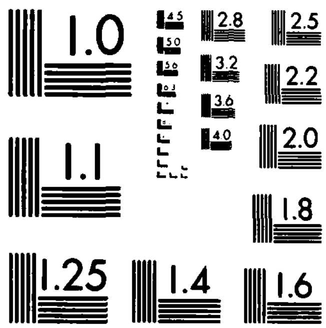
AD-A130 525 DIRECTIONALLY EFFICIENT ROBUST ESTIMATORS OF LOCATION 1/1
VIA EXPONENTIAL EMBEDDING(U) WISCONSIN UNIV-MADISON
MATHEMATICS RESEARCH CENTER W LOH MAY 83 MRC-TSR-2523

UNCLASSIFIED DAAG29-80-C-0041 F/G 12/1 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

ADA130525

②

MRC Technical Summary Report #2523

DIRECTIONALLY EFFICIENT ROBUST ESTIMATORS OF LOCATION VIA EXPONENTIAL EMBEDDING

Wei-Yin Loh

Mathematics Research Center
University of Wisconsin—Madison
610 Walnut Street
Madison, Wisconsin 53706

May 1983

(Received April 21, 1983)

Approved for public release
Distribution unlimited

Sponsored by

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

National Science Foundation
Washington, DC 20550

UNIVERSITY OF WISCONSIN - MADISON
MATHEMATICS RESEARCH CENTER

DIRECTIONALLY EFFICIENT ROBUST ESTIMATORS OF LOCATION VIA EXPONENTIAL EMBEDDING

Wei-Yin Loh*

Technical Summary Report #2523

May 1983

ABSTRACT

A general method is presented for constructing a location estimator which is asymptotically efficient at any two different location-scale families of symmetric distributions as well as at an appropriately defined class of distributions lying in between. The method works by embedding the two families in a comprehensive parametric model and identifying the estimator with the MLE. The case when the families are Normal and Double exponential is examined in detail.

AMS (MOS) Subject Classifications: Primary 62F35, Secondary 62G05

Key Words: Robust estimation, embedding, maximum likelihood, asymptotic efficiency.

Work Unit Number 4 - Statistics and Probability

* Department of Statistics and Mathematics Research Center, University of Wisconsin-Madison, Madison, WI 53705.

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041. This material is based upon work supported by the National Science Foundation under Grant No. MCS-7927062, Mod. 2.

SIGNIFICANCE AND EXPLANATION

This paper considers the following problem in the estimation of the center of a symmetric probability distribution. Suppose the statistician has a model F which he hopes is a good approximation to the true underlying distribution H generating his data. Further suppose that he has reason to believe that any deviation of H from F will probably be in the direction of another model G. A general procedure is presented for constructing an estimator which is asymptotically efficient at both F and G as well as at a suitably defined family of distributions lying in between. The case when F is Normal and G Double exponential is studied in detail via both asymptotic theory and Monte Carlo simulation (for finite sample sizes). The estimator is shown to compare favorably against nine other well known competitors. Computer programs are included.

The responsibility for the wording and views expressed in this descriptive summary lies with MRC, and not with the author of this report.

DIRECTIONALLY EFFICIENT ROBUST ESTIMATORS OF LOCATION VIA EXPONENTIAL EMBEDDING

Wei-Yin Loh*

1. Introduction

In robust estimation of the center of a symmetric distribution, we usually assume that we have a parametric model F (e.g. Normal location-scale family) which we hope is a good approximation to the true underlying distribution, but we do not assume it to be exactly right. A robust estimator is then desired, i.e. one which is efficient, or nearly efficient, at F and has reasonably good efficiency in a neighborhood of F. In the case that F is Normal, the neighborhood is quite often taken to consist of all symmetric distributions with tails ranging in thickness from the Normal to the Cauchy. Typically no particular distribution in the neighborhood (other than F) is preferred over the others, i.e. we do not require the estimator to be more efficient at some distributions than at others.

In this paper, the case is considered where the statistician has reason to believe that, if the true distribution were to deviate from F, it would probably (but not definitely) be towards a heavier-tailed model G. In such circumstances it would be desirable for the estimator to possess high, if not optimal, efficiency at G as well. Gastwirth (1966) and Crow and Siddiqui (1967) have studied this problem when G consists of one or more parametric families. Both these papers suppose it is known that the population sampled belongs to a set $\mathcal{F}$ of parametric families, like {Normal, Double

* Department of Statistics and Mathematics Research Center, University of Wisconsin-Madison, Madison, WI 53705.

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041. This material is based upon work supported by the National Science Foundation under Grant No. MCS-7927062, Mod. 2.

exponential} or {Normal, Double exponential, Cauchy}. They then search for an estimator, within various classes (like R- and L- estimators), which has maximin efficiency over $\mathcal{F}$. Since maximin estimators emphasize safety over efficiency however, they may be efficient (or even asymptotically efficient) nowhere in $\mathcal{F}$.

If asymptotic efficiency were the only criterion, of course solutions to our problem are available from the class of fully adaptive estimators (see e.g. Andrews et al. (1972), Sacks (1975), Stone (1975) and Beran (1978)), these being constructed to be asymptotically efficient at all sufficiently smooth distributions. But since these estimators make no use of our knowledge of F and G, it seems plausible that a semi-adaptive procedure which uses this information explicitly may perform better (at F and G) for small samples. Examples of such procedures have been suggested by Hogg and others. Hogg (1967) proposes an estimator which chooses among the sample mean, median, 25% trimmed mean, and mid-range according to the sample kurtosis. This estimator is asymptotically efficient at the Normal, Double exponential and uniform distributions, and empirical evidence (see e.g. Andrews et al. (1972) and Wegman and Carroll (1977)) suggests that its small sample performance at these distributions is better than many of the blatantly adaptive procedures. Further it is robust since it yields the sample median or trimmed mean whenever the sample kurtosis is large. Another procedure that applies maximum likelihood or Bayes rule to first select a family of distributions from within a predetermined set and then uses the optimal estimator for that family is considered in Hogg et al. (1972).

In this paper we introduce a general method for constructing a semi-adaptive estimator which is asymptotically efficient at any two parametric families F and G as well as at a family of distributions lying between them. The definition of this family and an explanation for its relevance are given in Section 2. Unlike the Hogg-type procedures which are based on trimmed means, our estimator is more like an M-estimator since it is the maximum likelihood estimator (MLE) corresponding to a genuine likelihood function. If the tails of F and G span a wide enough range, we expect this estimator to be robust. The particular example when F is Normal and G Double exponential is studied in detail in Sections 3-5. Section 3 derives the likelihood equations, section 4 deals with the asymptotic properties of the estimator, and section 5 contains empirical evidence on its small sample performance compared to some well-known adaptive and nonadaptive competitors.

2. Some embedding methods

Let $F = \{\sigma^{-1}f(\sigma^{-1}(x-\theta));\ -\infty < \theta < \infty,\ \sigma > 0\}$ and $G = \{\tau^{-1}g(\tau^{-1}(x-\theta));\ -\infty < \theta < \infty,\ \tau > 0\}$ be two location-scale families of densities on the real line. We assume that $f(x)$ and $g(x)$ are symmetric about 0 and want an estimator $\hat{\theta}$ of θ that is asymptotically efficient at F and G as well as those densities in between. There appears to be no universal agreement on what is meant by the set of distributions "between" two families. We will define it in the following way. Let H be a comprehensive parametric model parametrized by an additional parameter $\lambda \in [\lambda_1, \lambda_2]$ such that $\lambda = \lambda_1$ corresponds to F and $\lambda = \lambda_2$ corresponds to G. Then all distributions in H that are not in F or G will be considered as being in-between and G. In order for the estimation problem to be well-defined, the densities in H will have to be symmetric. Once a suitable embedding H is found, we will define $\hat{\theta}$ as the MLE over H of the center of symmetry θ.

There are at least three approaches to constructing such an embedding. The linear embedding consists of densities defined by

$$h(x;\theta,\sigma,\tau,\lambda) = \sigma^{-1}(1-\lambda)f\{\sigma^{-1}(x-\theta)\} + \tau^{-1}\lambda g\{\tau^{-1}(x-\theta)\}; \tag{2.1}$$
$$-\infty < \theta < \infty;\ \sigma,\tau > 0;\ 0 < \lambda < 1 \quad .$$

This construction has the advantage of being simple and allowing a physical interpretation for λ as a prior probability. Unfortunately, since σ and τ are unknown parameters, the likelihood function is unbounded at each data point $x_1,\ldots,x_n$ for $n > 2$. Hence MLEs do not exist. This difficulty can be avoided by including the restriction that σ/τ = constant. But besides drastically diminishing the size of the embedding, this seems to make the choice of the constant artificially important.

Another approach is to have the members of H be "F in the middle and G in the tails", i.e. for $k > 0$ define

$$h(x;\ \theta,\sigma,\tau,k) \propto \begin{cases} \sigma^{-1} f\{\sigma^{-1}(x-\theta)\} , & |x-\theta| < k \\ \tau^{-1} g\{\tau^{-1}(x-\theta)\} , & \text{otherwise} . \end{cases}$$

The M-estimator of Huber (1964) used this construction with F Normal and G Double exponential, and the additional requirement that h be continuously differentiable in x. The problem of estimating all the parameters (θ,σ,τ,k) via maximum likelihood does not appear to have been attempted, although Bell (1980) has investigated the question of adaptively choosing k from the data using the criterion of minimum estimated asymptotic variance.

A third construction is the <u>exponential</u> embedding where

$$h(x;\ \theta,\sigma,\tau,\lambda) = c(\sigma,\tau,\lambda) f^{1-\lambda}\{\sigma^{-1}(x-\theta)\} g^{\lambda}\{\tau^{-1}(x-\theta)\} ;$$
(2.2)
$$-\infty < \theta < \infty;\ \sigma,\tau > 0;\ 0 < \lambda < 1 ;$$

and $c(\sigma,\tau,\lambda)$ is a scaling factor. Cox (1961), Atkinson (1970), Brown (1971) and Weerahandi and Zidek (1978) have used this in various contexts. Recently, using the Kullback-Leibler information number as a measure of statistical distance, Loh (1983) showed that as λ ranges from 0 to 1, the distributions represented in (2.2) in fact constitute the shortest path between f and g in distribution-space. This result offers a justification for claiming that (2.2) yields densities in between F and G.

We adopt the method of exponential embedding in this paper and estimate θ with its MLE $\hat{\theta}$, regarding σ, τ and λ as nuisance parameters. (The fact that the nuisance parameters may not all be identifiable is not worrisome because we are only interested in estimating θ.) It is clear that $\hat{\theta}$ is location and scale equivariant, i.e. if we transform the data vector $\underline{x}$ to $a\underline{x} + b$ for some constants a and b, then

$$\hat{\theta}(a\underline{x}+b) = a\hat{\theta}(\underline{x}) + b .$$

If $\hat{\theta}_F$ and $\hat{\theta}_G$ are the MLEs for θ under the submodels F and G, the following theorem which will be used later gives conditions for $\hat{\theta}$ to lie between $\hat{\theta}_F$ and $\hat{\theta}_G$.

<u>Theorem 2.1</u>. Suppose that for each (σ,τ) the likelihood functions (of θ) under F and G are unimodal. Then the MLE $\hat{\theta}$ for (2.2), when σ, τ and λ are treated as nuisance parameters, always lies between $\hat{\theta}_F$ and $\hat{\theta}_G$.

<u>Proof</u>. This follows from the assumption of unimodality of the likelihoods and the fact that the values of $\hat{\theta}_F$ and $\hat{\theta}_G$ are unchanged whether σ and τ are known or not.

When F is Normal and G Double exponential, the MLEs are the sample mean $\bar{X}$ and median $\tilde{X}$ respectively. These two estimators are at the extremes of the family of symmetrically trimmed means, and the α-trimmed mean $\bar{X}_\alpha$ is often thought of as a compromise if the true underlying distribution is believed to have tails between those of the Normal and Double exponential. It is interesting to note that $\bar{X}_\alpha$ does not share the property of $\hat{\theta}$ in Theorem 2.1 in this case.

Since $\hat{\theta}$ is an MLE, classical theory suggests that under regularity conditions, it is asymptotically efficient when $0 < \lambda < 1$. When $\lambda = 0$ or 1, proofs of asymptotic normality are more difficult since the true parameter vector is now a boundary point. There appears to be no general theorems for such situations. In specific cases, a proof will probably depend on a combination of the results of Huber (1967) and ad hoc arguments. Robustness of $\hat{\theta}$ is likely to depend on the robustness of $\hat{\theta}_G$ if G is heavier-tailed than F. Intuitively this is because the estimated density (2.2) will tend to be close to some member in G if the true distribution is heavier-tailed than G. We illustrate these points with an example in the following sections.

3. Normal-Double exponential example: likelihood equations

We study here the case when F is Normal and G Double exponential. Equation (2.2) becomes

$$h(x;\ \theta,s,t) = c(s,t)\exp\{-\tfrac{1}{2}s^2(x-\theta)^2 - t|x-\theta|\} \tag{3.1}$$

$$\text{where } c(s,t) = \begin{cases} \frac{1}{2}s\phi(t/s)/\Phi(-t/s) & \text{if } s,t > 0 \\ s/(2\pi)^{1/2} & \text{if } t = 0,\ s > 0 \\ t/2 & \text{if } s = 0,\ t > 0 \end{cases}$$

and $\phi(\ )$, $\Phi(\ )$ are the standard Normal density and cumulative distribution functions. The scale factor $c(s,t)$ is defined on $\mathcal{D} = [0,\infty) \times [0,\infty) \setminus \{(0,0)\}$. It can be checked (using e.g. (3.6) below) that $c(s,t)$ is continuous on $\mathcal{D}$. Clearly (3.1) yields F when $t = 0$ and G when $s = 0$.

Let $\hat{\theta}$, $\hat{s}$, $\hat{t}$ be the MLEs for θ, s, t. The following theorem whose proof is sketched in the Appendix shows that the three-parameter minimization problem of determining the MLEs may be reduced to one involving only one parameter in $[0,\infty)$.

<u>Theorem 3.1</u>. Let $(x_1,\ldots,x_n)$ be an ordered sample of size n and $\bar{x}$, $\tilde{x} = (x_{[(n+1)/2]} + x_{[n/2+1]})$ be the sample mean and median respectively (here $[\ ]$ is the greatest integer function). Suppose that $\bar{x} < \tilde{x}$. If n is even and $x_{n/2} < \bar{x}$, then $\hat{\theta} = \bar{x}$. Otherwise the minimization of the likelihood corresponding to (3.1) can be reduced to the following one-dimensional problem: Let $v = t/s$, $w = t/s^2$ and let k_0 be the largest integer k such that $x_k < \bar{x}$. Define to be .e set of integers $\{k_0,\ldots,[(n-1)/2]\}$ and divide the interval $[0,\infty)$ into the subintervals $\{[p_{1k},p_{2k}],\ [q_{1k},q_{2k}];\ k \in N\}$ where $p_{1k_0} = 0$, $q_{2,[(n-1)/2]} = \infty$ and for $k = k_0+1,\ldots,[(n-3)/2]$,

$$p_{2k} = n(n-2k)^{-1}(x_{k+1}-\bar{x})\{(n-2k)^{-1}(x_{k+1}-\bar{x})\Sigma|x_i-x_{k+1}|$$

$$+ n^{-1}\Sigma(x_i-x_{k+1})^2\}^{-1/2} ,$$

(3.2) $$q_{1k} = p_{2k} ,$$

$$q_{2k} = n(n-2k-2)^{-1}(x_{k+1}-\bar{x})\{(n-2k-2)^{-1}(x_{k+1}-\bar{x})\Sigma|x_i-x_{k+1}|$$

$$+ n^{-1}\Sigma(x_i-x_{k+1})^2\}^{-1/2} ,$$

$$p_{1,k+1} = q_{2k} .$$

For each $k \in N$, define for $v \in [0,\infty)$ the function

(3.3) $$f_k(v) = \begin{cases} \log(w/v) + \log \Phi(-v) + v^2(2nw^2)^{-1}\{\sum_1^k (x_i-\theta-w)^2 \\ \qquad + \sum_{k+1}^n (x_i-\theta+w)^2\}, \text{ if } v \in [p_{1k},q_{2k}] \\ 0 \quad , \text{ otherwise} \end{cases}$$

where

(i) for $v \in [p_{1k},p_{2k}]$,

$$w = n^{-1}v^2 \sum_1^k (\bar{x}-x_i) + \{n^{-2}v^4\{\sum_1^k (\bar{x}-x_i)\}^2 + n^{-1}v^2\Sigma(x_i-\bar{x})^2\}^{1/2} ,$$

(3.4) $$\theta = \bar{x} + w(n-2k)/n ;$$

(ii) for $v \in [q_{1k},q_{2k}]$,

$$w = (2n)^{-1}v^2\Sigma|x_i-x_{k+1}| + \tfrac{1}{2}\{n^{-2}v^4(\Sigma|x_i-x_{k+1}|)^2$$

$$+ 4n^{-1}v^2\Sigma(x_i-x_{k+1})^2\}^{1/2} ,$$

(3.5) $$\theta = x_{k+1} .$$

Let $f(v) = \sum_{k \in N} f_k(v)$ and v^* minimize $f(v)$ over $[0,\infty)$. Then the $\hat{\theta}$ corresponding to v^* (given by (3.4) or (3.5)) is the MLE for θ for the density (3.1).

We are unable to prove or disprove that $\hat{\theta}$ is a.s. unique, although our experience with numerical examples suggests that this is the case. Should multiple roots occur, however, we can choose the root closest to the sample median. In view of the theorems in the next section, this will guarantee a consistent sequence of roots. To implement the method on a computer, we note that if v^* lies in $[q_{1,[(n-1)/2]},\infty)$, its value need not be computed exactly since $\hat{\theta}$ is independent of it. So we need only determine whether v^* lies in this interval. This search is greatly assisted by the following approximation which effectively reduces the interval to a finite one.

<u>Theorem 3.2</u>. Let $S_1 = \Sigma|x_i - x_{k+1}|$ and $S_2 = \Sigma(x_i - x_{k+1})^2$ where $k = [(n-1)/2]$. If $v > \max\{2\sqrt{2},\ 2S_1^{-1}\sqrt{nS_2},\ q_{1k}\}$, then $f(v) - \log(S_1/(n\sqrt{2\pi})) - 1 - v^{-2}\{nS_2/(2S_1^2) - 1\}$ is bounded above by $v^{-4}\{5/2 + (7/12)(nS_2/S_1^2)^2\}$ and below by $-v^{-4}\{(3/2)(nS_2/S_1^2)^2 - (\pi/2)(1 - \pi/8)\}$.

<u>Proof</u>. Expand $f(v)$ in powers of v^{-2} using Taylor's series and the inequalities (see e.g. Johnson and Kotz (1970) p. 279): for $x > 0$,

$$\{(x^2+8)^{1/2} + 3x\}/4 < \phi(x)/\Phi(-x) < \{(x^2+2\pi)^{1/2} + (\pi-1)x\}/\pi \quad . \tag{3.6}$$

A Fortran program that uses these two theorems is given in the Appendix. It uses a modified version of the function minimization routine FMIN in Forsythe, Malcolm and Moler (1977) and calls IMSL subroutines MDNORD and MSMRAT to compute $\Phi(x)$ and Mill's ratio.

4. Asymptotics

It is shown in this section that $\hat{\theta}$ is asymptotically efficient when the model (3.1) is correct. If (θ,s,t) is an interior point of the parameter space, standard methods can be used to prove consistency and asymptotic efficiency of $(\hat{\theta},\hat{s},\hat{t})$. However these methods are inapplicable when (θ,s,t) is a boundary point, as when the true underlying distribution is Normal or Double exponential. For these cases we use a thoerem in Huber (1967) to prove consistency and then resort to ad hoc methods to argue asymptotic efficiency. Incidentally, consistency of $\hat{\theta}$ alone is a consequence of Theorem 2.1 since it is sandwiched between $\bar{X}$ and $\tilde{X}$ both of which are consistent under (3.1). Consistency for the other MLEs (as well as $\hat{\theta}$) is shown in the following theorem.

Theorem 4.1. $\hat{\theta}$, $\hat{s}$ and $\hat{t}$ are consistent estimates of θ, s and t when (3.1) is correct.

Proof. The proof consists of checking that the conditions in Theorem 1 in Huber (1967) holds. These are called assumptions (A1) - (A5) in the paper to which we refer the reader for a precise statement. Let $\Omega = (-\infty,\infty) \times \mathcal{D}$, where $\mathcal{D}$ is defined in Section 3. Let $\alpha = (\theta,s,t) \in \Omega$, and $\alpha_0 = (\theta_0,s_0,t_0)$ be the parameter vector corresponding to the true distribution. Following the suggestion in Huber (1981, p. 130), we take pairs $y_n = (x_{2n-1},x_{2n})$ of the original data $(x_1,x_2,\ldots)$ as our new observations and define

$$\rho(y,\alpha) = -2 \log c(s,t) + \frac{1}{2} s^2\{(x_1-\theta)^2 + (x_2-\theta)^2\}$$
$$+ t\{|x_1-\theta| + |x_2-\theta|\} \quad .$$

Assumption (A-1) is immediate, and (A-2) follows from the continuity of $\rho(y,\alpha)$ as a function of α. Let $a(y) = \rho(y,\alpha_0)$ and note that $E\{\rho(y,\alpha) - a(y)\}$ is a Kullback-Leibler information number; hence it is nonnegative and well-defined (possible $+\infty$) for all $\alpha \in \Omega$, and vanishes only when $\alpha = \alpha_0$. (Here all expectations are taken with respect to α_0.) This implies assumptions (A-3) and (A-4).

Finally to verify (A-5), let $\underline{\infty}$ be the point at infinity in the one-point compactification of Ω. In our context (A-5) may be stated as follows: There is a continuous function $b(\alpha) > 0$ such that

(i) $\inf_\alpha \{\rho(y,\alpha) - a(y)\}/b(\alpha) > h(y)$ for some integrable h,

(ii) $\liminf_{\alpha \to \underline{\infty}} b(\alpha) > 0$, and

(iii) $E\{\liminf_{\alpha \to \underline{\infty}} [\rho(y,\alpha) - a(y)]/b(\alpha)\} > 1$.

Take $b(\alpha)$ to be identically 1 for all α. Then (ii) is immediate and (i) will follow if we show that $\inf_\alpha \rho(y,\alpha)$ is integrable. For each s and t, $\rho(y,\alpha)$ is minimized when $\theta = (x_1+x_2)/2$. Therefore writing $z = |x_1-x_2|/2$, we see that

$$\text{(4.1)} \qquad \inf_\alpha \rho(y,\alpha) = \inf_{s,t} \{-2 \log c(s,t) + s^2z^2 + 2tz\} \quad .$$

Now suppose (s,t) is an interior point, and make the change of variable $u = t/s$. The expression in parenthesis on the RHS of (4.1) can be rewritten as

$$\text{(4.2)} \qquad H(z,s,u) = -2 \log s + u^2 + 2 \log \Phi(-u) + s^2z^2 + 2suz \quad .$$

For fixed u this is minimized when $2sz = (u^2+4)^{1/2} - u$. Substituting this for s in (4.2) and differentiating with respect to u yields

$$dH/du = u - 2\phi(u)/\Phi(-u) + (u^2+4)^{1/2}$$

which is positive for all u (see Birnbaum (1942)). We therefore conclude that $\inf H(z,s,u)$ is attained at $u = 0$ and $s = z^{-1}$. Hence $\inf_\alpha \rho(y,\alpha) = 2 \log z + \text{constant}$, which is clearly integrable, and so (i) obtains. To verify (iii) it suffices to check that $\liminf_{\alpha \to \underline{\infty}} \rho(y,\alpha) = \infty$. There are two cases to consider, namely (a) $|\theta|$, s, t $\to \infty$, and (b) $|\theta| \to \infty$, s,t $\to$ 0. In both cases however, we have

$$\rho(y,\alpha) = \begin{cases} O(-\log s + s^2\theta^2/2 + t|\theta|) & \text{if } ts^{-1} \to \text{constant} \\ O(-\log t + s^2\theta^2/2 + t|\theta|) & \text{if } ts^{-1} \to \infty \end{cases} .$$

Obviously $\lim\inf \rho(y,\alpha) = \infty$ in either case. Thus (iii) is verified and the proof is ended.

We are now ready to deduce asymptotic normality and efficiency.

Theorem 4.2. $\hat{\theta}$ is asymptotically efficient when the underlying distribution F has a density given by (3.1).

Proof. When the true parameter vector (θ,s,t) is an interior point of the parameter space, the asymptotic efficiency of the MLEs follows easily from the standard theorems (see e.g. Lehmann (1983)). We therefore only prove the result when either s or t is zero.

(i) F is Normal. By the preceding theorem, $\hat{t} \xrightarrow{P} 0$ and $\hat{s} \xrightarrow{P} s$ for some $s > 0$. It is clear that equations (3.4) and (3.5) together define a monotone function of v in $[0,\infty)$. Therefore

$$\sqrt{n}\ |\hat{\theta}-\bar{x}| < \hat{w}|n-2k|/\sqrt{n} = \hat{t}|n-2k|/(\hat{s}^2\sqrt{n})$$

for some k between k_0 (defined in Theorem 3.1) and n/2. Here $|n-2k_0|$ is the difference between the number of deviations $\{x_i-\bar{x}\}$ with positive signs and the number with negative sign. Since $|n-2k_0| = O_P(\sqrt{n})$ (David (1962)), we see that $\sqrt{n}\ |\hat{\theta}-\bar{x}| < \hat{t}|n-2k_0|/(\hat{s}^2\sqrt{n}) \xrightarrow{P} 0$. Hence $\hat{\theta}$ has the same asymptotic variance as $\bar{x}$ which is efficient.

(ii) F is Double exponential. We assume without loss of generality that $\bar{x} < \tilde{x}$ and again use the notation of Theorem 3.1. We know from Theorem 4.1 that $\hat{s} \xrightarrow{P} 0$ and $\hat{t} \xrightarrow{P} t > 0$. Therefore $\hat{v} = \hat{t}/\hat{s} \xrightarrow{P} \infty$. Let $k_n(x)$ be the largest integer k such that q_{2k} (defined in (3.2)) satisfies $q_{2k} < \hat{v}$.

Then $k_0 < k_n < n/2$ and $n-2k_0 = O_p(\sqrt{n})$ (see Brown and Kildea (1979)). So $(n-2k_n)/\sqrt{n}$ converges in probability to some random variable Y. We will show that Y is degenerate at 0. First observe that $\sqrt{n}|x_{k_n} - \bar{x}| < \sqrt{n}|\tilde{x}-\bar{x}| = O_p(1)$ and $n^{-1}\sum_1^{k_n} (\bar{x}-x_i) = (2n)^{-1}\Sigma|x_i-\bar{x}| \xrightarrow{P}$ constant. Next note that q_{2k} can be written as

$$q_{2k} = n(n-2k-2)^{-1}(x_{k+1}-\bar{x})/\{2(n-2k-2)^{-1}(x_{k+1}-\bar{x}) \sum_1^{k+1} (\bar{x}-x_i)$$

$$+ n^{-1}\Sigma(x_i-\bar{x})^2\}^{1/2} .$$

Since $\hat{v} \xrightarrow{P} \infty$, this yields $q_{2k_n} \xrightarrow{P} \infty$. This implies that $Y = 0$. Therefore $k_n/n = \frac{1}{2} + R_n$ where $\sqrt{n}\, R_n \xrightarrow{P} 0$. Since $x_{k_n} < \hat{\theta} < \tilde{x}$, it follows that $\hat{\theta}$ has the same asymptotic variance as $\tilde{x}$ (Lehmann (1983), Chapter 5, Problem 3.5).

The next theorem shows that for heavy-tailed distributions like the Cauchy (or Tukey's "slash") $\hat{\theta}$ is asymptotically equivalent to the median. It is therefore robust against these distributions.

Theorem 4.3. Assume that the true underlying density has tails of order $|x|^{-2}$. Then $\hat{\theta}$ has the same asymptotic variance as $\tilde{X}$.

Proof. Assume as before that the x's are ordered and use the notations of Theorem 3.1. Brown and Tukey (1946) showed that for such distributions

$$\bar{X} = O_p(1),\ n^{-1}\Sigma|X_i| = O_p(1),\ n^{-1}\Sigma X_i^2 = O_p(n) \quad . \tag{4.3}$$

It is easy to check that

$$\hat{s}^{-2} = O_p(n^{-1}\Sigma(x_i-\hat{\theta})^2),\ \hat{v}/\hat{s} = O_p(n^{-1}\Sigma|x_i-\hat{\theta}|) \quad .$$

Since $\hat{\theta}$ lies between $\bar{X}$ and $\tilde{X}$, (4.3) implies that $\hat{v} = O_p(\sqrt{n})$. Now choose $\{k_n\}$ such that $k_n = \frac{1}{2}n + nR_n$ where $\sqrt{n}\, R_n \to 0$. This ensures that the k_n^{th} order statistic X_{k_n} has the same asymptotic variance as $\tilde{X}$. Putting $k = k_n$ in (3.2) we see also that

$$q_{2,k_n} = (\sqrt{n}\, R_n)^{-1}(x_{k_n} - \bar{x})/\{((\sqrt{n}\, R_n)^{-1}(x_{k_n} - \bar{x})(n^{-1}\, \Sigma|x_i - x_{k_n}|)$$

$$+ n^{-2}\, \Sigma(x_i - x_{k_n})^2\}^{1/2}$$

which in view of (4.3) is at most $O_p((\sqrt{n}\, R_n)^{-1})$. Clearly, we can choose k_n so that $\sqrt{n}\, R_n \to 0$ and $nR_n \to \infty$. Then for large n, $q_{2,k_n} \ll \hat{v}$ with high probability. This yields $X_{k_n} < \hat{\theta} < \tilde{X}$ and hence the result.

5. Small-sample behavior: sensitivity curve, breakdown bounds and Monte Carlo

This section contains computer-generated results on the small-sample performance of $\hat{\theta}$. Figure 5.1 shows its stylized sensitivity curve for sample size 20. This is seen to be very similar to that for a trimmed mean, and may be compared with those of other estimators given in Andrews et al. (1972). The curve is obtained by starting with a pseudo-sample of 19 expected normal order statistics, adding to this a moving point x, evaluating $\hat{\theta}$ from the combined sample, and plotting 20 times $\hat{\theta}$ as a function of x.

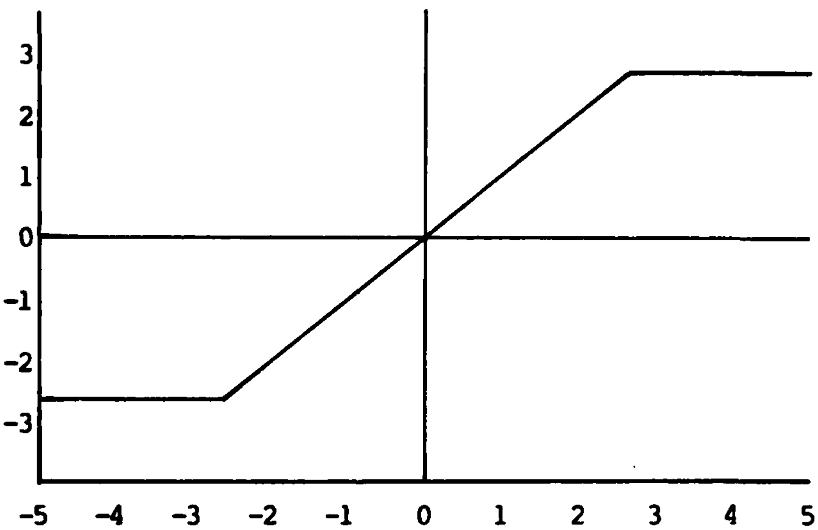


Figure 5.1. Sensitivity curve

Andrews et al. (1972) also used the concept of breakdown bounds to get some idea of the tolerance of an estimator to extremely aberrant data. For each sample size n, j sample points are taken to be 100, 200,...,j00 and the remaining n-j points are taken to be the n-j expected normal order statistics from a sample of size n-j. The numbers in Table 5.1 give the

Table 5.1. Breakdown Bounds

Largest % of contamination such that estimator < 3.

	Sample size n			
	5	10	20	40
$\hat{\theta}$	20	40	40	45
$\bar{X}$	0	0	0	2.5
$\bar{X}_{.10}$	0	10	10	10
$\bar{X}_{.25}$	20	20	25	25
$\tilde{X}$	40	40	45	47.5
JOH	40	20	20	22.5
TAK	20	10	15	17.5
JAE	20	20	25	25
HG1	20	30	35	37.5
HG2	40	30	35	37.5

largest j for which the estimator is less than 3 (the estimator is said to break down if it exceeds 3). For comparison the breakdown bounds for the following estmators are also given in the table:

Nonadaptive: $\bar{X}$ = sample mean

$\bar{X}_{.10}$ = 10% symmetrically trimmed mean

$\bar{X}_{.25}$ = 25% symmetrically trimmed mean

$\tilde{X}$ = sample median

<u>Adaptive</u>: JOH = John's adaptive estimator

TAK = Takeuchi's adaptive estimator

JAE = Jaeckel's adaptive trimmed mean

$\left.\begin{array}{l}\text{HG1}\\ \text{HG2}\end{array}\right\}$ Hogg-type adaptive estimators.

The definitions of JOH, TAK and JAE are given in Andrews et al. (1972) under the same names. HG1 was first suggested by Hogg (1974) and is defined to be $\bar{X}_{.125}$ if $Q < 1.81$, $\bar{X}_{.25}$ if $1.81 \leq Q \leq 1.87$, and $\bar{X}_{.375}$ if $Q > 1.87$. Here Q is the ratio $\{\bar{U}(.2) - \bar{L}(.2)\}/\{\bar{U}(.5) - \bar{L}(.5)\}$, where $\bar{U}(\beta)$ and $\bar{L}(\beta)$ are the averages of the largest and smallest $[(n+1)\beta]$ order statistics respectively. The estimator HG2 is a modification of HG1 to make it asymptotically efficient at the Normal and Double exponential distributions. It is defined to be $\bar{X}$ if $Q < 1.81$, $\bar{X}_{.25}$ if $1.81 \leq Q \leq 1.87$, and $\tilde{X}$ if $Q > 1.87$. From Table 5.1 it is clear that for $n = 10$, 20 or 40, the breakdown bounds for $\hat{\theta}$ are superior to all the others except those for the sample median.

Monte Carlo estimates of the variances (multiplied by n) of each of these estimators are given in Tables 5.2 - 5.4 for $n = 10$, 20 and 40 and the following distributions: (i) N(0,1) (Normal with mean 0 and variance 1), (ii) the density (2.2) with $f = N(0,1)$, $g =$ Double exponential with density $\frac{1}{2} e^{-|x|}$, and $\lambda = \frac{1}{2}$, (iii) Double exponential, (iv) contaminated Normal: 90% N(0,1) + 10% N(0,100), and (v) Cauchy. The first three distributions are picked for the study because they span the range in which $\hat{\theta}$ is efficient. Distributions (iv) and (v) are included to test its robustness properties. Estimates of the standard errors are given in parentheses, and the minimum estimated variance for each distribution is underlined. The simulations were done on a VAX/11/750 computer. The IMSL generator GGUW was used to generate uniform random numbers and the Box-Muller transformation applied to produce normal deviates. The Princeton swindle was used whenever possible. The number of replications ranged from 1000 - 5000.

It is immediately clear from these tables that in none of the sample size-distribution combinations considered does $\hat{\theta}$ beat all of the other estimators. To analyse them further, we can look at deficiencies. These are defined to be the ratios (estimated variance)/(minimum estimated variance), where the minimum is taken over the ten estimators compared. The coded deficiencies are shown in Tables 5.5 - 5.7. Only deficiencies greater than 1.5 appear as digits or x's. All the estimators (with the possible exception of the mean and median) seem to be equally good at the Normal and Double exponential distributions. The excellent performance of TAK at distribution (ii) also stands out.

Finally to compare the relative performance of the estimators over all situations combined we follow Tukey's (1979) suggestion to look at maxima and sums of deficiencies. For each sample size, let A(i) and B(i) be the maximum and total deficiency of the i^{th} estimator over a set of distributions. The estimators are then ranked according to the values of $\{A(i)\}$ and $\{B(i)\}$. These two criteria are denoted by "minimax" and "total" in Tables 5.8 - 5.9 where the estimators are ranked first for distributions (i) - (iii), and then again for all five distributions.

The following points may be made from these two tables:

(a) For $n = 20$ or 40, TAK and JOH appear hard to beat. For $n = 10$ however, the picture is quite different. Here JOH is somewhat below average when only distributions (i) - (iii) are considered, and TAK has a poor showing for all distributions combined. The reason may be that these two estimators are over-adapting at this sample size.

(b) There seems to be little to choose between $\hat{\theta}$ and HG2. Both are consistently good over all three sample sizes.

(c) HG2 is superior to HG1 for the distributions considered.

(d) The adaptive trimmed mean JAE trails $\hat{\theta}$ and HG2 almost every time.

(e) None of the nonadaptive trimmed means (including the mean and median) are competitive.

The above results encourage us to feel that for the kind of situation described in the introduction, our proposed procedure will produce viable estimators.

Table 5.2. Variance x(n=10)

Standard errors in parentheses

	N(0,1)	$\lambda = \frac{1}{2}$	DEXP	10%10N	Cauchy
$\hat{\theta}$	1.069	0.416	1.59	1.98	5.09
	(.004)	(.009)	(.02)	(.03)	(.16)
$\bar{X}$	<u>1.000</u>	0.347	1.99	10.17	768180.0
		(.007)	(.04)	(.40)	(73179.0)
$\bar{X}_{.10}$	1.051	0.460	1.61	2.78	14.63
	(.002)	(.009)	(.03)	(.13)	(3.51)
$\bar{X}_{.25}$	1.157	0.630	<u>1.41</u>	<u>1.57</u>	4.15
	(.006)	(.011)	(.03)	(.02)	(.15)
$\tilde{X}$	1.382	0.874	1.46	1.74	<u>3.36</u>
	(.015)	(.015)	(.02)	(.20)	(.08)
JOH	1.194	0.445	1.62	1.82	4.38
	(.007)	(.010)	(.03)	(.19)	(.23)
TAK	1.048	<u>0.307</u>	1.71	2.62	18.55
	(.002)	(.007)	(.03)	(.19)	(5.42)
JAE	1.081	0.428	1.56	1.87	6.53
	(.004)	(.009)	(.02)	(.13)	(.53)
HG1	1.119	0.513	1.48	1.67	4.18
	(.005)	(.010)	(.02)	(.06)	(.18)
HG2	1.094	0.411	1.60	1.91	4.73
	(.008)	(.009)	(.02)	(.16)	(.25)

Table 5.3. Variance x(n=20)

	N(0,1)	$\lambda = \frac{1}{2}$	DEXP	10%10N	Cauchy
$\hat{\theta}$	1.074	0.373	1.46	1.97	3.41
	(.005)	(.008)	(.02)	(.02)	(.09)
$\bar{x}$	1.000	0.349	1.92	11.40	19921.0
		(.007)	(.04)	(.41)	(14362.0)
$\bar{x}_{.10}$	1.056	0.476	1.52	2.08	7.81
	(.002)	(.009)	(.03)	(.08)	(.40)
$\bar{x}_{.25}$	1.190	0.689	1.30	1.52	3.27
	(.008)	(.013)	(.03)	(.01)	(.14)
$\tilde{x}$	1.494	1.012	1.31	1.83	2.79
	(.028)	(.018)	(.02)	(.02)	(.06)
JOH	1.127	0.300	1.43	1.44	2.88
	(.005)	(.008)	(.02)	(.01)	(.09)
TAK	1.048	0.218	1.55	1.42	3.70
	(.002)	(.006)	(.02)	(.02)	(.16)
JAE	1.102	0.404	1.41	1.48	3.58
	(.005)	(.009)	(.02)	(.01)	(.12)
HG1	1.121	0.519	1.35	1.58	2.85
	(.005)	(.010)	(.02)	(.06)	(.08)
HG2	1.072	0.376	1.47	1.72	2.98
	(.007)	(.008)	(.03)	(.02)	(.09)

Table 5.4. Variance x(n=40)

	N(0,1)	$\lambda = \frac{1}{2}$	DEXP	10%10N	Cauchy
$\hat{\theta}$	1.051	0.353	1.31	1.95	2.91
	(.004)	(.007)	(.02)	(.02)	(.05)
$\bar{X}$	<u>1.000</u>	0.350	1.99	10.49	235438.0
		(.007)	(.05)	(.33)	(118967.0)
$\bar{X}_{.10}$	1.062	0.488	1.51	1.59	6.14
	(.003)	(.010)	(.04)	(.03)	(.20)
$\bar{X}_{.25}$	1.199	0.727	1.25	1.50	2.89
	(.009)	(.014)	(.03)	(.01)	(.06)
$\tilde{X}$	1.513	1.166	<u>1.23</u>	1.87	2.62
	(.023)	(.021)	(.02)	(.02)	(.04)
JOH	1.097	0.217	1.27	1.43	2.50
	(.004)	(.005)	(.02)	(.01)	(.06)
TAK	1.035	<u>0.139</u>	1.42	<u>1.35</u>	2.73
	(.002)	(.004)	(.02)	(.01)	(.07)
JAE	1.077	0.362	1.36	1.47	3.06
	(.004)	(.008)	(.02)	(.01)	(.08)
HG1	1.107	0.528	1.27	1.61	<u>2.46</u>
	(.004)	(.010)	(.02)	(.02)	(.05)
HG2	1.067	0.356	1.36	1.74	2.57
	(.008)	(.007)	(.02)	(.02)	(.05)

Table 5.5. Deficiencies of estimators (n=10)

Variances divided by minimum variance among 10 estimators rounded to nearest integer. One's are suppressed, numbers > 9 are coded x.

	Normal	$\lambda = \frac{1}{2}$	DEXP	10%10N	Cauchy
$\hat{\theta}$	.	.	.	.	2
$\bar{X}$	.	.	.	6	x
$\bar{X}_{.10}$	.	.	.	2	4
$\bar{X}_{.25}$	.	2	.	.	2
$\tilde{X}$	.	3	.	.	3
JOH	.	.	.	.	.
TAK	.	.	.	2	6
JAE	.	.	.	.	2
HG1	.	2	.	.	2
HG2	.	.	.	.	.

Table 5.6. Deficiencies of estimators (n=20)

	Normal	$\lambda = \frac{1}{2}$	DEXP	10%10N	Cauchy
$\hat{\theta}$	.	2	.	.	.
$\bar{X}$	.	2	.	8	x
$\bar{X}_{.10}$	.	2	.	.	3
$\bar{X}_{.25}$	.	3	.	.	3
$\tilde{X}$	.	5	.	.	5
JOH	.	.	.	.	.
TAK	.	.	.	.	.
JAE	.	2	.	.	.
HG1	.	2	.	.	.
HG2	.	2	.	.	.

Table 5.7. Deficiencies of estimators (n=40)

	Normal	$\lambda=\frac{1}{2}$	DEXP	10%10N	Cauchy
$\hat{\theta}$	.	3	.	.	.
$\bar{X}$	.	3	2	8	x
$\bar{X}_{.10}$	.	4	.	.	2
$\bar{X}_{.25}$	.	5	.	.	.
$\tilde{X}$	2	8	.	.	.
JOH	.	2	.	.	.
TAK	.	.	.	.	.
JAE	.	3	.	.	.
HG1	.	4	.	.	.
HG2	.	3	.	.	.

Table 5.8. Ranks of estimators for dists. (i) - (iii)

n	Criterion	1	2	3	4	5	6	7	8	9	10
10	Minimax	TAK	HG2	$\hat{\theta}$	JAE	$\tilde{X}$	JOH	$\bar{X}_{.10}$	HG1	$\bar{X}_{.25}$	$\tilde{X}$
	Total	TAK	$\hat{\theta}$	$\bar{X}$	HG2	JAE	$\bar{X}_{.10}$	JOH	HG1	$\bar{X}_{.25}$	$\tilde{X}$
20	Minimax	TAK	JOH	$\bar{X}$	$\hat{\theta}$	HG2	JAE	$\bar{X}_{.10}$	HG1	$\bar{X}_{.25}$	$\tilde{X}$
	Total	TAK	JOH	$\hat{\theta}$	HG2	JAE	$\bar{X}$	$\bar{X}_{.10}$	HG1	$\bar{X}_{.25}$	$\tilde{X}$
40	Minimax	TAK	JOH	$\bar{X}$	$\hat{\theta}$	HG2	JAE	$\bar{X}_{.10}$	HG1	$\bar{X}_{.25}$	$\tilde{X}$
	Total	TAK	JOH	$\hat{\theta}$	HG2	JAE	$\bar{X}$	$\bar{X}_{.10}$	HG1	$\bar{X}_{.25}$	$\tilde{X}$

Table 5.9. Ranks of estimators for all 5 dists.

n	Criterion	1	2	3	4	5	6	7	8	9	10
10	Minimax	HG2	JOH	$\hat{\theta}$	HG1	JAE	$\bar{X}_{.25}$	$\tilde{X}$	$\bar{X}_{.10}$	TAK	$\bar{X}$
	Total	HG1	HG2	JOH	$\hat{\theta}$	$\bar{X}_{.25}$	JAE	$\tilde{X}$	$\bar{X}_{.10}$	TAK	$\bar{X}$
20	Minimax	TAK	JOH	$\hat{\theta}$	HG2	JAE	HG1	$\bar{X}_{.10}$	$\bar{X}_{.25}$	$\tilde{X}$	$\bar{X}$
	Total	TAK	JOH	HG2	JAE	$\hat{\theta}$	HG1	$\bar{X}_{.25}$	$\bar{X}_{.10}$	$\tilde{X}$	$\bar{X}$
40	Minimax	TAK	JOH	$\hat{\theta}$	HG2	JAE	$\bar{X}_{.10}$	HG1	$\bar{X}_{.25}$	$\tilde{X}$	$\bar{X}$
	Total	TAK	JOH	HG2	JAE	$\hat{\theta}$	HG1	$\bar{X}_{.10}$	$\bar{X}_{.25}$	$\tilde{X}$	$\bar{X}$

Appendix

Proof of Theorem 3.1. Let $(x_1,\dots,x_n)$ be an ordered sample of size n such that $\bar{x} < \tilde{x}$. Let $v = t/s$ and $w = t/s^2$. Then (3.1) can be rewritten as

$$h(x;\theta,v,w) = \{2\sqrt{2\pi}\, wv^{-1}\Phi(-v)\}^{-1}\exp\{-\tfrac{1}{2}v^2w^{-2}(|x-\theta| + w)^2\} \quad ,$$

$$-\infty < \theta < \infty \quad , \quad v,w > 0 \quad .$$

First observe that if n is even and $x_{n/2} < \bar{x}$, then $\hat{\theta} = \bar{x}$. So for n even, we need only consider the case when $\bar{x} < x_{n/2}$. For fixed v and w, the likelihood is maximized by that integer k and θ satisfying

$$x_k < \theta < x_{k+1} \tag{A.1}$$

which minimizes $\sum_1^n (x_i-\theta)^2 + 2w\sum_1^n |x_i-\theta|$. It is clear from the assumptions we have made and Theorem 2.1 that $2k < n$. This gives

$$\theta = \begin{cases} \bar{x} + w(n-2k)/n & \text{if } w \in I_k \\ x_{k+1} & \text{if } w \in J_k \end{cases} \tag{A.2}$$

where I_k is the interval $[\{n(x_k-\bar{x})/(n-2k)\}^+,\ n(x_{k+1}-\bar{x})/(n-2k)]$ and

$$J_k = \begin{cases} [n(x_{k+1}-\bar{x})/(n-2k),\ n(x_{k+1}-\bar{x})/(n-2k-2)] & \text{if } n-2k-2 > 0 \\ [n(x_{k+1}-\bar{x})/(n-2k),\infty) & \text{if } n-2k-2 < 0 \quad . \end{cases}$$

Now for fixed v, and (θ,k) given by (A.2), it can be verified that there is a unique w that minimizes the likelihood. This w is given by

$$w = \begin{cases} n^{-1}v^2\sum_1^k(\bar{x}-x_i) + \{n^{-2}v^4(\sum_1^k(\bar{x}-x_i))^2 + n^{-1}v^2\sum_1^n(x_i-\bar{x})^2\}^{1/2} & \text{if } w \in I_k \\ (2n)^{-1}v^2\sum_1^n|x_i-x_{k+1}| + \frac{1}{2}\{n^{-2}v^4(\sum_1^n|x_i-x_{k+1}|)^2 + 4n^{-1}v^2\sum_1^n(x_i-x_{k+1})^2\}^{1/2} & \\ & \text{if } w \in J_k \quad . \end{cases}$$

Inverting these equations yields the intervals $\{[p_{1k},p_{2k}],\ [q_{1k},q_{2k}];$ $k = k_0,\ldots,[(n-1)/2]\}$ given in (3.2).

Fortran programs

```
      SUBROUTINE TSUB(XDATA,X,N,TOL,THETA)
C  XDATA IS AN N-DIMENSIONAL VECTOR CONTAINING THE ORDERED X-VALUES
C  X IS AN N-DIMENSIONAL WORK VECTOR
C  TOL IS THE PRECISION REQUIRED IN FMIN
C  THETA CONTAINS THE COMPUTED MLE
      IMPLICIT DOUBLE PRECISION (A-H,O-Z)
      DOUBLE PRECISION  XDATA(N),X(N),VL(2,50),VR(2,50)
      DOUBLE PRECISION V0(2,50),SS(50),SA(50),SK(50),FV(2,50)
      COMPLEX ZSM,ZLG
      LOGICAL  GCASE
      COMMON DT1,RN,DRS,DRK
      EXTERNAL F1,F2
      DATA SQRT8/2.828427125D0/
      RN=DBLE(N)
      NHALF=N/2
      NHP1=NHALF+1
      M=(N-1)/2
      IFLAG=N-2*NHALF
      DO 5 I=1,N
5     X(I)=XDATA(I)
      S=0.0
      T1=0.0
      T2=0.0
      DO 10 I=1,NHALF
      T1=T1+X(NHP1-I)
      T2=T2+X(NHALF+I)
10    S=S+X(NHP1-I)**2+X(NHALF+I)**2
      IF(IFLAG .EQ. 0) THEN
      XMED=0.5*(X(NHALF)+X(NHP1))
      ELSE
      XMED=X(NHP1)
      T2=T2+X(N)
      S=S+X(N)**2
      END IF
      XBAR=(T1+T2)/RN
      AX2=S/RN
      TEMP=XBAR**2
      XVAR=AX2-TEMP
      SCEN=S-RN*TEMP
      IF(XBAR .EQ. XMED) THEN
      THETA=XBAR
      GCASE=.FALSE.
      GO TO 700
      END IF
      IF(XBAR  .LT. XMED) THEN
      GCASE=.FALSE.
      ELSE
      DO 20 I=1,NHALF
      TEMP=-X(I)
      X(I)=-X(N-I+1)
20    X(N-I+1)=TEMP
      IF(IFLAG .EQ. 1) X(NHP1)=-X(NHP1)
      GCASE=.TRUE.
      XBAR=-XBAR
```

```
      END IF
      KO=NHALF
30    IF(X(KO) .LE. XBAR) GO TO 35
      KO=KO-1
      GO TO 30
35    IF((IFLAG .EQ. 0) .AND. (KO .EQ. NHALF)) THEN
      THETA=XBAR
      GO TO 700
      END IF
      S1=0.0
      DO 50 I=KO,1,-1
50    S1=XBAR-X(I)+S1
      K=KO
      TEMP=0.0
60    KP1=K+1
      VL(1,K)=TEMP
      RM2K=DBLE(N-2*K)
      AS1=S1/RM2K
      DIFF=X(KP1)-XBAR
      VNUM=DIFF*RN/RM2K
      SK(K)=S1
      VR(1,K)=VNUM/SQRT(2.0*DIFF*AS1+XVAR)
      VL(2,K)=VR(1,K)
      S=0.0
      DO 70 I=1,K
      S=S+ABS(X(K+I)-X(KP1))+ABS(X(KP1-I)-X(KP1))
70    CONTINUE
      DO 80 I=2*K+1,N
80    S=S+ABS(X(I)-X(KP1))
      SA(K)=S
      ASQ=AX2+X(KP1)*(X(KP1)-2.0*XBAR)
      SS(K)=ASQ*RN
      IF(K .EQ. M) GO TO 100
      RM2KP=DBLE(N-2*K-2)
      VNUM2=DIFF*RN/RM2KP
      VR(2,K)=VNUM2/SQRT(DIFF*S/RM2KP+ASQ)
      TEMP=VR(2,K)
      K=K+1
      S1=S1+XBAR-X(K)
      GO TO 60
100   DRS=RN*SCEN
      DTEMP=0.5D0*LOG(DBLE(XVAR))
      DO 210 K=KO,M
      DT1=SK(K)
      DRK=DBLE(K)
      AA=VL(1,K)
      BB=VR(1,K)
      VX=FMIN(AA,BB,F1,TOL)
      VO(1,K)=VX
      FV(1,K)=F1(VX)+DTEMP
210   CONTINUE
      DTEMP2=LOG(5.013256548*RN)
      DO 220 K=KO,M-1
      DT1=SA(K)
      DRS=RN*SS(K)
      AA=VL(2,K)
```

```
          BB=VR(2,K)
          VX=FMIN(AA,BB,F2,TOL)
          VO(2,K)=VX
          FV(2,K)=F2(VX)-DTEMP2
 220      CONTINUE
          FVM=FV(1,K0)
          K1=K0
          I1=1
          DO 230 K=K0,M
          DO 230 I=1,2
          IF(FVM .GE. FV(I,K)) THEN
          FVM=FV(I,K)
          K1=K
          I1=I
          END IF
          IF(K .EQ. M) GO TO 235
 230      CONTINUE
 235      IF(K1 .EQ. M .AND. VO(1,M) .EQ. VR(1,M)) GO TO 300
          DT1=SA(M)
          DRS=RN*SS(M)
          AA=VR(1,M)
          BB=MAX(SQRT8,2.0*RN*SQRT(DBLE(ASQ))/DT1)
          VO(2,M)=FMIN(AA,BB,F2,TOL)
          FV(2,M)=F2(VO(2,M))-DTEMP2
          I=2
          K=M
          IF(FVM .GE. FV(2,M)) GO TO 300
          SRATIO=SS(M)*RN/SA(M)**2
          CON=LOG(SA(M)/2.506628275*RN)+1.0
          A=CON-FVM
          B=SRATIO/2.0-1.0
          C=0.9539460517-1.5*SRATIO**2
          AA=BB
          IF(B)240,250,260
 240      TEMP=5.0+7.0*SRATIO**2/6.0
          VINF=SQRT(AMAX1(-TEMP/B,8.0))
          YU=CON+B/VINF**2+0.5*TEMP/VINF**4
          IF(FVM .LE. (CON+B/8.0+C/64.0)) THEN
          GO TO 600
          ELSE IF (FVM .GE. YU) THEN
          GO TO 300
          ELSE
          CALL ZQADR(A,B,C,ZSM,ZLG,IER)
          IF(IER .NE. 0) WRITE(*,*) ' IER FROM ZQADR=',IER
          BB=SQRT(MAX(DBLE(ZSM),DBLE(ZLG)))
          GO TO 280
          END IF
 250      IF(A) 300,300,270
 260      IF(A) 300,265,270
 265      BB=-C/B
          GO TO 275
 270      CALL ZQADR(A,B,C,ZSM,ZLG,IER)
          IF(IER .NE. 0) WRITE(*,*) ' IER FROM ZQADR=',IER
          BB=SQRT(MAX(DBLE(ZSM),DBLE(ZLG)))
 275      IF(BB .LE. AA) GO TO 600
 280      VX=FMIN(AA,BB,F2,TOL)
```

```
      TEMP=F2(VX)-DTEMP2
      IF(F2(VX)-DTEMP2 .GT. FVM) GO TO 600
300   THETA=X(M+1)
      GO TO 700
600   IF(I1 .EQ. 1) THEN
      K=K1
      V=V0(1,K)
      TEMP=V**2/RN
      TEMP2=TEMP*SK(K)
      W0=TEMP2+SQRT(TEMP2**2+TEMP*SCEN)
      THETA=XBAR+W0*(RN-DBLE(2*K))/RN
      GO TO 700
      ELSE
      THETA=X(K1+1)
      END IF
700   IF(GCASE) THEN
      THETA=-THETA
      XBAR=-XBAR
      END IF
      RETURN
      END

      DOUBLE PRECISION FUNCTION F1(V)
      IMPLICIT DOUBLE PRECISION (A-H,O-Z)
      COMMON S1,RN,RS,RK
      U=V*S1
      D=U/SQRT(RS)
      D2=D**2
      A=SQRT(1.0+D2)+D
      CALL MDNORD(-V,P)
      F1=LOG(A*P)+0.5/A**2+2.0*(D/A+RK*(RN-RK)*(V/RN)**2)
      RETURN
      END

      DOUBLE PRECISION FUNCTION F2(V)
      IMPLICIT DOUBLE PRECISION (A-H,O-Z)
      REAL RM
      COMMON S1,RN,RS
      A=4.0*RS
      U=V*S1
      Y=U+SQRT(U**2+A)
      CALL MSMRAT(REAL(V),RM,IER)
      IF(IER .NE. 0) RM=(SQRT(V**2+8.D0)+3.0*V)/4.D0
      F2=LOG(Y/RM)+2.0*(U+RS/Y)/Y
      RETURN
      END

      DOUBLE PRECISION FUNCTION FMIN(AX.BX.F,TOL)
C  THIS IS A SLIGHTLY MODIFIED VERSION OF A PROGRAM BY THE
C  SAME NAME IN FORSYTHE, MALCOLM AND MOLER(1977)
      DOUBLE PRECISION AX,BX,F.TOL
      DOUBLE PRECISION A,B.C.D,E,EPS,XM.P,Q,R,TOL1.TOL2.U.V,W
      DOUBLE PRECISION FU,FV,FW.FX,X
      SAVE EPS
      DATA C/0.3819660113D0/
      IF(EPS .GT. 0.0D0) GO TO 15
```

```
      EPS=1.00
10    EPS=EPS/2.00
      TOL1=1.0+EPS
      IF(TOL1 .GT. 1.00) GO TO 10
      EPS=SQRT(EPS)
15    A=AX
      B=BX
      V=A+C*(B-A)
      W=V
      X=V
      E=0.0
      FX=F(X)
      FV=FX
      FW=FX
20    XM=0.5*(A+B)
      TOL1=EPS*ABS(X)+TOL/3.0
      TOL2=2.0*TOL1
      IF(ABS(X-XM) .LE. (TOL2 - 0.5*(B-A))) GO TO 90
      IF(ABS(E) .LE. TOL1) GO TO 40
      R=(X-W)*(FX-FV)
      Q=(X-V)*(FX-FW)
      P=(X-V)*Q-(X-W)*R
      Q=2.00*(Q-R)
      IF(Q .GT. 0.0) P=-P
      Q=ABS(Q)
      R=E
      E=D
30    IF(ABS(P) .GE. ABS(0.5*Q*R)) GO TO 40
      IF(P .LE. Q*(A-X)) GO TO 40
      IF(P .GE. Q*(B-X)) GO TO 40
      D=P/Q
      U=X+D
      IF((U-A) .LT. TOL2) D=SIGN(TOL1,XM-X)
      IF((B-U) .LT. TOL2) D=SIGN(TOL1,XM-X)
      GO TO 50
40    IF(X .GE. XM) E=A-X
      IF(X .LT. XM) E=B-X
      D=C*E
50    IF(ABS(D) .GE. TOL1) U=X+D
      IF(ABS(D) .LT. TOL1) U=X+SIGN(TOL1,D)
      FU=F(U)
      IF(FU .GT. FX) GO TO 60
      IF(U .GE. X) A=X
      IF(U .LT. X) B=X
      V=W
      FV=FW
      W=X
      FW=FX
      X=U
      FX=FU
      GO TO 20
60    IF(U .LT. X) A=U
      IF(U .GE. X) B=U
      IF(FU .LE. FW) GO TO 70
      IF(W .EQ. X) GO TO 70
      IF(FU .LE. FV) GO TO 80
```

```
      IF(V .EQ. X) GO TO 80
      IF(V .EQ. W) GO TO 80
      GO TO 20
70    V=W
      FV=FW
      W=U
      FW=FU
      GO TO 20
80    V=U
      FV=FU
      GO TO 20
90    TOL2=2.0*TOL
      IF(X-AX .LT. TOL2) THEN
      IF(F(X) .GE. F(AX)) X=AX
      ELSE IF (BX-X .LT. TOL2) THEN
      IF(F(X) .GE. F(BX)) X=BX
      END IF
      FMIN=X
      RETURN
      END
```

REFERENCES

1. Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H. and Tukey, J. W. (1972). Robust estimates of location. Princeton University Press.

2. Atkinson, A. C. (1970). A method for discriminating between models. J. Roy. Statist. Soc. Ser. B, 32, 323-353.

3. Bell, R. M. (1980). An adaptive choice of the scale parameter for M-estimators. Stanford University, Department of Statistics Technical Report No. 3.

4. Beran, R. (1978). An efficient and robust adaptive estimator of location. Ann. Statist. 6, 292-313.

5. Birnbaum, Z. W. (1942). An inequality for Mills' ratio. Ann. Math. Statist. 13, 245-246.

6. Brown, B. M. and Kildea, D. G. (1979). Outhier-detection tests and robust estimators based on signs of residuals. Commun. Statist. A8, 257-270.

7. Brown, G. W. and Tukey, J. W. (1946). Some distributions of sample means. Ann. Math. Statist. 17, 1-12.

8. Brown L. D. (1971). Non-local asymptotic optimality of appropriate likelihood ratio tests. Ann. Math. Statist. 42, 1206-1240.

9. Cox, D. R. (1961). Tests of separate families of hypotheses. Proc. Fourth Berkeley Symp. 1, 105-123.

10. Crow, E. L. and Siddiqui, M. M. (1967). Robust estimation of location. J. Amer. Statist. Assoc. 62, 353-389.

11. David, H. T. (1962). The sample mean among the moderate order statistics. Ann. Math. Statist. 33, 1160-1166.

12. Forsythe, G. E., Malcolm, M. A. and Moler, C. B. (1977). <u>Computer methods for mathematical computations</u>. Prentice Hall. New Jersey.

13. Gastwirth, J. L. (1966). On robust procedures. J. Amer. Statist. Assoc. 61, 929-948.

14. Hogg, R. V. (1967). Some observations on robust estimation. J. Amer. Statist. Assoc. 62, 1179-1185.

15. Hogg, R. V. (1974). Adaptive robust procedures: A partial review and some suggestions for future applications and theory. J. Amer. Statist. Assoc. 69, 909-927.

16. Hogg, R. V., Uthoff, V. A., Randles, R. H. and Davenport, A. S. (1972). On the selection of the underlying distribution and adaptive estimation. J. Amer. Statist. Assoc. 67, 597-600.

17. Huber, P. J. (1964). Robust estimation of a location parameter. Ann. Math. Statist. 35, 73-101.

18. Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. Proc. Fifth Berkeley Symp. 1, 221-234.

19. Huber, P. J. (1981). <u>Robust Statistics</u>. Wiley, New York.

20. Johnson, N. L. and Kotz, S. (1970). <u>Continuous univariate distributions</u> Vol. 2. Wiley, New York.

21. Lehmann, E. L. (1983). <u>Theory of Point Estimation</u>. Wiley, New York.

22. Loh, Wei-Yin (1983). A note on the geometry of Kullback-Leibler information numbers. University of Wisconsin-Madison. Mathematics Research Center Technical Summary Report #2506.

23. Sacks, J. (1975). An asymptotically efficient sequence of estimators of a location parameter. Ann. Statist. 3, 285-298.

24. Stone, C. J. (1975). Adaptive maximum likelihood estimators of a location parameter. Ann. Statist. 3, 267-284.

25. Tukey, J. W. (1979). Study of robustness by simulation: particularly improvement by adjustment and combination. In: Launer, R. L. and Wilkinson, G. N. (Eds.), Robustness in Statistics. Academic Press, New York.

26. Weerahandi, S. and Zidek, J. V. (1978). Pooling prior distributions. University of British Columbia. Inst. of Applied Math. and Statistics. Technical Report 78-34.

27. Wegman, E. J. and Carroll, R. J. (1977). A Monte Carlo study of robust estimators of location. Commun. Statist. A6, 795-812.

WYL/jvs

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER #2523	2. GOVT ACCESSION NO AD-A130525	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Directionally Efficient Robust Estimators of Location Via Exponential Embedding		5. TYPE OF REPORT & PERIOD COVERED Summary Report - no specific reporting period
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Wei-Yin Loh		8. CONTRACT OR GRANT NUMBER(s) MCS-7927062, Mod. 2. DAAG29-80-C-0041
9. PERFORMING ORGANIZATION NAME AND ADDRESS Mathematics Research Center, University of Wisconsin 610 Walnut Street Madison, Wisconsin 53706		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Work Unit Number 4 - Statistics and Probability
11. CONTROLLING OFFICE NAME AND ADDRESS See Item 18 below		12. REPORT DATE May 1983
		13. NUMBER OF PAGES 37
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

National Science Foundation
Washington, DC 20550

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Robust estimation, embedding, maximum likelihood, asymptotic efficiency.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A general method is presented for constructing a location estimator which is asymptotically efficient at any two different location-scale families of symmetric distributions as well as at an appropriately defined class of distributions lying in between. The method works by embedding the two families in a comprehensive parametric model and identifying the estimator with the MLE. The case when the families are Normal and Double exponential is examined in detail.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

END

FILMED

8-83

DTIC