K}\ \sup_{\lambda_0\le\lambda\le\lambda_1} |y(t,g,\lambda) + \tfrac{1}{2}\lambda| = 0.$$

Using the definition of Φ again, we see

$$F(\tfrac{1}{2}\lambda) - F(-\tfrac{1}{2}\lambda) \ge F(x(t,g,\lambda)) - F(y(t,g,\lambda)),$$

which, together with (3), gives

$$\begin{aligned}&\left|1/t\left[\left(F(y(t,g,\lambda)) - F(-\tfrac{1}{2}\lambda)\right) - \left(F(x(t,g,\lambda)) - F(\tfrac{1}{2}\lambda)\right)\right]\right|\\ &\quad\le |g(y(t,g,\lambda)) - g(-\tfrac{1}{2}\lambda)| + |g(x(t,g,\lambda)) - g(\tfrac{1}{2}\lambda)|.\end{aligned}$$

Therefore

$$\begin{aligned}&\left|1/t[\Phi(F+tg) - \Phi(F)](\lambda) - \left(g(\tfrac{1}{2}\lambda) - g(-\tfrac{1}{2}\lambda)\right)\right|\\ &\quad\le |g(x(t,g,\lambda)) - g(\tfrac{1}{2}\lambda)| + |g(y(t,g,\lambda)) - g(-\tfrac{1}{2}\lambda)|\\ &\qquad + \left|1/t\left[\left(F(y(t,g,\lambda)) - F(-\tfrac{1}{2}\lambda)\right) - \left(F(x(t,g,\lambda)) - F(\tfrac{1}{2}\lambda)\right)\right]\right|\\ &\quad\le 2|g(x(t,g,\lambda)) - g(\tfrac{1}{2}\lambda)| + 2|g(y(t,g,\lambda)) - g(-\tfrac{1}{2}\lambda)|.\end{aligned}$$

Now the following property of compact subsets of $(C(I), \|\cdot\|_\infty)$ is decisive:

$$\limsup_{\delta\to 0}\ \sup_{g\in K} \omega(g,\delta) = 0;$$

see, for example, Billingsley [(1968), page 221]. Because of (4) and (5), we can use this to obtain from the preceding estimate

$$\limsup_{t\to 0}\ \sup_{g\in K}\ \sup_{\lambda_0\le\lambda\le\lambda_1} \left|1/t[\Phi(F+tg) - \Phi(F)](\lambda) - (g(\lambda/2) - g(-\lambda/2))\right| = 0,$$

which is the assertion of the proposition. □

PROOF OF THEOREM 2. A linear transformation of a Gaussian process is Gaussian again; Theorem 2 now follows on combining Theorem 1, Proposition 5 (τ is taken to be the norm topology) and Proposition 8. □

3.4. We have $G(\lambda_0) < \alpha_0 < \alpha_1 < G(\lambda_1)$; note that G has a strictly positive derivative on (λ_0, λ_1). Define Ψ: $C([\lambda_0, \lambda_1]) \to D_l([\alpha_0, \alpha_1])$ by

$$\Psi(g)(\alpha) = \begin{cases} \inf\{\lambda: g(\lambda) \geq \alpha\}, & \text{if } \{\cdots\} \neq \varnothing, \\ \lambda_1, & \text{if } \{\cdots\} = \varnothing. \end{cases}$$

The Skorohod σ-field $\mathscr{H}$ on $D_l(I)$ is generated by the projections π_t, $t \in I$, where $\pi_t(h) = h(t)$ [Billingsley (1968), Theorem 14.5]. Since

$$\Psi(g)(\alpha) > \lambda \Leftrightarrow g(\mu) < \alpha \quad \text{for all } \mu \leq \lambda,$$

the sets $(\pi_t \circ \Psi)^{-1}((\lambda, \infty))$ are open in $(C([\lambda_0, \lambda_1]), \|\cdot\|_\infty)$; hence Ψ is $\mathscr{H}$-measurable.

PROPOSITION 9. *The function Ψ is compact differentiable at G and*

$$\Psi_G'(g)(\alpha) = -g(G^{-1}(\alpha))/G'(G^{-1}(\alpha)) \quad \textit{for all } g \in C([\lambda_0, \lambda_1]),\ \alpha \in [\alpha_0, \alpha_1].$$

A similar result for α fixed and $G(x) = x$ has already been given by Reeds [(1976), Section 6.2.4]; see Gill (1986) for more general G. In a slightly different framework this result has been obtained by Esty, Gillette, Hamilton and Taylor (1985), so we omit the proof of Proposition 9.

PROOF OF THEOREM 3. Using Proposition 5 with $B_2 = D_l([\alpha_0, \alpha_1])$, τ the Skorohod topology, we obtain

$$n^{1/2}(\tilde{H}_n - H) \to_{\mathscr{D}} \Psi_G'(Z^G).$$

This limit is easily seen to have the structure displayed in Theorem 3.

Using Theorem 4.1 of Billingsley (1968), we see that it remains to show

$$\rho(n^{1/2}(\hat{H}_n - \tilde{H}_n), 0) \to_P 0,$$

where ρ is a metric for τ; we may assume $\rho(f, g) \leq \|f - g\|_\infty$.

Let $\varepsilon, \delta > 0$ be given. Since $\tilde{F}_n$ is strictly increasing on $\{\hat{F}_n < 1\}$, $\tilde{G}_n$ is strictly increasing and $\tilde{H}_n$ thereby continuous; thus, $n^{1/2}(\tilde{H}_n - H)$ takes its values in $C = C([\alpha_0, \alpha_1])$. We can therefore find a compact set $K = K(\delta)$ in $(C, \|\cdot\|_\infty)$ such that

$$P(n^{1/2}(\tilde{H}_n - H) \in K) \geq 1 - \delta \quad \text{for all } n \in \mathbb{N}.$$

Further there exists an $\eta = \eta(\varepsilon, K)$ such that $\omega(f, 2\eta) < \frac{1}{2}\varepsilon$ for all $f \in K$. H is continuously differentiable in some open interval containing $[\alpha_0, \alpha_1]$, so there exists an $n_0 = n_0(\eta, \varepsilon)$ such that $n_0 > 1/\eta$ and $n^{1/2}\omega(H, 2/n) < \frac{1}{2}\varepsilon$ for all $n \geq n_0$. Lemma 6, together with some elementary manipulations, yields

$$\|n^{1/2}(\hat{H}_n - \tilde{H}_n)\|_\infty \leq \omega(n^{1/2}(\tilde{H}_n - H), 2/n) + n^{1/2}\omega(H, 2/n);$$

hence choosing n_0 as described, we obtain for all $n \geq n_0$,

$$P\left(\rho\left(n^{1/2}(\hat{H}_n - \tilde{H}_n), 0\right) > \varepsilon\right) < \delta. \qquad \square$$

4. Comparisons and extensions.

4.1. A common class of scale estimates in the symmetric case is based on $X_{n-r:n} - X_{r+1:n}$, the difference of two "symmetrically located" order statistics [Mosteller (1946), Section 3.B]; the semiinterquartile range, for example, is often used in connection with Cauchy distributions. It turns out that this procedure is asymptotically equivalent to our method in the sense that the limiting Gaussian distributions are the same if r depends on n such that $r/n \to (1-\alpha)/2$, $0 < \alpha < 1$, as $n \to \infty$. This seems to indicate that, for example, the right endpoint of the shorth and the upper quartile differ only by $o_P(n^{-1/2})$. However the center of the shorth has rate $n^{-1/3}$ [Rousseeuw (1983)], whereas the upper quartile behaves as $n^{-1/2}$.

Using this asymptotic equivalence and the results of Mosteller (1946), we obtain the behaviour of our estimators in the case of normal distributions: As an estimator for σ an appropriately scaled shortest α-fraction has asymptotic relative efficiency 0.37 with respect to the sample standard deviation if $\alpha = 0.5$; the best α is 0.86 with efficiency about 0.65. Linear combinations of these estimators with different α's yield even better values. The functional character of our limit result is important in this context; see Chernoff, Gastwirth and Johns [(1967), Section 3]. For the Cauchy family $\alpha = 0.5$ is optimal, with efficiency of 0.81 [use Haas, Bain and Antle (1970), Section 8].

Given an estimator T and a sample $\mathbf{x} = (x_1, \ldots, x_n)$, let $\beta(T, \mathbf{x}, p)$ denote the supremum of all distances $\|T(\mathbf{x}) - T(\mathbf{y})\|$, where the "corrupted" sample $\mathbf{y} = (y_1, \ldots, y_n)$ is obtained from $\mathbf{x}$ by changing at most a proportion p of the values $x_1, \ldots, x_n$. The (finite sample) breakdown point of T at $\mathbf{x}$ is defined by

$$\varepsilon_n^*(T, \mathbf{x}) = \inf\{p: \beta(T, \mathbf{x}, p) = \infty\}.$$

This notion is due to Hampel (1971) and has received some interest recently [see, for example, Rousseeuw (1983), Huber (1984) and Davies (1987)].

Scale estimators based on $X_{n-r:n} - X_{r+1:n}$ with $r \sim qn$ have $\varepsilon_n^* \sim q$. Our corresponding (i.e., same asymptotic variance) estimator is based on the length of the shortest α-fraction with $\alpha = 1 - 2q$, its breakdown point being $2q$ [see also Rousseeuw and Leroy (1988)]. For scale estimation a realistic notion of finite sample breakdown should also include the possibility of "implosions," i.e., of $T(\mathbf{x})$ becoming too small [Davies (1987), Section 4]; this means that we should take $q < 1/4$ or equivalently $\alpha > 1/2$.

4.2. Gill's concept of compact differentiation tangentially to a subspace [Gill (1986)] and a suitable modification of the notion of weak convergence [Pollard (1984)] can be used to avoid the introduction of the Skorohod topology. In this setting it is also possible to tackle the compound functional $\Psi \circ \Phi$ directly.

4.3. Our approach also carries over to nonsymmetric situations. Technically we only have to assume that the location of $(x(\lambda), y(\lambda))$ is marked enough (Section 3.3). Consider the following example: $X_1, \dots, X_n$ are independent with common density

$$f(x; \mu, \sigma) = \sigma^{-1}\exp[-\sigma^{-1}(x - \mu)]\, I_{(\mu, \infty)}(x),$$

where $\sigma > 0$ and $\mu \in \mathbb{R}$ are unknown and σ is the parameter of interest. The maximum likelihood estimator is

$$\hat{\sigma}_n(x_1, \dots, x_n) = n^{-1} \sum_1^n x_i - \min\{x_1, \dots, x_n\},$$

$n^{1/2}(\hat{\sigma}_n - \sigma) \to_{\mathscr{D}} N(0, \sigma^2)$ and $\{\hat{\sigma}_n\}$ is asymptotically efficient; see Lehmann (1983), Example 6.6.10. The robustness performance of this estimator is devastating; it is even more sensitive to outliers than the notoriously nonrobust location estimator, the mean.

The family $\{f(\cdot; \mu, \sigma)\}$ arises via affine transformations from a fixed density f which vanishes on $(-\infty, 0)$ and is strictly decreasing on $(0, \infty)$. Our approach applies to this situation as well, resulting in $n^{1/2}(\hat{H}_n - H) \to_{\mathscr{D}} Z^H$, where Z^H is Gaussian with mean 0 and $\operatorname{cov}(Z^H_\alpha, Z^H_\beta) = (\alpha \wedge \beta - \alpha\beta)/[f(F^{-1}(\alpha))f(F^{-1}(\beta))]$. In the preceding shifted exponential case the best asymptotic relative efficiency with respect to $\{\hat{\sigma}_n\}$ is obtained for $\alpha = 0.795$; it is 0.65 and the corresponding breakdown point is 0.205.

Acknowledgments. I would like to thank Laurie Davies for many interesting discussions and for bringing this problem to my attention. Charles Goldie and a referee provided helpful comments on the first draft of the paper.

REFERENCES

ANDREWS, D. F., BICKEL, P. J., HAMPEL, F. R., HUBER, P. J., ROGERS, W. H. and TUKEY, J. W. (1972). *Robust Estimation of Location: Survey and Advances*. Princeton Univ. Press, Princeton, N.J.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

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

DAVIES, P. L. (1987). Asymptotic behaviour of S-estimates of multivariate location parameters and dispersion matrices. *Ann. Statist.* **15** 1269–1292.

ESTY, W., GILLETTE, R., HAMILTON, M. and TAYLOR, D. (1985). Asymptotic distribution theory of statistical functionals: The compact derivative approach for robust estimators. *Ann. Inst. Statist. Math.* **37** 109–129.

FERNHOLZ, L. T. (1983). *von Mises Calculus for Statistical Functionals. Lecture Notes in Statist.* **19**. Springer, Berlin.

GILL, R. D. (1986). Non- and semi-parametric maximum likelihood estimators and the von Mises method (Part I). *Scand. J. Statist.* To appear.

HAAS, G., BAIN, L. and ANTLE, C. (1970). Inferences for the Cauchy distribution based on maximum likelihood estimators. *Biometrika* **57** 403–408.

HAMPEL, F. R. (1971). A general qualitative definition of robustness. *Ann. Math. Statist.* **42** 1887–1896.

HUBER, P. J. (1984). Finite sample breakdown of M- and P-estimators. *Ann. Statist.* **12** 119–126.
LEHMANN, E. L. (1983). *Theory of Point Estimation*. Wiley, New York.
MOSTELLER, F. (1946). On some useful "inefficient" statistics. *Ann. Math. Statist.* **17** 377–408.
POLLARD, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.
REEDS, J. A. (1976). On the definition of von Mises functionals. Ph.D. thesis, Harvard Univ.
REID, N. (1981). Influence functions for censored data. *Ann. Statist.* **9** 78–92.
ROUSSEEUW, P. J. (1983). Multivariate estimation with high breakdown point. *Fourth Pannonian Symposium on Mathematical Statistics*. Bad Tatzmannsdorf, Austria.
ROUSSEEUW, P. J. and LEROY, A. (1988). A robust scale estimator based on the shortest half. *Statist. Neerlandica*. To appear.
SHORACK, G. R. and WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.

DEPARTMENT OF MATHEMATICS
IMPERIAL COLLEGE OF SCIENCE AND TECHNOLOGY
HUXLEY BUILDING
180 QUEEN'S GATE
LONDON SW7 2BZ
UNITED KINGDOM