The Canadian Journal of Statistics
Vol. 26, No. 2, 1998, Pages 231–238
La Revue Canadienne de Statistique

On the trimmed mean and minimax-variance L-estimation in Kolmogorov neighbourhoods*

Douglas P. WIENS, Eden K.H. WU and Julie ZHOU

University of Alberta, Chinese University of Hong Kong and *Lakehead University*

Key words and phrases: L-estimation, Kolmogorov neighbourhood, robustness, trimmed mean, minimax variance.
AMS 1991 subject classifications: Primary 62F35, 62G30; secondary 62F10, 62F12.

ABSTRACT

We consider the properties of the trimmed mean, as regards minimax-variance L-estimation of a location parameter in a Kolmogorov neighbourhood $\mathcal{K}_\epsilon(\Phi)$ of the normal distribution:

$$\mathcal{K}_\epsilon(\Phi) = \{ F \big| |F(x) - \Phi(x)| \leq \epsilon, -\infty < x < \infty; \; F \text{ symmetric}\}.$$

We first review some results on the search for an *L-minimax estimator* in this neighbourhood, i.e. a linear combination of order statistics whose maximum variance in $\mathcal{K}_\epsilon(\Phi)$ is a minimum in the class of L-estimators. The natural candidate — the L-estimate which is efficient for that member of $\mathcal{K}_\epsilon(\Phi)$ with minimum Fisher information — is known not to be a saddlepoint solution to the minimax problem. We show here that it is not a solution at all. We do this by showing that a smaller maximum variance is attained by an appropriately trimmed mean. We argue that this trimmed mean, as well as being computationally simple — much simpler than the efficient L-estimate referred to above, and simpler than the minimax M- and R-estimators — is at least "nearly" minimax.

RÉSUMÉ

Nous considérons les propriétés de la moyenne tronquée, en ce qui concerne l'estimation-L de variance minimax d'un paramètre de position dans un voisinage Kolmogorov $\mathcal{K}_\epsilon(\Phi)$ de la distribution normale:

$$\mathcal{K}_\epsilon(\Phi) = \{ F \big| |F(x) - \Phi(x)| \leq \epsilon, -\infty < x < \infty; \; F \text{ symetrique}\}.$$

Nous passons premièrement en revue quelques résultats sur la recherche d'un estimateur L-minimax dans ce voisinage, c'est à dire une combinaison linéaire de statistiques d'ordre dont la variance maximale dans $\mathcal{K}_\epsilon(\Phi)$ est un minimum dans la classe des estimateurs-L. Le candidat naturel — l'estimation-L qui est efficace pour ce membre de $\mathcal{K}_\epsilon(\Phi)$ avec information Fisher minimale — est reconnu comme n'étant pas une solution point-selle au problème minimax. Nous montrons ici que ce n'est pas du tout une solution. Nous faisons cela en montrant qu'une variance maximum plus petite est atteinte par une moyenne tronquée de façon appropriée. Nous soutenons que cette moyenne tronquée, en plus d'être facile à calculer — beaucoup plus simple que l'estimation-L efficace à laquelle nous nous réferrons plus haut — et plus simple que les estimateurs minimax M et R — est au moins presque minimax.

*This research was supported by the Natural Sciences and Engineering Research Council of Canada (D.P.W. and J.Z.) and by a Hong Kong UGC Direct Grant for Research (E.K.H.W.).

1. INTRODUCTION AND SUMMARY

Let $X_{1:n} \leq \cdots \leq X_{n:n}$ be the order statistics from a location family, distributed as $F(x-\theta)$ for a symmetric distribution function F. An L-estimate (for *linear combination of order statistics*) of θ is given by

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

for a *weights-generating* function $m(t)$ with $m(t) = m(1-t)$ and $\int_0^1 m(t)\,dt = 1$. One of the simplest examples of an L-estimate is the α-trimmed mean, with weights-generating function

$$m_1(t) = \frac{I(\alpha \leq t \leq 1-\alpha)}{1-2\alpha}. \tag{1.1}$$

The estimate corresponding to $m_1(t)$ is the average of the inner order statistics after omitting the $[\alpha n]$ largest and $[\alpha n]$ smallest observations.

Under appropriate regularity conditions on m and F [see Serfling (1980) and references cited therein], $\sqrt{n}(T_n - \theta)$ is asymptotically normally distributed with mean zero and variance

$$V(m,F) = 2\int_{\frac{1}{2}}^{1} h_F^2(u;m)\,du,$$

where

$$h_F(u;m) = \int_{\frac{1}{2}}^{u} m(t)\,dF^{-1}(t) \tag{1.2}$$

and $F^{-1}(t) = \inf\{x \mid F(x) \geq t\}$.

Suppose that F has an absolutely continuous density f, that the score function

$$\psi_F(x) = -\frac{f'}{f}(x)$$

is absolutely continuous, and that Fisher information

$$I(F) = \int_{-\infty}^{\infty} \psi_F^2(x) f(x)\,dx$$

is positive and finite. Put

$$m_F(t) = \frac{\psi_F'(F^{-1}(t))}{I(F)}.$$

Then $V(m_F, F) = I(F)^{-1}$, i.e., the L-estimator generated by $m_F(\cdot)$ is asymptotically efficient.

If F is unknown, and is assumed only to belong to a certain class $\mathcal{F}$ of distributions in which Fisher information is minimized by a member F_0, then it is a plausible conjecture that $m_0 := m_{F_0}$ is *L-minimax*, i.e., that

$$\sup_{F\in\mathcal{F}} V(m_0,F) = \inf_m \sup_{F\in\mathcal{F}} V(m,F). \tag{1.3}$$

The main results of this paper are that:

(1) In the Kolmogorov neighbourhood of the normal distribution defined below, (1.3) is false.

(2) (1.3) is nearly — to within bounds of negligible practical importance — attained if m_0 is replaced by m_1 of (1.1), with α chosen to minimize the resulting left-hand member of (1.3). That is, the corresponding α-trimmed mean is "nearly" minimax.

We first give a brief history of the search for solutions to the problems represented by (1.3). When (1.3) holds, it is generally established by showing that (m_0, F_0) is a *saddlepoint*:

$$V(m_0, F) \leq V(m_0, F_0) = I(F_0)^{-1} \leq V(m, F_0) \qquad \forall m, \forall F \in \mathcal{F}. \tag{1.4}$$

The saddlepoint condition (1.4) is sufficient but not necessary for (1.3). Jaeckel (1971) established (1.4) for the *gross-errors* neighbourhood of the normal distribution:

$$\mathcal{G}_\epsilon(\Phi) = \{ F | F(x) = (1 - \epsilon)\Phi(x) + \epsilon H(x);\ H \text{ symmetric} \}.$$

The corresponding L-minimax estimator is an α-trimmed mean. Sacks and Ylvisaker (1972) showed that (1.4) *fails* in Kolmogorov neighbourhoods of the normal distribution:

$$\mathcal{K}_\epsilon(\Phi) = \{ F \,|\, |F(x) - \Phi(x)| \leq \epsilon,\ -\infty < x < \infty;\ F \text{ symmetric} \},$$

for $\epsilon \geq 0.07$ approximately. They did this by exhibiting an $F \in \mathcal{K}_\epsilon(\Phi)$ with $V(m_0, F) > I(F_0)^{-1}$. Collins and Wiens (1989) extended these results to all $\epsilon \in (0, 0.5)$, by employing a completely different method of proof (see Lemma 2.2 below). Their results apply as well in Lévy neighbourhoods

$$\mathcal{L}_{\epsilon,\delta}(G) = \{ F | G(x - \delta) - \epsilon \leq F(x) \leq G(x + \delta) + \epsilon \}$$

under conditions on G which imply the existence of a strongly unimodal density. (Note that $\mathcal{K}_\epsilon = \mathcal{L}_{\epsilon,0}$.) Wiens (1990) showed that in fact the saddlepoint property must fail in *all* neighbourhoods $\mathcal{K}_\epsilon(G)$ and $\mathcal{L}_{\epsilon,\delta}(G)$ if G is strictly increasing, with a sufficiently smooth score function and finite Fisher information.

None of these results for Kolmogorov and Lévy neighbourhoods answer the questions:

(1) Does the weights-generating function which is efficient for the minimum information distribution furnish an *L-minimax* estimator?

(2) If the answer to the previous question is negative, is there another L-estimate which is minimax, or at least approximately so?

We answer the first question above in the negative, for $\mathcal{K}_\epsilon(\Phi)$ and $\epsilon \in [0.029, 0.481]$. We do this by exhibiting distributions F_1, $F_2 \in \mathcal{K}_\epsilon(\Phi)$, such that with m_1 given by (1.1) and α chosen to minimize $V(m_1, F_1)$ we have

$$\sup_{F \in \mathcal{K}_\epsilon(\Phi)} V(m_1, F) = V(m_1, F_1) < V(m_0, F_2) \leq \sup_{F \in \mathcal{K}_\epsilon(\Phi)} V(m_0, F). \tag{1.5}$$

Thus the maximum variance of the α-trimmed mean is less than that of the L-estimate corresponding to m_0. We find that this α-trimmed mean furnishes an answer to the second question above, in that $\sup_{F \in \mathcal{K}_\epsilon(\Phi)} V(m_1, F)$ in (1.5) is only slightly larger than $I(F_0)^{-1}$. The difference is in fact so small as to indicate that the search for an L-minimax estimator for $\mathcal{K}_\epsilon(\Phi)$, although mathematically interesting, is now unlikely to

TABLE 1: Comparative variances: $V(m_0, F_0) = I(F_0)^{-1}$, maximum variance $V(m_1, F_1)$ of α_*-trimmed mean and $V^0(m_0, F_2)$.

ϵ	α_*	$I(F_0)^{-1}$	$V(m_1, F_1)$	$V^0(m_0, F_2)$
0.025	0.0851	1.3163	1.3361	1.3310
0.029	0.0926	1.3663	1.3906	1.3910
0.03	0.0943	1.3790	1.4045	1.4068
0.04	0.1101	1.5115	1.5493	1.5641
0.05	0.1235	1.6556	1.7066	1.7393
0.1	0.1741	2.6134	2.7538	2.9530
0.15	0.2155	4.1997	4.4852	4.9910
0.2	0.2547	6.9812	7.5041	8.5101
0.25	0.2936	12.241	13.164	14.610
0.30	0.3329	23.329	24.969	26.717
0.40	0.4137	144.22	150.84	153.63
0.45	0.4558	766.83	787.27	790.89
0.48	0.4819	6237.5	6313.7	6314.7
0.481	0.4828	6997.3	7079.0	7079.5
0.49	0.4908	29048	29239	29228

yield significant gains of any statistical importance. Some representative values of α, $I(F_0)^{-1}$ and $\sup_{F\in\mathcal{K}_\epsilon(\Phi)} V(m_1, F)$, for various choices of ϵ, are given in Table 1.

A related consequence of this work, of some interest to the statistical practitioner, is as follows. Note that if the saddlepoint property holds for a class $\mathcal{M}$ of location estimators with generic member m and in a class $\mathcal{F}$ of distributions, then

$$\inf_{\mathcal{M}} \sup_{\mathcal{F}} V(m,F) = I(F_0)^{-1}, \tag{1.6}$$

with the lower bound being *attained* by the m_0 which is efficient for F_0. Huber (1964) established (1.6) for the class of M-estimators and any vaguely compact, convex class $\mathcal{F}$. Collins (1983) showed that (1.6) holds for R-estimation in $\mathcal{K}_\epsilon(\Phi)$; this was extended to $\mathcal{L}_{\epsilon,\delta}(G)$, under certain conditions on G admitting the normal distribution, by Collins and Wiens (1989). In contrast, the current results imply that for L-estimation in $\mathcal{K}_\epsilon(\Phi)$,

$$I(F_0)^{-1} < \inf_{m} \sup_{\mathcal{K}_\epsilon(\Phi)} V(m,F) \le V(m_1, F_1). \tag{1.7}$$

It may be asked if this means that one should not bother with L-estimation in $\mathcal{K}_\epsilon(\Phi)$, opting instead for the minimax M- or R-estimator. However, as pointed out above, the difference between the first and last members of (1.7) is so slight as to be negligible for practical purposes. Furthermore, the minimax M- and R-estimators, and the L-estimator generated by m_0, are much more computationally intensive — none has an explicit closed-form expression or simple description — and much less intuitively pleasing than the trimmed mean. Thus, there is much to be gained and little to be lost by using the trimmed mean, now known to be at least "nearly" minimax in $\mathcal{K}_\epsilon(\Phi)$.

The construction of F_1 and F_2 is carried out in Section 2. We are motivated by two considerations:

(1) The trimmed mean is L-minimax in $\mathcal{G}_\epsilon(\Phi) \subset \mathcal{K}_\epsilon(\Phi)$. It might then also be expected to perform well in $\mathcal{K}_\epsilon(\Phi)$. It is an easy matter to maximize the variance of the trimmed mean in $\mathcal{K}_\epsilon(\Phi)$; this yields F_1.

(2) As shown in Lemma 2.2, there is a subclass of $\mathcal{K}_\epsilon(\Phi)$ in which $V(m_0, F)$ is *minimized* by F_0. A search of this class yielded F_2 satisfying (1.5).

We believe that there exist distributions satisfying (1.5) for all values of $\epsilon \in (0, 0.5)$, but such examples will evidently be rather more complicated than the F_2 exhibited here.

We do not know if the trimmed mean is L-minimax in $\mathcal{K}_\epsilon(\Phi)$. We do know that (m_1, F_1) is not a saddlepoint, since $V(m, F_1)$ is not minimized by m_1. In fact, we have found that $V(m_0, F_1) < V(m_1, F_1)$ at all of those values of ϵ at which these quantities were compared.

2. CONSTRUCTION OF F_1 AND F_2

For fixed $\epsilon \in (0, 0.5)$ and for $\alpha \in (\epsilon, 0.5)$ define a distribution function in $\mathcal{K}_\epsilon(\Phi)$ by

$$F(x;\alpha) = \begin{cases} 0.5, & -a \le x \le a := \Phi^{-1}(\epsilon + 0.5), \\ \Phi(x) - \epsilon, & a \le x \le b := \Phi^{-1}(1 - \alpha + \epsilon), \\ \Phi(x) + \epsilon, & -b \le x \le -a, \\ H(x;\alpha) = 1 - H(-x;\alpha), & x \ge b, \end{cases}$$

where $H(x;\alpha)$ is arbitrary, subject only to the requirement that $F(x;\alpha)$ be continuous and remain within the bounds $\Phi(x) \pm \epsilon$. See Figure 1.

Define $m(t;\alpha)$ as in (1.1), so that the L-estimator with weights generated by $m(\cdot;\alpha)$ is an α-trimmed mean.

THEOREM 2.1. *The asymptotic variance of the α-trimmed mean is maximized in $\mathcal{K}_\epsilon(\Phi)$ by $F(\cdot;\alpha)$, with*

$$V(m(\cdot;\alpha), F(\cdot;\alpha)) = \frac{1 - 2\alpha + 2a\phi(a) + 2b\{\alpha b - \phi(b)\}}{(1 - 2\alpha)^2}. \tag{2.1}$$

Proof. From (1.2) the α-trimmed mean has

$$h_F(u; m(\cdot;\alpha)) = \frac{1}{1 - 2\alpha} \cdot \begin{cases} F^{-1}(u), & \frac{1}{2} < u < 1 - \alpha, \\ F^{-1}(1 - \alpha), & 1 - \alpha < u < 1. \end{cases}$$

For $F \in \mathcal{K}_\epsilon(\Phi)$ we then have

$$\begin{aligned} h_F(u; m(\cdot;\alpha)) &\le \frac{1}{1 - 2\alpha} \cdot \begin{cases} \Phi^{-1}(u + \epsilon), & \frac{1}{2} < u < 1 - \alpha, \\ b, & 1 - \alpha < u < 1. \end{cases} \\ &= h_{F(\cdot;\alpha)}(u; m(\cdot;\alpha)). \end{aligned}$$

Thus

$$V(m(\cdot;\alpha), F) \le V(m(\cdot;\alpha), F(\cdot;\alpha)),$$

and we easily calculate (2.1). □

We now define

$$\begin{aligned} \alpha_* &= \operatorname*{argmin}_{(\epsilon, \frac{1}{2})} V(m(\cdot;\alpha), F(\cdot;\alpha)), \\ F_1(x) &= F(x;\alpha_*), \\ m_1(t) &= m(t;\alpha_*). \end{aligned}$$

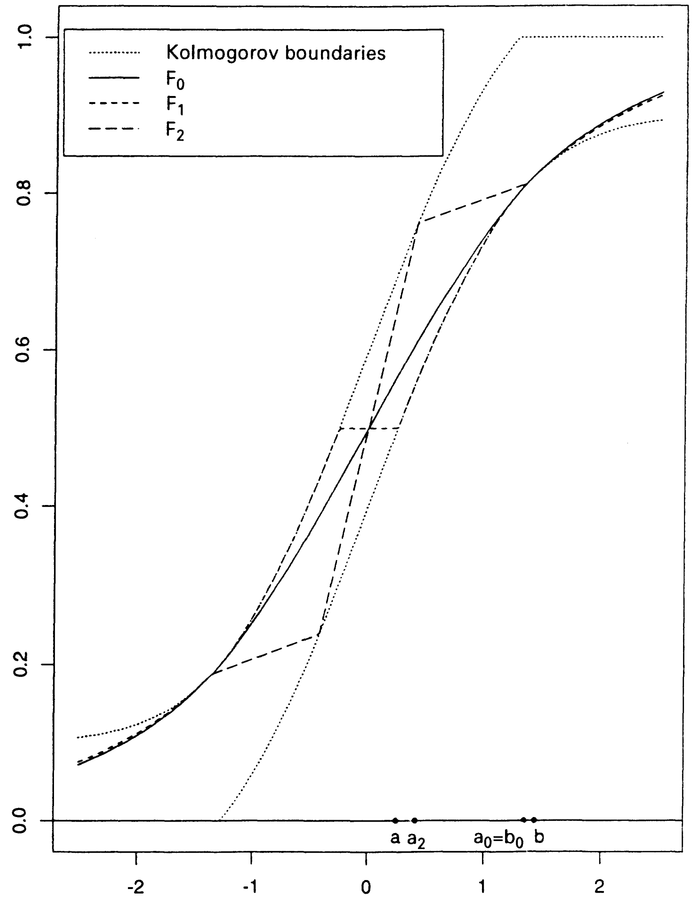

FIGURE 1: F_0, F_1 and F_2; $\epsilon = 0.1$.

Then the pair (m_1, F_1) satisfies the equality in (1.5). See Figure 1 for a plot of F_1.

The minimum-information distribution $F_0 \in \mathcal{K}_\epsilon(\Phi)$ was obtained by Huber (1964) for $\epsilon \le \epsilon_0 = 0.03033$, and by Sacks and Ylvisaker (1972) for $\epsilon_0 \le \epsilon < \frac{1}{2}$. For $\epsilon \le \epsilon_0$ the score function and density are given by

$$\psi_0(x) = \left\{ \lambda_1 \tan \frac{\lambda_1 x}{2},\ x,\ \lambda = b_0 \right\},$$

$$f_0(x) = \left\{ \phi(a_0) \frac{\cos^2(\lambda_1 x/2)}{\cos^2(\lambda_1 a_0/2)},\ \phi(x),\ \phi(b_0) e^{-\lambda(x-b_0)} \right\}$$

on intervals $[0, a_0]$, $[a_0, b_0]$, $[b_0, \infty)$ respectively. The constants $a_0 < b_0$ and λ_1 are determined in terms of ϵ by

(i) $F_0(a_0) = \Phi(a_0) - \epsilon$,
(ii) $F_0(\infty) = 1$,
(iii) $\psi_0(a_0 - 0) = a_0$.

Thus $F_0(x) = \Phi(x) - \epsilon$ on $[a_0, b_0]$, and $|F_0(x) - \Phi(x)| < \epsilon$ elsewhere in $[0, \infty)$. For $\epsilon > \epsilon_0$ the solution has $a_0 = b_0$, $\lambda = \lambda_1 \tan(\lambda_1 a_0/2)$ and the constants a_0, λ_1 are determined by (i) and (ii) above. See Figure 1 and Huber (1981, p. 90) for numerical values of the constants.

Our search for a distribution function F_2 satisfying (1.5) was motivated by the following result. Its proof is contained in that of Theorem 4 of Collins and Wiens (1989).

LEMMA 2.2. *Let $\mathcal{K}_\epsilon'(\Phi)$ be the subclass of $\mathcal{K}_\epsilon(\Phi)$ defined by*

$$\mathcal{K}_\epsilon'(\Phi) = \{ F \in \mathcal{K}_\epsilon(\Phi) |\, F(x) = F_0(x) \text{ for } |x| \geq a_0 \}.$$

Then $V(m_0, F)$ is non-constant on $\mathcal{K}_\epsilon'(\Phi)$, and

$$V(m_0, F_0) = \min_{F \in \mathcal{K}_\epsilon'(\Phi)} V(m_0, F).$$

For values δ and a_2 as specified below, define

$$F_2(x;\delta) = \begin{cases} \dfrac{1}{2} + \dfrac{F_0(a_0) - \frac{1}{2} - \delta}{a_2} x, & 0 \leq x \leq a_2, \\ F_0(a_0) - \delta \dfrac{a_0 - x}{a_0 - a_2}, & a_2 \leq x \leq a_0, \\ F_0(x), & x \geq a_0, \\ 1 - F_2(-x;\delta), & x < 0. \end{cases}$$

The value of δ is intended to be "small" — we will be letting $\delta \longrightarrow 0$. If $2\epsilon < \Phi(a_0) - \frac{1}{2}$, we choose $\delta \in (0, \Phi(a_0) - \frac{1}{2} - 2\epsilon)$ and put $a_2 = \Phi^{-1}(F_0(a_0) - \delta - \epsilon)$, so that $F_2(a_2;\delta) = \Phi(a_2) + \epsilon$. [This is the case for $F_2(\cdot; 0.05)$ in Figure 1, for which $\epsilon = 0.1$ and $V(m_0, F_2(\cdot; 0.05)) = 2.8255 > 2.7538 = V(m_1, F_1)$.] If $2\epsilon \geq \Phi(a_0) - \frac{1}{2}$, we set $a_2 = \delta$. We have $F_2 \in \mathcal{K}_\epsilon(\Phi)$ as long as $F_2(x;\delta) - \Phi(x) \leq \epsilon$ and $\geq -\epsilon$ on each of $[0, a_2]$ and $[a_2, a_0]$. Three of these four conditions can be shown to hold for all ϵ and sufficiently small δ; the requirement $F_2(x;\delta) - \Phi(x) \leq \epsilon$ on $[a_2, a_0]$ turns out to hold for $\epsilon \geq 0.015$, if δ is sufficiently small.

We require $\delta > 0$ so that F_2^{-1} is uniquely defined at a_0, where m_0 is discontinuous — see Huber (1981, p. 61). It is however sufficient for our purposes to show that

$$V^0(m_0, F_2) := \lim_{\delta \to 0} V(m_0, F_2(\cdot;\delta)) > V(m_1, F_1). \tag{2.2}$$

We find that

$$\begin{aligned} V^0(m_0, F_2) = \frac{2}{I(F_0)^2} \Bigg\{ & \left(\frac{c_1 a_2^0}{\lambda_1 (F_0(a_0) - \frac{1}{2})} \right)^2 \\ & \times \left(2a_0 \phi(a_0) + a_0^2 \psi(a_0)\phi(a_0) + \frac{c_1 a_0^3}{3} - 2\{F_0(a_0) - \tfrac{1}{2}\} \right) \\ & + (1 + c_2^2)\{\Phi(b_0) - \Phi(a_0)\} - 2c_2\{\phi(b_0) - \phi(a_0)\} \\ & - \{b_0\phi(b_0) - a_0\phi(a_0)\} + (c_2 + \lambda)^2\{1 - F_0(b_0)\} \Bigg\} \end{aligned}$$

where

$$c_1 = \frac{(\lambda_1^2/2)\phi(a_0)}{\cos^2(\lambda_1 a_0/2)},$$

$$c_2 = \frac{c_1 a_0 a_2^0}{F_0(a_0) - \frac{1}{2}} - \psi_0(a_0) + \frac{c_1(a_0 - a_2^0)}{\phi(a_0)},$$

$$a_2^0 = \lim_{\delta \to 0} a_2 = \{\Phi^{-1}(\Phi(a_0) - 2\epsilon)\}^+.$$

We have verified (2.2), thus completing the verification of (1.5), for $\epsilon \in [0.029, 0.481]$. See Table 1 for the numerical values.

ACKNOWLEDGEMENTS

We are grateful to John Collins and to an anonymous referee for their helpful comments.

REFERENCES

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

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. Statist.*, 42, 1020–1034.

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

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

Wiens, D.P. (1990). Minimax-variance L- and R-estimators of location. *Canad. J. Statist.*, 18, 47–57.

Received 25 September 1996
Revised 5 March 1997
Accepted 5 March 1997

Department of Mathematical Sciences
University of Alberta
Edmonton, Alberta
Canada T6G 2G1

email: wiens@stat.ualberta.ca

Department of Statistics
The Chinese University of Hong Kong
Shatin, N.T.
Hong Kong

Department of Mathematical Sciences
Lakehead University
Thunder Bay, Ontario
Canada P7B 5E1