The Annals of Statistics
1986, Vol. 14, No. 2, 724–732

MINIMAX VARIANCE *M*-ESTIMATORS OF LOCATION IN KOLMOGOROV NEIGHBOURHOODS[1]

By Doug Wiens

Dalhousie University

We exhibit those distributions with minimum Fisher information for location in various Kolmogorov neighbourhoods $\{F|\sup_x|F(x) - G(x)| \leq \varepsilon\}$ of a fixed, symmetric distribution G. The associated M-estimators are then most robust (in Huber's minimax sense) for location estimation within these neighbourhoods. The previously obtained solution of Huber (1964) for $G = \Phi$ and "small" ε is shown to apply to all distributions with strongly unimodal densities whose score functions satisfy a further condition. The "large" ε solution for $G = \Phi$ of Sacks and Ylvisaker (1972) is shown to apply under much weaker conditions. New forms of the solution are given for such distributions as "Student's" t, with nonmonotonic score functions. The general form of the solution is discussed.

1. Introduction and summary. Consider Huber's (1964) theory of robust M-estimation of a location parameter θ. Let $\hat{\theta}$ be defined as a zero of $\sum_1^n \psi(x_i - \cdot)$, for a suitably chosen ψ, where $X_i \sim F(x - \theta)$ and F is an unknown member of a convex, vaguely compact class $\mathscr{F}$ of distributions. Typically, $\sqrt{n}(\hat{\theta} - \theta)$ is asymptotically normally distributed. Let $V(\psi, F)$ denote the asymptotic variance functional. The choice ψ_0 is then *most robust*, in the minimax sense, if it minimizes $\sup_{\mathscr{F}} V(\psi, F)$.

In Huber (1964) and in particular in Chapter 4 of Huber (1981), general procedures are derived for finding most robust M-estimators. We briefly summarize what are, for us, the salient features. One first demands optimality only over that subclass $\mathscr{F}'$ of $\mathscr{F}$ whose members have finite Fisher information for location $I(F)$. Any $F \in \mathscr{F}'$ necessarily has an absolutely continuous, bounded density f, tending to 0 as $x \to \pm\infty$, and then $I(F) = \int (f'/f)^2 f\,dx$. There exists $F_0 \in \mathscr{F}'$ minimizing $I(F)$. If $I(F_0) > 0$, and f_0 has convex support, then F_0 is unique. Furthermore, $\psi_0 = -f_0'/f_0$ is most robust over $\mathscr{F}'$. If $\mathscr{F}'$ is vaguely dense in $\mathscr{F}$, and if ψ_0 is sufficiently regular—see Theorem 5 of Huber (1964)—then ψ_0 is optimal over the larger class. Necessary and sufficient for F_0 to minimize $I(F)$ is the condition

$$\int 2(f_0 - f)'\psi_0 + (f_0 - f)\psi_0^2\,dx \geq 0, \quad \text{all } F \in \mathscr{F}'. \tag{1.1}$$

In this paper we apply the above theory to cases in which $\mathscr{F}$, written K_ε, is a Kolmogorov neighbourhood of a fixed distribution G: $K_\varepsilon = \{F|\sup_x|F(x) - G(x)| \leq \varepsilon\}$. In the case $G = \Phi$, the normal cumulative, Huber (1964) obtained the most robust ψ_0 for $\varepsilon \leq 0.0303$, Sacks and Ylvisaker (1972) for $\varepsilon \geq 0.0303$.

Received August 1984; revised July 1985.

[1]Research supported by Natural Sciences and Engineering Research Council of Canada grant A-8603.

AMS 1980 *subject classifications*. Primary 62G35; secondary 62G05.

Key words and phrases. Robust estimation, M-estimators, minimax variance, Kolmogorov neighbourhood, minimum Fisher information.

Somewhat surprisingly, the general case of this problem seems not to have been addressed.

We will assume throughout that G is fully stochastic, symmetric, strictly increasing on $(-\infty, \infty)$, and has an absolutely continuous density g with respect to Lebesgue measure. The score function $\xi = -g'/g$ is assumed to be differentiable except possibly at zero. The assumption of symmetry implies that F_0 is symmetric [Huber (1981), page 89]. Although it is not assumed that $I(G) < \infty$, the continuity of G ensures that K_ε' is dense in K_ε [Vandelinde (1979), page 186].

Huber (1964) showed that the most robust ψ_0 has essentially the same form for all ε-contamination classes $\{F = (1-\varepsilon)G + \varepsilon H;\ G'$ symmetric, strongly unimodal$\}$. In contrast, we will show that the aforementioned "small ε" solution for K_ε does not extend in this way, but that it *does* apply in the presence of the requirement—strictly stronger than strong unimodality—that $J(\xi) = 2\xi' - \xi^2$ be decreasing on $[0, \infty)$. Note that, under the requisite regularity, (1.1) becomes $\int J(\psi_0)\,d(F - F_0) \geq 0$ by partial integration. Similarly, $I(G) = \int J(\xi)\,dG$.

It is our thesis that the form of ψ_0 may be inferred quite generally from the behaviour of $J(\xi)$. This approach was adopted by Collins and Wiens (1985) in determining general properties of least informative distributions in arbitrary ε-contamination classes. In Section 3 it is applied to such distributions as $G_\ell(x)$, with density proportional to $\exp(-|x|^\ell/\ell)$, and to "Student's" t-distribution. For distributions such as the t, ξ and $J(\xi)$ are nonmonotone, resulting in this case in *six* distinct forms of the solution, depending upon ε and the degrees of freedom. Five of these are rather unwieldy; the sixth coincides with the Sacks–Ylvisaker "large ε" form. This form is shown to apply for all sufficiently large Kolmogorov neighbourhoods, under very mild conditions on ξ.

2. Necessary and sufficient properties of F_0. In this section we exhibit some conditions which are necessary and sufficient in order that F_0 have minimum information in K_ε. These lead to some heuristic considerations of the general form of ψ which motivate the solutions given, in Section 3, for some special classes of distributions G.

Partition the support of f_0, in $(0, \infty)$, into disjoint sets

$$B_0 = \left\{x \mid \max\left(\tfrac{1}{2}, G(x) - \varepsilon\right) < F_0(x) < \min(1, G(x) + \varepsilon)\right\},$$
$$B_L = \left\{x \mid F_0(x) = G(x) - \varepsilon\right\},$$
$$B_U = \left\{x \mid F_0(x) = G(x) + \varepsilon\right\}.$$

Define a functional J on the set of continuous (except possibly at zero), piecewise continuously differentiable functions ψ by $J(\psi) = 2\psi' - \psi^2$. Extend $J(\psi)$ by left continuity where ψ' is discontinuous. If ψ is discontinuous at zero, set $J(\psi)(0) = \operatorname{sign}(\psi(0^+) - \psi(0^-)) \cdot \infty$, corresponding to use of the Schwarz derivative [Natanson (1960)].

It turns out that $J(\psi_0) \equiv$ constant on each component of B_0. We note that the only solution to $J(\psi_0) \equiv \lambda^2$ is of the form $\lambda \tan(\lambda(x - \omega)/2)$ for some parameter ω, and that those to $J(\psi_0) \equiv -\lambda^2$ are $\lambda \tanh(-\lambda(x - \omega)/2)$, λ, and $\lambda \coth(-\lambda(x - \omega)/2)$.

THEOREM 1. *If F_0 possesses the following properties, then it is the unique member of K_ε' minimizing $I(F)$ over K_ε.*

1. $F_0 \in K_\varepsilon$, F_0 *symmetric*, $F_0(\infty) = 1$.
2. F_0 *has an absolutely continuous density f_0 with respect to Lebesgue measure, and* $\psi_0 = -f_0'/f_0$ *is absolutely continuous on* $(-\infty, \infty)$.
3. *There exists a, possibly infinite, set of intervals* $[b_i, a_{i+1}]$, *with* $0 < b_1$, $a_i < b_i \le a_{i+1}$, $\limsup_{i\to\infty} a_i := a < \infty$; *and constants* λ_i, λ *such that*
   (i) $B_L \cup B_U = \bigcup_i [b_i, a_{i+1}]$;

$$\text{(ii) } J(\psi_0)(x) = \begin{cases} \lambda_1, & x \in [0, b_1], \\ \lambda_i, & x \in (a_i, b_i], \\ -\lambda^2, & x > a, \\ J(\xi)(x), & x \in \operatorname{Int}(B_L \cup B_U); \end{cases}$$

   (iii) *if* $x \in B_L$, *then* $J(\psi_0)(x) \ge J(\psi_0)(x+0)$,
   if $x \in B_U$, *then* $J(\psi_0)(x) \le J(\psi_0)(x+0)$.

PROOF. It follows from "$J(\psi_0) \equiv -\lambda^2$" on (a, ∞) that $\psi_0 \equiv \lambda > 0$ and $f_0(x) = f_0(a)\exp(\lambda(a-x))$ there. Here we use the fact that the tanh and coth solutions are both eventually negative (f_0 increasing). In particular, ψ_0 is bounded and $f_0(x) \to 0$ as $x \to \pm\infty$. On $\operatorname{Int}(B_L \cup B_U)$, $f_0(x) = g(x) > 0$. On B_0, $f_0 > 0$ since no f_0 corresponding to a solution to $J(\psi_0) \equiv \lambda$ can descend to zero with ψ_0 remaining bounded. Thus $f_0 > 0$ on $(-\infty, \infty)$, so that we need only verify (1.1), and that $0 < I(F_0) < \infty$. By partial integration, (1.1) becomes

$$\int J(\psi_0)\, d(F - F_0) \ge 0. \tag{2.1}$$

It suffices to check this for symmetric $F \in K_\varepsilon'$. Assume that "a" is the only possible accumulation point of $\{a_i\}$—the general case is similar. If we put $H = F - F_0$ and integrate $\int J(\xi)\, dH$ by parts on those nondegenerate intervals in $B_L \cup B_U$, (2.1) becomes

$$\begin{aligned} 0 &\le \lim_{n\to\infty} \int_0^{a_{n+1}} J(\psi_0)\, dH - \lambda^2 \int_a^\infty dH \\ &= \lim_{n\to\infty} \Bigg[\Bigg\{ \sum_{b_i < a_{i+1} \le a_{n+1}} \big(J(\psi_0)(b_i) - J(\psi_0)(b_i + 0)\big) H(b_i) \\ &\qquad + \sum_{b_i \le a_{i+1} \le a_n} \big(J(\psi_0)(a_{i+1}) - J(\psi_0)(a_{i+1} + 0)\big) H(a_{i+1}) \\ &\qquad - \lambda^2 H(\infty) - \sum_{b_i < a_{i+1} \le a_n} \int_{b_i}^{a_{i+1}} H(x) \frac{d}{dx} J(\xi)(x)\, dx \Bigg\} \\ &\qquad + \big\{ J(\psi_0)(a_{n+1}) H(a_{n+1}) - J(\psi_0)(a + 0) H(a) \big\} \Bigg]. \end{aligned} \tag{2.2}$$

By 1 and 3(iii), all terms within the first set of braces are nonnegative, as is the limit of the remaining term. Thus (1.1) is satisfied. That $0 < I(F_0) = \int \psi_0^2 \, dF_0 < \infty$ is obvious. □

It is also necessary that F_0 satisfy the conditions of Theorem 1. Since the necessity is not explicitly required, the proof (available from the author) is omitted. We note however that $F_0 \in K_\varepsilon$ forces, in turn, the additional necessary conditions

4(i) $f_0(x) = g(x)$, $x \in B_L \cup B_U$;
4(ii) $\psi_0(x) - \xi(x) \leq 0$ on $B_L (\geq 0$ on $B_U)$.

In Section 3 we exhibit the minimum information distributions F_0 for some particular Kolmogorov neighbourhoods. The general principle at work appears to be that for sufficiently small ε, ψ_0 should differ from ξ only near the local extrema of $J(\xi)$; and that here we should have $J(\psi_0) \equiv$ const, with this constant being less extreme than that attained by $J(\xi)$. In line with (2.1), we should have $f_0 > g$, $F_0 - G$ increasing from $-\varepsilon$ to ε, near the local minima of $J(\xi)$, $f_0 < g$, $F_0 - G$ decreasing from ε to $-\varepsilon$, near the local maxima. This is illustrated by Theorem 2 below. As ε increases, the regions of constancy of $J(\psi_0)$ coalesce. It is shown in Theorem 3 that for sufficiently large ε, the solution quite generally has $J(\psi_0) \equiv \lambda_1^2(\varepsilon)$ on $[-b(\varepsilon), b(\varepsilon)]$, $J(\psi_0) \equiv -\lambda^2(\varepsilon)$ elsewhere, with $F_0(b) = G(b) - \varepsilon$ and $b, \lambda_1, \lambda \to 0$ as $\varepsilon \uparrow \frac{1}{2}$. We conjecture that this "large ε" form of the solution is universally valid. We also give examples (Examples 2 and 3 below) of classes of distributions for which there are intermediary forms of the solution.

3. Some classes of solutions. The preceding discussion suggests that if $J(\xi)$ is decreasing on $[0, \infty)$, so that $\xi(0^+) \geq 0$ as well, we should have $B_L = [a, b]$, $B_U = \phi$, $0 < a < b < \infty$. Before proving this, we show that our monotonicity assumption implies that g is strongly unimodal.

LEMMA 1. *If $J(\xi)$ is strictly decreasing on $[0, \infty)$, and continuously differentiable on $(0, \infty)$, then ξ is positive and strictly increasing on $(0, \infty)$. The converse is false.*

PROOF. Under the stated conditions, any critical point x_0 of ξ must furnish a local maximum. Thus $\xi(x_0) > 0$, and in order that ξ not become negative on an unbounded interval there must exist an inflection point $x_1 > x_0$ at which $0 = \xi''(x_1) < \xi(x_1)\xi'(x_1) \leq 0$, a contradiction. Thus ξ is monotonic and nonnegative on $(0, \infty)$. From this observation the result is immediate.

Counterexamples to the converse are furnished by the distributions G_ℓ defined in the Introduction with $\ell > 2$. □

If $J(\xi)$ is merely decreasing on $(0, \infty)$, Lemma 1 fails for, say, $g(x) = (1 + 2|x|)\exp(-|x|)/6$. Some distributions satisfying the conditions of Lemma 1 are the logistic, and those G_ℓ with $1 < \ell \leq 2$.

THEOREM 2. *Under the conditions of Lemma* 1, *there exists* $\varepsilon_0 = \varepsilon_0(G)$ *such that for* $\varepsilon \in [0, \varepsilon_0]$, $I(F)$ *is minimized over* K_ε *by that* F_0 *with*

$$\psi_0(x) = \left\{\lambda_1 \tan\frac{\lambda_1 x}{2}, \xi(x), \lambda = \xi(b)\right\},$$

$$f_0(x) = \left\{\frac{g(a)\cos^2\frac{\lambda_1 x}{2}}{\cos^2\frac{\lambda_1 a}{2}}, g(x), g(b)\exp(-\xi(b)(x-b))\right\}$$

on $[0, a], [a, b], [b, \infty)$, *respectively. The constants* a, b, λ_1 *are determined by* (i) $F_0(a) = G(a) - \varepsilon$, (ii) $F_0(\infty) = 1$, *and* (iii) $\psi_0(a - 0) = \xi(a)$. *Thus* $B_L = [a, b]$, $B_{U} = \phi$. *Minimum information is*

$$I(F_0) = 2\left\{\lambda_1^2\left[G(a) - \varepsilon - \tfrac{1}{2}\right] - \lambda^2\left[1 - G(b) + \varepsilon\right] + \int_a^b J(\xi)\, dG\right\}.$$

The limiting values are $(\varepsilon, a, b, \lambda_1^2, -\lambda^2) \to (0, 0, \infty, J(\xi)(0), J(\xi)(\infty))$, *and* $\varepsilon_0(G)$ *is defined by* $a(\varepsilon_0) = b(\varepsilon_0)$.

PROOF. It is a straightforward matter—see Wiens (1985) for details—to establish the existence of constants a, b, λ_1 satisfying (i)–(iii) and

$$\xi(x) \geq \psi_0(x), \qquad x \in [0, a]; \ J(\xi)(a) < \lambda_1^2.$$

Integrating this first inequality shows that $f_0 \leq g$ on $[0, a]$. The monotonicity of ξ ensures that $\xi \geq \psi_0$ and $f_0 \geq g$ on $[b, \infty)$, and that $J(\xi)(b) > -\lambda^2$. The conditions of Theorem 1 are then satisfied, as long as $a \leq b$. □

We now establish sufficient conditions under which the "large ε" form of the solution is valid.

THEOREM 3. *Suppose that, on* $(0, \infty)$, $\xi(x)$ *satisfies* (A.1) $\xi(x) > 0$, $(x\xi(x))' > 0$, (A.2) $(\xi(x)/x)' \leq 0$, *and* (A.3) ξ *has no local minima in* (b_0, ∞), *where* b_0 *is defined by* $b_0\xi(b_0) = 1$. *Then there exists* $\varepsilon_1 = \varepsilon_1(G)$ *such that for* $\varepsilon \in (\varepsilon_1, 1/2]$, *the minimum information* $F_0 \in K_\varepsilon$ *is described by*

$$\psi_0(x) = \left\{\lambda_1 \tan\frac{\lambda_1 x}{2}, \lambda = \lambda_1 \tan\frac{\lambda_1 b}{2}\right\},$$

$$f_0(x) = \left\{\frac{g(b)\cos^2\frac{\lambda_1 x}{2}}{\cos^2\frac{\lambda_1 b}{2}}, g(b)\exp(-\lambda(x-b))\right\}$$

on $[0, b]$ *and* $[b, \infty)$, *respectively. The constants* λ_1, b *satisfy* (i) $F_0(b) =$

$G(b) - \varepsilon$, (ii) $F_0(\infty) = 1$, *and* (iii) $\psi_0(b) < \xi(b)$. *Thus* $B_L = \{b\}$, $B_U = \phi$. *Minimum information is*

$$I(F_0) = 2(\lambda_1^2 + \lambda^2)(G(b) - \tfrac{1}{2} - \varepsilon) - \lambda^2.$$

The limiting values are $(\varepsilon, b, \lambda_1^2, -\lambda^2) \to (\frac{1}{2}, \infty, 0, 0)$.

PROOF. The identity $(xg(x))' = g(x)(1 - x\xi(x))$, together with (A.1), implies that $\lim_{x \to \infty} xg(x) = 0$, $\lim_{x \to 0} x\xi(x) < 1$, $\lim_{x \to \infty} x\xi(x) > 1$; hence the existence of a unique point b_0 as in (A.3). It is easily checked that if (i)–(iii) are satisfied, and if $F_0 \in K_\varepsilon$, then the conditions of Theorem 1 are met.

Similar to the development in Sacks and Ylvisaker (1972), (A.1) implies the existence of $\varepsilon_*(G) < \frac{1}{2}$ such that (i)–(iii) are satisfiable for $\varepsilon > \varepsilon_*$. Then (A.2) ensures that on $[0, b]$, $\xi(x)$ remains above the line segment joining $(0,0)$ to $(b, \psi_0(b))$, hence above the convex function ψ_0. As in Theorem 2, this implies that $g \geq f_0$ on $[0, b]$, so that $G \geq F_0 \geq G - \varepsilon$ there. Alternatively, this may be established under the conditions of Lemma 1. Now (A.3) implies the existence of $\varepsilon_1(G) \in [\varepsilon_*, \frac{1}{2}]$ such that for $\varepsilon \geq \varepsilon_1$, F_0 remains within the boundaries of the Kolmogorov strip on (b, ∞). As in Theorem 2, if (A.3) is replaced by the stronger

$$\text{(A.3)}': \xi'(x) \geq 0 \text{ on } (0, \infty),$$

then we may take $\varepsilon_1 = \varepsilon_*$. See Wiens (1985) for the details. □

COROLLARY 1. *Under the conditions of Lemma* 1, *the least informative* $F_0 \in K_\varepsilon$ *is as described in Theorems* 2 *and* 3, *with* $\varepsilon_0(G) = \varepsilon_1(G)$.

EXAMPLE 1. Those distributions G_ℓ, $1 < \ell \leq 2$, are covered by Corollary 1. Huber (1964) and Sacks and Ylvisaker (1972) obtained $\varepsilon_1(G_2) \approx 0.0303$. Working through the numerical details of Theorem 3 extends the result to the Laplace distribution ($\ell > 1$), with $\varepsilon_1(G_1) = 0$.

Theorem 3 applies to those G_ℓ with $\ell < 1$, and we find $\varepsilon_1(G_{0.5}) \approx 0.0355$, $\varepsilon_1(G_{0.75}) \approx 0.0153$. Although assumption (A.2) of Theorem 3 fails for G_ℓ if $\ell > 2$, a sligh. modification to the proof shows that the conclusions apply to these cases as well, with $\varepsilon_1 = \varepsilon_*$.

EXAMPLE 2. Denote by $G_r(x)$ the "Student's" t distribution on r d.f., with $\xi_r(x) = (r + 1)x/(r + x^2)$. Theorem 3 applies, but Theorem 2 does not. The function $J(\xi_r)(x) = (r + 1)(2r - (r + 3)x^2)/(r + x^2)^2$ attains a positive maximum at 0, decreases to a negative minimum at $(r(r + 7)/(r + 3))^{1/2} := M_r$, then increases to 0 at ∞. The discussion in Section 2 then suggests that for sufficiently small ε, say $\varepsilon \leq \varepsilon_I(r)$, there should exist points a, b, c, d, with $0 < a < b < M_r < c < d$ such that F_0 has $B_L = [a, b]$, $B_U = [c, d]$. More precisely, this

"Stage I" solution is given by

$$\psi_0(x) = \left\{\lambda_1 \tan\frac{\lambda_1 x}{2}, \xi(x), \lambda_2 \tanh\left(\frac{-\lambda_2}{2}(x-\omega)\right), \xi(x), \lambda = \xi(d)\right\},$$

$$f_0(x) = \left\{\frac{g(x)\cos^2\frac{\lambda_1 x}{2}}{\cos^2\frac{\lambda_1 a}{2}}, g(x), \frac{g(b)\cosh^2\left(\frac{-\lambda_2}{2}(x-\omega)\right)}{\cosh^2\left(\frac{-\lambda_2}{2}(b-\omega)\right)},\right.$$

$$\left. g(x), g(d)\exp(-\xi(d)(x-d))\right\}$$

on $[0, a]$, $[a, b]$, $[b, c]$, $[c, d]$, $[d, \infty)$, respectively. See Figure 1. The seven constants are determined by the conditions $F_0(a) = G(a) - \varepsilon$, $F_0(c) = G(c) + \varepsilon$, $F_0(\infty) = 1$, and continuity of f_0 at c and of ψ_0 at a, b, c. Given the existence of such constants, the conditions of Theorem 1 are easily verified.

For $r = 1$, some numerical values of the constants are given in Table 1 below for this, and the three subsequent stages. Stage II differs from I in that $a = b$ and $\psi_0(b) < \xi(b)$, and is valid for $\varepsilon \in [\varepsilon_{\mathrm{I}}(1), \varepsilon_{\mathrm{II}}(1)] = [0.00573, 0.02515]$. Stage III has as well $c = d$ and $\psi_0(c) > \xi(c)$, for $\varepsilon_{\mathrm{II}}(1) \le \varepsilon \le 0.0377 = \varepsilon_{\mathrm{III}}(1)$. Stage IV is then as described in Theorem 3, and is obtained by letting $\omega \to \infty$ in Stage III.

Since Theorem 2 becomes applicable at $r = \infty$, it is clear that this sequence of stages cannot hold for all r. Numerical investigations have shown that it is in fact only valid for $r = 1$. For $r \ge 2$, Stage II is altered by requiring $c = d$, $a < b$, $\psi_0(c) > \xi(c)$. On a range $2 \le r \le R$, Stage III then has $a = b$, $\psi_0(b) < \xi(b)$. For $r > R$, it has instead $a < b$, $c = d$, $F_0(c) < G(c) + \varepsilon$. In each of Stages I–III, $B_U \downarrow \phi$ as $r \to \infty$.

Collins and Wiens (1985) obtained the most robust ψ_0 for an ε-contamination neighbourhood G_ε of G_r, and found it to be of the form exhibited in Figure 1, without the "tan" and "constant" portions. This reflects the fact that in K_ε, maxima of $J(\xi)$ may be dampened by removing mass from g, whereas in G_ε only minima may be handled, by adding mass to $(1 - \varepsilon)g$.

EXAMPLE 3. If ξ is positive, decreasing, and convex on $(0, \infty)$, then $J(\xi)$ is negative and increasing there but $J(\xi)(0) = +\infty$. Examples are the distributions $G_\ell(x)$, $\ell < 1$, for which Theorem 3 applies.

In view of the Dirac delta in $J(\xi)$ at 0, we expect that for small values of ε, $J(\psi_0)$ is constant on three contiguous intervals symmetric about zero, and in neighbourhoods of $\pm\infty$. As at 3(iii) of Theorem 1, F_0 cannot remain on the lower boundary of the Kolmogorov strip, in $(0, \infty)$. The "small ε" solution should then be obtained from those same equations defining the Stage II, $r = 1$ solution of Example 2. It is then easy to see that the remaining two stages must be as for the Cauchy distribution. For $G_{0.5}$, see Wiens (1985) for some numerical values.

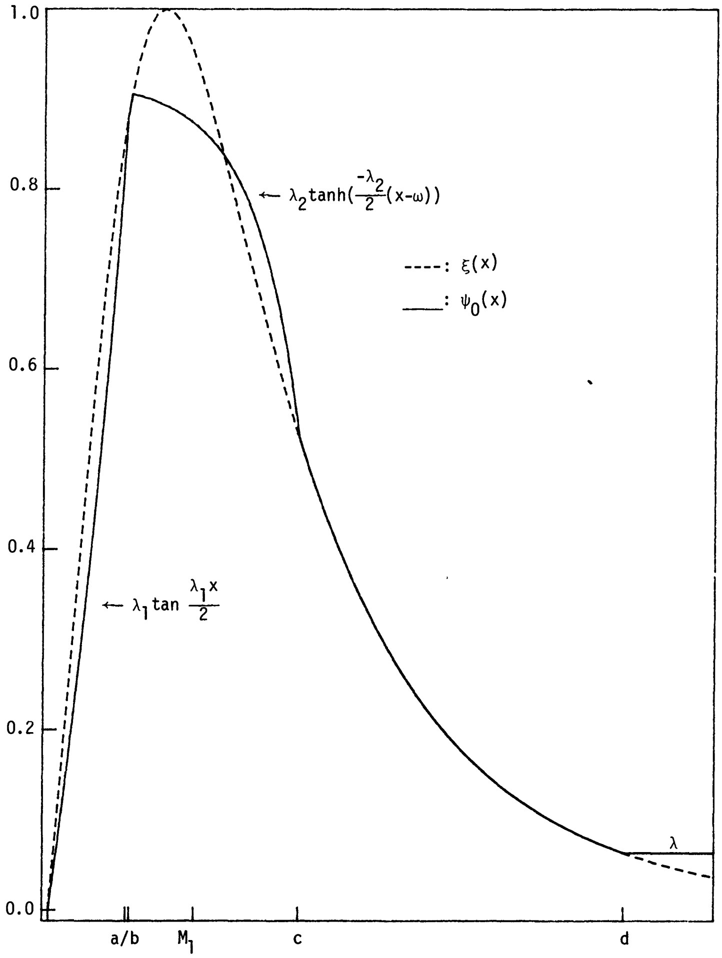

FIG. 1. *Most robust ψ_0 for a Kolmogorov neighbourhood of the Cauchy distribution, with $\varepsilon = 0.005$ (Stage I). The constants are given in Table* 1, *and the horizontal axis is* $\ln(1 + x)$.

REMARKS. 1. Some feeling for the geometry of a Kolmogorov neighbourhood is given by the "infinitesimal loss of Fisher information" $d/d\varepsilon\, I(F_0)_{|\varepsilon=0}$. In general, if F_0 is determined by equations (i)–(iii) of Theorem 2, or by (i) and (ii) of Theorem 3, then

$$\frac{d}{d\varepsilon} I(F_0) = -2\big(\lambda_1^2(\varepsilon) + \lambda^2(\varepsilon)\big).$$

For G_l, $1 \le l \le 2$, this varies monotonically from $-\infty$ at $\varepsilon = 0$ to 0 at $\varepsilon = \frac{1}{2}$. For the logistic, it varies from -4 to 0.

2. Consider the $\mathscr{L}^1$ neighbourhood of G, defined by $\mathscr{L}_\varepsilon^1 = \{F \mid \int |f - g|\, dx \le \varepsilon\}$. If F_0 is determined as in Theorem 2, or as in Theorem 3 with (A.3′) holding,

TABLE 1
Least informative F_0 in Kolmogorov neighbourhoods of the Cauchy distribution

Stage	ε	a	b	c	d	λ_1	λ_2	λ	ω	$1/I(F_0)$
I	0	0	$\sqrt{2}$	$\sqrt{2}$	∞	2	1.155	0		2
	0.001	0.411	0.800	2.66	159.15	1.81	1.04	0.013	4.09	2.05
	0.005	0.599	0.635	3.55	31.81	1.64	0.94	0.063	4.88	2.22
	0.00573	0.620	0.620	3.66	27.75	1.63	0.93	0.072	4.99	2.26
II	0.006	0.622		3.70	26.50	1.62	0.92	0.075	5.03	2.27
	0.010	0.657		4.26	15.87	1.53	0.87	0.125	5.57	2.45
	0.025	0.765		6.19	6.26	1.29	0.71	0.312	7.54	3.19
	0.02515	0.766		6.22	6.22	1.29	0.71	0.313	7.56	3.20
III	0.026	0.772		6.24		1.28	0.70	0.32	7.67	3.24
	0.030	0.797		6.31		1.23	0.67	0.39	8.31	3.45
	0.035	0.825		6.21		1.17	0.62	0.51	9.93	3.72
	0.0377	0.839		6.08		1.14	0.59	0.59	∞	3.86
IV	0.0377	0.839				1.14	0.59			3.86
	0.0535	0.946				1.02	0.54			4.71
	0.0608	1.00				0.98	0.51			5.17
	0.1512	1.56				0.58	0.28			16.74
	0.3366	3.85				0.16	0.05			484.12
	0.5	∞				0	0			∞

then $F_0 \in \mathscr{L}^1_{4\varepsilon}$. Since the symmetric (hence less informative) subclass $\mathscr{L}^{(1)}_{4\varepsilon}$ of $\mathscr{L}^1_{4\varepsilon}$ is contained in K_ε, F_0 minimizes information over $\mathscr{L}_{4\varepsilon}$ as well. Note that for $\varepsilon \geq \frac{1}{2}$, $\inf\{I(F) | F \in \mathscr{L}^1_{4\varepsilon}\} = 0$.

REFERENCES

COLLINS, J. R. and WIENS, D. P. (1985). Minimax variance M-estimators in ε-contamination models. *Ann. Statist.* **13** 1078–1096.

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

HUBER, P. J. (1981). *Robust Statistics*. Wiley, New York.

NATANSON, I. P. (1960). *Theory of Functions of a Real Variable* **2**. Ungar, New York.

SACKS, J. and YLVISAKER, D. (1972). A note on Huber's robust estimation of a location parameter. *Ann. Math. Statist.* **43** 1068–1075.

VANDELINDE, V. D. (1979). Robust techniques in communication. In *Robustness in Statistics*. (R. Launer and G. Wilkinson, eds.) 177–200. Academic, New York.

WIENS, D. P. (1985). Minimax variance M-estimators of location in Kolmogorov neighbourhoods. Research report DALTRS-85-6, Dalhousie Univ.

DEPARTMENT OF MATHEMATICS, STATISTICS,
AND COMPUTING SCIENCE
DALHOUSIE UNIVERSITY
HALIFAX, NOVA SCOTIA B3H 4H8
CANADA