The Annals of Statistics
1983, Vol. 11, No. 4, 1190–1195

ON THE MINIMAX PROPERTY FOR *R*-ESTIMATORS OF LOCATION[1]

BY JOHN R. COLLINS

University of Calgary

Consider the problem of estimating the unknown location parameter θ based on a random sample from $F(x-\theta)$, where F is an unknown member of the class of distribution functions $\mathscr{F} = \{F: F$ is symmetric about 0 and $\sup_x |F(x) - \Phi(x)| \le \varepsilon\}$, where Φ denotes the standard normal distribution function. Huber (1964) showed that M-estimation has a minimax property for this model, whereas Sacks and Ylvisaker (1972) showed that L-estimation fails to have the minimax property. It is shown here that R-estimation *does* have the minimax property for this model.

1. Introduction and summary. Consider the robust estimation problem of Huber (1964). Random samples are obtained from $F(x-\theta)$, where θ is an unknown parameter to be estimated and F is an unknown member of a convex and vaguely compact class $\mathscr{F}$ of distributions symmetric about 0. Under suitable regularity conditions, consistent and asymptotically normal estimates of θ can be obtained using the class of M-estimators (maximum likelihood type estimators), L-estimators (linear combinations of order statistics), or R-estimators (estimators based on rank tests). For definitions of these classes of estimators see, e.g., Chapter 3 of Huber (1981).

We briefly summarize Huber's asymptotic minimax theory. Let $\mathscr{S}$ be a class of estimators (either the $M-$, $L-$, or R-estimators), let T denote a functional representing a member of $\mathscr{S}$, and let $V(T, F)$ denote the corresponding asymptotic variance functional (where T ranges over $\mathscr{S}$ and F ranges over $\mathscr{F}$). Let F_0 denote the member of $\mathscr{F}$ minimizing $I(F)$, the Fisher information for location; and let T_0 denote the member of $\mathscr{S}$ which is asymptotically efficient when F_0 is the true error distribution (i.e., $V(T_0, F_0) = 1/I(F_0)$). Suppose that $\mathscr{S}$ and $\mathscr{F}$ are such that

$$\sup\{V(T_0, F) \colon F \in \mathscr{F}\} \le V(T_0, F_0). \tag{1.1}$$

Then we say that the *minimax property* holds, for an immediate consequence of (1.1) is that $\inf_{T\in\mathscr{S}} \sup_{F\in\mathscr{F}} V(T, F)$ is equal to $1/I(F_0)$, with the infimum attained at T_0. In cases where (1.1) fails, the minimax variance problem does not have a saddlepoint and, in fact, it can then be easily seen that $\inf_{T\in\mathscr{S}} \sup_{F\in\mathscr{F}} V(T, F) > 1/I(F_0)$.

Table 1 summarizes the results to date on the minimax property for M-, L-, and R-estimators for two important models for $\mathscr{F}$, each representing close neighborhoods of the standard normal error distribution: the gross errors model and the Kolmogorov model. Huber (1964) showed that the minimax property for M-estimators holds quite generally and for the gross errors and Kolmogorov models in particular. Jaeckel (1971) verified that the minimax property holds for both L- and R-estimators in the gross errors model. (See also Section 4.7 of Huber, 1981.) Sacks and Ylvisaker (1972) showed that the minimax property fails for L-estimators in the Kolmogorov model, at least when $\varepsilon > 0.07$. In this paper we settle the open question of whether the minimax property holds for R-estimators in the Kolmogorov model.

Received June 1982.

[1] This research was supported by the Natural Sciences and Engineering Research Council of Canada under Grant A-4499, while the author was visiting the Department of Statistics, University of California, Berkeley.

AMS 1970 *subject classifications.* Primary 62G35; secondary 62G05.

Key words and phrases. Robust estimation of location, R-estimators, minimax property.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Statistics.* www.jstor.org

TABLE 1.
Does the minimax property hold?

Class of Distributions $\mathscr{F}$	**M-estimators**	**L-estimators**	**R-estimators**
$\{F: F = (1-\varepsilon)\Phi + \varepsilon G$ for some G symmetric about $0\}$ (*gross errors model*).	YES [Huber, 1964]	YES [Jaeckel, 1971]	YES [Jaeckel, 1971]
$\{F: F$ is symmetric about 0 and $\sup_x \lvert F(x) - \Phi(x) \rvert \le \varepsilon\}$ (*Kolmogorov model*)	YES [Huber, 1964]	NO (when $\varepsilon > .07$) [Sacks-Ylvisaker, 1972]	OPEN PROBLEM

In Section 2, we give required definitions and notation. In Section 3 we show that the minimax property *does* hold for R-estimators in the Kolmogorov model and present a proof valid for all values of ε, $0 < \varepsilon < 1$.

The result is somewhat surprising for two reasons. One reason is, as noted by Huber (1981), that unlike the case for M-estimators, $1/V(F, T)$ is not a convex function of F in the case of R-estimators. Another reason is that Sacks and Ylvisaker (1982) have recently constructed an example of a convex class $\mathscr{F}$ for which the minimax property fails for R-estimators.

A conclusion, stated roughly, is that—according to the minimax variance criterion—one can do as well estimating a location parameter with R-estimators as with M-estimators, at least in two error distribution models (gross errors and Kolmogorov) of practical importance. Of course Sacks and Ylvisaker (1982) found a model $\mathscr{F}$ for which M-estimators do strictly better than R-estimators (and also L-estimators). However, their example is somewhat artificial and does not seem to have the same degree of practical importance as the gross errors and Kolmogorov models.

The question raised by Sacks and Ylvisaker (1982) of whether there are general conditions on $\mathscr{F}$ which guarantee the minimax property for L- or R-estimators remains open. An ad-hoc case-by-case check of the minimax property seems to be necessary.

2. Definitions and notation. For a fixed value of ε, $0 < \varepsilon < 1$, define the set of distribution functions $\mathscr{F}_\varepsilon$ by

$$\mathscr{F}_\varepsilon = \{F: F \text{ is symmetric about 0 and } \sup_x \lvert F(x) - \Phi(x) \rvert \le \varepsilon\} \tag{2.1}$$

where Φ is the standard normal distribution function, i.e., $\Phi(x) = \int_{-\infty}^{x} \phi(t)\,dt$ where $\phi(t) = (2\pi)^{-1/2}\exp(-t^2/2)$. Also denote by $\mathscr{F}'_\varepsilon$ the sub-class of F's in $\mathscr{F}_\varepsilon$ which have an absolutely continuous density f.

Let f_0 be the (necessarily) absolutely continuous density of the unique F_0 in $\mathscr{F}_\varepsilon$ with minimum Fisher information, and define ψ_0 by $\psi_0(x) = -f'_0(x)/f_0(x)$. Then there are positive constants C_0, C_1, ω, λ, k_0 and k_1 (depending on ε) such that

$$\begin{aligned} f_0(x) = f_0(-x) &= C_0\cos^2(\tfrac{1}{2}\omega x) && 0 \le x \le k_0 \\ &= \phi(x) && k_0 \le x \le k_1 \\ &= C_1 e^{-\lambda x} && x \ge k_1, \end{aligned} \tag{2.2}$$

and

$$\begin{aligned} \psi_0(x) = -\psi_0(-x) &= \omega \tan(\tfrac{1}{2}\omega x) && 0 \le x \le k_0 \\ &= x && k_0 \le x \le k_1 \\ &= \lambda && x \ge k_1. \end{aligned} \tag{2.3}$$

There are two cases to consider. In Case A, $\varepsilon < \varepsilon_0 \cong 0.0303$ [Huber, 1964], we have

$k_0 < k_1$; in Case B, $\varepsilon > \varepsilon_0$ [Sacks and Ylvisaker, 1972], we have $k_0 = k_1$. For the determination of the values for the constants C_1, C_1, ω, λ, k_0 and k_1, see Huber (1981), page 86, example 5.3. We list some properties of the solution that will be used in the next section:

(P1) The distribution function F_0 is strictly monotone increasing.

(P2) The density f_0 is absolutely continuous with $\int_{-\infty}^{\infty} f_0(x)\,dx = 1$.

(P3) The function $\psi_0 = -f_0'/f_0$ is absolutely continuous with piecewise continuous derivative ψ_0'.

(P4) For all $x \in [k_0, k_1]$, the identity $F_0(x) = \Phi(x) - \varepsilon$ holds. [In Case B, this just reduces to $F_0(k_0) = \Phi(k_0) - \varepsilon$.]

(P5) In Case A, we have $\psi_0'(k_0 - 0) \geq \psi_0'(k_0 + 0)$. [To see this, note that $\psi_0(0) = 0$, $\psi_0(k_0) = k_0$, and ψ_0' is positive and increasing on $(0, k_0)$. This forces $\psi_0(x) < x$ for $0 < x < k_0$, so that $\psi_0'(k_0 - 0) > 1 = \psi_0'(k_0 + 0)$.]

Let F be a member of $\mathscr{F}_\varepsilon$ with density f symmetric about 0. Consider the problem of estimating θ based on random samples from $F(x - \theta)$ using an R-estimator generated by a score function J (with $J(1 - t) = -J(t)$ for all $t \in [0, 1]$) as defined in Section 3.4 of Huber (1981). Then under suitable regularity conditions on J and F, the R-estimator of θ is consistent and asymptotically normal, with asymptotic variance given by

$$V(J, F) = \frac{\int J^2[F(x)]f(x)\,dx}{\left\{\int J'[F(x)]f^2(x)\,dx\right\}^2} \tag{2.4}$$

When the true underlying distribution is F_0, then any choice of J satisfying $J(F_0(x)) = c\psi_0(x)$ for some $c \neq 0$ yields an asymptotically efficient R-estimator [see Section 3.5 of Huber, 1981]. For then $V(J, F_0) = 1/I(F_0)$ where $I(F_0) = \int[(f_0')^2/f_0]\,dx$, the Fisher information. To be specific, we set $c = 1$ and define (without loss of generality) a specific asymptotically efficient score function J_0 as follows:

$$J_0(F_0(x)) = \psi_0(x). \tag{2.5}$$

It will be useful to note that if $F \in \mathscr{F}_\varepsilon'$, then both $J_0[F(x)]$ and $f(x)$ are absolutely continuous with derivatives $J_0'[F(x)]f(x)$ and $f'(x)$, respectively, so that integration-by-parts yields

$$V(J_0, F) = \frac{\int J_0^2[F(x)]f(x)\,dx}{\left[-\int J_0[F(x)]f'(x)\,dx\right]^2}. \tag{2.6}$$

3. The minimax result. We now show that the minimax property holds; i.e., in the notation of Section 2, that $\sup\{V(J_0, F): F \in \mathscr{F}_\varepsilon\} = V(J_1, F_0)$. One technical difficulty that arises is that there is no closed form expression for $J_0(t)$ as a function of $t \in [0, 1]$. However, a tractible expression for $J_0[F(x)]$ is obtained by the following device: write $J_0 \circ F(x) = J_0 \circ F_0 \circ F_0^{-1} \circ F(x) = \psi_0[F_0^{-1} \circ F](x)$.

THEOREM. *Let ε, $0 < \varepsilon < 1$, be fixed and let $\mathscr{F}_\varepsilon$ be defined by* (2.1). *Let F_0 be the distribution minimizing $I(F)$ over $\mathscr{F}_\varepsilon$ and let J_0 be the asymptotically efficient score function corresponding to F_0, as defined in Section* 2. *Then the minimax property holds*, i.e.,

$$\sup\{V(J_0, F): F \in \mathscr{F}_\varepsilon\} = V(J_0, F_0). \tag{3.1}$$

PROOF. Since we clearly have that

$$\sup\{V(J_0, F): F \in \mathscr{F}'_\varepsilon\} = \sup\{V(J_0, F): F \in \mathscr{F}_\varepsilon\}, \tag{3.2}$$

it suffices to show that the left-hand side of (3.2) is equal to $V(J_0, F_0)$. Let F be an arbitrary member of $\mathscr{F}'_\varepsilon$, with absolutely continuous density f. We will show that $V(J_0, F) \le V(J_0, F_0)$. From formula (2.6) for $V(J_0, F)$, it follows that it suffices to show that both

$$\int J_0^2[F(x)]f(x)\,dx \le \int J_0^2[F_0(x)]f_0(x)\,dx \tag{3.3}$$

and

$$-\int J_0[F(x)]f'(x)\,dx \ge -\int J_0[F_0(x)]f'_0(x)\,dx. \tag{3.4}$$

Note that the right-hand side of each of (3.3) and (3.4) is equal to $I(F_0) > 0$.

To prove (3.3), first note that by property (P1), the change of variables $t = F_0(x)$ yields that $\int J_0^2[F_0(x)]f_0(x)\,dx = \int_0^1 J_0^2(t)\,dt$. Some more care is required in handling the expression $\int J_0^2[F(x)]f(x)\,dx$, since there may exist intervals on which F is constant, and in fact F may be substochastic (we know only that $1 - \varepsilon \le F(\infty) \le 1$.) So let S be a subset of the real line, obtained by deleting intervals on which $f(x) \equiv 0$, with the property that the change of variables $t = F(x)$ yields a strictly increasing map of S onto $(F(-\infty), F(\infty))$. Then we have that

$$\begin{aligned}\int J_0^2[F(x)]f(x)\,dx &= \int_S J_0^2[F(x)]f(x)\,dx = \int_{F\{S\}} J_0^2(t)\,dt\\ &\le \int_0^1 J_0^2(t)\,dt = \int J_0^2[F_0(x)]f_0(x)\,dx,\end{aligned} \tag{3.5}$$

proving (3.3).

To prove (3.4), first note that by symmetry it is enough to show that

$$-\int_0^\infty J_0[F(x)]f'(x)\,dx \ge \tfrac{1}{2} I(F_0). \tag{3.6}$$

Note that it follows from property (P1) that F_0^{-1} is strictly monotone increasing and that $F_0 \circ F_0^{-1}(x) \equiv x$. We define the function q by

$$q(x) = F_0^{-1} \circ F(x), \tag{3.7}$$

and note that q is a continuous and monotone non-decreasing map of $[0, \infty]$ onto $[0, F_0^{-1} \circ F(\infty)]$ ($q(0) = 0$ since $F(0) = F_0(0) = \tfrac{1}{2}$ by symmetry). It follows that q' exists a.e. with $q'(x) \ge 0$ a.e. x. Note that $F(x) = F_0[q(x)]$, that $J_0[F(x)] = \psi_0[q(x)]$, and that $f(x) = f_0[q(x)]\cdot q'(x)$ a.e. x.

Rewriting (3.6) using the notation (3.7), it remains to establish that

$$-\int_0^\infty \psi_0[q(x)]\,f'(x)\,dx \ge \tfrac{1}{2} I(F_0). \tag{3.8}$$

Note that both ψ_0 and q are continuous and monotone non-decreasing, so that $\psi_0 \circ q$ is a continuous monotone non-decreasing map of $[0, \infty]$ onto $[0, \psi_0(q(\infty))]$. It follows that $\psi_0 \circ q$ is absolutely continuous on $[0, A]$ for all $A > 0$. Furthermore, the total variation of $\psi_0 \circ q$ over $[0, \infty)$ is $\le \lambda < \infty$. It follows that $\psi_0 \circ q(x) = \int_0^x [\psi_0 \circ q]'(y)\,dy$ for all $x > 0$. Also, by the definition of $\mathscr{F}'_\varepsilon$, we have $f(x) - f(0) = \int_0^x f'(y)\,dy$ for all $x > 0$. Hence it is valid to integrate by parts to obtain

$$
\begin{aligned}
-\int_0^\infty \psi_0[q(x)]f'(x)\,dx &= -\lim_{x\to\infty}\psi_0[q(x)]\,\lim_{x\to\infty}f(x) + \psi_0[q(0)]f(0) \\
&\quad + \int_0^\infty [\psi_0 \circ q]'(x)f(x)\,dx \\
&= \int_0^\infty \psi_0'[q(x)]q'(x)f_0[q(x)]q'(x)\,dx \\
&= \int_0^\infty [\psi_0' f_0](q(x))[q'(x)]^2\,dx.
\end{aligned}
\tag{3.9}
$$

Define the function p by $q(x) = x + p(x)$, so that $q'(x) = 1 + p'(x)$. Then from (3.9), we have

$$
\begin{aligned}
-\int_0^\infty \psi_0[q(x)]f'(x)\,dx &= \int_0^\infty [\psi_0' f_0](q(x))[1 + p'(x)]^2\,dx \\
&= \int_0^\infty [\psi_0' f_0](q(x))q'(x)\,dx \\
&\quad + \int_0^\infty [\psi_0' f_0](q(x))\cdot p'(x)[1 + p'(x)]\,dx \\
&= \int_0^{F_0^{-1}\circ F(\infty)} [\psi_0' f_0](y)\,dy + \int_0^\infty [\psi_0' f_0](q(x))p'(x)\,dx \\
&\quad + \int_0^\infty [\psi_0' f_0](q(x))[p'(x)]^2\,dx.
\end{aligned}
\tag{3.10}
$$

Now straightforward calculation from (2.2) and (2.3) yields that in Case A,

$$
\begin{aligned}
[\psi_0' f_0](x) &= \tfrac{1}{2} C_0\omega^2, & 0 \le x < k_0 \\
&= \phi(x), & k_0 < x < k_1 \\
&= 0, & x > k_1;
\end{aligned}
\tag{3.11}
$$

and in Case B,

$$
\begin{aligned}
[\psi_0' f_0](x) &= \tfrac{1}{2} C_0\omega^2, & 0 \le x < k_0 = k_1 \\
&= 0 & x > k_0 = k_1.
\end{aligned}
\tag{3.12}
$$

In either Case A or Case B, it follows from property (P4) that $F(\infty) \ge F(k_1) \ge \Phi(k_1) - \varepsilon = F_0(k_1)$, and so from (3.11) or (3.12), it follows that

$$
\int_0^{F_0^{-1}\circ F(\infty)} [\psi_0' f_0](y)\,dy = \int_0^\infty \psi_0'(y)f_0(y)\,dy = \tfrac{1}{2} I(F_0).
\tag{3.13}
$$

In view of (3.10) and (3.13), the inequality (3.8) is equivalent to

$$
\int_0^\infty [\psi_0' f_0](q(x))p'(x)\,dx + \int_0^\infty [\psi_0' f_0](q(x))[p'(x)]^2\,dx \ge 0.
\tag{3.14}
$$

But since $\psi_0'(x)f_0(x) \ge 0$ for all $x > 0$, the second term is non-negative; so in order to complete the proof of the theorem, we need only show that

$$\int_0^\infty [\psi_0' f_0](q(x))p'(x)\,dx \geq 0. \tag{3.15}$$

Consider first Case A, where $\varepsilon < \varepsilon_0$ and $k_0 < k_1$. Then $F(x) \geq \Phi(x) - \varepsilon = F_0(x)$ for all $x \in [k_0, k_1]$. Since $F(x) = F_0(x + p(x)) \geq F_0(x)$ for $x \in [k_0, k_1]$, we have that $p(x) \geq 0$ for all $x \in [k_0, k_1]$. Let $x_1 = \inf\{x : x + p(x) = k_0\}$ and $x_2 = \sup\{x : x + p(x) = k_1\}$. Then from (3.11), we calculate that

$$\begin{aligned}\int_0^\infty [\psi_0' f_0](q(x))p'(x)\,dx &= \tfrac{1}{2}\, C_0\omega^2[p(x_1) - p(0)] + \int_{x1}^{x_2} \phi[q(x)]p'(x)\,dx \\ &= \tfrac{1}{2}\, C_0\omega^2 p(x_1) + p(x_2)\phi(k_1) - p(x_1)\phi(k_0) \\ &\quad + \int_{x_1}^{x_2} q(x)\phi[q(x)]q'(x)p(x)\,dx.\end{aligned} \tag{3.16}$$

Thus to prove that (3.15) holds in Case A, it suffices to show that both

$$p(x) \geq 0 \quad \text{for all} \quad x \in [x_1, x_2]; \tag{3.17}$$

and

$$\tfrac{1}{2}\, C_0\omega^2 - \phi(k_0) \geq 0. \tag{3.18}$$

To obtain (3.17), first recall that $p(x) \geq 0$ for all $x \in [k_0, k_1]$. Since $q(x) = x + p(x)$ is monotone non-decreasing in x, and since $x_1 + p(x_1) = k_0$ and $k_0 + p(k_0) \geq k_0$, we must have that $x_1 \leq k_0$ and hence $p(x_1) \geq 0$. Similarly it is easily seen that $x_2 \leq k_1$ and $p(x_2) \geq 0$. To finish the verification of (3.17), we must show that if y satisfies $x_1 < y < k_0$, then $p(y) \geq 0$. But since $x_1 + p(x_1) = k_0$, we have $y + p(y) \geq k_0$, so that $p(y) \geq k_0 - y \geq 0$.

To prove (3.18) in Case A, note that it follows from property (P5) and continuity of f_0 at k_0 that $[\psi_0' f_0](k_0 - 0) \geq [\psi_0' f_0](k_0 + 0)$. But this inequality is exactly (3.18). This completes the proof of (3.15) in Case A.

Consider now Case B, where $\varepsilon > \varepsilon_0$ and $k_0 = k_1$. Then from (3.12), we calculate that

$$\int_0^\infty [\psi_0' f_0](q(x))p'(x)\,dx = \tfrac{1}{2}\, C_0\omega^2 p(x_1) \tag{3.19}$$

where $x_1 = \inf\{x : x + p(x) = k_0\}$. But we know from property (P4) that $p(k_0) \geq 0$. Hence $x_1 \leq k_0$ and $p(x_1) \geq 0$, so that inequality (3.15) holds in Case B. This completes the proof of the theorem. □

REFERENCES

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 estimation of location: symmetry and asymmetric contamination. *Ann. Math. Statist.* **42** 1024–1034.

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.

DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF CALGARY
2500 UNIVERSITY DRIVE N.W.
CALGARY, ALBERTA
CANADA T2N 1N4