The Canadian Journal of Statistics
Vol. 18, No. 1, 1990, Pages 47–57
La Revue Canadienne de Statistique

Minimax-variance *L*- and *R*-estimators of location*

Douglas P. WIENS

University of Alberta

Key words and phrases: Robust estimation, L- and R-estimators, minimax variance, minimum Fisher information.
AMS 1985 subject classifications: Primary 62F35, secondary 62F12.

ABSTRACT

We consider the problem of minimax-variance, robust estimation of a location parameter, through the use of L- and R-estimators. We derive an easily checked necessary condition for L-estimation to be minimax, and a related sufficient condition for R-estimation to be minimax. Those cases in the literature in which L-estimation is known not to be minimax, and those in which R-estimation is minimax, are derived as consequences of these conditions. New classes of examples are given in each case. As well, we answer a question of Scholz (1974), who showed essentially that the asymptotic variance of an R-estimator never exceeds that of an L-estimator, if both are efficient at the same strongly unimodal distribution. Scholz raised the question of whether or not the assumption of strong unimodality could be dropped. We answer this question in the negative, theoretically and by examples. In the examples, the minimax property fails both for L-estimation and for R-estimation, but the variance of the L-estimator, as the distribution of the observation varies over the given neighbourhood, remains unbounded. That of the R-estimator is unbounded.

RÉSUMÉ

On étudie le problème de l'estimation robuste de variance minimax, d'un paramètre de position, en utilisant les L et R-estimateurs. On obtient une condition nécessaire, facile à vérifier, pour qu'un L-estimateur soit minimax, et une condition apparentée qui est suffisante pour qu'un R-estimateur soit minimax. Les cas connus où un L-estimateur n'est pas minimax, et ceux où un R-estimateur est minimax, découlent de ces conditions. De nouvelles classes d'exemples sont donnés pour chaque cas. Scholz (1974) démontra essentiellement que la variance asymptotique d'un R-estimateur n'excède jamais celle d'un L-estimateur si les deux sont efficaces pour la même densité fortement unimodale. Il souleva aussi la question à savoir si l'hypothèse d'unimodalité forte pouvait être relâchée. On répond par la négative à cette question, cela tant théoriquement que par des exemples. Dans ces exemples, la propriété minimax n'est pas vérifiée autant pour le L-estimateur que pour le R-estimateur. Par contre, la variance du L-estimateur demeure bornée lorsque la distribution des observations varie dans un voisinage donné, tandis que celle du R-estimateur ne l'est pas.

1. INTRODUCTION AND SUMMARY

Let $X_{1:n} \leq \cdots \leq X_{n:n}$ be the order statistics from a location family, distributed as $F(x-\theta)$. Consider the L-estimator of θ given by

$$T_L = n^{-1} \sum_{i=1}^{n} m\left(\frac{i}{n+1}\right) X_{i:n},$$

*Research supported by the Natural Sciences and Engineering Research Council of Canada.

where $m(\cdot)$ is a weights-generating function chosen by the statistician, and the R-estimator T_R obtained by inverting a one-sample rank test, with absolutely continuous scores-generating function J. See Huber (1981), Lehmann (1983) for basic properties of such estimators. Under appropriate conditions—see Serfling (1980) and references cited therein, in particular Chernoff and Savage (1958), Chernoff, Gastwirth, and Johns (1967)—both estimators are consistent and asymptotically normal:

$$\sqrt{n}(T-\theta) \to_d \mathrm{N}(0, E_F[\mathrm{IC}_T^2(X;F)]) \tag{1.1}$$

where IC_T represents the influence curve of either T_L or T_R. The influence curve of T_L is

$$\mathrm{IC}_L(x;F) = -\int_0^1 \{I[F(x) \le t] - t\} m(t)\, dF^{-1}(t), \tag{1.2}$$

where $F^{-1}(t) = \inf\{x \mid F(x) \ge t\}$. For symmetric, absolutely continuous F, that of T_R is

$$\mathrm{IC}_R(x;F) = \frac{J \circ F(x)}{D_R(F)}, \qquad \text{where} \quad D_R(F) = \int_{-\infty}^{\infty} J' \circ F(x) f^2(x)\, dx. \tag{1.3}$$

If F is asymmetric, or is not absolutely continuous, then IC_R is considerably more complex—see Chapter 3 of Huber (1981).

Now suppose that F is an unknown member of a convex class $\mathcal{F}$ of distributions, in which the Fisher information for location $I(F)$ is minimized at a member F_0, symmetric about θ ($= 0$, without loss of generality). Assume that $0 < I(F_0) < \infty$, so that (Huber, 1981, Section 4.4) $F_0(x)$ has an absolutely continuous density $f_0(x)$, tending to zero as $x \to \pm\infty$. Assume that $f_0(x) > 0$ on $0 < F_0(x) < 1$. Put $\psi_0(x) = -f_0'(x)/f_0(x)$, and assume that $\psi_0(x)$ is absolutely continuous, with a piecewise continuous derivative $\psi_0'(x)$.

Choose the weights- and scores-generating functions

$$\begin{aligned} m_0(t) &= \psi_0' \circ F_0^{-1}(t)/I(F_0), \\ J_0(t) &= \psi_0 \circ F_0^{-1}(t)/I(F_0). \end{aligned} \tag{1.4}$$

Denote by $V_L(m,F)$ and $V_R(J,F)$ the asymptotic variances of $\sqrt{n}(T_L-\theta)$ and $\sqrt{n}(T_R-\theta)$. Then (1.4), together with results of Stone (1974) for location and scale equivariant estimators of location, gives

$$\begin{aligned} \inf_m V_L(m, F_0) &= V_L(m_0, F_0) = \frac{1}{I(F_0)}, \\ \inf_J V_R(J, F_0) &= V_R(J_0, F_0) = \frac{1}{I(F_0)}. \end{aligned} \tag{1.5}$$

In this paper, we investigate whether or not the *saddle-point* property holds, i.e. (1.5) combined with

$$\sup_{\mathcal{F}} V_L(m_0, F) = \frac{1}{I(F_0)} \quad \text{or} \quad \sup_{\mathcal{F}} V_R(J_0, F) = \frac{1}{I(F_0)}. \tag{1.6}$$

This is of interest because it implies the *minimax* property—that the supremum, over $\mathcal{F}$, of the asymptotic variance is minimized, over the class of L- or R-estimators, by the appropriate choice in (1.4).

For a brief history of this problem, and a survey of results in specific neighbourhoods, see Section I of Collins and Wiens (1989). The results to date are restricted to neighbourhoods in which either the minimum information distribution is strongly unimodal, or there is a strongly unimodal "target" distribution around which the neighbourhood is formed. Analyses of cases in which neither condition holds are conspicuously absent from the literature. One aim of this paper is to fill this gap.

Of central importance is the behaviour of the functions

$$L_0(x) = \frac{2\psi_0'(x) - \psi_0^2(x)}{I(F_0)}, \qquad -\infty < x < \infty,$$

$$K_0(u) = \frac{\psi_0' \circ F_0^{-1}(u) f_0 \circ F_0^{-1}(u)}{I(F_0)}, \qquad 0 < u < 1.$$

Note that $\int_{-\infty}^{\infty} L_0(x)\, dF_0(x) = \int_0^1 K_0(u)\, dF_0^{-1}(u) = 1$. Consider the conditions

$$\int_{-\infty}^{\infty} L_0(x)\, dF(x) \geq 1, \qquad \text{all } F \in \mathcal{F} \text{ with } I(F) < \infty, \tag{1.7}$$

$$\int_0^1 K_0(u)\, dF^{-1}(u) \leq 1, \qquad \text{all } F \in \mathcal{F} \text{ with } I(F) < \infty. \tag{1.8}$$

The condition (1.7) is necessary and sufficient for F_0 to minimize $I(F)$ in $\mathcal{F}$ —see Huber (1964). It will be shown that (1.8) is a necessary condition for L-estimation to satisfy (1.6). Furthermore, it is necessary that equality in (1.8) be attained only by F_0 and by members of $\mathcal{F}$ equivalent to F_0 in a sense made precise by Theorem 2.1 below and in the examples of Section 3. If F_0 is strongly unimodal, i.e. $\psi_0(x)$ nondecreasing, then (1.8) is a sufficient condition for R-estimation to be minimax in $\mathcal{F}$.

These statements are proven in Section 2 of the paper. In Section 3 they are applied to give simple and straightforward proofs of (1.6), or of its failure, in those cases currently in the literature and in other classes. In particular, in Kolmogorov or Lévy neighbourhoods of a strictly increasing d.f., it is shown that (1.6) always fails for L-estimation, and holds for R-estimation if and only if F_0 is strongly unimodal. For ϵ-contamination neighbourhoods we give a partial converse to the results of Jaeckel (1971), who showed that both L- and R-estimation are minimax in ϵ-contamination neighbourhoods of strongly unimodal df's.

In some cases in which F_0 is not strongly unimodal we construct subneighbourhoods $\mathcal{F}_1$, containing F_0, of $\mathcal{F}$ in which both $V_L(m_0, F)$ and $V_R(J_0, F)$ are *minimized* at F_0, and with $\sup_{\mathcal{F}_1} V_L(m_0, F) < \sup_{\mathcal{F}_1} V_R(J_0, F) = \infty$. This answers in the negative a question raised by Scholz (1974), who showed that $V_R(J_0, F) \leq V_L(m_0, F)$ if F_0 is strongly unimodal, with F being symmetric, strictly increasing on its support, and such that (1.1) holds. He then asked if the assumption of strong unimodality could be dropped. Our negative answers apply in particular to Kolmogorov and ϵ- contamination neighbourhoods of the Cauchy distribution.

Throughout the rest of the paper we write $V_L(F)$ for $V_L(m_0, F)$, and $V_R(F)$ for $V_R(J_0, F)$.

2. MAIN RESULTS

2.1. Minimax-Variance L-Estimation

Recall (1.2), and that the asymptotic variance of T_L is $V_L(F) = E_F[\mathrm{IC}_L^2(X; F)]$. For continuous F, a useful alternate form is

$$V_L(F) = E_U[\mathrm{IC}_L^2(F^{-1}(U); F)], \tag{2.1}$$

where

$$\mathrm{IC}_L(F^{-1}(u);F) = -\int_0^1 \{I[u \le t] - t\} m_0(t)\, dF^{-1}(t),$$

and where U is a r.v. uniformly distributed on [0, 1].

We are aware of only two sets of *sufficient* conditions implying the saddle-point property (1.6), both of them pointwise:

(1) all $F \in \mathcal{F}$ continuous, with

$$|\mathrm{IC}_L(F^{-1}(u);F)| \le |\mathrm{IC}_L(F_0^{-1}(u);F_0)| \qquad \text{a.e. } u \in (0,1), \tag{2.2}$$

or

(2) both

$$|\mathrm{IC}_L(x;F)| \le |\mathrm{IC}_L(x;F_0)| \; [= |\psi_0(x)/I(F_0)|] \qquad \text{a.e. } x, \quad \text{all } F, \tag{2.3a}$$

and

$$\int_{-\infty}^{\infty} \psi_0^2(x)\, dF(x) \le \int_{-\infty}^{\infty} \psi_0^2(x)\, dF_0(x) \; [= I(F_0)]. \tag{2.3b}$$

That (2.2) and (2.3) each imply (1.6) is trivial. See Section 2.3 and Example 3.2 for applications.

The following theorem generalizes Theorem 4 of Collins and Wiens (1989), where a similar result was established for Lévy neighbourhoods of certain strongly unimodal distributions.

THEOREM 2.1. *Assume that T_L satisfies* (1.1) (*Asymptotic normality*) *and* (1.6) (*the saddle-point property*) *for $F \in \mathcal{F}$. It is then necessary that* (1.8) *hold. Furthermore, if*

$$\mathcal{F}_0 = \{F \in \mathcal{F} \mid F \textit{ strictly increasing on } 0 < F(x) < 1,$$
$$I(F) < \infty, \qquad \int_0^1 K_0(u)\, dF^{-1}(u) = 1\}$$

then for $F \in \mathcal{F}_0$ we have either

(i) $V_L(F) \gneqq V_L(F_0)$, or

(ii) In each interval I_i, on which $\psi_0'(x)$ is continuous and a.e. nonzero, we have $F_0(x) = F(x + c_i)$ for some constant $c_i(F)$ and all $x \in I_i$.

Proof. If (1.6) holds and $I(F) < \infty$, then F is absolutely continuous and $V_L(F) < \infty$; hence (2.1) applies and we have the identity

$$\begin{aligned} V_L(F) =& V_L(F_0) + 2[\mathcal{C}ov_U\{\mathrm{IC}_L(F^{-1}(U);F), \mathrm{IC}_L(F_0^{-1}(U);F_0)\} - V_L(F_0)] \\ &+ E_U[\{\mathrm{IC}_L(F^{-1}(U);F) - \mathrm{IC}(F_0^{-1}(U);F_0)\}^2] \end{aligned} \tag{2.4}$$

Applying Fubini's theorem gives $E_U[\mathrm{IC}_L(F^{-1}(U);F)] = E_U[\mathrm{IC}_L(F_0^{-1}(U);F_0)] = 0$, as is required by the definition of the influence curve. Also, $\mathrm{IC}_L(F_0^{-1}(u);F_0) = \psi_0 \circ F_0^{-1}(u)/I(F_0)$, whence another application of Fubini's theorem gives

$$\begin{aligned} &\mathcal{C}ov_U[\mathrm{IC}_L(F^{-1}(U);F), \mathrm{IC}(F_0^{-1}(U);F_0)] \\ &= -\int_0^1 \int_0^1 \frac{(I[u \le t] - t) m_0(t) \psi_0 \circ F_0^{-1}(u)\, dF^{-1}(t)\, du}{I(F_0)} \\ &= -\int_0^1 m_0(t) \int_0^1 \frac{\{I[u \le t] - t\} \psi_0 \circ F_0^{-1}(u)\, du\, dF^{-1}(t)}{I(F_0)} \\ &= \int_0^1 \frac{K_0(t)\, dF^{-1}(t)}{I(F_0)}. \end{aligned} \tag{2.5}$$

Now (2.5) in (2.4) yields the necessity of the condition (1.8), and that if equality holds in (1.8) we must further have $\mathrm{IC}_L(F^{-1}(u); F) = \mathrm{IC}_L(F_0^{-1}(u); F_0)$ a.e. $u \in [0,1]$, i.e.

$$\int_0^u t m_0(t)\, dF^{-1}(t) - \int_u^1 (1-t) m_0(t)\, dF^{-1}(t) = \frac{\psi_0 \circ F_0^{-1}(u)}{I(F_0)} \qquad \text{a.e. } u. \tag{2.6}$$

For $F \in \mathcal{F}_0$ the left side of (2.6) is a continuous function of u, as is the right side; hence equality holds throughout (0,1). Differentiating (2.6) gives $d/duF^{-1}(u) = d/duF_0^{-1}(u)$ at every continuity point of $m_0(u)$ at which $m_0(u) \neq 0$; hence $F^{-1}(u) - F^{-1}(u_0) = F_0^{-1}(u) - F_0^{-1}(u_0)$ whenever m_0 is continuous and a.e. nonzero on $[u_0, u]$. Then if ψ_0' is continuous and a.e. nonzero on an interval $I_i \supset [a_i, x]$, the last equality, with $u = F_0(x), u_0 = F_0(a_i)$, becomes $F^{-1} \circ F_0(x) = x + F^{-1} \circ F_0(a_i) - a_i = x + c_i$, say; hence $F_0(x) = F(x + c_i)$ on I_i. Q.E.D.

2.2. Minimax Variance R-Estimation

Assume that every $F \in \mathcal{F}$ is symmetric about $\theta = 0$, and has finite Fisher information, so that (1.3) applies. Note that $\int_{-\infty}^{\infty} J_0^2 \circ F(x) f(x)\, dx = 1/I(F_0)$. Make the substitution $u = F(x)$ in $D_R(F)$. Assume that $D_R(F) > 0$—otherwise, as in the proof of Theorem 2.2 below, $\sup_{0 \le \lambda \le 1} V_R((1-\lambda)F_0 + \lambda F) = \infty$. We then have that the saddle-point property (1.6) holds iff, for all $F \in \mathcal{F}$,

$$D_R(F) = \int_0^1 m_0(u) r_F(u)^{-1}\, du \ge 1, \qquad \text{where} \quad r_F(u) = \frac{f_0 \circ F_0^{-1}(u)}{f \circ F^{-1}(u)}. \tag{2.7}$$

THEOREM 2.2. *Assume that T_R satisfies* (1.1) *(asymptotic normality) for $F \in \mathcal{F}$. In order that the saddle-point property* (1.6) *hold, it is sufficient that $\psi_0(x)$ be nondecreasing and that the condition* (1.8) *hold. The requirement that ψ_0 be nondecreasing is necessary in the following sense. Suppose that there is an interval [a,b] throughout which ψ_0 is strictly decreasing, and on which $0 < f_0(x) < \infty$. Define*

$$\mathcal{F}_1 = \{F \in \mathcal{F} \mid F \equiv F_0 \text{ on } [(a,b) \cup (-b,-a)]^c\}.$$

If $\mathcal{F}$ is sufficiently rich that $\sup_{\mathcal{F}_1} \int_a^b f^2(x)\, dx = \infty$, then $\sup_{\mathcal{F}_1} V_R(F) = \infty$.

Proof. If ψ_0 is nondecreasing, then $m_0(u)$ is a density on [0,1] and by Jensen's inequality

$$\int_0^1 m_0(u) r_F(u)^{-1}\, du \ge \left(\int_0^1 m_0(u) r_F(u)\, du\right)^{-1} = \left(\int_0^1 K_0(u)\, dF^{-1}(u)\right)^{-1}.$$

Then (1.8) implies (2.7) for all $F \in \mathcal{F}$. Now let $[a,b]$ and $\mathcal{F}_1$ be as above. It suffices to show that under the stated conditions there exists $F_1 \in \mathcal{F}_1$ for which $D_R(F_1) < 0$. Then with $F_\lambda = (1-\lambda)F_0 + \lambda F_1$ and $\phi(\lambda) = D_R(F_\lambda)$, we have $\phi(0) > 0, \phi(1) < 0$, and, since $\|F_\lambda - F_{\lambda'}\| = |\lambda - \lambda'|\, \|F_0 - F_1\|, \phi(\lambda)$ is continuous with respect to any norm $\|\cdot\|$ on $\mathcal{F}_1$. It must then assume arbitrarily small positive values.

To establish the existence of F_1, note that

$$D_R(F) = 2\int_A m_0(u) r_F(u)^{-1}\, du + 2\int_{F_0(a)}^{F_0(b)} f \circ F^{-1}(u) J_0'(u)\, du,$$

where

$$A = \left[\frac{1}{2}, 1\right] \cap (F_0(a), F_0(b))^c.$$

For $F \in F_1$, $r_F(u) \equiv 1$ on $(F_0(a), F_0(b))^c$. The assumptions on ψ_0 imply that $0 > \sup_{[F_0(a),F_0(b)]} J_0'(u) = -C^2$, say, so that

$$\begin{aligned} D_R(F) &< 2\int_A m_0(u)\,du - 2C^2 \int_{F_0(a)}^{F_0(b)} f \circ F^{-1}(u)\,du \\ &= 1 - 2\int_{F_0(a)}^{F_0(b)} m_0(u)\,du - 2C^2 \int_a^b f^2(x)\,dx; \end{aligned}$$

hence $\inf_{\mathcal{F}_1} D_R(F) = -\infty$. Q.E.D.

2.3. Minimax L-Estimation versus Minimax R-Estimation.

In the class $\mathcal{F} = \{F \mid F$ symmetric and absolutely continuous, $\int x^2\,dF(x) \le 1\}$, the sample mean is the minimax L-estimator. This follows from (2.3). See Mason (1983) for cases in which this L-estimator is also optimal with respect to a different minimax criterion. Chernoff and Savage (1958) [see also Gastwirth and Wolff (1968)] showed that the variance of the normal scores estimator—the R-estimator efficient at the minimum-information normal distribution—never exceeds that of the sample mean in $\mathcal{F}$; hence it too is minimax.

Under the conditions leading to (2.5) and (2.7), we have

$$\frac{V_R(F)}{V_L(F)} = \frac{\mathrm{Corr}_U^2[\mathrm{IC}_L(F^{-1}(U);\, F),\ \mathrm{IC}_L(F_0^{-1}(U);\, F_0)]}{[\int_0^1 m_0(u) r_F(u)\,du \int_0^1 m_0(u) r_F(u)^{-1}\,du]^2}. \tag{2.8}$$

If ψ_0 is nondecreasing, then $m_0(u) \ge 0$ and Jensen's inequality asserts that the denominator of (2.8) is ≥ 1. This proves:

THEOREM 2.3. *If $\mathcal{F}$ is the class of symmetric distributions, with finite Fisher information, for which* (1.1) *holds for T_L and for T_R, and if ψ_0 is nondecreasing, then $V_R(F) \le V_L(F)$ for $F \in \mathcal{F}$. Thus if ψ_0 is nondecreasing and* (1.6) *is to hold for T_L, it is necessary that it hold for T_R.*

REMARK. The first statement of Theorem 2.3 was proven by Scholz (1974), under the additional assumption that each F is strictly increasing on its support. Froda (1986) extended Scholz's result to possibly discontinuous, but still nondecreasing, ψ_0. As an application of Theorem 2.2, we show in the remarks following Theorems 3.1, 3.2 that the assumption that ψ_0 is nondecreasing cannot, in general, be dropped. Without it, even the weaker statement $\sup_{\mathcal{F}} V_R(F) \le \sup_{\mathcal{F}} V_L(F)$ can fail.

3. EXAMPLES

Recall the conditions (1.7) and (1.8). In "regular" classes $\mathcal{F}$, (1.8) implies (1.7). This is seen by noting that (1.8) implies that the Gateaux derivative

$$\frac{d}{d\lambda} \int_0^1 K_0\{F_0(u) + \lambda(F_1 - F_0)(u)\}\,du|_{\lambda=0}$$

is ≤ 0 for all $F_1 \in \mathcal{F}$; this becomes (1.7) after a calculation. At least when $\mathcal{F}^{-1} = \{F^{-1} \mid F \in \mathcal{F}\}$ is convex, as is the case of Kolmogorov or Lévy neighbourhoods,

(1.7) implies (1.8). This is seen by writing $I(F)$ as a functional of F^{-1} before taking the Gateaux derivative. We do not make these arguments rigorous here, since (1.7) is part of the definition of F_0 and (1.8), where required, is more easily verified directly.

In particular, (1.8) holds in those cases in which F_0 has been obtained for the important contamination classes

$$\mathcal{G}_\epsilon(G) = \{F \mid F = (1-\epsilon)G + \epsilon H;\ G \text{ symmetric and fixed,}$$

$$H \text{ symmetric and arbitrary}\}$$

(ϵ-contamination neighbourhood),

$$\mathcal{K}_\epsilon(G) = \{F \mid \sup_x |F(x) - G(x)| \le \epsilon;\ G \text{ symmetric}\}$$

(Kolmogorov neighbourhood), and

$$\mathcal{L}_{\epsilon,\delta}(G) = \{F \mid G(x-\delta) - \epsilon \le F(x) \le G(x+\delta) + \epsilon,\ \text{all } x;\ G \text{ symmetric}\}$$

(Lévy neighbourhood). The minimum-information distributions were obtained for $\mathcal{G}_\epsilon(G)$ by Huber (1964) for strongly unimodal G and by Collins and Wiens (1985) in more general situations; for $\mathcal{K}_\epsilon(G)$ see Huber (1964) and Sacks and Ylvisaker (1972) if $G = \Phi$ and Wiens (1986) for general G; for $\mathcal{L}_{\epsilon,\delta}(G)$ see Collins and Wiens (1989).

The solutions obtained by the above authors all satisfy

$$F_0(\infty) = 1, \qquad \lim_{u\to 1} K_0(u) \le 0, \tag{3.1}$$

and

$L_0(x)$ is piecewise continuously differentiable and

$$\text{nondecreasing on } \{\sup_{\mathcal{F}} F(x) = F_0(x)\},$$

$$\text{nonincreasing on } \{\inf_{\mathcal{F}} F(x) = F_0(x)\},$$

$$\text{constant on } \{\inf_{\mathcal{F}} F(x) < F_0(x) < \sup_{\mathcal{F}} F(x)\}. \tag{3.2}$$

For $\mathcal{K}_\epsilon(G)$ and $\mathcal{L}_{\epsilon,\delta}(G)$, (3.1) and (3.2) are necessary features of F_0, under the assumption

$$G \text{ is strictly increasing on } (-\infty,\infty), \text{ with } I(G) < \infty$$

$$\text{and } -g'/g \text{ twice continuously differentiable.} \tag{3.3}$$

See Section 2 of Wiens (1986) for $\mathcal{K}_\epsilon$; the extension to $\mathcal{L}_{\epsilon,\delta}$ is straightforward. In any $\mathcal{F}$, if (3.1) and (3.2) hold, then so does (1.8). Using $K_0'(u) = \frac{1}{2}L_0' \circ F_0^{-1}(u)$ and then integrating by parts gives

$$\begin{aligned} 1 - \int_0^1 K_0(u)\, dF^{-1}(u) &= \lim_{u\to 0} K_0(u)\{F^{-1}(u) - F_0^{-1}(u)\} \\ &\quad - \lim_{u\to 1} K_0(u)\{F^{-1}(u) - F_0^{-1}(u)\} \\ &\quad + \frac{1}{2}\int_{-\infty}^{\infty} \{F^{-1} \circ F_0(x) - x\} f_0(x)\, dL_0(x) \ge 0. \end{aligned} \tag{3.4}$$

EXAMPLE 3.1 (Kolmogorov and Lévy neighbourhoods). Sacks and Ylvisaker (1972) showed that the saddle-point property fails for L-estimation in $\mathcal{K}_\epsilon(\Phi)$ if $\epsilon \geq 0.07$; Collins and Wiens (1989) extended this result to $\epsilon > 0$ and to more general, but still strongly unimodal, G in Lévy as well as Kolmogorov neighbourhoods. Collins (1983) established that R-estimation *is* minimax in $\mathcal{K}_\epsilon(\Phi)$; see Collins and Wiens (1989) for generalizations to $\mathcal{L}_{\epsilon,\delta}(G)$, G strongly unimodal. The following consequence of Theorems 2.1 and 2.2 subsumes all of these results, and greatly simplifies their proofs.

THEOREM 3.1. *Under the assumption* (3.3), *the saddle-point property* (1.6) *fails for L-estimation in every neighbourhood* $\mathcal{F} = \mathcal{K}_\epsilon(G)$ *or* $\mathcal{F} = \mathcal{L}_{\epsilon,\delta}(G)$. *The saddle-point property holds for R-estimation in these neighbourhoods if, and only if,* F_0 *is strongly unimodal. If* F_0 *is not strongly unimodal, then* $\sup_{\mathcal{F}} V_R(F) = \infty$.

Proof. Under (3.3), further necessary features of F_0 can be shown, as in Section 2 of Wiens (1986), to be:

(a) There is a set I of finite, symmetrically placed open intervals on each of which $F_0(x)$ is strictly between the boundaries defining $\mathcal{F}$, and on each of which $L_0(x)$ is constant.

(b) There is a set J of finite, symmetrically placed closed intervals on which $F_0(x)$ is on one of the boundaries.

(c) $(I \cup J)^c$ is of the form $(-\infty, -b) \cup (b, \infty)$; on these intervals $\psi_0(x)$ is constant; i.e., F_0 has exponential tails.

Define $\mathcal{F}_0' \subseteq \mathcal{F}$ to be those strictly increasing F, with $I(F) < \infty$, which agree with F_0 on J. For $x \in I$, we have $K_0' \circ F_0(x) = \frac{1}{2}L_0'(x) = 0$, so that K_0 is constant on $F_0\{I\}$. On $F_0\{J\}$, $F^{-1} \equiv F_0^{-1}$; and on $F_0\{(I \cup J)^c\}$, $K_0 \equiv 0$. It follows that $\int_0^1 K_0(u)\, dF^{-1}(u) = \int_0^1 K_0(u)\, dF_0^{-1}(u) = 1$ for $F \in \mathcal{F}_0'$, so that $\mathcal{F}_0' \subseteq \mathcal{F}_0$, where $\mathcal{F}_0$ is as in Theorem 2.1. Continuity considerations now dictate that each $c_i(F)$ there be zero. Note that the set I of (a) above must contain a neighbourhood of the origin (since F_0 is necessarily symmetric) and that $\psi_0'(x)$ is continuous and nonzero in this neighbourhood—the constant solution to $L_0(x) = \text{const}$, satisfying as well $\psi_0(0) = 0$, is clearly untenable.

Thus, by Theorem 2.1, F_0 minimizes $V_L(F)$ over $\mathcal{F}_0$, there exists $F \in \mathcal{F}_0$ violating (ii) of Theorem 2.1, and $V_L(F)$ strictly exceeds $V_L(F_0)$ at any such F.

For R-estimation, the statements of the theorem follow directly from Theorem 2.2 and (3.4). Q.E.D.

REMARKS. We can now show that the first conclusion of Theorem 2.3 can fail if ψ_0 is nonmonotone. In $\mathcal{K}_\epsilon(G)$, with G the Cauchy d.f. and $\epsilon < 0.0377$, there is an interval $(a,b) \in I$ on which ψ_0 is a strictly decreasing solution to $L_0(x) = \text{const}$, and f_0 is decreasing and positive. See Wiens (1986, Example 2). Define $\mathcal{F}_1$ as in Theorem 2.2; then $\sup_{\mathcal{F}_1} V_R(F) = \infty$. In contrast, $V_L(F)$ is bounded in $\mathcal{F}_1$. As in the proof of Theorem 2.1, we have $\int_0^1 K_0(u)\, dF^{-1}(u) = 1$ for $F \in \mathcal{F}_1$. It then follows from (2.4), (2.5), and an easy calculation that $V_L(F) = V_L(F_0) + 4\alpha^4 \int_0^1 A^2(u)\, du$, where $K_0(u) \equiv -\alpha^2$ on $(F_0(a), F_0(b))$ and

$$A(u) = \int_{F_0(a)}^{F_0(b)} \frac{I[u \leq t] - t}{f_0 \circ F_0^{-1}(t)}\, d\{F_0^{-1}(t) - F^{-1}(t)\}$$

has $|A(u)| \leq 4(b-a)/f_0(b) < \infty$. Thus for $\epsilon < 0.0377$,

$$V_L(F_0) = \inf_{\mathcal{F}_1} V_L(F) < \sup_{\mathcal{F}_1} V_L(F) < \sup_{\mathcal{F}_1} V_R(F) = \infty. \tag{3.5}$$

For $\epsilon > 0.0377$, F_0 is strongly unimodal and R-estimation *is* minimax.

EXAMPLE 3.2 (ϵ-Contamination neighbourhoods). For $\mathcal{G}_\epsilon(G)$, with G strongly unimodal and satisfying (3.3), Jaeckel (1971) showed that L-estimation is minimax, by verifying (2.2). In this case, ψ_0 is nondecreasing, so that by Theorem 2.3, R-estimation is minimax as well. This was shown directly by Jaeckel (1971).

In this class, $F_0 = (1-\epsilon)G + \epsilon H_0$ places all contaminating mass H_0 on intervals on which $\psi_0(x)$ is the constant solution to $L_0(x) = \text{const}$. A partial converse is then given by

THEOREM 3.2. *Let $\mathcal{F} = \mathcal{G}_\epsilon(G)$, with G satisfying* (3.3) *and $F_0 = (1-\epsilon)G + \epsilon H_0$ the minimum-information distribution in $\mathcal{F}$. If there is an interval $[a,b]$ with $H_0' = h_0(x) > 0$ on (a,b) and $\psi_0(x)$ nonconstant on $[a,b]$, then the saddle-point property fails for L-estimation. If there exists such an interval on which $\psi_0(x)$ is strictly decreasing, then the saddle-point property fails for R-estimation as well, and* (3.5) *holds, where $\mathcal{F}_1$ is as in Theorem* 2.2.

Proof. On any interval $[a,b]$, with $h_0(x) > 0$ on (a,b), $\psi_0(x)$ is a continuously differentiable solution to $L_0(x) = \text{const}$. [See Theorem 3 of Collins and Wiens (1985).] Then $K_0(u)$ is constant on $(F_0^{-1}(a), F_0^{-1}(b))$, so that if $\mathcal{F}_1$ is as in Theorem 2.2, we have $\int_0^1 K_0(u)\, dF^{-1}(u) = 1$. Thus $\mathcal{F}_1 \subseteq \mathcal{F}_0$, with $\mathcal{F}_0$ as in Theorem 2.1, and we conclude that F_0 minimizes $V_L(F)$ over $\mathcal{F}_1$. Any $F \neq F_0$ which places all of its contaminating mass on $(-b,-a)\cup(a,b)$ has $V_L(F) \gneqq V_L(F_0)$. For R-estimation, Theorem 2.2 applies directly. Now (3.5) follows exactly as before. Q.E.D.

REMARK. We note that (3.5) holds, for all $\epsilon > 0$, if G is a Student's t-distribution. This follows from Theorem 3.2, together with Example 3.2 of Collins and Wiens (1985), where it is shown that $h_0(x)$ is of the required form.

EXAMPLE 3.3. This example shows that:

(a) Even if ψ_0 is strictly increasing, (1.8) is not a *necessary* condition for R-estimation to be minimax.

(b) The Hodges-Lehmann estimator is the minimax-variance R-estimator in the largest convex class in which the logistic distribution $F_0(x) = (1+e^{-x})^{-1}$ minimizes the Fisher information; and $\sup V_L(F) = \infty$ in this class.

For this $F_0(x)$, we have $\psi_0(x) = \tanh \frac{1}{2}x$, $L_0(x) = 3(1 - 2\tanh^2 \frac{1}{2}x)$, $K_0(u) = 6u^2(1-u)^2$. Then as at (1.7), any convex class in which $I(F_0) = \min$ is a subset of $\mathcal{F}_L = \{F \mid E_F[\tanh^2 \frac{1}{2}X] \leq \frac{1}{3}\}$. It is easy to see that $\sup_{\mathcal{F}_L} \int_0^1 K_0(u)\, dF^{-1}(u) = \infty$; hence by (2.4) and (2.5), $\sup_{\mathcal{F}_L} V_L(F) = \infty$. See also the remark on p. 72 of Huber (1981).

The efficient R-estimator at F_0 is the Hodges-Lehmann estimator, with $J_0(t) = 3(2t-1)$ and $D_R(F) = 6\int_{-\infty}^{\infty} f^2(x)\, dx$. Put $F_\lambda = (1-\lambda)F_0 + \lambda F_1$. Then $D_R(F_\lambda)$ is a convex function of λ; hence (2.7) is equivalent to "$(d/d\lambda)\, D_R(F_\lambda)|_{\lambda=0} \geq 0$, all $F_1 \in \mathcal{F}_L$". This becomes exactly the definition of $\mathcal{F}_L$, after a calculation.

EXAMPLE 3.4. The robustness of the R-estimator of Example 3.3 is destroyed if the score function is truncated, say by replacing it by $\psi_*(x) = \{\tanh \frac{1}{2}x, \tanh \frac{1}{2}a, \tanh -(\frac{1}{2}a)\}$ on $\{|x| \leq a,\ x \geq a,\ x \leq -a\}$ respectively. The corresponding F_* has density $f_*(x) = \{f_*(0)\operatorname{sech}^2 \frac{1}{2}x, f_*(a)e^{a-|x|}\}$ on $\{|x| \leq a,\ |x| \geq a\}$ respectively, and F_* has minimum

information in

$$\mathcal{F}_* = \left\{ F \,\middle|\, \int_{-a}^{a} [2f_*(x) - f_*(a)]\, d(F - F_*)(x) \geq 0 \right\}.$$

Sacks and Ylvisaker (1982) constructed an F_1 for which equality is attained in the definition of $\mathcal{F}_*$ and with $D_R(F_1) < 1$, so that R-estimation fails to satisfy (1.6) in $\{F = (1-\lambda)F_* + \lambda F_1;\ 0 \leq \lambda \leq 1\}$, a convex class in which F_* minimizes information. As shown by Sacks and Ylvisaker, or now by appealing to Theorem 2.3, L-estimation also fails to satisfy (1.6) in this class.

In general, suppose that ψ_0 is strictly increasing on a finite interval $[-a, a]$, and constant on $x \geq a$ and $x \leq -a$. Suppose there is an $F_1 \in \mathcal{F}$, strictly increasing on $[-a, a]$, whose restriction to $[-a, a]$ satisfies

(a) $\int_{-a}^{a} dF_1 < \int_{-a}^{a} dF_0$,
(b) $\int_{-\infty}^{\infty} L_0(x)\, d(F_1 - F_0)(x)\ [= \int_{-a}^{a} \{L_0(x) + \psi_0^2(a)/I(F_0)\}\, d(F_1 - F_0)(x)] = 0$.

Suppose also

(c) The structure of $\mathcal{F}$ places no restrictions on the behaviour of its members in $|x| > a$.

Then R-estimation fails to satisfy (1.6) in $\mathcal{F}$. This is because

$$\begin{aligned} D_R(F) &= \int_{F(-a)}^{F(a)} J_0'(u) f \circ F^{-1}(u)\, du + 2 \int_{F(a)}^{F_0(a)} J_0'(u) \circ F^{-1}(u)\, du \\ &= \phi_1(F) + \phi_2(F), \end{aligned}$$

say. A calculation gives

$$\frac{d}{d\lambda}\phi_1(F_\lambda)\bigg|_{\lambda=0} = \frac{1}{2} \int_{-\infty}^{\infty} L_0(x)\, d(F_1 - F_0)(x) + \frac{2\psi_0'(a^-)(F_1 - F_0)(a)}{I(F_0)} < 0$$

by (a) and (b), so that $\phi_1(F_\lambda) < \phi_1(F_0) = 1$ for sufficiently small $\lambda > 0$. Since (a) and (b) restrict F_λ only in $[-a, a]$, we may now, by (c), extend any such F_λ in $|x| > a$ so as to make $\phi_2(F_\lambda)$ sufficiently small that $D_R(F_\lambda) > 1$, violating (2.7) and hence (1.6).

Thus, any $\mathcal{F}$ in which the Fisher information is minimized at a strongly unimodal F_0 with exponential tails can be embedded in a neighbourhood $\mathcal{F}'$ in which F_0 continues to have minimum information, but in which the saddle-point property fails for both L- and R-estimation.

REFERENCES

Chernoff, H., and Savage, I.R. (1958). Asymptotic normality and efficiency of certain non-parametric test statistics. *Ann. Math. Statist.*, 29, 972–994.

Chernoff, H., Gastwirth, J.L., and Johns, M.V., Jr. (1967). Asymptotic distribution of linear combinations of order statistics, with applications to estimation. *Ann. Math. Statist.*, 38, 52–72.

Collins, J.R. (1983). On the minimax property for R-estimators of location. *Ann. Statist.*, 11, 1190–1195.

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

Collins, J.R., and Wiens, D.P. (1989). Minimax properties of M-, R- , and L-estimators of location in Lévy neighbourhoods. *Ann. Statist.*, 17, 327–336.

Froda, S. (1986). Comparison of efficient L- and R-estimators of location. *Canad. J. Statist.*, 14, 75–79.

Gastwirth, J.L., and Wolff, S. (1968). An elementary method for obtaining lower bounds on the asymptotic power of rank tests. *Ann. Math. Statist.*, 39, 2128–2130.

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.

Jaeckel, L.A. (1971). Robust estimates of location: Symmetry and asymmetric contamination. *Ann. Math. Statist.*, 42, 1020–1034.

Lehmann, E.L. (1983). *Theory of Point Estimation*. Wiley, New York.

Mason, D.M. (1983). A minimax criterion for choosing weight functions for L-estimates of location. *Ann. Statist.*, 11, 317–325.

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

Sacks, J., and Ylvisaker, D. (1982). L- and R-estimation and the minimax property. *Ann. Statist.*, 10, 643–645.

Scholz, F.-W. (1974). A comparison of efficient location estimators. *Ann. Statist.*, 2, 1323–1326.

Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

Stone, C.J. (1974). Asymptotic properties of estimators of a location parameter. *Ann. Statist.*, 2, 1127–1137.

Wiens, D. (1986). Minimax variance M-estimators of location in Kolmogorov neighbourhoods. *Ann. Statist.*, 14, 724–732.

Received 22 June 1989
Revised 24 October 1989
Accepted 26 October 1989

Department of Statistics and Applied Probability
University of Alberta
434 Central Academic Building
Edmonton, AB T6G 2G1