The Annals of Statistics
1976, Vol. 4, No. 1, 68–85

ROBUST ESTIMATION OF A LOCATION PARAMETER IN THE PRESENCE OF ASYMMETRY[1]

BY JOHN R. COLLINS

Purdue University

Huber's theory of robust estimation of a location parameter is adapted to obtain estimators that are robust against a class of asymmetric departures from normality. Let F be a distribution function that is governed by the standard normal density on the set $[-d, d]$ and is otherwise arbitrary. Let $X_1, \cdots, X_n$ be a random sample from $F(x-\theta)$, where θ is the unknown location parameter. If ψ is in a class of continuous skew-symmetric functions Ψ_c which vanish outside a certain set $[-c, c]$, then the estimator T_n, obtained by solving $\sum \psi(X_i - T_n) = 0$ by Newton's method with the sample median as starting value, is a consistent estimator of θ. Also $n^{\frac{1}{2}}(T_n - \theta)$ is asymptotically normal. For a model of symmetric contamination of the normal center of F, an asymptotic minimax variance problem is solved for the optimal ψ. The solution has the form $\psi(x) = x$ for $|x| \leqq x_0$, $\psi(x) = x_1 \tanh[\frac{1}{2}x_1(c-|x|)] \operatorname{sgn}(x)$ for $x_0 \leqq |x| \leqq c$, and $\psi(x) = 0$ for $|x| \geqq c$. The results are extended to include an unknown scale parameter in the model.

1. Introduction and summary. A general theory of robust estimation of a location parameter was developed by Huber (1964). Huber derived estimators which are robust against symmetric departures from a symmetric model distribution. In this paper Huber's theory is adapted to allow for asymmetric departures from the symmetric model.

Let $X_1, \cdots, X_n$ be independent identically distributed (i.i.d.) random variables with distribution function $F((x-\theta)/\sigma)$, where θ is the unknown location parameter and σ is a scale parameter. Sections 2 and 3 consider the problem of estimating θ when σ is known, and Section 4 extends the results to scale invariant estimators of θ for the case of unknown scale. Consider the scale known case, taking $\sigma = 1$ without loss of generality. Huber proposed estimating θ by solving

$$\sum_{i=1}^{n} \psi(X_i - T_n) = 0. \tag{1.1}$$

Under regularity conditions on F and ψ, T_n is a consistent estimator of θ and $n^{\frac{1}{2}}(T_n - \theta)$ converges in distribution to the normal distribution with mean 0 and variance

$$V(\psi, F) = \frac{\int \psi^2 \, dF}{(\int \psi' \, dF)^2}\,. \tag{1.2}$$

Received July 1974; revised June 1975.

[1] Much of this paper is based on the author's Ph. D. dissertation at the University of California, Berkeley, 1973. This research was partially supported by a Hughes Aircraft Company Staff Doctoral Fellowship and partially by the National Science Foundation under Grant GP-38485.

AMS 1970 *subject classifications.* Primary 62G35; Secondary 62G05, 62G20.

Key words and phrases. Robust estimation, asymmetry, location parameter, M-estimators, minimax variance.

Estimators T_n defined by (1.1) are called maximum-likelihood-type estimators or M-estimators, since if F has a smooth density f, then the maximum likelihood estimator of θ is given by (1.1) with $\psi = -f'/f$.

In robust estimation theory, F is not known but is assumed to lie in some appropriate neighborhood of distributions $\mathscr{F}$. One can define the most robust M-estimator as the solution to Huber's minimax variance problem: find the ψ that minimizes $\sup\{V(\psi, F) : F \in \mathscr{F}\}$.

To obtain consistency of T_n for all F in $\mathscr{F}$, one requires

$$(1.3) \qquad \int \psi \, dF = 0$$

for all $F \in \mathscr{F}$. The natural way to satisfy (1.3) is to restrict $\mathscr{F}$ to symmetric distributions and to consider skew-symmetric ψ's. For example, consider the class $\mathscr{F}_\varepsilon$ of ε-contaminated normal distributions: fix ε, $0 < \varepsilon < 1$, and say that F is in $\mathscr{F}_\varepsilon$ if

$$(1.4) \qquad F(x) = (1 - \varepsilon)\Phi(x) + \varepsilon H(x) ,$$

where $\Phi(x) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{x} \exp(-t^2/2) \, dt$. In this case one requires the unknown contaminating distribution H to be symmetric. This is a stringent requirement, since in practice one would not expect departures from normality to be symmetric. However if the assumption of symmetry of H were removed, then the parameter θ would not be identifiable in the ε-contamination model. Since the estimators of θ would not be consistent, asymptotic variance would no longer be a reasonable criterion for judging the performance of the estimators. Further discussion of this problem is found in Huber (1964).

In Section 2, we carry out the idea of restricting the class of ψ's so that the resulting M-estimators are consistent in the presence of asymmetry of F. The model is the following: specify a number $d > 0$, and say that F is in $\mathscr{F}$ if F is governed by the normal density $\varphi(x) = (2\pi)^{-\frac{1}{2}} \exp(-x^2/2)$ on $[-d, d]$. The distribution outside the interval $[-d, d]$ is arbitrary. The model $\{F((x - \theta)/\sigma), F \in \mathscr{F}\}$ reflects the type of departure from normality that is typically detected in large samples of data: the data appears to be normal in a central region out to some number d of standard deviations, but nonnormal and asymmetric in the tails.

The appropriate M-estimators for this model are obtained by restricting ψ to a class Ψ_c of continuous skew-symmetric functions that vanish outside a certain set $[-c, c]$, where $c < d$. Then (1.3) holds if $F \in \mathscr{F}$ and $\psi \in \Psi_c$. Note that symmetry in the central region of F is required. The use of such M-estimators was proposed by F. R. Hampel, who observed that the asymptotic behavior should depend only on the center and not on the tails of F. Some examples of ψ's which vanish outside an interval appear in Andrews et al. (1972).

In Section 2 it is shown that when $F \in \mathscr{F}$ and $\psi \in \Psi_c$, then T_n is consistent and asymptotically normal. One complication that arises is that when ψ is in Ψ_c, equation (1.1) has multiple roots with probability 1, even asymptotically.

This difficulty is resolved by defining T_n to be the particular solution of (1.1) obtained by Newton's method with the sample median of the observations as the starting value.

In Section 3 we consider a model of symmetric contamination of the normal center of F, and solve the minimax variance problem. The solution has the form

$$\begin{aligned} \psi(x) &= x && |x| \leqq x_0 \\ &= x_1 \tanh\left[\tfrac{1}{2}x_1(c - |x|)\right] \operatorname{sgn}(x) && x_0 \leqq |x| \leqq c \\ &= 0 && |x| \geqq c\,. \end{aligned} \tag{1.5}$$

Two other optimality criteria based on asymptotic variances are also considered, and the corresponding variational problems are solved to obtain the optimal ψ's.

In Section 4 the results are extended to the scale unknown case.

2. The model and class of estimators. Let α be a fixed number, $0 < \alpha < \frac{1}{2}$, and let

$$d = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right). \tag{2.1}$$

Let $\mathscr{F}$ be the class of distribution functions F satisfying the following condition:

There exists $\gamma \in (-\alpha/2, \alpha/2)$ such that $F(x) = \gamma + \Phi(x)$ for all $x \in [-d, d]$.

Let $X_1, \cdots, X_n$ be i.i.d. random variables with distribution function $F(x - \theta)$, where F is in $\mathscr{F}$ but is otherwise unknown. The problem is to estimate the unknown parameter θ, the center of symmetry of the symmetric part of F.

Define

$$k = \Phi^{-1}\left(\frac{1}{2} + \frac{\alpha}{2}\right) \tag{2.2}$$

and

$$c = d - k\,. \tag{2.3}$$

Note that $0 < \alpha < \frac{1}{2}$ implies that $c > 0$. Also, denoting the median of F by $m(F)$, note that $F \in \mathscr{F}$ implies that $m(F) \in (-k, k)$.

Let Ψ_c denote the class of functions ψ which map the real line R into R and satisfy:

(i) ψ is continuous and has a continuous derivative ψ'.
(ii) $\psi(x) = -\psi(-x)$ for all x.
(iii) $\psi(x) = 0$ when $|x| > c$.
(iv) $\psi \geqq 0$ on $[0, c]$, and $\psi(x) > 0$ for some $x \in [0, c]$.

Note that $\psi(-c) = \psi(0) = \psi(c) = 0$ and that both ψ and ψ' are uniformly continuous on R.

We propose to estimate θ by solving

$$\textstyle\sum_{i=1}^{n} \psi(X_i - \theta) = 0 \tag{2.4}$$

for θ, where $\psi \in \Psi_c$. However, since (2.4) has no unique solution, we define the estimator to be essentially the solution of (2.4) obtained by Newton's method with M_n, the sample median of $(X_1, \cdots, X_n)$, as the starting value.

DEFINITION 2.1. Consider the sequence $\{\hat\theta_j\}$ given by $\hat\theta_0 = M_n$ and

$$\hat\theta_{j+1} = \hat\theta_j + \frac{\sum_{i=1}^n \psi(X_i - \hat\theta_j)}{\sum_{i=1}^n \psi'(X_i - \hat\theta_j)}, \qquad j = 0, 1, 2, \cdots. \tag{2.5}$$

The estimator $T_n = T_n(X_1, \cdots, X_n)$ is defined by

$$\begin{aligned} T_n &= \lim_{j\to\infty} \hat\theta_j \quad \text{if } \lim_{j\to\infty} \hat\theta_j \text{ exists} \\ &= M_n \quad \text{otherwise.} \end{aligned} \tag{2.6}$$

REMARK 2.1. In practice one does not try to determine the existence of $\lim_{j\to\infty} \hat\theta_j$. An appropriate approximate algorithm is to specify an integer $N_0 > 0$ and a number $\varepsilon_0 > 0$, and take T_n to be $\hat\theta_{j_0}$ if j_0 is the first $j \leqq N_0$ for which $|\hat\theta_j - \hat\theta_{j-1}| < \varepsilon_0$; otherwise, if no such $j \leqq N_0$ exists, take T_n to be M_n.

REMARK 2.2. Since T_n is a translation invariant estimator, we shall assume without loss of generality that $\theta = 0$.

THEOREM 2.1. *Let $\psi \in \Psi_c$, and let $X_1, X_2, \cdots$ be* i.i.d. *with distribution function $F \in \mathscr{F}$. Then*

$$T_n \to_P 0 . \tag{2.7}$$

REMARK 2.3. The motivation of the proof is as follows. We define

$$\lambda(t) = E_F \psi(X_1 - t) = \int \psi(x - t)\, dF(x) \tag{2.8}$$

and

$$\lambda_n(t) = \lambda_n(X_1, \cdots, X_n; t) = \frac{1}{n} \sum_{i=1}^n \psi(X_i - t) . \tag{2.9}$$

The general consistency proof of Huber (1967) does not apply here because the condition that $\lambda(t)$ have a unique zero at $t = 0$ is not satisfied. A special argument is needed to show that the Newton's method solution of $\lambda_n(t) = 0$ with starting value M_n is consistent. This is made plausible by the fact that the Newton's method solution of $\lambda(t) = 0$ with starting value $m(F)$ is 0 (a direct consequence of the following lemma).

LEMMA 2.1. *Let $\psi \in \Psi_c$ and $F \in \mathscr{F}$. Then*

(i) $\lambda(t) = \int_{-c+t}^{c+t} \psi(x - t)\varphi(x)\, dx$ *when* $|t| \leqq k$,
(ii) $\lambda(t) = -\lambda(-t)$ *when* $|t| \leqq k$,
(iii) $\sup \{\lambda'(t) : |t| \leqq k\} < 0$,
(iv) $\sup \left\{ \left| \frac{\lambda(t)}{t\lambda'(t)} \right| : |t| \in (0, k] \right\} < 2$.

PROOF. (i) follows immediately from the definition of $\mathscr{F}$, c, and Ψ_c.

If $|t| \leqq k$, then

$$\lambda(t) = \int_{-c}^{c} \psi(x)\varphi(x + t)\, dx = \int_0^c \psi(x)[\varphi(x + t) - \varphi(x - t)]\, dx , \tag{2.10}$$

so that (ii) follows by symmetry of φ and skew-symmetry of ψ.

Note that $0 < k < \Phi^{-1}(\frac{3}{4})$. When $|t| < k$, we have

$$-\lambda'(t) = \int_0^c \psi(x)[(x+t)\varphi(x+t) + (x-t)\varphi(x-t)]\,dx$$

$$(2.11) \qquad = 2\exp(-t^2/2) \int_0^c \left\{ \sum_{j=0}^{\infty} \left[\frac{t^{2j}}{(2j)!} - \frac{t^{2j+2}}{(2j+1)!} \right] x^{2j} \right\} x\psi(x)\varphi(x)\,dx$$

$$\geqq 2[1 - t^2] \cdot [\exp(-t^2/2)] \cdot \int_0^c x\psi(x)\varphi(x)\,dx > 0\,,$$

proving (iii).

Similarly, when $t \in (0, k]$, one obtains

$$(2.12) \qquad -\lambda(t) \leqq 2t[\cosh(kc)][\exp(-t^2/2)] \int_0^c x\psi(x)\varphi(x)\,dx\,,$$

so that

$$(2.13) \qquad \frac{\lambda(t)}{t\lambda'(t)} \leqq \frac{\cosh(kc)}{1-t^2} \leqq \frac{\cosh(kc)}{1-k^2} < 2$$

for all $t \in (0, k]$, proving (iv). □

LEMMA 2.2. *Let* $\psi \in \Psi_c$ *and let* $X_1, X_2, \cdots$ *be* i.i.d. *with distribution function* $F \in \mathscr{F}$. *Then*

(i) $\sup\{|\lambda_n(t) - \lambda(t)| : |t| \leqq k\} \to_P 0$.

(ii) $\sup\{|\lambda_n'(t) - \lambda'(t)| : |t| \leqq k\} \to_P 0$.

PROOF. (i) First note that λ is a continuous function on $[-k, k]$ and that the process λ_n has continuous paths on $[-k, k]$. By Theorems 8.1 and 8.2 of Billingsley (1968), or by an obvious direct argument, it is sufficient to show:

(A) $\lambda_n(t) \to_P \lambda(t)$ for each $t \in [-k, k]$.

and

(B) For every $\varepsilon > 0$,

$$(2.14) \qquad \lim_{\delta\to 0} \limsup_n P[\sup\{|\lambda_n(t_1) - \lambda_n(t_2)| : |t_1| \leqq k, |t_2| \leqq k, |t_1 - t_2| < \delta\} \geqq \varepsilon] = 0\,.$$

The weak law of large numbers proves (A). To prove (B), let $\varepsilon > 0$. By uniform continuity of ψ, there is a $\delta > 0$ such that $|\psi(x_1) - \psi(x_2)| < \varepsilon$ whenever $|x_1 - x_2| < \delta$. Thus for $|t_1 - t_2| < \delta$,

$$(2.15) \qquad |\lambda_n(t_1) - \lambda_n(t_2)| \leqq \frac{1}{n} \sum_{i=1}^n |\psi(X_i - t_1) - \psi(X_i - t_2)| < \varepsilon\,,$$

so that

$$(2.16) \qquad P[\sup\{|\lambda_n(t_1) - \lambda_n(t_2)| : |t_1| \leqq k, |t_2| \leqq k, |t_1 - t_2| < \delta\} \geqq \varepsilon\} = 0$$

for all $n \geqq 1$.

(ii) By the weak law of large numbers,

$$(2.17) \qquad \lambda_n'(t) = -\frac{1}{n} \sum_{i=1}^n \psi'(X_i - t) \to_P -E_F\psi'(X_1 - t)\,.$$

The conditions on ψ ensure that

$$(2.18) \qquad \lambda'(t) = -E_F\psi'(X_1 - t)$$

holds when $|t| \leqq k$. The rest of the proof of (ii) follows that of (i). □

PROOF OF THEOREM 2.1. Let

$$\gamma' = \sup\left\{\left|\frac{\lambda(t)}{t\lambda'(t)}\right| : 0 < |t| \leqq k\right\} \tag{2.19}$$

and choose γ and ε satisfying $\gamma' < \gamma < 2$ and $0 < \varepsilon < \min\{(k - |m(F)|)/4, (2-\gamma)/\gamma\}$. Consider the event $E_{1,n}$ defined by

$$\sup\{\lambda_n'(t) : |t| \leqq k\} < 0\,, \qquad \lambda_n(-k) > 0\,, \qquad \lambda_n(k) < 0\,. \tag{2.20}$$

The event $E_{1,n}$ implies that λ_n is a strictly monotone decreasing function on $[-k, k]$ with 0 in its range, so that the set $\lambda_n^{-1}\{0\} \cap [-k, k]$ has exactly one member. Define the random variable Z_n by

$$\begin{aligned} Z_n &= \lambda_n^{-1}\{0\} \cap [-k, k] \qquad &&\text{if } E_{1,n} \text{ obtains} \\ &= 0 &&\text{otherwise.} \end{aligned} \tag{2.21}$$

Let $I = \{t : |t| \leqq (|m(F)| + k)/2\}$, and define the process $h_n(t)$ on I by

$$\begin{aligned} h_n(t) &= 1 \qquad &&\text{if } t = 0 \\ &= \left|\frac{\lambda_n(t + Z_n)}{t\lambda_n'(t + Z_n)}\right| &&\text{if } t \neq 0 \text{ and } E_{1,n} \text{ obtains} \\ &= 1 &&\text{otherwise.} \end{aligned} \tag{2.22}$$

Consider the events

$$E_{2,n} : |M_n| < |m(F)| + (k - |m(F)|)/4\,,$$

$$E_{3,n} : |Z_n| < \varepsilon\,,$$

and

$$E_{4,n} : \sup\{|h_n(t)| : t \in I\} < \gamma\,.$$

The event $\bigcap_{i=1}^4 E_{i,n}$ implies that

$$\left|\frac{\lambda_n(\hat\theta_j)}{\lambda_n'(\hat\theta_j)}\right| < \gamma|\hat\theta_j - Z_n| \tag{2.23}$$

for all $\hat\theta_j$ in the sequence defined by (2.5), forcing $T_n = Z_n$. Hence it is sufficient to show that

$$P(\textstyle\bigcap_{i=1}^4 E_{i,n}) \to 1 \qquad \text{as } n \to \infty\,, \tag{2.24}$$

for then we can conclude that

$$P[|T_n| < \varepsilon] \geqq P[T_n = Z_n, |Z_n| < \varepsilon] \to 1 \qquad \text{as } n \to \infty\,. \tag{2.25}$$

Since $M_n \to_P m(F)$, $P(E_{2,n}) \to 1$.

Lemmas 2.1 (iii) and 2.2 (ii) imply that $P[\sup\{\lambda_n'(t) : |t| \leqq k\} < 0] \to 1$. Also $\lambda_n(-k) \to_P \lambda(-k) > 0$ and $\lambda_n(k) \to_P \lambda(k) < 0$. Thus $P(E_{1,n}) \to 1$. It also follows easily from the lemmas that $P(E_{3,n}) \to 1$.

To show $P(E_{4,n}) \to 1$, define

$$\begin{aligned} h(t) &= \left|\frac{\lambda(t)}{t\lambda'(t)}\right| \qquad &&\text{if } t \in I \text{ and } t \neq 0 \\ &= 1 &&\text{if } t = 0\,. \end{aligned} \tag{2.26}$$

Note that h is continuous on I and that the process h_n has continuous paths on I. (Continuity at $t = 0$ follows from (2.28) and (2.29) below). By (iv) of Lemma 2.1 it is sufficient to show that

$$P[\sup\{|h_n(t) - h(t)| : t \in I\} < \varepsilon] \to 1 \qquad \text{as } n \to \infty . \tag{2.27}$$

A sketch of the proof of (2.27) follows. In a neighborhood of $t = 0$, given that $E_{1,n}$ holds, use the mean value theorem to write

$$h_n(t) = \left|\frac{\lambda_n(Z_n) + t\lambda_n'(Z_n + \eta)}{t\lambda_n'(Z_n + t)}\right| = \left|\frac{\lambda_n'(Z_n + \eta)}{\lambda_n'(Z_n + t)}\right| \tag{2.28}$$

for some η, $|\eta| \leqq |t|$. Also

$$h(t) = \left|\frac{\lambda(0) + t\lambda'(\xi)}{t\lambda'(t)}\right| = \left|\frac{\lambda'(\xi)}{\lambda'(t)}\right| \tag{2.29}$$

for some ξ, $|\xi| \leqq |t|$. Using Lemmas 2.1 and 2.2 and the fact that $P(\bigcap_{i=1}^{3} E_{i,n}) \to 0$, find a $\delta > 0$ such that

$$P[\sup\{|h_n(t) - h(t)| : |t| \leqq \delta\} < \varepsilon] \to 1 \qquad \text{as } n \to \infty . \tag{2.30}$$

Then for those $t \in I$ for which $|t| > \delta$, we have

$$|h_n(t) - h(t)| \leqq \frac{1}{\delta}\left|\frac{\lambda_n(t + Z_n)}{\lambda_n'(t + Z_n)} - \frac{\lambda(t)}{\lambda'(t)}\right| , \tag{2.31}$$

so that one can show from the lemmas that

$$P[\sup\{|h_n(t) - h(t)| : t \in I, |t| > \delta\} < \varepsilon] \to 1 \qquad \text{as } n \to \infty . \tag{2.32}$$

Then combine (2.30) and (2.32) to obtain (2.27). □

If ψ is a measurable function, and f is an absolutely continuous density, we define

$$V(\psi, f) = (\textstyle\int \psi^2 f)/(\int \psi f')^2 . \tag{2.33}$$

Note that if $K \neq 0$, then $V(K\psi, f) = V(\psi, f)$.

THEOREM 2.2. *Let $X_1, X_2, \cdots$ be* i.i.d. *with distribution function G, where G is governed on $[-c, c]$ by an absolutely continuous density g. Suppose that $\psi \in \Psi_c$, $T_n \to_P 0$, and $P[T_n = M_n] \to 0$ as $n \to \infty$. Then*

$$n^{\frac{1}{2}} T_n \to_{\mathscr{D}} N(0, V(\psi, g)) . \tag{2.34}$$

PROOF. The hypotheses of Lemma 5 of Huber (1964) hold, except that the condition that $\sum \psi(X_i - T_n) = 0$ with probability 1 does not hold for our definition of T_n. However Huber's proof goes through with the weaker condition

$$n^{\frac{1}{2}} \sum \psi(X_i - T_n) \to_P 0 . \tag{2.35}$$

By Definition 2.1, either $T_n = M_n$ or $\sum \psi(X_i - T_n) = 0$. Since $P[T_n = M_n] \to 0$, $P[\sum \psi(X_i - T_n) = 0] \to 1$ as $n \to \infty$. So

$$P[n^{\frac{1}{2}}|\sum \psi(X_i - T_n)| < \varepsilon] \geqq P[\sum \psi(X_i - T_n) = 0] \to 1 . \tag{2.36}$$ □

We remark that the condition $P[T_n = M_n] \to 0$ in the hypothesis is redundant when $m(G) \neq 0$.

By Theorem 2.1, we immediately have:

COROLLARY 2.1. *If $F \in \mathscr{F}$ and $\psi \in \Psi_c$, then*

$$n^{\frac{1}{2}} T_n \to_{\mathscr{D}} N(0, V(\psi, \varphi)) . \tag{2.37}$$

Note that the asymptotic variance $V(\psi, \varphi) = (\int_{-c}^{c} \psi^2 \varphi)/(\int_{-c}^{c} \psi \varphi')^2$ is independent of $F \in \mathscr{F}$.

3. Robust estimators: criteria and derivations. A natural criterion for an optimal estimator for our model is to minimize $V(\psi, \varphi)$.

LEMMA 3.1. *The infimum of $V(\psi, \varphi)$ as ψ ranges over Ψ_c is*

$$1/\int_{-c}^{c} x^2 \varphi(x)\, dx .$$

PROOF. By the Cauchy–Schwarz inequality, ψ minimizes $(\int_{-c}^{c} \psi^2 \varphi)/(\int_{-c}^{c} \psi \varphi')^2$ in the class of all Lebesgue measurable functions if and only if $\psi(x) = K\psi^*(x)$ for almost all $x \in [-c, c]$, where $K \neq 0$ and

$$\begin{aligned} \psi^*(x) &= x \quad x \in (-c, c) \\ &= 0 \quad \text{otherwise.} \end{aligned} \tag{3.1}$$

The minimum value is $V(\psi^*, \varphi) = 1/\int_{-c}^{c} x^2 \varphi(x)\, dx$. Now ψ^* is not in Ψ_c, but if we consider any dominated sequence $\{\psi_j\} \subset \Psi_c$ satisfying $\psi_j(x) \to \psi^*(x)$ a.e. as $j \to \infty$, then $\int_{-c}^{c} \psi_j^2 \varphi \to \int_{-c}^{c} (\psi^*)^2 \varphi$ and $\int_{-c}^{c} \psi_j \varphi' \to \int_{-c}^{c} \psi^* \varphi'$, so that $V(\psi_j, \varphi) \to V(\psi^*, \varphi)$. □

Lemma 3.1 shows that, for the model of distributions with normal center and unknown tails, the estimator T_n based on ψ^* is optimal (only formally, however, since $\psi^* \notin \Psi_c$). It is shown below that, when the model is extended to allow for symmetric contamination of the normal center, then discontinuous ψ's such as ψ^* can not be optimal according to an asymptotic variance criterion.

Specifically, in analogy to Huber's ε-contamination model, we fix ε, $0 < \varepsilon < 1$, and consider the class of distributions $\mathscr{P}_\varepsilon$ defined as follows: $G \in \mathscr{P}_\varepsilon$ if G is governed on $[-c, c]$ by a density $g(x) = (1 - \varepsilon)\varphi(x) + \varepsilon h(x)$, where h is assumed to be absolutely continuous, symmetric about 0, and satisfies $0 \leqq \int_{-c}^{c} h(x)\, dx \leqq 1$, but is otherwise unknown. The distribution outside the interval $[-c, c]$ is completely arbitrary.

A natural optimality problem for this model is to find the ψ for which $\sup\{V(\psi, g) : G \in \mathscr{P}_\varepsilon\}$ is minimized.

LEMMA 3.2. $\sup\{V(\psi^*, g) : G \in \mathscr{P}_\varepsilon\} = \infty$.

PROOF. As G ranges over $\mathscr{P}_\varepsilon$, $\int_{-c}^{c} [\psi^*(x)]^2 g(x)\, dx$ is bounded above by $(1 - \varepsilon) \int_{-c}^{c} x^2 \varphi(x)\, dx + \varepsilon c^2$. So it is sufficient to show that, given $\delta > 0$, there exists a $G \in \mathscr{P}_\varepsilon$ with density $g_0 = (1 - \varepsilon)\varphi + \varepsilon h_0$ on $[-c, c]$, such that

$$|\int_0^c \psi^*(x) g_0'(x)\, dx| < \delta . \tag{3.2}$$

Since $0 > \int_0^c x\varphi'(x)\,dx = c\varphi(c) - \int_0^c \varphi(x)\,dx$, then the number z defined by

$$(1-\varepsilon)\int_0^c x\varphi'(x)\,dx + \varepsilon[cz - \int_0^c \varphi(x)\,dx] = 0 \tag{3.3}$$

satisfies $z > \varphi(c)$. Let $\delta' = \delta/(2\varepsilon z)$, and choose h_0 so that $h_0(x) = \varphi(x)$ for $x \in [0, c-\delta']$, and $\sup\{h_0(x) : x \in [0, c]\} = h_0(c) = z$. Then

$$\begin{aligned}|\textstyle\int_0^c x[(1-\varepsilon)\varphi'(x) + \varepsilon h_0'(x)]\,dx| \\ &= |(1-\varepsilon)\textstyle\int_0^c x\varphi'(x)\,dx + \varepsilon[cz - \int_0^c h_0(x)\,dx]| \\ &\leqq \varepsilon|\textstyle\int_0^c (\varphi(x) - h_0(x))\,dx| \leqq \varepsilon\delta' z < \delta\,. \qquad \Box\end{aligned} \tag{3.4}$$

A modification of the above proof shows that the result holds for all discontinuous ψ's, so that the optimal ψ must be continuous.

Define the class Ψ_c' by replacing condition (i) in the definition of Ψ_c by

(i′) ψ is continuous and has a piecewise continuous derivative ψ'.

Note that $\Psi_c' \supset \Psi_c$.

THEOREM 3.1. (1) *If* $\varepsilon/(1-\varepsilon) \geqq 2c\varphi(0) - 2\Phi(c) + 1$, *then*

$$\inf\{\sup[V(\psi, g) : G \in \mathscr{P}_\varepsilon] : \psi \in \Psi_c'\} = \infty\,. \tag{3.5}$$

(2) *If* $\varepsilon/(1-\varepsilon) < 2c\varphi(0) - 2\Phi(c) + 1$, *then*

(i) $\int_{-c}^{c} (g'/g)^2 g$ *is minimized among* $G \in \mathscr{P}_\varepsilon$ *by any* G_0 *with density on* $[-c, c]$ *given by*

$$\begin{aligned} g_0(x) &= (1-\varepsilon)\varphi(x)\,, & |x| \leqq x_0 \\ &= \frac{(1-\varepsilon)\varphi(x_0)}{\cosh^2[\frac{1}{2}x_1(c - x_0)]}\cosh^2[\tfrac{1}{2}x_1(c - |x|)]\,, & x_0 \leqq |x| \leqq c\,, \end{aligned} \tag{3.6}$$

where x_0 *and* x_1 *satisfy* $0 < x_0 < c$ *and* $x_0 < x_1$, *and are uniquely determined by*

$$x_0 = x_1 \tanh[\tfrac{1}{2}x_1(c - x_0)] \tag{3.7}$$

and

$$\int_{-c}^{c} [g_0(x) - (1-\varepsilon)\varphi(x)]\,dx = \varepsilon\,. \tag{3.8}$$

(ii) $\psi \in \Psi_c'$ *minimizes* $\sup\{V(\psi, g) : G \in \mathscr{P}_\varepsilon\}$ *if and only if* ψ *is a nonzero multiple of*

$$\begin{aligned} \psi_0(x) &= -g_0'(x)/g_0(x) & x \in [-c, c] \\ &= 0 & \textit{otherwise}\,; \end{aligned} \tag{3.9}$$

or equivalently,

$$\begin{aligned} \psi_0(x) &= x\,, & |x| \leqq x_0 \\ &= x_1 \tanh[\tfrac{1}{2}x_1(c - |x|)]\,\mathrm{sgn}\,(x)\,, & x_0 \leqq |x| \leqq c \\ &= 0\,, & |x| \geqq c\,. \end{aligned} \tag{3.10}$$

PROOF. (1) If h is defined on $[-c, c]$ by

$$(1-\varepsilon)\varphi(0) = (1-\varepsilon)\varphi(x) + \varepsilon h(x)\,, \tag{3.11}$$

then h is nonnegative on $[-c, c]$, and the inequality in the hypothesis implies that $\int_{-c}^{c} h(x)\,dx \leqq 1$. So $g_1(x) = (1-\varepsilon)\varphi(0)$ is the density on $[-c, c]$ of a $G_1 \in \mathscr{P}_\varepsilon$. So for any $\psi \in \Psi_c'$, $0 < \int \psi^2 g_1 < \infty$ and $\int \psi g_1' = 0$, so that $\sup\{V(\psi, g) : G \in \mathscr{P}_\varepsilon\} = \infty$.

(2) We apply the minimax theory developed in Section 8 of Huber (1964). Define $I_c(G) = \int_{-c}^{c} (g'/g)^2 g$ for $G \in \mathscr{P}_\varepsilon$. To show the existence of a unique $G_0 \in \mathscr{P}_\varepsilon$ minimizing $I_c(G)$, we extend $\mathscr{P}_\varepsilon$ to the class $\mathscr{P}_\varepsilon'$ consisting of distributions of the form $G = (1-\varepsilon)F + \varepsilon H$, where $F \in \mathscr{F}$ and H is an arbitrary distribution symmetric on $[-c, c]$. We extend I_c to $\mathscr{P}_\varepsilon'$ by defining, for $G \in \mathscr{P}_\varepsilon'$,

$$I_c(G) = \sup\{[(\int \psi' \, dG)^2 / \int \psi^2 dG] : \psi \in \Psi_c'\} . \tag{3.12}$$

Since $\mathscr{P}_\varepsilon'$ is convex and vaguely compact, and $\inf\{I_c(G) : G \in \mathscr{P}_\varepsilon\} < \infty$, there is a unique $G^* \in \mathscr{P}_\varepsilon'$ minimizing $I_c(G)$ by Huber's Theorem 4. By Huber's Theorem 3, we necessarily have $G^* \in \mathscr{P}_\varepsilon$.

We now show that there exists x_0 and x_1, where $0 < x_0 < c$, $x_0 < x_1$, satisfying (3.7) and (3.8). For a fixed $x_0 \in (0, c)$, x_0/x is a strictly decreasing map of $[x_0, \infty)$ onto $(0, 1]$ and $\tanh[\frac{1}{2}x(c - x_0)]$ is a strictly increasing map of $[x_0, \infty)$ onto $[\tanh \frac{1}{2}[x_0(c - x_0)], 1)$. Hence there is a unique $x_1 > x_0$ satisfying $x_0/x_1 = \tanh[\frac{1}{2}x_1(c - x_0)]$. Now denote the left hand side of (3.8) by $J(x_0)$, where x_1 is determined from x_0 by (3.7) and g_0 is determined by (3.6). Then $J(x_0)$ is a continuous function of x_0 on $(0, c)$, $J(x_0) \to 0$ as $x_0 \to c$ and $J(x_0) \to 2(1-\varepsilon)c\varphi(0) - (1-\varepsilon)\int_{-c}^{c} \varphi(x)\, dx$ as $x_0 \to 0$. Thus by the inequality of the hypothesis, there is an $x_0 \in (0, c)$ satifying $J(x_0) = \varepsilon$.

By (3.6) and (3.7), g_0, g_0' and ψ_0 are continuous at $x = x_0$. Also it is easily checked that $g_0(x) - (1-\varepsilon)\varphi(x) \geqq 0$ on $[-c, c]$, so that h_0, defined by

$$g_0(x) = (1-\varepsilon)\varphi(x) + \varepsilon h_0 , \qquad x \in [-c, c] , \tag{3.13}$$

satisfies $h_0(x) \geqq 0$ on $[-c, c]$. By (3.8), $\int_{-c}^{c} h_0(x)\, dx = 1$. So $G_0 \in \mathscr{P}_\varepsilon$. Also, since $\psi_0(c) = 0$, we have $\psi_0 \in \Psi_c'$.

To show that the unique $G \in \mathscr{P}_\varepsilon$ minimizing $I_c(G)$ is G_0, it is sufficient by Huber's Theorem 2 to show that

$$\int_0^c (2\psi_0'(x) - \psi_0^2(x))(g_1(x) - g_0(x))\, dx \geqq 0 , \tag{3.14}$$

where g_1 is the density on $[0, c]$ of any $G_1 \in \mathscr{P}_\varepsilon$. Since $x_1 > x_0$, we have

$$x_1^2 + 2\psi_0'(x) - \psi_0^2(x) \geqq 0 , \qquad x \in [0, x_0] . \tag{3.15}$$

Also

$$x_1^2 + 2\psi_0'(x) - \psi_0^2(x) = 0 , \qquad x \in [x_0, c] . \tag{3.16}$$

Since $g_1 \geqq g_0$ in the interval $[0, x_0]$, and since

$$\int_0^c g_1(x)\, dx \leqq (1-\varepsilon)\int_0^c \varphi(x)\, dx + (\varepsilon/2) = \int_0^c g_0(x)\, dx , \tag{3.17}$$

so that $\int_0^c (g_1 - g_0)\, dx \leqq 0$, it follows that

$$\begin{aligned} &\int_0^c (2\psi_0' - \psi_0^2)(g_1 - g_0)\, dx \\ &\quad = \int_0^{x_0} (x_1^2 + 2\psi_0' - \psi_0^2)(g_1 - g_0)\, dx - x_1^2 \int_0^c (g_1 - g_0)\, dx \geqq 0 . \end{aligned} \tag{3.18}$$

This proves (i), and (ii) follows immediately from Huber's Theorem 2. The minimum value of $\sup\{V(\psi, g) : G \in \mathscr{P}_\varepsilon\}$ is

$$(3.19)\quad \begin{aligned}\sup\{V(\psi_0, g) &: G \in \mathscr{P}_\varepsilon\}\\ &= V(\psi_0, g_0) = 1/I_c(G_0)\\ &= 1\Big/\Big[(1-\varepsilon)\Big(2\Phi(x_0) - 1 - 2x_0\varphi(x_0)\\ &\quad + x_1\varphi(x_0)\frac{\sinh[x_1(c-x_0)] - x_1(c-x_0)}{\cosh^2[\frac{1}{2}x_1(c-x_0)]}\Big)\Big].\end{aligned} \quad \square$$

REMARK 3.1. I have learned of the prior discovery of the above ψ_0 through personal communication with P. Bickel and F. Hampel. In 1972 Hampel found ψ_0 as the solution to a different optimality problem (see Hampel (1973), page 98), and P. Huber found later in the same year that ψ_0 solves the minimax problem.

REMARK 3.2. If the normal density φ on $[-c, c]$ in the formulation of the problem is replaced by a known strongly unimodal density f on $[-c, c]$, then we obtain minimax solutions of the form

$$(3.20)\quad \begin{aligned}\psi_0(x) &= -f'(x)/f(x) & |x| &\leqq x_0\\ &= x_1\tanh[\tfrac{1}{2}x_1(c-|x|)]\operatorname{sgn}(x) & x_0 \leqq |x| &\leqq c\\ &= 0 & |x| &\geqq c\end{aligned}$$

where

$$(3.21)\quad -f'(x_0)/f(x_0) = x_1\tanh[\tfrac{1}{2}x_1(c-x_0)].$$

REMARK 3.3. Define $\varepsilon(c)$ by

$$(3.22)\quad \varepsilon(c)/(1-\varepsilon(c)) = 2c\varphi(0) - 2\Phi(c) - 1.$$

If $\varepsilon < \varepsilon(c)$, then part (2) of the theorem gives the form of all minimax solutions. If $\varepsilon \geqq \varepsilon(c)$, then by part (1) all ψ's in Ψ_c' are minimax (in a degenerate way). Table 1 gives some values of α with the corresponding values of the parameters d, k, c and $\varepsilon(c)$. Values of $V(\psi^*, \varphi)$ of Lemma 3.1 are also tabled.

TABLE 1

Values of parameters corresponding to selected values of the tail probability α

α	$d = \Phi^{-1}(1-\alpha/2)$	$k = \Phi^{-1}(\frac{1}{2}+\alpha/2)$	$c = d - k$	$V(\psi^*, \varphi) = 1/\int_{-c}^{c} x^2\varphi(x)\,dx$	$\varepsilon(c)$ (see formula (3.22))
.001	3.2905	.0012	3.2893	1.0129	.6191
.005	2.8070	.0063	2.8008	1.0519	.5535
.01	2.5758	.0125	2.5633	1.0952	.5135
.02	2.3264	.0251	2.3013	1.1784	.4617
.05	1.9600	.0627	1.8973	1.4452	.3637
.10	1.6449	.1257	1.5192	2.0450	.2542
.15	1.4395	.1891	1.2504	3.0092	.1728
.20	1.2816	.2534	1.0282	4.7040	.1105
.30	1.0364	.3853	0.6511	15.4445	.0333

REMARK 3.4. Table 2 gives minimax solutions for some selected cases. For selected values of x_0, corresponding values of x_1 and ε were obtained by solving (3.7) and (3.8). The latter can be rewritten as

$$(3.23)\qquad \frac{\varepsilon}{1-\varepsilon} = \frac{\varphi(x_0)}{x_1 \cosh^2[\frac{1}{2}x_1(c-x_0)]}[\sinh(x_1(c-x_0)) + x_1(c-x_0)] - 2\Phi(c) + 2\Phi(x_0)\,.$$

TABLE 2
Minimax solutions for selected cases

$\alpha = .001,\ c = 3.2893$				$\alpha = .01,\ c = 2.5633$				$\alpha = .10,\ c = 1.5192$			
x_0	x_1	ε	$V(\psi_0, g_0)$	x_0	x_1	ε	$V(\psi_0, g_0)$	x_0	x_1	ε	$V(\psi_0, g_0)$
0^+	0^+	.6191	∞	0^+	0^+	.5135	∞	0^+	0^+	.2542	∞
0.4	0.5826	.4261	14.5201	0.4	0.6556	.3363	15.7132	0.2	.5631	.1879	78.8818
0.8	0.9610	.2311	3.4978	0.8	1.0800	.1767	3.9508	0.4	.8783	.1290	19.5185
1.2	1.3515	.1006	1.8138	1.2	1.5368	.0748	2.0555	0.6	1.1978	.0812	8.8998
1.6	1.7696	.0375	1.3263	1.6	2.0921	.0256	1.4739	0.8	1.5669	.0456	5.2899
2.0	2.2369	.0125	1.1447	2.0	2.9422	.0061	1.2378	1.0	2.0517	.0217	3.6576
2.4	2.8240	.0036	1.0683	2.4	5.6052	.0004	1.1243	1.2	2.8327	.0075	2.7852
2.8	3.8209	.0007	1.0331	2.5633^-	—	0	1.0952	1.4	4.9151	.0010	2.2630
3.2893^-	—	0	1.0129					1.5192^-	—	0	2.0450

REMARK 3.5. To give a meaningful statistical interpretation to the formal result of the theorem, the domain of ψ and G must be restricted so that

$$(3.24)\qquad n^{\frac{1}{2}}T_n \to_{\mathscr{L}} N(0, V(\psi, g))$$

is satisfied. We first note that, given $\delta > 0$, there exists $\psi_\delta \in \Psi_c$ such that $\sup\{V(\psi_\delta, g): G \in \mathscr{P}_\varepsilon\} < V(\psi_0, g_0) + \delta$, so that if we restrict ψ to the class Ψ_c, δ-minimax solutions can be found. Next note that if $\psi \in \Psi_c$ and $G \in \mathscr{P}_\varepsilon$ satisfy $T_n \to_P 0$ and $P[T_n = M_n] \to 0$, then (3.24) holds. For $\psi \in \Psi_c$ and $G \in \mathscr{P}_\varepsilon$, define $\lambda_G(t) = \int \psi(x-t)\,dG(x)$. To show $T_n \to_P 0$ and $P[T_n = M_n] \to 0$, it suffices to show that $\lambda_G(t)$ satisfies (i)—(iv) of Lemma 2.1. If we change the definition of k from $k = \Phi^{-1}(\frac{1}{2} + \alpha/2)$ to $k = \Phi^{-1}(1/(2(1-\varepsilon)) + \alpha/2)$, then (i) and (ii) are satisfied: i.e., $\lambda_G(t) = \int_{-c+t}^{c-t} \psi(x-t)g(x)\,dx$ and $\lambda_G(t) = -\lambda_G(-t)$ when $|t| \leqq k$, $\psi \in \Psi_c$ and $G \in \mathscr{P}_\varepsilon$. Conditions (iii) and (iv) can be checked for particular pairs (ψ, G), but there appears to be no simple characterization of the subset of $\Psi_c \times \mathscr{P}_\varepsilon$ for which (iii) and (iv) hold.

An alternative criterion for a robust estimator is due to Hampel (1968): minimize $V(\psi, \varphi)$ subject to an upper bound on

$$(3.25)\qquad B_0(\psi) = \sup_x \frac{|\psi(x)|}{|\int_{-c}^{c} \psi(t)\varphi'(t)\,dt|}\,.$$

Hampel proved that in the class of continuous skew symmetric ψ's, the solution has the form

$$\psi(x) = x \qquad |x| \leqq K$$
$$= K \operatorname{sgn}(x) \quad |x| > K, \tag{3.26}$$

where K is determined by the upper bound on $B_0(\psi)$. $B_0(\psi)$ is called the "sensitivity" of the M-estimator: a complete rationale for the definition is given in Hampel (1968). For the restricted class Ψ_c', it seems reasonable to require a measure of sensitivity B to satisfy:

(i) $0 < B(\psi) < \infty$ for all $\psi \in \Psi_c'$.
(ii) $B(K\psi) = B(\psi)$ if $\psi \in \Psi_c'$, $K \neq 0$.
(iii) If $\{\psi_j\}$ is a sequence in Ψ_c' satisfying $\psi_j(x) \to \psi^*(x)$ a.e., then $B(\psi_j) \to \infty$.

We consider two measures of sensitivity satisfying (i), (ii) and (iii):

$$B_1(\psi) = \sup \frac{|\psi'(x)|}{|\int_{-c}^{c} \psi(t)\varphi'(t)\,dt|}, \tag{3.27}$$

where the supremum is taken over all x for which ψ' exists; and

$$B_2(\psi) = \frac{\int_{-c}^{c} [\psi'(x)]^2\varphi(x)\,dx}{[\int_{-c}^{c} \psi(x)\varphi'(x)\,dx]^2}. \tag{3.28}$$

For members of the class Ψ_c', $B_1(\psi)$ coincides with Hampel's "local shift sensitivity" (Hampel (1973), page 98; (1974), page 389).

We define the class of "Hamdels" (Andrews et al. (1972)): for $0 < a \leqq b < c$, define

$$\begin{aligned} \psi_{a,b,c}(x) &= x & 0 \leqq |x| < a \\ &= \operatorname{sgn}(x) \cdot a & a \leqq |x| < b \\ &= \operatorname{sgn}(x) \cdot \frac{c - |x|}{c - b}\, a & b \leqq |x| < c \\ &= 0 & |x| \geqq c. \end{aligned} \tag{3.29}$$

THEOREM 3.2. *In the class* Ψ_c':

(i) $\psi_{c/2,c/2,c}$ *minimizes* $B_1(\psi)$.

(ii) *If* $K \geqq B_1(\psi_{c/2,c/2,c}) = 1/(4\Phi(c/2) - 2\Phi(c) - 1)$, *then there is an* $a \in [c/2, c)$ *such that* $B_1(\psi_{a,a,c}) = K$, *and* $\psi_{a,a,c}$ *minimizes* $V(\psi, \varphi)$ *subject to* $B_1(\psi) \leqq K$.

PROOF. Let a satisfy $c/2 < a < c$. To show that $\psi_{a,a,c}$ minimizes $V(\psi, \varphi)$ subject to $B_1(\psi) \leqq B_1(\psi_{a,a,c})$, it suffices to show that $\psi_{a,a,c}$ minimizes $\int_{-c}^{c} \psi^2\varphi$ subject to

$$\int_{-c}^{c} \psi\varphi' = \int_{-c}^{c} \psi_{a,a,c}\varphi' \tag{3.30}$$

and

$$\sup |\psi'(x)| \leqq \sup |\psi'_{a,a,c}(x)| = a/(c - a). \tag{3.31}$$

If $\psi \in \Psi_c'$ satisfies (3.30) and (3.31), then

$$\begin{aligned}(3.32)\qquad &\textstyle\int_{-c}^{c} (x - \psi(x))^2\varphi(x)\,dx \\ &= \textstyle\int_{-c}^{c} x^2\varphi(x)\,dx + \int_{-c}^{c} \psi^2(x)\varphi(x)\,dx - 2\int_{-c}^{c} x\psi(x)\varphi(x)\,dx \\ &= \textstyle\int_{-c}^{c} x^2\varphi(x)\,dx + \int_{-c}^{c} \psi^2(x)\varphi(x)\,dx + 2\int_{-c}^{c} \psi_{a,a,c}(x)\varphi'(x)\,dx\,,\end{aligned}$$

so that it is sufficient to show that $\psi_{a,a,c}$ minimizes $\int_{-c}^{c} (x - \psi(x))^2\varphi(x)\,dx$ subject to $\psi \in \Psi_c'$ satisfying (3.30) and (3.31). (The above argument was taken from Hampel (1968).)

Suppose there is a $\psi_0 \in \Psi_c'$ satisfying (3.30) and (3.31) and $\int_{-c}^{c} (x - \psi_0(x))^2\varphi(x)\,dx < \int_{-c}^{c} (x - \psi_{a,a,c}(x))^2\varphi(x)\,dx$. Then there is a point $x_0 \in (a, c)$ such that $(x_0 - \psi_0(x_0))^2 < (x_0 - \psi_{a,a,c}(x_0))^2$. Since $x_0 > \psi_{a,a,c}(x_0) > 0$, we must have $\psi_0(x_0) > \psi_{a,a,c}(x_0)$. Let $x_1 = \inf\{x : x > x_0 \text{ and } \psi_0(x) \leqq \psi_{a,a,c}(x)\}$. Since ψ_0 is continuous and $\psi_0(c) = 0$, we have $x_0 < x_1 \leqq c$ and $\psi_0(x_1) = \psi_{a,a,c}(x_1)$. Since ψ_0' is piecewise continuous, there is a point $x_2 \in (x_0, x_1)$ such that $\psi_0(x_2) > \psi_{a,a,c}(x_2)$ and ψ_0' is continuous on $[x_2, x_1)$. By the mean value theorem, there is a point $x_3 \in (x_2, x_1)$ such that

$$(3.33)\qquad \psi_0'(x_3) = \frac{\psi_0(x_1) - \psi_0(x_2)}{x_1 - x_2} = -\frac{\psi_0(x_2) - \psi_{a,a,c}(x_1)}{x_1 - x_2} < -\frac{a}{c - a}\,.$$

This contradicts (3.31). So when $a \in (c/2, c)$, $\psi_{a,a,c}$ minimizes $V(\psi, \varphi)$ subject to $B_1(\psi) \leqq B_1(\psi_{a,a,c})$.

To show (i), it suffices to show that $\psi_{c/2,c/2,c}$ minimizes $\sup|\psi'(x)|$ subject to

$$(3.34)\qquad \textstyle\int_{-c}^{c} \psi\varphi' = \int_{-c}^{c} \psi_{c/2,c/2,c}\varphi'\,.$$

Suppose $\psi_0 \in \Psi_c'$ satisfies (3.34) and that $\sup|\psi_0'(x)| < \sup|\psi_{c/2,c/2,c}'(x)| = 1$. Then by the same type of argument as above, $\psi_0(x) < \min\{x, c - x\} = \psi_{c/2,c/2,c}(x)$ for $x \in (0, c)$. Hence $\int_{-c}^{c} x\psi_0(x)\varphi(x)\,dx < \int_{-c}^{c} x\psi_{c/2,c/2,c}(x)\varphi(x)\,dx$, contradicting (3.34).

To complete the proof of (ii), note that $B_1(\psi_{a,a,c})$ is a continuous function of a, and that $B_1(\psi_{a,a,c}) \to \infty$ as $a \to c$. Hence for any $K \geqq B_1(\psi_{c/2,c/2,c})$, there is a point $a \in [c/2, c)$ such that $B_1(\psi_{a,a,c}) = K$. □

REMARK 3.6. A generalization of the above proof yields the following: If f is a symmetric strongly unimodal density, and if f'/f has a piecewise continuous derivative on $[-c, c]$, then

$$(3.35)\qquad \begin{aligned}\psi_a(x) &= -f'(x)/f(x) && |x| \leqq a \\ &= \frac{-f'(a)}{f(a)}\,\frac{c - |x|}{c - a}\,\operatorname{sgn}(x) && a \leqq |x| \leqq c \\ &= 0 && |x| \geqq c\end{aligned}$$

minimizes $V(\psi, f)$ subject to $\psi \in \Psi_c'$ and

$$(3.36)\qquad \sup\frac{|\psi'(x)|}{|\int_{-c}^{c} \psi(t)f'(t)\,dt|} \leqq \sup\frac{|\psi_a'(x)|}{|\int_{-c}^{c} \psi_a(t)f'(t)\,dt|}\,,$$

provided that $a \in (0, c)$ is large enough so that

$$\sup |\psi_a'(x)| = \frac{-f'(a)}{f(a)} \frac{1}{c-a}. \tag{3.37}$$

REMARK 3.7. Hampels $\psi_{a,b,c}$ with $a < b$ can be obtained as solutions if we also impose an upper bound on $\sup |\psi(x)|/|\int_{-c}^{c} \psi\varphi'|$. Specifically, if $0 < a \leqq b < c$ and $a/(c-b) \geqq 1$, then $\psi_{a,b,c}$ minimizes $V(\psi, \varphi)$ subject to $\psi \in \Psi_c'$ and

$$\sup \frac{|\psi(x)|}{|\int_{-c}^{c} \psi\varphi'|} \leqq \frac{a}{|\int_{-c}^{c} \psi_{a,b,c}\varphi'|} \tag{3.38}$$

and

$$\sup \frac{|\psi'(x)|}{|\int_{-c}^{c} \psi\varphi'|} \leqq \frac{a/(c-b)}{|\int_{-c}^{c} \psi_{a,b,c}\varphi'|}. \tag{3.39}$$

To minimize $V(\psi, \varphi)$ subject to $\psi \in \Psi_c'$ and $B_2(\psi) \leqq K$, it suffices to minimize $\int_0^c \psi^2\varphi$ subject to $\psi \in \Psi_c'$, $\int_0^c \psi\varphi' = -1$, and $\int_0^c [\psi']^2\varphi \leqq K$. Using Lagrange multipliers λ_1 and λ_2, we obtain the Euler equation

$$2\psi(x)\varphi(x) + \lambda_1\varphi'(x) - 2\lambda_2[\psi'(x)\varphi'(x) + \psi''(x)\varphi(x)] = 0, \tag{3.40}$$

which, after substituting $-x\varphi(x) = \varphi'(x)$ and $t = 1/\lambda_2$, becomes

$$\psi''(x) - x\psi'(x) - t\psi(x) = -\frac{t\lambda_1}{2} x. \tag{3.41}$$

The solution to (3.41) subject to $\psi \in \Psi_c'$ has the form $C\psi_t$, where $C \neq 0$ and

$$\begin{aligned} \psi_t(x) &= x - \frac{c}{\phi_t(c)} \phi_t(|x|) \operatorname{sgn}(x) \quad & |x| \leqq c \\ &= 0 & |x| > c \end{aligned} \tag{3.42}$$

where

$$\phi_t(x) = \sum_{m=1}^{\infty} \frac{(1+t)(3+t)\cdots(2m-1+t)}{(2m+1)!} x^{2m+1} \tag{3.43}$$

A complete proof of the theorem below is found in Collins (1973).

THEOREM 3.3. *In the class* Ψ_c':

(i) ψ_0 *minimizes* $B_2(\psi)$.

(ii) *If* $B_2(\psi_0) < K < \infty$, *then there is a* $t > 0$ *such that* $B_2(\psi_t) = K$, *and* ψ_t *minimizes* $V(\psi, \varphi)$ *subject to* $B_2(\psi) \leqq K$.

4. Extension to the case of unknown scale. Let $X_1, \cdots, X_n$ be i.i.d. random variables with distribution function $F((x-\theta)/\sigma)$, where $F \in \mathscr{F}$ and θ and σ are unknown. The problem is to estimate θ, and we assume without loss of generality that $\theta = 0$.

Let F_n be the empirical distribution function, and define

$$\hat{\sigma}_n = \frac{F_n^{-1}(1-\alpha) - F_n^{-1}(\alpha)}{\Phi^{-1}(1-\alpha) - \Phi^{-1}(\alpha)} \tag{4.1}$$

where

$$F_n^{-1}(t) = \inf\{x : F_n(x) \geqq t\}, \qquad 0 < t < 1. \tag{4.2}$$

Also let

$$b = \frac{\Phi^{-1}(1 - (\alpha/2)) - \Phi^{-1}(3\alpha/2)}{\Phi^{-1}(1 - \alpha) - \Phi^{-1}(\alpha)}. \tag{4.3}$$

LEMMA 4.1. *There is a number* $\beta \in [1, b)$, *depending on* $F \in \mathscr{F}$, *such that*

$$\hat{\sigma}_n \to_P \beta\sigma. \tag{4.4}$$

PROOF. There is a $\gamma \in (-\alpha/2, \alpha/2)$ such that $F(x) = -\gamma + \Phi(x)$ whenever $\Phi(x) \in [\alpha/2, \alpha/2]$. So $F^{-1}(y) = \Phi^{-1}(y + \gamma)$ if $y + \gamma \in [\alpha/2, 1 - \alpha/2]$. In particular

$$F^{-1}(1 - \alpha) - F^{-1}(\alpha) = \Phi^{-1}(1 - \alpha + \gamma) - \Phi^{-1}(\alpha + \gamma). \tag{4.5}$$

So

$$\hat{\sigma}_n \to_P \sigma \frac{F^{-1}(1 - \alpha) - F^{-1}(\alpha)}{\Phi^{-1}(1 - \alpha) - \Phi^{-1}(\alpha)} = \beta_\gamma \sigma, \tag{4.6}$$

where

$$\beta_\gamma = \frac{\Phi^{-1}(1 - \alpha + \gamma) - \Phi^{-1}(\alpha + \gamma)}{\Phi^{-1}(1 - \alpha) - \Phi^{-1}(\alpha)}. \tag{4.7}$$

Note that $\beta_0 = 1$, $\beta_{-\gamma} = \beta_\gamma$ for $\gamma \in (-\alpha/2, \alpha/2)$, and that $\sup\{\beta_\gamma : \gamma \in (-\alpha/2, \alpha/2)\} = \beta_{\alpha/2} = b$, since $\varphi(x)$ is symmetric about 0 and strictly decreasing in x when $x \geqq 0$. □

We remark that the biasing factor b is typically small: e.g., when $\alpha = .10$, $b = 1.046$.

Define

$$c' = (d - k)/b, \tag{4.8}$$

and define $\Psi_{c'}$ by replacing c by c' in the definition of Ψ_c. For $\psi \in \Psi_{c'}$, we define the estimator $T_{n,\hat{\sigma}_n}$ by:

$$\begin{aligned} T_{n,\hat{\sigma}_n} &= \lim_{j\to\infty} \hat{\theta}_j \qquad \text{if } \lim_{j\to\infty} \hat{\theta}_j \text{ exists} \\ &= M_n \qquad \text{otherwise} \end{aligned} \tag{4.9}$$

where $\hat{\theta}_0 = M_n$ and

$$\hat{\theta}_{j+1} = \hat{\theta}_j + \hat{\sigma}_n \cdot \frac{\sum_{i=1}^n \psi\left(\frac{X_i - \hat{\theta}_j}{\hat{\sigma}_n}\right)}{\sum_{i=1}^n \psi'\left(\frac{X_i - \hat{\theta}_j}{\hat{\sigma}_n}\right)}, \qquad j = 0, 1, 2, \cdots \tag{4.10}$$

THEOREM 4.1. *Let* $\psi \in \Psi_{c'}$, $F \in \mathscr{F}$, $\sigma > 0$, *and let* $X_1, X_2, \cdots$ *be* i.i.d. *with* df $F(x/\sigma)$. *Then* $T_{n,\hat{\sigma}_n} \to_P 0$.

PROOF. Without loss of generality, assume $\sigma = 1$. There is a $\beta \in [1, b)$ such that $\hat{\sigma}_n \to_P \beta$. Define

$$\lambda_\beta(t) = \int \psi\left(\frac{x-t}{\beta}\right) dF(x) \tag{4.11}$$

and

$$\lambda_{n,\hat{\sigma}_n}(t) = \frac{1}{n} \sum_{i=1}^n \psi\left(\frac{X_i - t}{\hat{\sigma}_n}\right). \tag{4.12}$$

If $|t| \leqq k$, then

$$\lambda_\beta(t) = \int_{\{x:|(x-t)/\beta| \leqq c'\}} \psi\left(\frac{x-t}{\beta}\right) \varphi(x)\, dx, \tag{4.13}$$

since $|t| \leqq k$ and $|(x-t)/\beta| \leqq c'$ imply $|x| \leqq d$. It is easily seen that λ_β also satisfies properties (ii), (iii) and (iv) of Lemma 2.1.

By the argument of Lemma 2.2 (i),

$$\sup\{|\lambda_{n,\beta}(t) - \lambda_\beta(t)| : |t| \leqq k\} \to_P 0, \tag{4.14}$$

where

$$\lambda_{n,\beta}(t) = \frac{1}{n} \sum_{i=1}^n \psi\left(\frac{X_i - t}{\beta}\right). \tag{4.15}$$

To obtain the desired analogue of Lemma 2.2 (i), namely

$$\sup\{|\lambda_{n,\hat{\sigma}_n}(t) - \lambda_\beta(t)| : |t| \leqq k\} \to_P 0, \tag{4.16}$$

it suffices to show that

$$\sup\{|\lambda_{n,\hat{\sigma}_n}(t) - \lambda_{n,\beta}(t)| : |t| \leqq k\} \to_P 0. \tag{4.17}$$

But this follows easily, since $\hat{\sigma}_n \to_P \beta$ and ψ is continuous and vanishes outside $[-c', c']$. The analogue of Lemma 2.2 (ii) similarly follows.

The rest of the proof now follows that of Theorem 2.1 with λ_n, λ and T_n replaced by $\lambda_{n,\hat{\sigma}_n}$, λ_β and $T_{n,\hat{\sigma}_n}$, respectively. □

By a standard type of argument using a Taylor expansion of $\sum \psi((X_i - T_{n,\hat{\sigma}_n})/\hat{\sigma}_n)$ about $T_{n,\hat{\sigma}_n} = 0$ and $\hat{\sigma}_n = \beta$, asymptotic normality of $n^{\frac{1}{2}} T_{n,\hat{\sigma}_n}$ is obtained. The proof is found in Collins (1973).

THEOREM 4.2. *Under the assumptions of Theorem* 4.1,

$$n^{\frac{1}{2}} T_{n,\hat{\sigma}_n} \to_{\mathscr{D}} N(0, \sigma^2 V_\beta(\psi, \varphi)), \tag{4.18}$$

where

$$V_\beta(\psi, \varphi) = \frac{\int_{-c'}^{c'} \psi^2(y)\varphi(\beta y)\, dy}{\beta[\int_{-c'}^{c'} \psi(y)\varphi'(\beta y)\, dy]^2} \tag{4.19}$$

and where $\beta = \beta(F)$ *is the number in* $[1, b]$ *that satisfies*

$$\hat{\sigma}_n \to_P \beta\sigma. \tag{4.20}$$

We remark that, although β is unknown, $V_\beta(\psi, \varphi)$ changes very little as β ranges over the set $[1, b)$ when α is reasonably small. If we replace $\{V(\psi, \varphi) : \psi \in \Psi_c'\}$ in the scale-known theory by $\{V_1(\psi, \varphi) : \psi \in \Psi_{c'}'\}$, then the optimal ψ's found in Section 3, with c replaced by c', determine optimal estimators $T_{n,\hat{\sigma}_n}(\psi)$ in the scale unknown case.

Acknowledgment. I would like to thank Professor Peter J. Bickel for his generous guidance and encouragement.

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 Univ. Press.

[2] BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

[3] COLLINS, J. R. (1973). Robust estimation of a location parameter in the presence of asymmetry. Ph. D. dissertation, Univ. of California, Berkeley.

[4] HAMPEL, F. R. (1968). Contributions to the theory of robust estimation. Ph. D. dissertation, Univ. of California, Berkeley.

[5] HAMPEL, F. R. (1973). Robust estimation: a condensed partial survey. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **27** 87-104.

[6] HAMPEL, F. R. (1974). The influence curve and its role in robust estimation. *J. Amer. Statist. Assoc.* **69** 383-393.

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

[8] HUBER, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **1** 221-233. Univ. of California Press.

DEPARTMENT OF STATISTICS
PURDUE UNIVERSITY
MATHEMATICAL SCIENCES BUILDING
WEST LAFAYETTE, INDIANA 47907