The Annals of Statistics
1978, Vol. 6, No. 2, 292–313

AN EFFICIENT AND ROBUST ADAPTIVE ESTIMATOR OF LOCATION[1]

BY RUDOLF BERAN

University of California, Berkeley

A nonparametric minimum Hellinger distance estimator of location is introduced and shown to be asymptotically efficient at every symmetric density with finite Fisher information. Under small, possibly asymmetric, perturbations in such a density, the estimator is asymptotically robust in a technical sense which extends Hájek's concept of "regularity." A numerical example illustrates the computational feasibility of the estimator and its resistance to an arbitrary single outlier.

1. Introduction. Random variables $X_1, X_2, \cdots, X_n$ are observed. As a model for the data, it is postulated that the $\{X_i\}$ are independent identically distributed with density belonging to the location family $\{f(x-\theta): -\infty < \theta < \infty\}$, f symmetric about zero and absolutely continuous with $I(f) = \int [f'(x)]^2/f(x)\,dx < \infty$. Apart from the requirements of symmetry and finite $I(f)$, f is unspecified. An estimator $\hat{\theta}_n$ of θ is sought that possesses two desirable properties:

(1) $\hat{\theta}_n$ is asymptotically efficient under the model. Technically,

$$n^{\frac{1}{2}}(\hat{\theta}_n - \theta) = n^{-\frac{1}{2}} I^{-1}(f) \sum_{i=1}^{n} -f'(X_i - \theta)/f(X_i - \theta) + o_p(1)$$

under every symmetric density $f(x-\theta)$ belonging to the model.

(2) $\hat{\theta}_n$ is robust under small departures from the model. The distribution of $\hat{\theta}_n$ does not change much if the distribution of each X_i is deformed from an initial symmetric shape into an arbitrary nearby shape.

Historically, the notion of a robust location estimator has been defined in several different ways, one of which amounts to property (1) above (see Huber (1972) for a review of robustness definitions). However, since property (1) says nothing about the behavior of $\hat{\theta}_n$ under even slightly asymmetric distributions, we view (1) as an asymptotic efficiency property under the model of symmetry rather than as a robustness property. Our preferred robustness concept, property (2), is a local stability in distribution requirement first formalized by Hampel (1971). The aim here is to guard against substantially wrong inferences based upon $\hat{\theta}_n$ if the assumed model is not quite correct. The mathematical formulation of robustness used in this paper is a strengthened version of Hájek's (1970) concept of regularity.

Received August 1976; revised January 1977.

[1] This research was supported in part by National Science Foundation Grant MPS 75-10376 and was completed while the author was visiting the University of Melbourne.

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

Key words and phrases. Minimum Hellinger distance, adaptive location estimator, asymptotically efficient, robust estimator, nonparametric estimator, contiguity.

Estimators of θ which satisfy requirement (1) have been the goal of several investigations since Stein (1956) suggested the possibility of constructing such fully efficient adaptive estimators. Papers concerned with this specific problem include van Eeden (1970), Weiss and Wolfowitz (1970), Takeuchi (1971), Fabian (1973), Beran (1974), Johns (1974), Sacks (1975), and Stone (1975). The analogous two-sample location shift problem, which does not require symmetry of the data distribution, has been treated by Bhattacharya (1967), Weiss and Wolfowitz (1970), Beran (1974), Samanta (1974), and the two-sample scale change problem has been studied by Wolfowitz (1974).

The first group of papers cited vary considerably in their assumptions and, to a lesser extent, in their results. Most of the constructions use additional assumptions on the density beyond those needed to state the efficiency results. Two of the papers (Beran (1974), Stone (1975)) have shown that there exist location estimators which possess property (1) without any further regularity assumptions on f. Moreover, Stone proved for his estimator the desirable additional property that $n^{\frac{1}{2}}(\hat{\theta}_n - \theta) \to_p 0$ under every symmetric distribution that does not have finite Fisher information.

Little is known theoretically about the robustness or distributions of one-sample adaptive location estimators under asymmetric distributions. A priori, it seems likely that the efficiency property (1) need not entail the robustness property (2). Huber (1974) has commented on the possible nonconvergence of the Jaeckel (1971 b) adaptive estimator under certain asymmetric distributions. However, this difficulty can be avoided with many of the adaptive estimators cited above.

Any good location estimator, whether adaptive or not, can be expressed as a sample-size dependent functional of the empirical cdf, this being a sufficient statistic. Such an estimator's rate of convergence to its asymptotics will depend upon the nature of the corresponding functional sequence and upon the nature of the actual distribution; however, this dependence has not been analyzed to date. Monte Carlo results, such as those given by Stone (1975), suggest that the distribution of a well-designed adaptive location estimator could converge fairly quickly to its limiting form over an interesting subset of the set of all symmetric densities.

This paper presents a new, intuitively appealing, adaptive estimator $\hat{\theta}_n$ which possesses property (1) without further assumptions on the density. Moreover, asymptotic calculations indicate that $\hat{\theta}_n$ is robust under small perturbations of the density. The results of these calculations include the asymptotic distributions of $\hat{\theta}_n$ under fixed, possibly asymmetric, densities as well as the asymptotic distributions of $\hat{\theta}_n$ under sequences of fairly arbitrary densities contiguous to a given symmetric shape. A numerical example at the end of the paper illustrates the computational feasibility of the estimator and its resistance to an arbitrary single outlier.

Let $\hat{g}_n(x)$ denote a suitable estimator of the density of X_i (a precise definition of $\hat{g}_n$ will be given in Section 2). The location estimator $\hat{\theta}_n$ is defined as that value (or values) which minimizes over all real t the Hellinger distance between $\hat{g}_n(x)$ and $\hat{g}_n(-x+2t)$. Note that $\hat{g}_n(-x+2t)$ is simply the reflection of $\hat{g}_n(x)$ about the point t on the real line and that the square of the Hellinger distance in question is $\int [\hat{g}_n^{\frac{1}{2}}(-x+2t) - \hat{g}_n^{\frac{1}{2}}(x)]^2\,dx$. Equivalently, $\hat{\theta}_n$ maximizes over all real t the inner product $\int \hat{g}_n^{\frac{1}{2}}(-x+2t)\hat{g}_n^{\frac{1}{2}}(x)\,dx$.

The significance of the Hellinger metric in the definition of $\hat{\theta}_n$ may be appreciated by considering the location estimator θ_n^* defined analogously in terms of the Kolmogorov metric: θ_n^* minimizes $\sup_x |1 - \hat{G}_n((-x+2t)^-) - \hat{G}_n(x)|$ over all real t, where $\hat{G}_n$ is the empirical cdf. The limiting distribution of $n^{\frac{1}{2}}(\theta_n^* - \theta)$ under a symmetric density, established in implicit form by Rao, Schuster and Littel (1975), is complicated and differs from that of $n^{\frac{1}{2}}(\hat{\theta}_n - \theta)$; unlike $\hat{\theta}_n$, θ_n^* is not asymptotically efficient under symmetric densities with finite Fisher information.

There is an interesting link between the adaptive estimator $\hat{\theta}_n$ and the class of parametric estimators considered in Beran (1977b). Let $\{g_\beta : \beta \in B \subset R^k\}$ denote a parametric family of densities, the functional form of g_β being known. The minimum Hellinger distance estimator $\hat{\beta}_n$ of β is defined as that value (or values) of s which minimizes $\int [g_s^{\frac{1}{2}}(x) - \hat{g}_n^{\frac{1}{2}}(x)]^2\,dx$ over all $s \in B$, $\hat{g}_n(x)$ being a density estimator. This parametric procedure admits formal extensions to models having a countable infinity of unknown real parameters. For instance, let the parameter space B consist of all pairs $\beta = (\theta, f)$, where $-\infty < \theta < \infty$ and f is a density symmetric about zero. Define the family $\{g_\beta : \beta \in B\}$ through the requirement $g_\beta(x) \equiv f(x-\theta)$ when $\beta = (\theta, f)$. It is easily checked that the choices of real θ and symmetric density f which minimize the Hellinger distance between g_β and $\hat{g}_n$ are, respectively, the adaptive estimator $\hat{\theta}_n$ and the symmetric density which is proportional to $[\hat{g}_n^{\frac{1}{2}}(x+\hat{\theta}_n) + \hat{g}_n^{\frac{1}{2}}(-x+\hat{\theta}_n)]^2$.

Estimators analogous to $\hat{\theta}_n$ can be defined for location and scale change in nonparametric two-sample location-scale models and for regression parameters in nonparametric linear models. The asymptotic efficiency and robustness of these estimators remains to be checked.

2. The estimator and consistency. The density estimator used in this paper to construct the adaptive location estimator $\hat{\theta}_n$ is a smoothly truncated version of a standard window estimator. To be specific, let $b(x)$ be a function satisfying the following assumption:

A1. $b(x)$ has range $[0, 1]$, is symmetric about zero with $b(0) = 1$, vanishes outside $[-1, 1]$, and is twice absolutely continuous with $b''(x)$ bounded on the real line.

An example of such a function is provided by $b(x) = (1 - x^2)^2$ for $|x| \leq 1$ and $b(x) = 0$ otherwise. Let $\{a_n : n \geq 1\}$ be a sequence of positive constants and define $b_n(x)$ by

$$b_n(x) = \begin{cases} 1 & \text{if } |x| \leqq a_n \\ b(x - a_n) & \text{if } a_n \leqq x \leqq a_n + 1 \\ b(x + a_n) & \text{if } -a_n - 1 \leqq x \leqq -a_n \\ 0 & \text{otherwise.} \end{cases} \tag{2.1}$$

Let $w(x)$ be a function which has the following properties:

A2. $w(x)$ is a nonvanishing density, symmetric about zero, absolutely continuous, and the ratio $w'(x)/w(x)$ is bounded over the real line.

For instance, $w(x)$ might be the double-exponential or Cauchy densities. Let $\{c_n : n \geqq 1\}$ be a sequence of positive constants and define $\hat{s}_n(x)$ by

$$\hat{s}_n(x) = [(nc_n)^{-1} \textstyle\sum_{i=1}^n w(c_n^{-1}(x - X_i))]^{\frac{1}{2}} . \tag{2.2}$$

We will suppose throughout that the $\{X_i : 1 \leqq i \leqq n\}$ are independent identically distributed random variables with density g. Let $\hat{\mu}_n$ denote a preliminary location estimator that has the following properties:

A3. For every density, g, there exists a location functional μ such that $n^{\frac{1}{2}}(\hat{\mu}_n - \mu(g)) = O_p(1)$ when the $\{X_i\}$ are distributed according to g. Moreover $\mu(g)$ is the center of symmetry whenever g is symmetric.

Location estimators that satisfy this assumption can be found readily. Some examples are discussed in Section 4 following assumption A5.

Finally, define the density estimator $\hat{g}_n(x)$ by

$$\hat{g}_n(x) = \hat{A}_n \hat{h}_n^2(x) , \tag{2.3}$$

where

$$\hat{h}_n(x) = \hat{s}_n(x) b_n(x - \hat{\mu}_n) \tag{2.4}$$
$$\hat{A}_n^{-1} = \int \hat{h}_n^2(x)\, dx .$$

For later use, it is convenient to define the analogous functions

$$s_n(x) = [c_n^{-1} \int w(c_n^{-1}(x - y)) g(y)\, dy]^{\frac{1}{2}} \tag{2.5}$$

and

$$h_n(x) = s_n(x) b_n(x - \mu(g)) . \tag{2.6}$$

The location estimator $\hat{\theta}_n$ is taken to be that value (or values) of real t which maximizes $\int \hat{g}_n^{\frac{1}{2}}(-x + 2t)\hat{g}_n^{\frac{1}{2}}(x)\, dx$ or, equivalently, maximizes $\int \hat{h}_n(-x + 2t)\hat{h}_n(x)\, dx$. It remains to verify that this definition is possible. Let $\mathscr{C}$ denote the set of all nonnegative square integrable functions which are continuous a.e. and have nonzero L_2-norm. Let $\|\cdot\|$ denote the L_2-norm. Define a functional T by the requirement that for every $k \in \mathscr{C}$

$$\int k(-x + 2T(k))k(x)\, dx = \max_t \int k(-x + 2t)k(x)\, dx . \tag{2.7}$$

The sense of this definition is clarified by the following result.

Lemma 1. *For $k \in \mathscr{C}$, the set of real values $T(k)$ which satisfy (2.7) is nonempty and compact. If k is symmetric, $T(k)$ is uniquely the center of symmetry.*

Proof. Set $r(t) = \int k(-x+2t)k(x)\,dx$. For every sequence $\{t_n : t_n \in R,\ t_n \to t\}$

$$|r(t_n) - r(t)|^2 \leqq \|k\|^2 \int [k(-x+2t_n) - k(-x+2t)]^2\,dx\,. \tag{2.8}$$

By Vitali's theorem and the continuity a.e. of k, the right side of (2.8) converges to zero. Hence r is continuous and achieves a maximum on every compact subset of the real line. Moreover, $\lim_{t\to\pm\infty} r(t) = 0$ and $r(t) \equiv 0$ is not possible for any $k \in \mathscr{C}$. Hence, $r(t)$ achieves a maximum over all real t and the maximizing value or values lie within a compact subset of the real line. By continuity of r, the values of $T(k)$ form a closed set and, therefore, a compact set.

The second assertion of the lemma is immediate.

It is not known what functions k, other than symmetric k, determine the value of $T(k)$ uniquely. An example of an asymmetric k which has unique $T(k)$ occurs when k is the square root of a chi-squared density with two or more degrees of freedom.

Since $\hat{\theta}_n = T(\hat{h}_n)$ and $\hat{h}_n$ is continuous under assumptions A1, A2, the definition of $\hat{\theta}_n$ is not vacuous but may not be unique. This possible nonuniqueness does not cause much difficulty when the distribution is symmetric, but it is awkward, theoretically, for asymmetric distributions. The uniqueness problem could be resolved by replacing $\hat{\theta}_n$ with the related iterated estimator (as will be done computationally in Section 5). However, this solution has two drawbacks: it loses the intuitive appeal of the present estimator and it leads to messier asymptotics under asymmetric densities. We will persist, therefore, with $\hat{\theta}_n$.

The next result shows that the functional T possesses a form of continuity on its domain $\mathscr{C}$. Let $T^*(k)$ denote the set whose elements are the values of $T(k)$. Let $d(x, y)$ denote Euclidean distance between real x and y and, for every $k \in \mathscr{C}$ and every $\varepsilon > 0$, let

$$T^*(k, \varepsilon) = \{x \in R : d(x, y) < \varepsilon \ \text{ for some } \ y \in T^*(k)\}\,. \tag{2.9}$$

Lemma 2. *For every sequence $\{k_n \in \mathscr{C}\}$ converging to $k \in \mathscr{C}$ in the L_2-metric and for every $\varepsilon > 0$, there exists n_0 such that $T(k_n) \in T(k, \varepsilon)$ for every $n \geqq n_0$. In particular, if $T(k)$ is uniquely defined, every value of $T(k_n)$ converges to $T(k)$ as $n \to \infty$.*

Proof. Set $r_n(t) = \int k_n(-x+2t)k_n(x)\,dx$ and define $r(t)$ as in the proof of Lemma 1. By the L_2-convergence of $\{k_n\}$,

$$\lim_{n\to\infty} \sup_t |r_n(t) - r(t)| = 0\,, \tag{2.10}$$

which implies that $\max_t r_n(t) - \max_t r(t) \to 0$ or, equivalently, $r_n[T(k_n)] \to r[T(k)]$ for every value of $T(k_n)$ and $T(k)$ respectively. Since (2.10) also implies that $r_n[T(k_n)] - r[T(k_n)] \to 0$, we conclude that

$$\lim_{n\to\infty} r[T(k_n)] = r[T(k)]\,. \tag{2.11}$$

If the first lemma assertion is false, there exists a sequence $\{\theta_n : \theta_n$ is a value of $T(k_n)\}$ and an $\varepsilon > 0$ such that $\{\theta_n\}$ fails to remain in $T^*(k, \varepsilon)$ for all sufficiently large n. The sequence $\{\theta_n\}$ does, however, remain within some compact set $K \subset R$; if it did not, there would exist a subsequence $\{\theta_m\} \subset \{\theta_n\}$ such that $\theta_m \to \pm\infty$, in which event $r(\theta_m) \to 0$. This would contradict (2.11), since $r[T(k)] > 0$.

Since K is compact and $T^*(k, \varepsilon)$ is open, there exists another subsequence $\{\theta_j\} \subset \{\theta_n\}$ such that $\theta_j \to \theta \in K$ and $\theta \notin T^*(k, \varepsilon)$ for some $\varepsilon > 0$. By continuity of r and (2.11),

$$\lim_{j\to\infty} r(\theta_j) = r(\theta) < r[T(k)] = \lim_{j\to\infty} r(\theta_j) . \tag{2.12}$$

This contradiction establishes the first lemma assertion; the second assertion is immediate.

The following theorem describes a consistency property of the estimator $\hat{\theta}_n$.

THEOREM 1. *Suppose assumptions* A1, A2, A3 *are fulfilled, the density g is continuous, and*

$$\lim_{n\to\infty} c_n = 0 , \qquad \lim_{n\to\infty} a_n = \infty , \qquad \lim_{n\to\infty} n^{\frac{1}{2}}c_n = \infty . \tag{2.13}$$

Then for every $\varepsilon > 0$,

$$\lim_{n\to\infty} P_g[T^*(\hat{h}_n) \subset T^*(g^{\frac{1}{2}}, \varepsilon)] = 1 . \tag{2.14}$$

In particular, if $T(g^{\frac{1}{2}})$ is uniquely defined, then every value of $\hat{\theta}_n \to_p T(g^{\frac{1}{2}})$ under g as $n \to \infty$.

PROOF. Let $\hat{G}_n$ be the empirical cdf of the $\{X_i\}$, let G be the cdf of g, and let $B_n(x) = n^{\frac{1}{2}}(\hat{G}_n(x) - G(x))$. Since

$$|\hat{s}_n^2(x) - s_n^2(x)|b_n^2(x - \hat{\mu}_n) \leqq n^{-\frac{1}{2}}c_n^{-1} \sup_t |B_n(t)| \int |w'(z)|\, dz \tag{2.15}$$

and

$$\begin{aligned} &|b_n^2(x - \hat{\mu}_n) - b_n^2(x - \mu(g))|s_n^2(x) \\ &\quad \leqq 2|\hat{\mu}_n - \mu(g)| \sup_t |b_n'(t)| \int w(z)g(x - c_n z)\, dz , \end{aligned} \tag{2.16}$$

there exist versions of the $\{\hat{h}_n\}$, defined on a suitable probability space, such that $\sup_x |\hat{h}_n^2(x) - h_n^2(x)| \to 0$ w.p. 1 as $n \to \infty$. A construction of such versions: Let $v: R \to [0, 1]$ be a continuous, strictly monotone function and let $d_n(t) = (\hat{h}_n \cdot v^{-1}(t))^2 - (h_n \cdot v^{-1}(t))^2$ on $[0, 1]$. The processes $\{d_n(t)\}$ converge weakly in $C[0, 1]$ to the zero process because by (2.15) and (2.16), $\sup_t |d_n(t)| \to 0$ in probability. Thus, there exist versions of the $\{d_n(t)\}$, say $\{d_n^*(t)\}$, for which $\sup_t |d_n^*(t)| \to 0$ w.p. 1 (Skorokhod (1956), Theorem 3.1.1, page 281). The desired version of $\hat{h}_n$ is defined as the positive square root of $d_n^* \cdot v(x) + h_n^2(x)$.

Evidently,

$$\begin{aligned} h_n^2(x) - g(x) &= b_n^2(x - \mu(g)) \int [g(x - c_n z) - g(x)]w(z)\, dz \\ &\quad - [1 - b_n^2(x - \mu(g))]g(x) \end{aligned} \tag{2.17}$$

tends to zero for every x as $n \to \infty$. Hence, for the above versions,

$P_g[\lim_{n\to\infty} \hat{h}_n(x) = g^{\frac{1}{2}}(x)$ for every $x] = 1$. Since $||\hat{h}_n|| \leqq 1 = ||g^{\frac{1}{2}}||$, Vitali's theorem implies that $\lim_{n\to\infty} ||\hat{h}_n - g^{\frac{1}{2}}|| = 0$ w.p. 1 for these versions.

From this and Lemma 2,

$$\text{(2.18)}\qquad \limsup_n P_g[\{T^*(\hat{h}_n) \subset T^*(g^{\frac{1}{2}}, \varepsilon)\}^c] \leqq P_g[\limsup_n \{T^*(\hat{h}_n) \subset T^*(g^{\frac{1}{2}}, \varepsilon)\}^c] = 0$$

for every $\varepsilon > 0$. The theorem follows.

It is clear from the proof that Theorem 1 holds under weaker assumptions than A1, A2, A3. Since these assumptions are harmless from a practical point of view and since their full strength will be used to establish the asymptotic normality of $\hat{\theta}_n$, the present formulation of Theorem 1 is convenient.

3. Asymptotic normality. Two related theorems are proved in this section: the first establishes asymptotic normality of the adaptive estimator $\hat{\theta}_n$ under densities which are not necessarily symmetric; the second theorem specializes the first result to symmetric densities (under improved assumptions) and thereby verifies that $\hat{\theta}_n$ has the efficiency property (1) discussed in the introduction.

The notation of Section 2 is retained, with two additions: let $s(x) = g^{\frac{1}{2}}(x)$ and let

$$\text{(3.1)}\qquad w_n^2(g) = \sup_{|x-\mu(g)|\leqq a_n+1} \sup_y \left\{\frac{w[c_n^{-1}(-x-y+2T(h_n))]}{w[c_n^{-1}(x-y)]}\right\}.$$

The following assumption on the density g will be used:

A4. g is absolutely continuous with finite Fisher information $I(g) = \int [g'(x)]^2/g(x)\,dx$.

This assumption entails absolute continuity of $s(x)$, with $s'(x) = g'(x)/(2g^{\frac{1}{2}}(x))$ a.e.

THEOREM 2. *Suppose assumptions* A1, A2, A3, A4 *are fulfilled*, $T(g)$ *is uniquely defined*,

$$\text{(3.2)}\qquad \lim_{n\to\infty} c_n = 0\,, \qquad \lim_{n\to\infty} a_n = \infty\,, \qquad \lim_{n\to\infty} (nc_n^4)^{-1}a_n^2 = 0\,,$$

and

$$\text{(3.3)}\qquad \lim_{n\to\infty} (nc_n^4)^{-1}a_n^2 w_n^2(g) = 0\,.$$

Then the limiting distribution of $n^{\frac{1}{2}}(\hat{\theta}_n - T(h_n))$ *under* g *is* $N(0, \sigma^2(g))$ *where*

$$\text{(3.4)}\qquad \sigma^2(g) = [2\int s'(-x+2T(g))s'(x)\,dx]^{-2}\int [s'(x)]^2\,dx\,.$$

Although both $\hat{\theta}_n = T(\hat{h}_n)$ and $T(h_n)$ may have several values, both converge to the uniquely defined $T(g)$ because of Theorem 1. By $\hat{\theta}_n - T(h_n)$ we mean any one of the possible differences. Assumption (3.3) is probably unnecessary, but our proof uses it at one point. It can be replaced by less awkward assumptions for specific choices of smoothing window w. For instance, if $w(x) = 2^{-1}\exp(-|x|)$, then

$$\text{(3.5)}\qquad \sup_y \left\{\frac{w[c_n^{-1}(-x-y+2T(h_n))]}{w[c_n^{-1}(x-y)]}\right\} \leqq \exp[2c_n^{-1}(|x| + |T(h_n)|)]\,.$$

Since $\lim_{n\to\infty} T(h_n) = T(g)$ is finite, as is $\mu(g)$, (3.3) can be replaced by a simpler requirement which does not depend upon g:

$$\lim_{n\to\infty} (nc_n^4)^{-1} a_n^2 \exp(3c_n^{-1}a_n) = 0 . \tag{3.6}$$

If g is symmetric about θ, then $\mu(g) = \theta$ by assumption A3 and h_n is symmetric about θ for every n; hence $T(h_n) = \theta$ uniquely. By specializing Theorem 2 and modifying the proof slightly to avoid (3.3), we obtain the following result, which corresponds to property (1) in the introduction.

THEOREM 3. *Suppose g is symmetric about θ, assumptions* A1, A2, A3, A4 *are fulfilled, and* (3.2) *holds. Then*

$$n^{\frac{1}{2}}(\hat{\theta}_n - \theta) = n^{-\frac{1}{2}} I^{-1}(g) \sum_{i=1}^n -g'(X_i)/g(X_i) + o_p(1) \tag{3.7}$$

and the limiting distribution of $n^{\frac{1}{2}}(\hat{\theta}_n - \theta)$ under g is therefore $N(0, I^{-1}(g))$.

The proofs for Theorems 2 and 3 are elementary, but require a large number of careful approximations. The following three lemmas will be needed. It is assumed without further mention that the $\{X_i\}$ are i.i.d. with density g.

LEMMA 3. *Suppose assumptions* A1, A2, A3 *are fulfilled and* $\lim_{n\to\infty} c_n = 0$, $\lim_{n\to\infty} a_n = \infty$. *Then*

$$\begin{aligned} \int [\hat{h}_n(x) - h_n(x)]^2\,dx &= O_p[(nc_n)^{-1}a_n] \\ \int [\hat{h}_n'(x) - h_n'(x)]^2\,dx &= O_p[(nc_n^3)^{-1}a_n] . \end{aligned} \tag{3.8}$$

PROOF. Let $\hat{f}_n(x) = \hat{s}_n^2(x)$ and $f_n(x) = s_n^2(x)$. By calculation

$$\begin{aligned} E[\hat{s}_n(x) - s_n(x)]^2 &\leqq f_n^{-1}(x) E[\hat{f}_n(x) - f_n(x)]^2 \\ &\leqq (nc_n^2)^{-1} f_n^{-1}(x) \int w^2[c_n^{-1}(x-y)] g(y)\,dy \\ &\leqq (nc_n)^{-1} \sup_z w(z) . \end{aligned} \tag{3.9}$$

Hence

$$\int [\hat{s}_n(x) - s_n(x)]^2 b_n^2(x - \mu(g))\,dx = O_p[(nc_n)^{-1}a_n] . \tag{3.10}$$

Moreover

$$\begin{aligned} \int \hat{s}_n^2(x)[b_n(x - \hat{\mu}_n) - b_n(x - \mu(g))]^2\,dx &\leqq (\hat{\mu}_n - \mu(g))^2 \sup_x [b'(x)]^2 \\ &= O_p(n^{-1}) . \end{aligned} \tag{3.11}$$

The first bound in (3.8) follows from (3.10) and (3.11), using the definitions (2.4), (2.6) of $\hat{h}_n$ and h_n.

By Minkowski's inequality,

$$\begin{aligned} \int [\hat{h}_n'(x) - h_n'(x)]^2\,dx \leqq\ & 4 \int [\hat{s}_n'(x)]^2 [b_n(x - \hat{\mu}_n) - b_n(x - \mu(g))]^2\,dx \\ & + 4 \int [\hat{s}_n'(x) - s_n'(x)]^2 b_n^2(x - \mu(g))\,dx \\ & + 4 \int \hat{s}_n^2(x)[b_n'(x - \hat{\mu}_n) - b_n'(x - \mu(g))]^2\,dx \\ & + 4 \int [\hat{s}_n(x) - s_n(x)]^2 [b_n'(x - \mu(g))]^2\,dx \\ =\ & \textstyle\sum_{i=1}^4 T_{in} , \quad \text{say.} \end{aligned} \tag{3.12}$$

Evidently, $T_{3n} = O_p(n^{-1})$ and, using (3.9), $T_{4n} = O_p[(nc_n)^{-1}a_n]$. Since

(3.13) $$|\hat{f}_n'(x)|/\hat{f}_n(x) \leqq c_n^{-1} \sup_z |w'(z)|/w(z) ,$$

the term $T_{1n} = O_p[(nc_n)^{-1}]$. Moreover,

(3.14) $$\begin{aligned} &\hat{s}_n'(x) - s_n'(x) \\ &= (2s_n(x))^{-1}[\hat{f}_n'(x) - f_n'(x)] \\ &\quad - (2\hat{f}_n(x))^{-1}\hat{f}_n'(x)[s_n^{-1}(x)(\hat{s}_n(x) - s_n(x))^2 + \hat{s}_n(x) - s_n(x)] \end{aligned}$$

and

(3.15) $$f_n^{-1}(x)E[\hat{f}_n'(x) - f_n'(x)]^2 \leqq (nc_n^3)^{-1} \sup_z |w'(z)|^2/w(z)$$

and

(3.16) $$\begin{aligned} f_n^{-1}(x)E[\hat{s}_n(x) - s_n(x)]^4 &\leqq f_n^{-1}(x)E[\hat{f}_n(x) - f_n(x)]^2 \\ &\leqq (nc_n)^{-1} \sup_z w(z) . \end{aligned}$$

Applying (3.9), (3.13), (3.15) and (3.16) to (3.14) yields $T_{2n} = O_p[(nc_n^3)^{-1}a_n]$. The second bound on (3.8) now follows from (3.12).

LEMMA 4. *Suppose assumption* A4 *is fulfilled. Then*

(3.17) $$\lim_{t\to 0} \int [s'(x + t) - s'(x)]^2\, dx = 0 .$$

Suppose also that $\lim_{n\to\infty} c_n = 0$, $\lim_{n\to\infty} a_n = \infty$. *Then*

(3.18) $$\lim_{n\to\infty} \int [h_n'(x) - s'(x)]^2\, dx = 0$$

and for every real sequence $\{t_n\}$ *converging to zero as* $n \to \infty$,

(3.19) $$\lim_{n\to\infty} \int [t_n^{-1}(h_n(x + t_n) - h_n(x)) - h_n'(x)]^2\, dx = 0 .$$

PROOF. Since $s' \in L_2$, there exists for every $\varepsilon > 0$ a differentiable function $\psi_\varepsilon \in L_2$ such that $\psi_\varepsilon' \in L_2$ and $||s' - \psi_\varepsilon|| \leqq \varepsilon$, where $||\cdot||$ denotes the L_2-norm. Thus, for every $\varepsilon > 0$,

(3.20) $$\begin{aligned} &|\int \psi_\varepsilon(x + t)s'(x)\, dx - \int \psi_\varepsilon(x)s'(x)\, dx| \\ &\leqq ||s'||\{\int [\psi_\varepsilon(x + t) - \psi_\varepsilon(x)]^2\, dx\}^{\frac{1}{2}} \leqq |t|\, ||\psi_\varepsilon'||\, ||s'|| \end{aligned}$$

by Cauchy–Schwarz, the fundamental theorem of calculus, and Fubini's theorem. Moreover,

(3.21) $$\begin{aligned} |\int \psi_\varepsilon(x)s'(x)\, dx - \int [s'(x)]^2\, dx| &\leqq \varepsilon||s'|| \\ |\int \psi_\varepsilon(x + t)s'(x)\, dx - \int s'(x + t)s'(x)\, dx| &\leqq \varepsilon||s'|| \end{aligned}$$

for every real t. Inequalities (3.20) and (3.21) imply that

(3.22) $$\lim_{t\to 0} \int s'(x + t)s'(x)\, dx = \int [s'(x)]^2\, dx ,$$

from which (3.17) follows.

By Cauchy–Schwarz,

(3.23) $$[f_n'(x)]^2 \leqq f_n(x) \int w(z)g^{-1}(x - c_n z)[g'(x - c_n z)]^2\, dz$$

and therefore

$$\int [s_n'(x)]^2\,dx \leqq \int [s'(x)]^2\,dx\,. \tag{3.24}$$

For every $\varepsilon > 0$,

$$|\int s_n'(x)\psi_\varepsilon(x)\,dx - \int s'(x)\psi_\varepsilon(x)\,dx| \leqq ||s_n - s||\,||\psi_\varepsilon'|| \tag{3.25}$$

which tends to zero as $n \to \infty$ since g is continuous. Moreover, using (3.24),

$$\begin{gathered} |\int s'(x)\psi_\varepsilon(x)\,dx - \int [s'(x)]^2\,dx| \leqq \varepsilon||s'|| \\ |\int s_n'(x)\psi_\varepsilon(x)\,dx - \int s_n'(x)s'(x)\,dx| \leqq \varepsilon||s'|| \end{gathered} \tag{3.26}$$

for every n. Inequalities (3.25) and (3.26) imply

$$\lim_{n\to\infty} \int s_n'(x)s'(x)\,dx = \int [s'(x)]^2\,dx\,. \tag{3.27}$$

From this and (3.24), it follows that

$$\lim_{n\to\infty} \int [s_n'(x) - s'(x)]^2\,dx = 0\,. \tag{3.28}$$

To establish (3.18), observe that

$$\begin{aligned} \int [h_n'(x) - s'(x)]^2\,dx \leqq {} & 4 \int [s_n'(x) - s'(x)]^2 b_n^2(x - \mu(g))\,dx \\ & + 4 \int [s'(x)]^2[1 - b_n(x - \mu(g))]^2\,dx \\ & + 4 \int [s_n(x) - s(x)]^2[b_n'(x - \mu(g))]^2\,dx \\ & + 4 \int s^2(x)[b_n'(x - \mu(g))]^2\,dx\,. \end{aligned} \tag{3.29}$$

Use (3.28) for the first integral on the right, dominated convergence for the second one, L_2-convergence of s_n to s for the third term, and the vanishing of $b_n'(x)$ when x lies outside $[a_n, a_n + 1]$ or $[-a_n - 1, -a_n]$ for the final integral.

By Cauchy-Schwarz,

$$\begin{aligned} & \int [t_n^{-1}(h_n(x + t_n) - h_n(x)) - h_n'(x)]^2\,dx \\ & \quad = \int [t_n^{-1} \int_0^{t_n} (h_n'(x + u) - h_n'(x))\,du]^2\,dx \\ & \quad \leqq t_n^{-1} \int_0^{t_n} \int [h_n'(x + u) - h_n'(x)]^2\,dx\,du \end{aligned} \tag{3.30}$$

which tends to zero as $n \to \infty$, giving (3.19).

LEMMA 5. *Suppose assumptions* A1, A2, A3, A4 *are fulfilled*, $T(g)$ *is uniquely defined, and* (3.2) *holds. Then, under* g,

$$\begin{aligned} n^{\frac{1}{2}}(\hat\theta_n - T(h_n)) = {} & [-\int s'(-x + 2T(g))s'(x)\,dx]^{-1} n^{\frac{1}{2}} \\ & \times \int h_n'(-x + 2T(h_n))[\hat h_n(x) - h_n(x)]\,dx + o_p(1)\,. \end{aligned} \tag{3.31}$$

PROOF. For every n, $\hat h_n(x)$ is absolutely continuous with $\int [\hat h_n'(x)]^2\,dx < \infty$. By a standard argument,

$$\lim_{u\to 0} \int [u^{-1}(\hat h_n(x + u) - \hat h_n(x)) - \hat h_n'(x)]^2\,dx = 0 \tag{3.32}$$

for every n, which implies

$$\begin{aligned} & \lim_{u\to 0} u^{-1} \int [\hat h_n(-x + 2t + u) - \hat h_n(-x + 2t)]\hat h_n(x)\,dx \\ & \quad = \int \hat h_n'(-x + 2t)\hat h_n(x)\,dx = \int \hat h_n(-x + 2t)\hat h_n'(x)\,dx\,. \end{aligned} \tag{3.33}$$

Since the estimator $\hat{\theta}_n$ maximizes $\int \hat{h}_n(-x + 2t)\hat{h}_n(x)\,dx$ over all real t, it follows that $\int \hat{h}_n(-x + 2\hat{\theta}_n)\hat{h}_n'(x)\,dx = 0$ for every n. Since $T(h_n)$ maximizes $\int h_n(-x + 2t)h_n(x)\,dx$ over all real t, a similar argument shows that $\int h_n(-x + 2T(h_n))h_n'(x)\,dx = 0$ for every n.

For notational convenience, write θ_n in place of $T(h_n)$. By Lemma 2, $\lim_{n\to\infty} \theta_n = T(g^{\frac{1}{2}})$ since $\lim_{n\to\infty} ||h_n - g^{\frac{1}{2}}|| = 0$. It follows from Theorem 1 that $\hat{\theta}_n - \theta_n \to_p 0$ under g as $n \to \infty$. This fact and (3.19) of Lemma 4 justify the expansion

$$h_n(-x + 2\hat{\theta}_n) = h_n(-x + 2\theta_n) + 2(\hat{\theta}_n - \theta_n)h_n'(-x + 2\theta_n) + 2(\hat{\theta}_n - \theta_n)v_n(x) \tag{3.34}$$

where $||v_n|| \to_p 0$ as $n \to \infty$. Thus, for $r_n(x) \equiv \hat{h}_n(x) - h_n(x)$,

$$\begin{aligned} 0 &= \int \hat{h}_n'(x)\hat{h}_n(-x + 2\hat{\theta}_n)\,dx \\ &= \int [h_n'(x) + r_n'(x)][h_n(-x + 2\hat{\theta}_n) + r_n(-x + 2\hat{\theta}_n)]\,dx \\ &= \int [h_n'(x) + 2r_n'(x)]h_n(-x + 2\hat{\theta}_n)\,dx + \int r_n'(x)r_n(-x + 2\hat{\theta}_n)\,dx\,, \end{aligned} \tag{3.35}$$

the last step using the property

$$\int r_n'(x)h_n(-x + 2\hat{\theta}_n)\,dx = \int r_n(-x + 2\hat{\theta}_n)h_n'(x)\,dx\,, \tag{3.36}$$

which follows from the analogoues of (3.33) for h_n and r_n.

Moreover, from (3.34) and the fact $\int h_n(-x + 2\theta_n)h_n'(x)\,dx = 0$,

$$\begin{aligned} &\int [h_n'(x) + 2r_n'(x)]h_n(-x + 2\hat{\theta}_n)\,dx \\ &\quad = 2\int h_n'(-x + 2\theta_n)r_n(x)\,dx \\ &\quad\quad + 2(\hat{\theta}_n - \theta_n)\{\int h_n'(-x + 2\theta_n)h_n'(x)\,dx + R_n\} \end{aligned} \tag{3.37}$$

where

$$R_n = \int [h_n'(x)v_n(x) + 2r_n'(x)v_n(x) + 2r_n'(x)h_n'(-x + 2\theta_n)]\,dx\,. \tag{3.38}$$

Equations (3.35) and (3.37) yield an expression for $n^{\frac{1}{2}}(\hat{\theta}_n - \theta_n)$ which reduces to (3.31) by virtue of Lemmas 3 and 4 and (3.2) and the fact that $n^{\frac{1}{2}} \int h_n'(-x + 2\theta_n)[\hat{h}_n(x) - h_n(x)]\,dx = O_p(1)$; this last property is established in the course of the following proof.

Proof of Theorem 2. The result is established in two stages. First, expansion (3.31) for $n^{\frac{1}{2}}(\hat{\theta}_n - T(h_n))$ is reduced to the simpler approximation (3.46). Secondly, the right side of (3.46) is shown to have the desired limiting distribution.

By definition,

$$\begin{aligned} &n^{\frac{1}{2}} \int h_n'(-x + 2\theta_n)[\hat{h}_n(x) - h_n(x)]\,dx \\ &\quad = n^{\frac{1}{2}} \int h_n'(-x + 2\theta_n)[\hat{s}_n(x) - s_n(x)]b_n(x - \mu(g))\,dx \\ &\quad\quad + n^{\frac{1}{2}} \int h_n'(-x + 2\theta_n)\hat{s}_n(x)[b_n(x - \hat{\mu}_n) - b_n(x - \mu(g))]\,dx\,. \end{aligned} \tag{3.39}$$

The second integral on the right side of (3.39) can be expressed as the sum of two terms:

$$(3.40)\qquad \begin{aligned} S_{1n} &= -n^{\frac{1}{2}}(\hat{\mu}_n - \mu(g)) \int h_n'(-x + 2\theta_n)\hat{s}_n(x)b_n'(x - \mu(g))\, dx \\ S_{2n} &= n^{\frac{1}{2}}(\hat{\mu}_n - \mu(g))^2 \int h_n'(-x + 2\theta_n)\hat{s}_n(x)\xi_n(x)\, dx \end{aligned}$$

where $\xi_n(x) = (\hat{\mu}_n - \mu(g))^{-2} \int_{\mu(g)}^{\hat{\mu}_n} \int_{\mu(g)}^{t} b_n''(x - u)\, du\, dt$. In view of the assumptions and (3.18) of Lemma 4, $S_{2n} = O_p(n^{-\frac{1}{2}})$. On the other hand,

$$(3.41)\qquad \begin{aligned} &\int h_n'(-x + 2\theta_n)\hat{s}_n(x)b_n'(x - \mu(g))\, dx \\ &\quad = \int h_n'(-x + 2\theta_n)[\hat{s}_n(x) - s(x)]b_n'(x - \mu(g))\, dx \\ &\qquad + \int h_n'(-x + 2\theta_n)s(x)b_n'(x - \mu(g))\, dx\,, \end{aligned}$$

where $s(x) = g^{\frac{1}{2}}(x)$. By a simple argument, $||\hat{s}_n - s|| \to_p 0$ as $n \to \infty$. Moreover, $b_n'(x)$ vanishes for x outside $[a_n, a_n + 1]$ or $[-a_n - 1, -a_n]$. Hence $S_{1n} = o_p(1)$.

The first integral on the right side of (3.39) can be written as the sum of two terms

$$(3.42)\qquad \begin{aligned} V_{1n} &= n^{\frac{1}{2}} \int h_n'(-x + 2\theta_n)(2s_n(x))^{-1}[\hat{f}_n(x) - f_n(x)]b_n(x - \mu(g))\, dx \\ V_{2n} &= -n^{\frac{1}{2}} \int h_n'(-x + 2\theta_n)(2s_n(x))^{-1}[\hat{s}_n(x) - s_n(x)]^2 b_n(x - \mu(g))\, dx \end{aligned}$$

where $\hat{f}_n(x) = \hat{s}_n^2(x)$ and $f_n(x) = s_n^2(x)$. Since, for $|x - \mu(g)| \leqq a_n + 1$,

$$(3.43)\qquad \begin{aligned} &|s_n'(-x + 2\theta_n)|s_n^{-1}(x) \\ &\quad = [2^{-1}|f_n'(-x + 2\theta_n)|f_n^{-1}(-x + 2\theta_n)][s_n(-x + 2\theta_n)s_n^{-1}(x)] \\ &\quad \leqq (2c_n)^{-1}w_n(g) \sup_z |w'(z)|/w(z) \end{aligned}$$

by (3.1) and the analogue of (3.13) for $f_n(x)$, it follows with the aid of (3.9) that $V_{2n} = O_p[(n^{\frac{1}{2}}c_n^2)^{-1}a_n w_n(g)]$ and therefore, by (3.3), that $V_{2n} = o_p(1)$.

The term V_{1n} can be decomposed further into the sum of two terms

$$(3.44)\qquad \begin{aligned} W_{1n} &= n^{\frac{1}{2}} \int s_n'(-x + 2\theta_n)(2s_n(x))^{-1}[\hat{f}_n(x) - f_n(x)]\phi_{1n}(x)\, dx \\ W_{2n} &= 2^{-1}n^{\frac{1}{2}} \int [\hat{f}_n(x) - f_n(x)]\phi_{2n}(x)\, dx \end{aligned}$$

where $\phi_{1n}(x) = b_n(x - \mu(g))b_n(-x + 2\theta_n - \mu(g))$ and $\phi_{2n}(x) = b_n(x - \mu(g)) \times b_n'(-x + 2\theta_n - \mu(g))s_n(-x + 2\theta_n)s_n^{-1}(x)$. Now

$$(3.45)\qquad \begin{aligned} E(2W_{2n}^2) &= c_n^{-2} \operatorname{Var} [\int \phi_{2n}(x)w[c_n^{-1}(x - X_i)]\, dx] \\ &\leqq c_n^{-2}E[\int \phi_{2n}(x)w[c_n^{-1}(x - X_i)]\, dx]^2 \\ &\leqq \int \phi_{2n}^2(x)f_n(x)\, dx \\ &\leqq \int [b_n'(x - \mu(g))]^2 f_n(x)\, dx\,, \end{aligned}$$

the second last step using the Cauchy–Schwarz inequality. Since $b_n'(x)$ is bounded and vanishes for x not in $[a_n, a_n + 1]$ or $[-a_n - 1, -a_n]$ and since $\lim_{n\to\infty} \int |f_n(x) - g(x)|\, dx = 0$, $W_{2n} = o_p(1)$. The results of the last three paragraphs combined with Lemma 5 yield

$$(3.46)\qquad n^{\frac{1}{2}}(\hat{\theta}_n - \theta_n) = [-\int s'(-x + 2T(g))s'(x)\, dx)]^{-1}W_{1n} + o_p(1)\,.$$

By a calculation analogous to (3.45),

$$(3.47)\qquad \begin{aligned} \operatorname{Var}(2W_{1n}) &\leqq \int [s_n'(-x + 2\theta_n)]^2\, dx \\ &\leqq \int [s'(x)]^2\, dx\,, \end{aligned}$$

the last step using (3.24). Let $\theta \equiv T(g^{\frac{1}{2}})$ and let $U_n = n^{-\frac{1}{2}} \sum_{i=1}^n s'(-X_i + 2\theta)s^{-1}(X_i)$. Evidently $E(U_n) = 0$ and

$$\operatorname{Var}(U_n) = \int [s'(x)]^2 \, dx \,. \tag{3.48}$$

To complete the proof of Theorem 2, it suffices to show that

$$\lim_{n\to\infty} \operatorname{Cov}(2W_{1n}, U_n) = \int [s'(x)]^2 \, dx \,. \tag{3.49}$$

Indeed, (3.47), (3.48) and (3.49) imply that $\lim_{n\to\infty} E[2W_{1n} - U_n]^2 = 0$, from which the theorem follows in view of (3.46).

Now

$$\begin{aligned} \operatorname{Cov}(2W_{1n}, U_n) &= \int \left[\int s_n'(-x + 2\theta_n)\phi_{1n}(x)s_n^{-1}(x)c_n^{-1}w[c_n^{-1}(x-y)]\, dx\right] \\ &\quad \times s'(-y+2\theta)s(y)\, dy \\ &= \int s_n'(-x + 2\theta_n)\phi_{1n}(x)s_n^{-1}(x) \int c_n^{-1}w[c_n^{-1}(x-y)] \\ &\quad \times s'(-y+2\theta)s(y)\, dy\, dx \,, \end{aligned} \tag{3.50}$$

the interchange in order of integration being justified by the Tonelli and Fubini theorems (cf. (3.52) below for a method of checking the existence of the required absolute integral). For notational convenience, put

$$d_n(x) = s_n^{-1}(x) \int c_n^{-1}w[c_n^{-1}(x-y)]s'(-y+2\theta)s(y)\, dy \tag{3.51}$$

and define $d_{n,\varepsilon}$ analogously by replacing s' in (3.51) with the function ψ_ε used in the proof of Lemma 4.

By Cauchy-Schwarz and the definition of s_n,

$$\begin{aligned} \int d_n^2(x)\, dx &\leqq \iint c_n^{-1}w[c_n^{-1}(x-y)](s'(-y+2\theta))^2\, dy\, dx \\ &= ||s'||^2 \,. \end{aligned} \tag{3.52}$$

Since also $||d_{n,\varepsilon}|| \leqq ||\psi_\varepsilon||$ and $\lim_{n\to\infty} d_{n,\varepsilon}(x) = \psi_\varepsilon(-x+2\theta)$ under the theorem assumptions, it follows by Vitali's theorem that

$$\lim_{n\to\infty} \int \psi_\varepsilon(-x+2\theta_n)d_{n,\varepsilon}(x)\, dx = ||\psi_\varepsilon||^2 \,. \tag{3.53}$$

An argument analogous to (3.52) establishes the inequality $||d_{n,\varepsilon} - d_n|| \leqq ||\psi_\varepsilon - s'|| \leqq \varepsilon$; hence

$$\left|\int \psi_\varepsilon(-x+2\theta_n)d_{n,\varepsilon}(x)\, dx - \int s'(-x+2\theta_n)d_n(x)\, dx\right| \leqq 2\varepsilon||s'|| \,. \tag{3.54}$$

This inequality and (3.53) imply that

$$\lim_{n\to\infty} \int s'(-x+2\theta_n)d_n(x)\, dx = ||s'||^2 \,, \tag{3.55}$$

since ε can be chosen arbitrarily small. Finally, (3.28), (3.50), (3.52) and (3.55) yield the desired limit (3.49).

Proof of Theorem 3. The proof is by specialization of the proof for Theorem 2 with one exception: the treatment of the term V_{2n} defined in (3.42) can be simplified and improved. Since g is symmetric, θ_n equals the center of symmetry, $s_n(-x+2\theta_n) = s_n(x)$, and therefore $\sup_x |s_n'(-x+2\theta_n)|s_n^{-1}(x) = O(c_n^{-1})$ by

(3.13). Using this bound in place of (3.43) yields $V_{2n} = O_p[(n^{\frac{1}{2}}c_n^2)^{-1}a_n]$, which is $o_p(1)$ under the assumption (3.2).

4. Asymptotic robustness. This section establishes for $\hat{\theta}_n$ a mathematical property which corresponds, in large samples, to robustness in distribution as described by requirement (2) of the introduction. The basic idea is to examine the limiting behavior of $\hat{\theta}_n$ under fairly arbitrary sequences of densities contiguous to the sequence $\{\prod_{i=1}^n g(x_i); n \geqq 1\}$, where g is symmetric with finite Fisher information $I(g)$. It turns out that even under such local perturbations of the density, robust and nonrobust estimators exhibit distinguishable behavior; moreover the adaptive estimator $\hat{\theta}_n$ falls into the robust category. Jaeckel (1971 a) has examined the asymptotic biases induced in M-estimators of location by certain local sequences of asymmetric perturbations of a given symmetric distribution. Apart from the similarity in starting point, his work is unrelated to the argument in this section.

For every density g on the real line, let $\mathscr{T}(g, \beta)$ denote the set of all sequences of absolutely continuous densities $\{g^{(n)}\}$ such that

$$\lim_{n\to\infty} \int [n^{\frac{1}{2}}((g^{(n)}(x))^{\frac{1}{2}} - (g(x))^{\frac{1}{2}}) - \beta(x)]^2\, dx = 0 \tag{4.1}$$

for some β in L_2 and

$$\lim_{n\to\infty} I(g^{(n)}) = I(g)\,. \tag{4.2}$$

Note that since g and $\{g^{(n)}\}$ are densities, (4.1) cannot hold unless β is orthogonal in L_2 to $g^{\frac{1}{2}}$. The logical independence of (4.1) and (4.2) is evident: a density g satisfying A4 may be truncated abruptly in such a way that the truncated density remains close to g in the Hellinger metric, yet has much larger Fisher information than g.

For every density pair $(g^{(n)}, g)$, there exist $\alpha_n \in [0, \pi/2]$ and $\delta^{(n)} \in L_2$ such that $||\delta^{(n)}|| = 1$, $\delta^{(n)} \perp g^{\frac{1}{2}}$ and

$$[g^{(n)}(x)]^{\frac{1}{2}} = \cos(\alpha_n) g^{\frac{1}{2}}(x) + \sin(\alpha_n)\delta^{(n)}(x)\,. \tag{4.3}$$

The function $[g^{(n)}]^{\frac{1}{2}}$ is absolutely continuous if both $g^{\frac{1}{2}}$ and $\delta^{(n)}$ are absolutely continuous. Conditions (4.1) and (4.2) are equivalent to the requirements $\lim_{n\to\infty} n^{\frac{1}{2}}\alpha_n = ||\beta||$, $\lim_{n\to\infty} ||\delta^{(n)} - ||\beta||^{-1}\beta|| = 0$, and $\lim_{n\to\infty} \alpha_n^2 \int [\delta^{(n)\prime}(x)]^2\, dx = 0$. Thus, a density sequence in $\mathscr{T}(g, \beta)$ represents fairly general smooth contamination of g in approximate direction $||\beta||^{-1}\beta$. Robustness of $\hat{\theta}_n$ will be established under such perturbations of symmetric g. The practical consequences of (4.2) as a limitation to the robustness of $\hat{\theta}_n$ are not clear.

The perturbation sequences defined by (4.1) and (4.2) are not suitable for studying the effect of round-off errors in the data upon the distribution of the estimator. This limitation could be overcome by redefining the estimator so as to include a possible round-off operation. We will not do so because $\hat{\theta}_n$ is already complicated. Moreover, the continuity of $\hat{\theta}_n$, or of its iterated version, provides some assurance that small round-off errors in the data will not affect the estimate seriously.

For random variables $\{X_i: 1 \leqq i \leqq n\}$ which are i.i.d. with joint density $\prod_{i=1}^n g(x_i)$ and for $\{g^{(n)}\} \in \mathscr{F}(g, \beta)$, let L_n denote the log-likelihood ratio

$$(4.4) \qquad L_n = \log [\prod_{i=1}^n g^{(n)}(X_i)/g(X_i)] \quad \text{w.p. } 1 .$$

For every density g, every $\beta \perp g^{\frac{1}{2}}$, and every sequence $\{g^{(n)}\} \in \mathscr{F}(g, \beta)$, L_n can be approximated stochastically by

$$(4.5) \qquad L_n = 2n^{-\frac{1}{2}} \sum_{i=1}^n \beta(X_i) g^{-\frac{1}{2}}(X_i) - 2||\beta||^2 + o_p(1)$$

(see Le Cam (1969) for a similar expansion in a parametric setting). Thus the limiting distribution of L_n under g is $N(-2||\beta||^2, 4||\beta||^2)$, which implies that the sequence of distributions $\{\prod_{i=1}^n g^{(n)}(x_i)\}$ is contiguous to the sequence $\{\prod_{i=1}^n g(x_i)\}$ for every $\{g^{(n)}\} \in \mathscr{F}(g, \beta)$ and every $\beta \perp g^{\frac{1}{2}}$.

To ensure robustness of the estimator $\hat{\theta}_n$, it is necessary that the preliminary location estimator $\hat{\mu}_n$ (used in constructing the density estimator $\hat{g}_n$) be robust as well. Technically, the following smoothness assumption on the centering functional μ of assumption A3 suffices:

A5. For every density g satisfying A4, every $\beta \perp g^{\frac{1}{2}}$, and every sequence $\{g^{(n)}\} \in \mathscr{F}(g, \beta)$, $n^{\frac{1}{2}}(\mu(g^{(n)}) - \mu(g)) = 0(1)$.

There exist many estimators which satisfy both A3 and A5. For example, suppose that ψ is a bounded function defined on the real line which is strictly monotone increasing, odd, and has continuous bounded derivative ψ' (e.g., $\psi(x) = \arctan(x)$). Define the M-estimator $\hat{\mu}_n$ as the unique solution to the equation $\sum_{i=1}^n \psi(X_i - \hat{\mu}_n) = 0$ and, for every density g, define the functional μ through the equation $\int \psi(x - \mu(g))g(x)\,dx = 0$. Then A5 is satisfied because

$$\lim_{n\to\infty} n^{\frac{1}{2}}(\mu(g^{(n)}) - \mu(g)) = \int \sigma_g(x)\beta(x)\,dx ,$$

where

$$(4.6) \qquad \sigma_g(x) = [\int \psi'(t - \mu(g))g(t)\,dt]^{-1} 2\psi(x - \mu(g))g^{\frac{1}{2}}(x)$$

(Beran (1977a)); A3 is satisfied because the limiting distribution of $n^{\frac{1}{2}}(\hat{\mu}_n - \mu(g))$ under g is $N(0, 4^{-1}\int \sigma_g^2(x)\,dx)$ (Huber (1964)).

By analogy with the notation of Sections 2 and 3, which is retained, let $s^{(n)}(x) = [g^{(n)}(x)]^{\frac{1}{2}}$, $s_n^{(n)}(x) = c_n^{-1} \int w[c_n^{-1}(x - y)] g^{(n)}(y)\,dy$, and $h_n^{(n)}(x) = s_n^{(n)}(x) b_n(x - \mu(g^{(n)}))$.

THEOREM 4. *Suppose g is symmetric about θ, assumptions* A1, A2, A3, A4, A5 *are fulfilled, and* (3.2) *holds. Then the limiting distribution of $n^{\frac{1}{2}}(\hat{\theta}_n - T(h_n^{(n)}))$ under $\{\prod_{i=1}^n g^{(n)}(x_i)\}$ is $N(0, I^{-1}(g))$ for every sequence $\{g^{(n)}\} \in \mathscr{F}(g, \beta)$ and every β orthogonal to $g^{\frac{1}{2}}$.*

A standard contiguity argument based upon (3.7) and (4.5) shows that the limiting distribution of $n^{\frac{1}{2}}(\hat{\theta}_n - \theta)$ under $\{g^{(n)}\}$ is $N(-4I^{-1}(g)\int s'(x)\beta(x)\,dx, I^{-1}(g))$. Thus, the distinctive feature in Theorem 4 is the possibility of using $T(h_n^{(n)})$ as the centering parameter in the asymptotics; this possibility is equivalent to the property

(4.7) $$\lim_{n\to\infty} n^{\frac{1}{2}}(T(h_n^{(n)}) - \theta) = -4I^{-1}(g) \int s'(x)\beta(x)\,dx\,.$$

The proof of (4.7), and therefore Theorem 4, is given at the end of this section.

Let g be a fixed symmetric density with finite Fisher information. Since $\hat{\theta}_n$ is centered in the same way in both Theorems 2 and 4, Theorem 4 indicates, roughly speaking, that the convergence to the limiting distributions in Theorem 2 occurs uniformly over all sufficiently local (in Hellinger metric and Fisher information) perturbations of g in approximate direction $||\beta||^{-1}\beta$; moreover, the limit laws under these local perturbations are all approximately $N(0, I^{-1}(g))$. Both assertions hold for every $\beta \perp g^{\frac{1}{2}}$. From these facts and continuity at g of the centering parameter (see (4.7)), it is reasonable to conclude that sufficiently small, fairly arbitrary perturbations of g do not affect the exact distribution of $\hat{\theta}_n$ greatly, at least for sufficiently large sample sizes. This conclusion is an asymptotic version of the qualitative robustness property (2) described in the introduction.

Technically, the argument for asymptotic robustness of $\hat{\theta}_n$ at every symmetric density with finite Fisher information rests upon the following fact: when $\hat{\theta}_n$ is centered as in Theorem 2, its limiting distribution under the circumstances of Theorem 4 does not depend upon β or the particular perturbation sequence in $\mathscr{F}(g, \beta)$. Note that this property is much stronger but mathematically analogous to Hájek's (1970) concept of a regular estimator in a parametric model. From our viewpoint, therefore, the classical phenomenon of superefficiency in estimation is just one special form of nonrobustness.

An example of a nonrobust estimator illustrates further the relevance to robustness of the technical property just discussed. Let $\hat{m}_n$ denote the sample mean of the $\{X_i\}$, let $\mathscr{M}$ be the set of all densities on the real line with finite mean and variance, and let m, v denote the mean and variance functionals on $\mathscr{M}$. If the $\{X_i\}$ are i.i.d. with density g in $\mathscr{M}$, the limiting distribution of $n^{\frac{1}{2}}(\hat{m}_n - m(g))$ is $N(0, v(g))$. We examine the behavior of $\hat{m}_n$ under small perturbations within $\mathscr{M}$ of the data density.

Let g be a fixed density in $\mathscr{M}$ satisfying A4 and define a sequence of densities $\{g^{(n)}\}$ converging to g as follows: in (4.3), take $\alpha_n = n^{-\frac{1}{2}}$ and $\{\delta^{(n)}\}$ orthogonal to $g^{\frac{1}{2}}$ such that $||\delta^{(n)}|| = 1$,

$$0 < \int x[\delta^{(n)}(x)]^2\,dx < \infty \qquad \text{for every} \quad n\,,$$
$$\lim_{n\to\infty} n^{-\frac{1}{2}} \int x[\delta^{(n)}(x)]^2\,dx = \infty\,,$$

$\limsup_n ||\delta^{(n)\prime}|| < \infty$ and $\delta^{(n)}$ converges in L_2 to some function β of unit norm. There are many possible constructions of such a sequence $\{\delta^{(n)}\}$.

The sequence of densities $\{g^{(n)}\}$ so defined lies in $\mathscr{M}$ and satisfies both (4.1) and (4.2) for the function β just described. By contiguity, using (4.5), the limiting distribution of $n^{\frac{1}{2}}(\hat{m}_n - m(g))$ under $\{g^{(n)}\}$ is $N(2 \int x\beta(x)g^{\frac{1}{2}}(x)\,dx, v(g))$.

However,

$$\begin{aligned} &\lim_{n\to\infty} n^{\frac{1}{2}}(m(g^{(n)}) - m(g)) \\ (4.8) \qquad &= \lim_{n\to\infty} n^{\frac{1}{2}} \int x[\sin(2\alpha_n)\delta^{(n)}(x)g^{\frac{1}{2}}(x) + \sin^2(\alpha_n)(\delta^{(n)}(x))^2]\,dx \\ &= \infty, \end{aligned}$$

which implies that $n^{\frac{1}{2}}(\hat{m}_n - m(g^{(n)})) \to_p -\infty$ under $\{g^{(n)}\}$. This striking result reflects the fact that the distribution of the sample mean $\hat{m}_n$ under $g^{(n)}$ becomes severely skewed for large n, with the bulk of the probability mass to the left of $m(g^{(n)})$ and a long tail to the right. Thus, even very small perturbations within $\mathscr{M}$ of any fixed data density g in $\mathscr{M}$ can affect the distribution of the sample mean dramatically. In this sense, the sample mean is robust nowhere.

As noted earlier, Theorems 2 and 4 indicate, roughly, that the convergence in law of the centered estimator $\hat{\theta}_n$ is locally uniform at every density g which is symmetric about some θ and satisfies A4, a circumstance which enables us to use the asymptotic distributions near g as reasonable approximations to the exact distributions. Since the limiting distribution of $n^{\frac{1}{2}}(\hat{\theta}_n - \theta)$ under any sequence $\{g^{(n)}\} \in \mathscr{T}(g, \beta)$ is $N(-4I^{-1}(g) \int s'(x)\beta(x)\,dx, I^{-1}(g))$, the main approximate effect of a small perturbation of g in direction $||\beta||^{-1}\beta$ is to shift the distribution of $n^{\frac{1}{2}}(\hat{\theta}_n - \theta)$ by the amount $-4I^{-1}(g) \int s'(x)\beta(x)\,dx$. Quantitative robustness of $\hat{\theta}_n$ at symmetric g can be assessed by comparing this shift with the corresponding shifts induced in the distributions of competitive, robust, center-of-symmetry estimators.

A natural class $\mathscr{R}$ of competitive location estimators θ_n^* is defined by the following requirements:

(i) θ_n^* is location-invariant and

$$n^{\frac{1}{2}}(\theta_n^* - U_n(g)) = n^{-\frac{1}{2}} \sum_{i=1}^n \rho_g(X_i)s^{-1}(X_i) + o_p(1)$$

under every density satisfying A4;

(ii) $\rho_g \in L_2$ and $\rho_g \perp s$;

(iii) U_n is a location-invariant functional whose value at every symmetric g is the center of symmetry;

(iv) $\lim_{n\to\infty} n^{\frac{1}{2}}(U_n(g^{(n)}) - \theta) = 2 \int \rho_g(x)\beta(x)\,dx$ for every sequence $\{g^{(n)}\} \in \mathscr{T}(g, \beta)$, every symmetric g satisfying A4, every $\beta \perp s$.

The class $\mathscr{R}$ contains many estimators, including the M-estimators discussed earlier in this section following the introduction of assumption A5. Note that the location invariance of θ_n^* and U_n entails the property $2 \int \rho_g(x)s'(x)\,dx = -1$. (Consider the effect of contiguous location-shift alternatives upon the asymptotic distribution of θ_n^*.)

By Theorem 2 and (4.7), the adaptive estimator $\hat{\theta}_n$ nearly belongs to the class $\mathscr{R}$, with

$$(4.9) \qquad \rho_g(x) = [-2 \int s'(-t + 2T(g))s'(t)\,dt]^{-1}s'(-x + 2T(g))$$

whenever $T(g)$ is uniquely defined; this last proviso is what keeps $\hat{\theta}_n$ outside $\mathscr{R}$.

For symmetric g, ρ_g defined in (4.9) reduces to $\rho_g(x) = -2I^{-1}(g)s'(x)$. Analogues to Theorems 2 and 4 hold for $\theta_n^* \in \mathscr{R}$ and θ_n^* is, accordingly, asymptotically robust in the same qualitative sense as $\hat{\theta}_n$. The main approximate effect of a small perturbation of symmetric g in direction $||\beta||^{-1}\beta$ is to shift the centered distribution of θ_n^* by the amount $2 \int \rho_g(x)\beta(x)\,dx$.

Let $B(\rho, \delta) = |\int \rho(x)\delta(x)\,dx|$. Subject to the constraints $\rho,\ \delta \in L_2,\ \rho \perp s$, $\int \rho(x)s'(x)\,dx = -1$, $\delta \perp s$, $||\delta|| = 1$, it can be shown that

$$\max_\delta \min_\rho B(\rho, \delta) = B(\rho_0, \delta_0) = \min_\rho \max_\delta B(\rho, \delta)\,, \tag{4.10}$$

the saddle-point being $\rho_0(x) = -4I^{-1}(g)s'(x)$ and $\delta_0(x) = -2I^{-\frac{1}{2}}(g)s'(x)$ (see Theorem 5 in Beran (1977a)). This result is directly applicable to the comparison of asymptotic biases induced in robust location estimators by asymmetric contamination of a symmetric data density.

Since $||\beta||^2 = \lim_{n\to\infty} n||(g^{(n)})^{\frac{1}{2}} - g^{\frac{1}{2}}||^2$ under (4.1), the quantity $||\beta||$ reflects the level of contamination of g represented by any sequence $\{g^{(n)}\} \in \mathscr{F}(g, \beta)$. The saddle-point property (4.10) shows that, relative to estimators in $\mathscr{R}$, the adaptive estimator $\hat{\theta}_n$ is minimax asymptotically biased by small perturbations of symmetric g in arbitrary directions but of fixed level $||\beta||$. In this local sense, $\hat{\theta}_n$ is quantitatively most robust at every symmetric g satisfying A4. Upon reflection, it is intuitively plausible that an estimator with this property should also be asymptotically efficient at every symmetric g with finite Fisher information.

Proof of Theorem 4. As noted after the statement of the theorem, it suffices to establish (4.7). The argument parallels the proof of Theorems 2 and 3 in structure, with some differences in detail which will be considered below. There is an expansion for $n^{\frac{1}{2}}(T(h_n^{(n)}) - \theta)$ analogous to the one for $n^{\frac{1}{2}}(\hat{\theta}_n - \theta)$ and there are counterparts to Lemmas 3 and 5 with $h_n^{(n)}$ in place of $\hat{h}_n$ and $T(h_n^{(n)})$ in place of $\hat{\theta}_n$. These counterpart lemmas are based upon

$$\begin{aligned} &\int [s_n^{(n)}(x) - s_n(x)]^2\,dx = O(n^{-1}) \\ &\lim_{n\to\infty} \int [s_n^{(n)\prime}(x) - s_n'(x)]^2 = 0\,, \end{aligned} \tag{4.11}$$

which implies, with the aid of A5,

$$\begin{aligned} &\int [h_n^{(n)}(x) - h_n(x)]^2\,dx = O(n^{-1}) \\ &\lim_{n\to\infty} \int [h_n^{(n)\prime}(x) - h_n'(x)]^2\,dx = 0\,. \end{aligned} \tag{4.12}$$

The first bound in (4.11) is proved by Minkowski's inequality, which gives

$$\begin{aligned} &|s_n^{(n)}(x) - s_n(x)| \\ &= |\{\int [s^{(n)}(x - c_n z)]^2 w(z)\,dz\}^{\frac{1}{2}} - \{\int [s(x - c_n z)]^2 w(z)\,dz\}^{\frac{1}{2}}| \\ &\leq \{\int [s^{(n)}(x - c_n z) - s(x - c_n z)]^2 w(z)\,dz\}^{\frac{1}{2}}\,. \end{aligned} \tag{4.13}$$

Therefore, interchanging the order of integration and using (4.1),

$$\int [s_n^{(n)}(x) - s_n(x)]^2\,dx \leq \int [s^{(n)}(x) - s(x)]^2\,dx = O(n^{-1})\,. \tag{4.14}$$

To establish the second part of (4.11), it is enough to show that

$$\text{(4.15)} \qquad \lim_{n\to\infty} \int [s_n^{(n)\prime}(x) - s'(x)]^2\, dx = 0$$

because combining this with (3.28) gives the desired result. As in (3.24),

$$\text{(4.16)} \qquad \begin{aligned}\limsup_n \int [s_n^{(n)\prime}(x)]^2\, dx &\leq \limsup_n \int [s^{(n)\prime}(x)]^2\, dx \\ &= \int [s'(x)]^2\, dx\,,\end{aligned}$$

the last step using (4.2). On the other hand, $\lim_{n\to\infty} ||s_n^{(n)} - s|| = 0$ because of (4.11) and the fact $\lim_{n\to\infty} ||s_n - s|| = 0$. Thus, an argument parallelling (3.25) to (3.27) gives

$$\text{(4.17)} \qquad \lim_{n\to\infty} \int s_n^{(n)\prime}(x) s'(x)\, dx = \int [s'(x)]^2\, dx\,.$$

Finally, (4.16) and (4.17) imply (4.15).

As described at the start of the proof, (4.11) yields, in analogy to Lemma 5, the result

$$\text{(4.18)} \qquad \begin{aligned} &n^{\frac{1}{2}}(T(h_n^{(n)}) - \theta) \\ &\quad = -[\int [s'(x)]^2\, dx]^{-1} n^{\frac{1}{2}} \int h_n'(x)[h_n^{(n)}(x) - h_n(x)]\, dx + o_p(1)\,.\end{aligned}$$

The reduction of (4.18) to (4.7) parallels the proof of Theorem 3 with A5 being used in place of A3. The essential step is to show that

$$\text{(4.19)} \qquad \lim_{n\to\infty} n^{\frac{1}{2}} \int s_n'(x)(2s_n^{-1}(x))[f_n^{(n)}(x) - f_n(x)]\, dx = \int s'(x)\beta(x)\, dx\,,$$

where $f_n^{(n)}(x) = [s_n^{(n)}(x)]^2$, and that

$$\text{(4.20)} \qquad \lim_{n\to\infty} n^{\frac{1}{2}} \int s_n'(x) s_n^{-1}(x)[s_n^{(n)}(x) - s_n(x)]^2\, dx = 0\,.$$

The limit (4.20) follows from (4.11) and the bound $\sup_x |s_n'(x)|/s_n(x) = O(c_n^{-1})$. The left side of (4.19) can be written as the sum of two terms, putting $\beta_n(x) = n^{\frac{1}{2}}(s^{(n)}(x) - s(x))$ for brevity:

$$\text{(4.21)} \qquad \begin{aligned} Y_{1n} &= \int s_n'(x) s_n^{-1}(x) \int \beta_n(x - c_n z) s(x - c_n z) w(z)\, dz\, dx \\ Y_{2n} &= (2n^{\frac{1}{2}})^{-1} \int s_n'(x) s_n^{-1}(x) \int \beta_n^2(x - c_n z) w(z)\, dz\, dx\,.\end{aligned}$$

Evidently, $Y_{2n} = O[(n^{\frac{1}{2}}c_n)^{-1}]$.

Since $\lim_{n\to\infty} ||\beta_n - \beta|| = 0$, it follows by Cauchy-Schwarz, using (3.24), that

$$\text{(4.22)} \qquad Y_{1n} = \int s_n'(x) s_n^{-1}(x) \int \beta(x - c_n z) s(x - c_n z) w(z)\, dz\, dx + o(1)\,.$$

Since $\beta \in L_2$, there exists for every $\varepsilon > 0$ a differentiable function $\phi_\varepsilon \in L_2$ such that $\phi_\varepsilon' \in L_2$ and $||\beta - \phi_\varepsilon|| \leq \varepsilon$. Thus

$$\text{(4.23)} \qquad |Y_{1n} - \int s_n'(x) s_n^{-1}(x) \int \phi_\varepsilon(x - c_n z) s(x - c_n z) w(z)\, dz\, dx| \leq \varepsilon ||s'|| + o(1)\,.$$

By dominated convergence, for every $\varepsilon > 0$,

$$\text{(4.24)} \qquad \lim_{n\to\infty} \int w(z) \int [\phi_\varepsilon(x - c_n z) - \phi_\varepsilon(x)]^2\, dx\, dz = 0\,.$$

Equations (4.23) and (4.24) imply, for every $\varepsilon > 0$,

$$\text{(4.25)} \qquad |Y_{1n} - \int s_n'(x)\phi_\varepsilon(x)\, dx| \leq \varepsilon ||s'|| + o(1)$$

because $\phi_\varepsilon(x)s_n^{-1}(x) \int s(x - c_n z)w(z)\,dz$ converges in L_2 to $\phi_\varepsilon(x)$, by Vitali's theorem. Hence $\lim_{n\to\infty} Y_{1n} = \int s'(x)\beta(x)\,dx$, which gives (4.19).

5. Numerical example. The estimator $\hat{\theta}_n$ is defined as that value (or values) of real t which maximizes $\int \hat{h}_n(-x + 2s)\hat{h}_n(x)\,dx$ for $\hat{h}_n$ described in Section 2. If $\hat{\theta}^{(0)}$ denotes a reasonable initial guess at $\hat{\theta}_n$, Newton's method yields the iterative algorithm

$$\hat{\theta}^{(m+1)} = \hat{\theta}^{(m)} - [2 \int \hat{h}_n'(x)\hat{h}_n'(-x + 2\hat{\theta}^{(m)})\,dx]^{-1} \times \int \hat{h}_n'(x)\hat{h}_n(-x + 2\hat{\theta}^{(m)})\,dx\,, \quad m \geqq 0\,, \tag{5.1}$$

where the integrals are to be evaluated numerically.

To check the computational feasibility and finite sample behavior of $\hat{\theta}_n$, a simple numerical experiment was performed. A pseudo-random sample of size 40 was drawn by computer from a $N(0, 1)$ distribution. In the first stage of the experiment, a calibration trial, the estimate $\hat{\theta}_n$ was calculated for various values of c_n. For practical reasons, some small modifications were made in applying the iterative scheme (5.1). Specifically:

(i) The smoothing window used was $w(x) = (\frac{15}{16})(1 - x^2)^2$ for $|x| \leqq 1$. This window does not satisfy A2 because $|w'(x)|/w(x)$ is unbounded; however a numerically insignificant modification to the tails of w would remedy this.

(ii) $\hat{h}_n(x)$ was replaced by

$$\tilde{h}_n(x) = \{(nc_n\hat{\sigma})^{-1} \textstyle\sum_{i=1}^n w[(c_n\hat{\sigma})^{-1}(x - X_i)]\}^{\frac{1}{2}}\,,$$

where $\hat{\sigma}$ is a scale estimate. If $\hat{\theta}^{(0)}$ is a scale invariant-location estimate, the introduction of $\hat{\sigma}$ makes $\hat{\theta}_n^{(m)}$ scale invariant for every $m \geqq 1$.

(iii) The choices $\hat{\theta}^{(0)} = \text{median}\{X_i\}$ and $\hat{\sigma} = (.674)^{-1}\,\text{median}\{|X_i - \hat{\theta}^{(0)}|\}$ were made. Both are robust at the normal distribution and, under normality, are root-n consistent estimates of the distribution mean and standard deviation respectively.

(iv) Each integral in (5.1) was approximated by the trapezoidal rule, using a grid of 100 points spaced equally over the support of the integrands. Note that the chosen window w makes both integrands continuous functions, a prerequisite for accurate numerical integration.

(v) Iteration of the algorithm was continued until convergence to six significant figures had been achieved (this was the number of figures provided by the PDP 11/45 computer used).

For every case examined during the experiment, no more than three iterations were required to achieve the convergence criterion (v). The third iterate was taken as the value of $\hat{\theta}_n$. In each case, the negative sign of $\int \tilde{h}_n'(x)\tilde{h}_n'(-x+2\hat{\theta}_n)\,dx$ showed that a local maximum had been attained; it is not known whether the respective global maxima were achieved. The mean and median of the sample were .158 and .0926 respectively. With $c_n = 3.2$, the values of $\hat{\theta}_n$ was .158, the same as the sample mean. As c_n increased from 3 to 4, the corresponding values of $\hat{\theta}_n$ increased monotonically from .156 to .162. Not enough estimate values were computed to establish a pattern for choices of c_n outside [3, 4]; however, the monotone dependence on c_n persists further.

The second stage of the experiment examined the response of $\hat{\theta}_n$ to a single outlier moving towards infinity. Specifically, the observation nearest zero in the data set, $X_{22} = -.0192038$, was replaced by a series of increasing positive values, the other 39 observations being left unchanged. For each contaminated sample so generated, the estimate $\hat{\theta}_n$ was recomputed with $c_n = 3.2$, the value of c_n which, on the original data, best aligns $\hat{\theta}_n$ with the sample mean. This choice of c_n simplifies comparisons between the two estimates as X_{22} is varied. While the asymptotic equivalence under normality of $\hat{\theta}_n$ and the sample mean (see Theorem 3) suggests that this value of c_n may be reasonable for other normal samples of size 40, the theoretical question of how to choose c_n well remains open. Table 1 records the results for contaminated samples.

TABLE 1

Effect on $\hat{\theta}_n$ of varying X_{22} when $c_n = 3.2$; $\hat{\theta}_n$ remains unchanged for $X_{22} \geqq 7$

X_{22}	$-.0192038$	1	2	3	4	5	6	7
Sample mean	.158	.184	.209	.234	.259	.284	.309	.334
Adaptive estimate $\hat{\theta}_n$ ($c_n = 3.2$)	.158	.185	.210	.235	.225	.191	.166	.164

For values of X_{22} consistent with the assumption that the entire sample is drawn from a $N(0, 1)$ distribution, the estimate $\hat{\theta}_n$ follows the sample mean very closely. But for $X_{22} \geqq 4$ in Table 1, the adaptive estimator recognizes X_{22} as a possible outlier and begins to discount it smoothly. For $X_{22} \geqq 7$ the value of $\hat{\theta}_n$ remains unchanged because the estimate is now computed entirely from the other 39 observations; this follows from our choice of w and c_n. The behavior of $\hat{\theta}_n$ in this example accords with what we might expect from an efficient and robust estimator.

REFERENCES

BERAN, R. (1974). Asymptotically efficient adaptive rank estimates in location models. *Ann. Statist.* **2** 63–74.

BERAN, R. (1977a). Robust location estimates. *Ann. Statist.* **5** 431–444.

BERAN, R. (1977b). Minimum Hellinger distance estimates for parametric models. *Ann. Statist.* **5** 445–463.

BHATTACHARYA, P. K. (1967). Efficient estimation of a shift parameter from grouped data. *Ann. Math. Statist.* **38** 1770–1787.

FABIAN, U. (1973). Asymptotically efficient stochastic approximation: the RM case. *Ann. Statist.* **1** 486–495.

HÁJEK, J. (1970). A characterization of limiting distributions of regular estimates. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **14** 323–330.

HAMPEL, F. R. (1971). A general qualitative definition of robustness. *Ann. Math. Statist.* **42** 1887–1896.

HOGG, R. V. (1974). Adaptive robust procedures; a partial review and some suggestions for future applications and theory. *J. Amer. Statist. Assoc.* **69** 909–925.

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

HUBER, P. J. (1972). Robust statistics: a review. *Ann. Math. Statist.* **43** 1041–1067.

HUBER, P. J. (1974). Comment on "Adaptive robust procedures ···" by R. H. Hogg. *J. Amer. Statist. Assoc.* **69** 926–927.

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

JAECKEL, L. A. (1971 b). Some flexible estimates of location. *Ann. Math. Statist.* **42** 1540–1552.

JOHNS, M. V., JR. (1974). Nonparametric estimation of location. *J. Amer. Statist. Assoc.* **69** 453–460.

LE CAM, L. (1969). *Théorie Asymptotique de la Décision Statistique*. Les Presses de l'Université de Montréal.

RAO, P. V., SCHUSTER, E. F. and LITTEL, R. C. (1975). Estimation of shift and center of symmetry based on Kolmogorov-Smirnov statistics. *Ann. Statist.* **3** 862–873.

SACKS, J. (1975). An asymptotically efficient sequence of estimators of a location parameter. *Ann. Statist.* **3** 285–298.

SAMANTA, M. (1974). Efficient nonparametric estimation of a shift parameter. *Sankhyā Ser. A* **36** 273–292.

SKOROKHOD, A. V. (1956). Limit theorems for stochastic processes. *Theor. Probability Appl.* **1** 261–290.

STEIN, C. J. (1956). Efficient nonparametric testing and estimation. *Proc. Third Berkeley Symp. Math. Statist. Prob.* **1** 187–196.

STONE, C. J. (1975). Adaptive maximum likelihood estimators of a location parameter. *Ann. Statist.* **3** 267–284.

TAKEUCHI, K. (1971). A uniformly asymptotically efficient estimator of a location parameter. *J. Amer. Statist. Assoc.* **66** 292–301.

VAN EEDEN, C. (1970). Efficiency-robust estimation of location. *Ann. Math. Statist.* **41** 172–181.

WEISS, L. and WOLFOWITZ, J. (1970). Asymptotically efficient non-parametric estimators of location and scale parameters. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **16** 134–150.

WOLFOWITZ, J. (1974). Asymptotically efficient non-parametric estimators of location and scale parameters II. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **30** 117–128.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720