The Annals of Statistics
1975, Vol. 3, No. 2, 267–284

ADAPTIVE MAXIMUM LIKELIHOOD ESTIMATORS OF A LOCATION PARAMETER[1]

BY CHARLES J. STONE

University of California, Los Angeles

Consider estimators $\hat{\theta}_n$ of the location parameter θ based on a sample of size n from $\theta + X$, where the random variable X has an unknown distribution F which is symmetric about the origin but otherwise arbitrary. Let $\mathscr{I}$ denote the Fisher information on θ contained in $\theta + X$. We show that there is a nonrandomized translation and scale invariant adaptive maximum likelihood estimator $\hat{\theta}_n$ of θ which does not depend on F such that $\mathscr{L}(n^{\frac{1}{2}}(\hat{\theta}_n - \theta)) \to N(0, 1/\mathscr{I})$ as $n \to \infty$ for *all* symmetric F.

1. Introduction and summary of results. Let X be a random variable having distribution F. Let $\mathscr{I}$ denote the Fisher information on θ contained in $\theta + X$. If X has an absolutely continuous density f such that $\int (f'/f)^2 f\,dx < \infty$, then $\mathscr{I} = \int (f'/f)^2 f\,dx$. Otherwise $\mathscr{I} = \infty$.

Suppose we observe a sample $\theta + X_1, \cdots, \theta + X_n$ of size n from $\theta + X$ and wish to estimate the unknown location parameter θ from this sample. A possibly randomized estimator $\hat{\theta}_n$ of θ can be written as $\hat{\theta}_n = \psi(\theta + X_1, \cdots, \theta + X_n, W)$, where the random variable W is independent of $X_1, \cdots, X_n$. Such an estimator is *translation invariant* if $\psi(\theta + x_1, \cdots, \theta + x_n, w) \equiv \theta + \psi(x_1, \cdots, x_n, w)$ and *scale invariant* if $\psi(ax_1, \cdots, ax_n, w) \equiv a\psi(x_1, \cdots, x_n, w)$ for $a > 0$.

In Stone [18] we verified the existence under no regularity conditions of a nonrandomized translation invariant estimator $\hat{\theta}_n$ such that

$$\mathscr{L}(n^{\frac{1}{2}}(\hat{\theta}_n - \theta)) \to N(0, 1/\mathscr{I}) \quad \text{as} \quad n \to \infty\,. \tag{1.1}$$

If $\mathscr{I} < \infty$ the existence of such an estimator also follows from Proposition 6 of Le Cam [11]. If $\mathscr{I} = \infty$, then (1.1) is of course equivalent to

$$n^{\frac{1}{2}}(\hat{\theta}_n - \theta) \to 0 \quad \text{in probability as} \quad n \to \infty\,. \tag{1.2}$$

It follows from [18] or from Hájek [6] that, in various respects, it is impossible to improve upon (1.1). For example, if $\tilde{\theta}_n$ is a translation invariant estimator of θ such that $\mathscr{L}(n^{\frac{1}{2}}(\tilde{\theta}_n - \theta)) \to N(0, \sigma^2)$, then $\sigma^2 \geqq 1/\mathscr{I}$.

The estimators $\hat{\theta}_n$ constructed in [18] and [11] that satisfy (1.1) depend on the shape of F. It is obviously desirable to find an estimator of θ that satisfies (1.1) but is independent of F. Unless F is somehow restricted this is impossible, since we can change θ and F simultaneously so as to leave the distribution of $\theta + X$ unaltered. To avoid this difficulty we make the usual assumption that F is

Received September 1973; revised June 1974.

[1] This research was sponsored by the National Science Foundation under Grant Number GP-33431X.

AMS 1970 *subject classifications.* Primary 62F10; Secondary 62G35.

Key words and phrases. Location parameter, adaptive estimators.

symmetric about the origin, i.e., that X and $-X$ have the same distribution. In this paper we will show that *no restrictions other than symmetry* need be placed on F. Specifically we will verify the following result.

THEOREM 1.1. *There is a nonrandomized translation and scale invariant estimator $\hat{\theta}_n$ of θ which does not depend on F and which satisfies* (1.1) *for all symmetric F.*

The need for the estimator $\hat{\theta}_n$ of θ to be translation and scale invariant is easily seen. Suppose, for example, that we want to estimate the true temperature based on n temperature measurements. If our estimator yields a value $\hat{\theta}_n$ when applied to measurements $\theta + X_i$, $1 \leqq i \leqq n$, in Fahrenheit degrees it should yield the value $(\frac{5}{9})(\hat{\theta}_n - 32)$ when applied to the corresponding measurements $(\frac{5}{9})(\theta + X_i - 32)$, $1 \leqq i \leqq n$, in centigrade degrees.

Stein [17] first suggested that results along the lines of Theorem 1.1 might hold. Under a variety of further regularity conditions on F such results have previously been achieved by using rank estimators (van Eeden [21], Beran [2]), linear combination of order statistics (Takeuchi [20], Johns [10], Sacks [16]), and stochastic approximation (Fabian [4]). Corresponding results in the two-sample problem have been obtained by Bhattacharya [3], van Eeden [21], and Beran [2]. Weiss and Wolfowitz [22] considered simultaneous estimators of location and scale parameters in the two-sample problem and Wolfowitz [23] continued this work for scale parameters. (Further research involving these and other parameters appears quite promising.) There are of course many other approaches to the problem of estimating a location parameter (see, for example, Andrews *et al.*, [1], Miké [12] and [13], Huber [9], and Switzer [19]).

Beran's elegant paper [2], which appeared after the present paper was originally written, contains a result which is close to Theorem 1.1. He showed that there is a nonrandomized translation invariant adaptive rank estimator which satisfies (1.1) for all symmetric F which have finite Fisher information. His estimator appears not to be scale invariant. He obtained a similar result for the two-sample problem. Undoubtedly the appropriate analogy of Theorem 1.1 also holds for the two-sample problem.

Pfanzagl [14] obtained some interesting negative results. He showed that in the one-sample problem without any assumptions of symmetry one cannot expect asymptotically to improve upon the sample median in estimating the true median of a distribution of unknown shape.

We will now describe the adaptive maximum likelihood estimators that will be shown to satisfy (1.1). Firstly, it is easy to construct nonrandomized translation and scale invariant estimators $\bar{\theta}_n$ of θ which do not depend on F and which satisfy

$$n^{\frac{1}{2}}(\bar{\theta}_n - \theta) = O_p(1) \qquad \text{as} \quad n \to \infty\,, \tag{1.3}$$

provided of course that F is symmetric. For example, let J be a fixed continuously differentiable probability density having compact support in (0, 1) and such that

$J(1-x) \equiv J(x)$. Set $\bar{\theta}_n = J_n^{-1} \sum_1^n J(i/(n+1))(\theta + X_{(i)})$, where $\theta + X_{(i)}$, $1 \leqq i \leqq n$, are the order statistics from the sample and $J_n = \sum_1^n J(i/(n+1))$. In Chapter III of Huber [8] it is shown that $\mathscr{L}(n^{\frac{1}{2}}(\bar{\theta}_n - \theta)) \to N(0, \sigma^2)$ as $n \to \infty$ for some σ. This implies that (1.3) holds.

In order to motivate our adaptive estimators, let us briefly look at maximum likelihood estimators. Suppose that F has a known density f that satisfies appropriate regularity conditions which guarantee, in particular, that $\mathscr{I} < \infty$. Set $L(x) = f'(x)/f(x)$. We can find an estimator $\hat{\theta}_n$ satisfying (1.1) by solving the maximum likelihood equation

$$\sum_{i=1}^n L(\theta + X_i - \hat{\theta}_n) = 0 . \tag{1.4}$$

Alternatively we can think of $\bar{\theta}_n$ as an approximate solution to (1.4) and, using a Newton–Raphson approach, define a better approximation by

$$\hat{\theta}_n = \bar{\theta}_n + \frac{\frac{1}{n} \sum_{i=1}^n L(\theta + X_i - \bar{\theta}_n)}{\frac{1}{n} \sum_{i=1}^n L'(\theta + X_i - \bar{\theta}_n)} . \tag{1.5}$$

We would expect that the denominator in (1.5) converges to $EL'(X) = -\mathscr{I}$ in probability as $n \to \infty$. Thus we are led from (1.5) to the estimator

$$\hat{\theta}_n = \bar{\theta}_n - \frac{1}{n\mathscr{I}} \sum_{i=1}^n L(\theta + X_i - \bar{\theta}_n) . \tag{1.6}$$

We can modify this estimator to control for poor behavior of the tails of L. Let g be a twice continuously differentiable symmetric function having support in $[-1, 1]$ and such that $0 \leqq g \leqq 1$ and $g(0) = 1$. For $c > 0$ set $A(c) = \int L^2(x)g(x/c)f(x)\,dx$. Then an appropriate modification to (1.6) is

$$\hat{\theta}_n = \bar{\theta}_n - \frac{1}{nA(c_n)} \sum_{i=1}^n L(\theta + X_i - \bar{\theta}_n) g((\theta + X_i - \bar{\theta}_n)/c_n) , \tag{1.7}$$

where $c_n \to \infty$ slowly as $n \to \infty$.

In actuality F need not have a smooth density. To take care of this possibility let W be a standard normal random variable having density $\varphi(x) \equiv (2\pi)^{-\frac{1}{2}} e^{-x^2/2}$. For $r > 0$ set $\varphi(x, r) = r^{-1}\varphi(x/r)$. Suppose that W is independent of X. Then $X + rW$ has a smooth density $f(\cdot; r)$ given by $f(x; r) = \int \varphi(x-y; r)F(dy)$. Set $f'(x; r) = (\partial/\partial x)f(x; r)$, $L(x; r) = f'(x; r)/f(x; r)$, and $A(r, c) = \int L^2(x; r)g(x/c) \times f(x; r)\,dx$. Let $W_1, \cdots, W_n$ be independent standard normal random variables which also are independent of the X_i's. Let r_n be positive numbers that approach zero slowly as $n \to \infty$. Then an appropriate modification to (1.7) is

$$\hat{\theta}_n = \bar{\theta}_n - \frac{1}{nA(r_n, c_n)} \sum_{i=1}^n L(\theta + X_i + r_n W_i - \bar{\theta}_n; r_n) \times g((\theta + X_i + r_n W_i - \bar{\theta}_n)/c_n) . \tag{1.8}$$

This estimator is randomized since it depends on the W_i's. We can get a better nonrandomized estimator by taking the conditional expectation with respect to the X_i's, obtaining

$$\hat{\theta}_n = \bar{\theta}_n - \frac{1}{nA(r_n, c_n)} \sum_{i=1}^n \int L(\theta + X_i + x - \bar{\theta}_n; r_n) \times g((\theta + X_i + x - \bar{\theta}_n)/c_n)\varphi(x; r_n)\, dx\,. \tag{1.9}$$

It is interesting to note that if we take $g \equiv 1$ (or equivalently $c_n = \infty$) in (1.9), then when the X_i's are normally distributed with mean zero (1.9) reduces to $\theta_n = n^{-1} \sum_1^n (\theta + X_i)$, which is the maximum likelihood estimator in this case.

The estimator in (1.9) depends through L on knowledge of F. To eliminate this dependence we need to estimate $f(x; r)$. There is a very large literature on nonparametric estimators of densities (the bibliography in Good and Gaskins [5] contains 100 items). The estimator we will use, based upon the assumed symmetry of F, is

$$\hat{f}_n(x; r) = \frac{1}{2n} \sum_{j=1}^n [\varphi(x + \bar{\theta}_n - \theta - X_j; r) + \varphi(-x + \bar{\theta}_n - \theta - X_j; r)]\,.$$

Observe that $\hat{f}_n(\cdot; r)$ is a symmetric probability density. Set $\hat{f}_n'(x; r) = (\partial/\partial x)\hat{f}_n(x; r)$, $\hat{L}_n(x; r) = \hat{f}_n'(x; r)/\hat{f}_n(x; r)$, and

$$\hat{A}_n(r, c) = \int \hat{L}_n^2(x; r) g(x/c)\hat{f}_n(x; r)\, dx\,. \tag{1.10}$$

The appropriate modification of (1.9) is

$$\hat{\theta}_n = \bar{\theta}_n - \frac{1}{n\hat{A}_n(r_n, c_n)} \sum_{i=1}^n \int \hat{L}_n(\theta + X_i + x - \bar{\theta}_n; r_n) \times g((\theta + X_i + x - \bar{\theta}_n)/c_n)\varphi(x; r_n)\, dx\,. \tag{1.11}$$

To choose r_n and c_n we first choose s_n to be positive random variables depending only on $X_n, \cdots, X_n$ (and not on F). We will need to assume that

$$n^{\frac{1}{2}}\left(\frac{s_n}{s} - 1\right) = O_p(1) \qquad \text{as} \quad n \to \infty\,, \tag{1.12}$$

where the nonrandom positive number s can depend on F. Let t_n and d_n be positive constants that approach zero and infinity respectively as $n \to \infty$. Set $r_n = s_n t_n$ and $c_n = s_n d_n$. Then the estimator given by (1.11) is nonrandomized, translation invariant, and does not depend on F.

According to Theorem 5.3, *if*

$$\frac{d_n^2}{n^{1-\varepsilon} t_n^6} = O(1) \qquad \textit{as} \quad n \to \infty \quad \textit{for some} \quad \varepsilon > 0\,, \tag{1.13}$$

then the estimator θ_n *given by* (1.11) *satisfies* (1.1).

In order to complete the proof of Theorem 1.1 we need to see how $\hat{\theta}_n$ can be chosen to be scale invariant. For theoretical reasons it is convenient to separate continuous distributions F from those that are discontinuous. If the values of $\theta + X_1, \cdots, \theta + X_n$ are not all distinct, we know that F has a discontinuous distribution. It is easy to find a nonrandomized translation and scale invariant estimator $\hat{\theta}_n$ of θ which does not depend on F and which is such that $P(\theta_n \neq \theta) \to 0$ as $n \to \infty$ whenever F is a symmetric discontinuous distribution. Such an estimator clearly satisfies (1.2) and hence (1.1) since $\mathscr{I} = \infty$ whenever F is discontinuous. If the values of $\theta + X_1, \cdots, \theta + X_n$ are all distinct, we assume that F has a continuous distribution. The probability of our being wrong in this respect approaches zero as $n \to \infty$ regardless of F.

Thus in verifying the conclusion of Theorem 1.1 we can assume that F has a continuous distribution. We can also assume that $n \geqq 2$ (take $\hat{\theta}_1 = \theta + X_1$). Write the random variable s_n as $s_n = \psi(\theta + X_1, \cdots, \theta + X_n)$. It is not hard to choose ψ such that $\psi(x_1, \cdots, x_n) > 0$ whenever $x_1, \cdots, x_n$ are all distinct, $\psi(\theta + x_1, \cdots, \theta + x_n) \equiv \theta + \psi(x_1, \cdots, x_n)$, $\psi(ax_1, \cdots, ax_n) \equiv a\psi(x_1, \cdots, x_n)$ for $a > 0$, and (1.12) holds whenever F has a continuous distribution. With this choice of s_n the estimator given by (1.11) is scale invariant. Thus Theorem 1.1 follows from Theorem 5.3.

In Section 2 we discuss properties of $f(x; r)$. In Section 3 we discuss properties of

$$f_n(x; r) = \frac{1}{n} \sum_{i=1}^n \varphi(x - X_1; r) .$$

This is relevant for two reasons. Firstly

$$\hat{f}_n(x: r) = \tfrac{1}{2}(f_n(x + \bar{\theta}_n - \theta; r) + f_n(-x + \bar{\theta}_n - \theta; r)) .$$

Secondly, the estimators given by (1.9) and (1.11) can be rewritten respectively as

$$(1.14) \qquad \hat{\theta}_n = \bar{\theta}_n - \frac{1}{A(r_n, c_n)} \int L(x; r_n) g(x/c_n) f_n(x + \bar{\theta}_n - \theta; r_n)\, dx$$

and

$$(1.15) \qquad \hat{\theta}_n = \bar{\theta}_n - \frac{1}{\hat{A}_n(r_n, c_n)} \int \hat{L}_n(x; r_n) g(x/c_n) f_n(x + \bar{\theta}_n - \theta; r_n)\, dx .$$

In Sections 4 and 5 it is assumed that $\bar{\theta}_n$ satisfies (1.3) and that s_n satisfies (1.12). We denote the estimator given by (1.9) or (1.14) as $\tilde{\theta}_n$ (the definition of $\tilde{\theta}_n$ is actually modified in Section 4 to cover the asymmetric case). In Theorem 4.2 we show that if (1.13) holds, then $\mathscr{L}(n^{\frac{1}{2}}(\tilde{\theta}_n - \theta)) \to N(0, 1/\mathscr{I})$ as $n \to \infty$. In Theorem 5.2 we show that if (1.13) holds, then $n^{\frac{1}{2}}(\tilde{\theta}_n - \hat{\theta}_n) \to 0$ in probability as $n \to \infty$. In Theorem 5.3 we put these two facts together to conclude that if (1.13) holds, then $\hat{\theta}_n$ satisfies (1.1).

In order to get some feeling for the actual variance of the estimator given by (1.11) for specific values of n and F a small scale Monte Carlo study was performed. (Random numbers were generated as suggested in [1] and 3000 trials were made.) We used $n = 40$ and five densities. They were
normal:

$$f(x) = \varphi(x) .$$

contaminated normal:

$$f(x) = .9\varphi(x) + \frac{.1}{3}\varphi\left(\frac{x}{3}\right).$$

double exponential:

$$f(x) = \tfrac{1}{2}e^{-|x|}.$$

Cauchy:

$$f(x) = \frac{1}{\pi(1+x^2)}.$$

rectangular:

$$\begin{aligned} f(x) &= \tfrac{1}{2}, \qquad -1 \leqq x \leqq 1, \\ &= 0, \qquad |x| > 1. \end{aligned}$$

Note that the rectangular density has infinite Fisher information.

After some preliminary experimentation it seemed appropriate to set $\bar{\theta}_n = \text{median}(\theta + X_1, \cdots, \theta + X_n)$ and $s_n = \text{median}(|\theta + X_1 - \bar{\theta}_n|, \cdots, |\theta + X_n - \bar{\theta}_n|)$. We used $t_{40} = .6$. The function g was chosen as

$$\begin{aligned} g(x) &= 1, \qquad -1 \leqq x \leqq 1, \\ &= 0, \qquad |x| > 1. \end{aligned}$$

For the distributions studied the variances seemed to be constant (slightly decreasing for the Cauchy distribution) in d_{40} for $10 \leqq d_{40} \leqq 20$, so we chose $d_{40} = 20$. The integrals in (1.10) and (1.11) were satisfactorily evaluated by dividing the range into 60 intervals of equal length and calculating the corresponding Riemann sums.

The efficiency of the estimator is obtained by comparing its variance to that of the Pitman estimator P_n for the given distribution. The Pitman estimator in the normal case is the sample mean. In the rectangular case it is the average of the minimum value and the maximum value in the sample. In the three other cases it was readily computed numerically. Since P_n and $\hat{\theta}_n - P_n$ are orthogonal random variables,

$$\text{Eff}_{\hat{\theta}_n} = \frac{\text{Var}\, P_n}{\text{Var}\, P_n + \text{Var}\,(\hat{\theta}_n - P_n)}.$$

The resulting efficiencies along with those of the median, midmean (average of the middle $n/2$ observations), and mean are as follows:

Estimator	Normal	C.N.	Dexp.	Cauchy	Rect.
$\hat{\theta}_n$	.93	.91	.89	.86	.15
median	.67	.73	.96	.82	.05
midmean	.85	.91	.93	.77	.07
mean	1.00	.71	.59	.00	.14

The observation following (1.9) explains why the estimator $\hat{\theta}_n$ should work particularly well for nearly normal distributions even if t_n is not small.

2. Properties of the smoothed density. Let r_n, $n \geqq 1$, be positive random variables that approach zero in probability as $n \to \infty$. We write $g(x; r_n) = O_p(1)$ if $\sup_x |g(x; r_n)| = O_p(1)$.

Let $\varphi(x; r)$ and $f(x; r)$ be as in the Introduction. Let $\varphi^{(\nu)}(x; r)$ and $f^{(\nu)}(x; r)$ denote differentiation with respect to x. Then $\varphi^{(\nu)}(x; r) = r^{-(\nu+1)}\varphi^{(\nu)}(x/r)$ and $f^{(\nu)}(x; r) = \int \varphi^{(\nu)}(x - y; r)F(dy)$. By Schwarz's inequality

$$(f^{(\nu)}(x; r))^2 \leqq \int (\varphi^{(\nu)}(x - y; r))^2 F(dy) .$$

Consequently

$$\int (f^{(\nu)}(x; r_n))^2 \, dx = O(1) r_n^{-(2\nu+1)} . \tag{2.1}$$

It is easily seen that

$$\frac{\partial}{\partial r} \varphi^{(\nu)}(x; r) = r\varphi^{(\nu+2)}(x; r) . \tag{2.2}$$

Consequently

$$\frac{\partial}{\partial r} f^{(\nu)}(x; r) = r f^{(\nu+2)}(x; r) . \tag{2.3}$$

It follows from Hölder's inequality that

$$\int_{-c}^{c} f^{1-\varepsilon}(x; r)\, dx \leqq (2c)^{\varepsilon} , \qquad c > 0 \text{ and } 0 \leqq \varepsilon \leqq 1 . \tag{2.4}$$

PROPOSITION 2.1. *For* $0 < q \leqq r$

$$\frac{q}{r} f(x; q) \leqq f(x; r) \leqq \frac{q^{q^2/r^2}}{r} (f(x; q))^{q^2/r^2} .$$

PROOF. The first inequality is immediate and the second one follows from Hölder's inequality.

PROPOSITION 2.2. *For* $\varepsilon > 0$ *and* ν *a positive integer*

$$f^{(\nu)}(x; r_n) = O_p(1) r_n^{-(\nu+\varepsilon)}(f(x; r_n) + f^{1-\varepsilon}(x; r_n)) .$$

PROOF. Observe that $\varphi^{(\nu)}(x) = P_\nu(x)\varphi(x)$, where P_ν is a polynomial of degree ν with leading coefficient $(-1)^\nu$. There is an $L_0 > 0$ such that for $L \geqq L_0$

$$|\varphi^{(\nu)}(x)| \leqq |\varphi^{(\nu)}(L)| , \qquad x \geqq L ,$$

$$|\varphi^{(\nu)}(x)| \leqq |\varphi^{(\nu)}(-L)| , \qquad x \leqq -L ,$$

and

$$|P_\nu(x)| \leqq 2L^\nu , \qquad |x| \leqq L .$$

We conclude that for $L \geqq L_0$

$$|\varphi^{(\nu)}(x)| \leqq 2L^\nu(\varphi(x) + \varphi(L)) , \qquad -\infty < x < \infty .$$

Consequently there is a $\delta_0 > 0$ such that for $0 < \delta \leqq \delta_0$

$$|\varphi^{(\nu)}(x)| \leqq 2\delta^{-\varepsilon}(\varphi(x) + \varphi(\delta^{-\varepsilon/\nu})) . \qquad -\infty < x < \infty .$$

We can assume that $\varphi(\delta^{-\varepsilon/\nu}) \leqq \delta^2$ for $0 < \delta \leqq \delta_0$. Then for $0 < \delta \leqq \delta_0$

$$|\varphi^{(\nu)}(x)| \leqq 2\delta^{-\varepsilon}(\varphi(x) + \delta^2) , \qquad -\infty < x < \infty ,$$

and hence

$$|\varphi^{(\nu)}(x; r_n)| \leqq 2r_n^{-\nu}\delta^{-\varepsilon}(\varphi(x; r_n) + \delta^2 r_n^{-1}), \qquad -\infty < x < \infty .$$

Thus for $0 < \delta \leqq \delta_0$

$$|f^{(\nu)}(x; r_n)| \leqq 2r_n^{-\nu}\delta^{-\varepsilon}(f(x; r_n) + \delta^2 r_n^{-1}), \qquad -\infty < x < \infty .$$

If $r_n \leqq \delta_0$, we can take $\delta = \min(r_n, f(x; r_n))$ in the preceding inequality to conclude that

$$|f^{(\nu)}(x; r_n)| \leqq 4r_n^{-\nu} f(x; r_n)(r_n^{-\varepsilon} + f^{-\varepsilon}(x; r_n)), \qquad -\infty < x < \infty .$$

The conclusion of Proposition 2.2 now follows easily.

PROPOSITION. 2.3. *Suppose that* $1/nr_n^3 = O_p(1)$. *Then for* $\varepsilon > 0$ *and* $M > 0$

$$\max_{|t| \leqq Mn^{-\frac{1}{2}}} f(x + t; r_n) = O_p(1)(f(x; r_n) + f^{1-\varepsilon}(x; r_n)) .$$

PROOF. Set $a_n = r_n^2 n^{\frac{1}{2}} M^{-1}$. If $L \geqq 1$, $|t| < Mn^{-\frac{1}{2}}$ and $|x| \leqq a_n \log L$, then

$$\frac{\varphi(x + t; r_n)}{\varphi(x; r_n)} = \exp\left[-\frac{2xt + t^2}{2r_n^2}\right] \leqq \exp\left[-\frac{xt}{r_n^2}\right] \leqq \exp[\log L] = L .$$

We conclude that for $L \geqq 1$

$$\max_{|t| \leqq Mn^{-\frac{1}{2}}} \varphi(x + t; r_n) \leqq L\varphi(x; r_n) , \qquad |x| \leqq a_n \log L ,$$

and hence

$$\begin{aligned} f(x + t; r_n) &= \int \varphi(x + t - y; r_n)F(dy) \\ &\leqq Lf(x; r_n) + \int_{|x-y| \geqq a_n \log L} \varphi(x + t - y; r_n)F(dy) . \end{aligned}$$

Choose $\delta > 0$ and $L_0 > 1$. Then for n sufficiently large, except on a set having probability at most δ,

$$\begin{gathered} \varphi(x + t - y; r_n) \leqq L^{-1/\varepsilon} \quad \text{for} \quad L \geqq L_0 , \\ |x - y| \geqq a_n \log L , \qquad \text{and} \qquad |t| \leqq Mn^{-\frac{1}{2}} . \end{gathered}$$

Thus for n sufficiently large, except on a set having probability at most δ,

$$\max_{|t| \leqq Mn^{-\frac{1}{2}}} f(x + t; r_n) \leqq Lf(x; r_n) + L^{-1/\varepsilon} , \qquad L \geqq L_0 \text{ and } -\infty < x < \infty .$$

By setting $L = \max(f^{-\varepsilon}(x; r_n), 2^\varepsilon)$, we conclude that for n sufficiently large, except on a set having probability at most δ,

$$\max_{|t| \leqq Mn^{-\frac{1}{2}}} f(x + t; r_n) \leqq 2^\varepsilon f^{1-\varepsilon}(x; r_n) + (1 + 2^\varepsilon)f(x; r_n) .$$

Thus the conclusion of the proposition holds.

3. Properties of the empirical smoothed density. Let $X_1, X_2, \cdots$ be independent random variables each having distribution F. Let F_n be the empirical distribution given by

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n 1_{\{X_i \leqq x\}} .$$

Let $f_n(\cdot; r)$ be the random density defined by

$$f_n(x; r) = \frac{1}{n} \sum_{i=1}^n \varphi(x - X_i; r) = \int \varphi(x - y; r)F_n(dy) .$$

Then

$$f_n^{(\nu)}(x; r) = \frac{1}{n} \sum_{i=1}^n \varphi^{(\nu)}(x - X_i; r) = \int \varphi^{(\nu)}(x - y; r) F_n(dy) .$$

It follows from Hölder's inequality that

$$\int_{-c}^{c} f_n^{1-\varepsilon}(x; r)\, dx \leqq (2c)^\varepsilon , \qquad c > 0 \text{ and } 0 \leqq \varepsilon \leqq 1 . \tag{3.1}$$

It follows from (2.2) that

$$\frac{\partial}{\partial r} f_n^{(\nu)}(x; r) = r f_n^{(\nu+2)}(x; r) . \tag{3.2}$$

It follows as in the proof of Proposition 2.1 that

$$\frac{q}{r} f_n(x; q) \leqq f_n(x; r) \leqq \frac{q^{q^2/r^2}}{r} (f_n(x; q))^{q^2/r^2} \qquad \text{for } 0 < q \leqq r . \tag{3.3}$$

It follows as in the proof of Proposition 2.2 that for $\varepsilon > 0$ and ν a positive integer

$$f_n^{(\nu)}(x; r_n) = O_p(1) r_n^{-(\nu+\varepsilon)} (f_n(x; r_n) + f_n^{1-\varepsilon}(x; r_n)) . \tag{3.4}$$

It follows as in the proof of Proposition 2.3 that if $1/nr_n^3 = O_p(1)$, then for $\varepsilon > 0$ and $M > 0$

$$\max_{|t| \leqq Mn^{-\frac{1}{2}}} f_n(x + t; r_n) = O_p(1)(f_n(x; r_n) + f_n^{1-\varepsilon}(x; r_n)) . \tag{3.5}$$

Let r be a nonrandom positive number. Then $Ef_n^{(\nu)}(x; r) = f^{(\nu)}(x; r)$. Also

$$\operatorname{Var} f_n^{(\nu)}(x; r) = \frac{1}{n} \operatorname{Var} \varphi^{(\nu)}(x - X_i; r) \leqq \frac{1}{n} E(\varphi^{(\nu)}(x - X_i; r))^2 . \tag{3.6}$$

Let K_ν be the maximum value of $(\varphi^{(\nu)}(x))^2/\varphi(x)$. Then

$$(\varphi^{(\nu)}(x; r))^2 \leqq \frac{K_\nu}{r^{2\nu+1}} \varphi(x; r) . \tag{3.7}$$

It follows from (3.6) and (3.7) that

$$E(f_n^{(\nu)}(x; r) - f^{(\nu)}(x; r))^2 \leqq \frac{K_\nu f(x; r)}{nr^{2\nu+1}} . \tag{3.8}$$

We conclude from (3.8) that

$$E \int_{-\infty}^{\infty} (f_n^{(\nu)}(x; r) - f^{(\nu)}(x; r))^2\, dx \leqq \frac{K_\nu}{nr^{2\nu+1}} \tag{3.9}$$

and that

$$E \int_{-c}^{c} \frac{(f_n^{(\nu)}(x; r) - f^{(\nu)}(x; r))^2}{f(x; r)}\, dx \leqq \frac{2K_\nu c}{nr^{2\nu+1}} . \tag{3.10}$$

Since $K_0 = (2\pi)^{-\frac{1}{2}}$ it follows from (3.8) that

$$E \int_{-c}^{c} \frac{f_n^2(x; r)}{f(x; r)}\, dx \leqq 1 + \frac{2c}{nr} . \tag{3.11}$$

It also follows from (3.8) that for $\gamma \geqq 0$

$$E \int_{-c}^{c} \frac{f_n^2(x; r)}{f(x; r)} 1_{\{n^\gamma f(x;r)<1\}} dx \leqq 2c\left(\frac{1}{nr} + \frac{1}{n^\gamma}\right). \tag{3.12}$$

It follows from (3.11) and Hölder's inequality that for $0 \leqq \varepsilon \leqq 2$, $\delta \geqq 0$, and $\gamma \geqq 0$

$$E \int_{-c}^{c} \frac{f_n^{2-\varepsilon}(x; r)}{f^{1+\delta}(x; r)} 1_{\{n^\gamma f(x;r)\geqq 1\}} dx \leqq n^{\gamma\delta+(\gamma\varepsilon/2)}(2c)^{\varepsilon/2}\left(1 + \frac{2c}{nr}\right). \tag{3.13}$$

If follows from (3.1), (3.11), and Hölder's inequality that for $\varepsilon \geqq 0$, $\delta \geqq 0$, and $\delta + 2\varepsilon \leqq 1$

$$E \int_{-c}^{c} \frac{f_n^{1-\delta}(x; r)}{f^\varepsilon(x; r)} dx \leqq (2c)^{\delta+\varepsilon}\left(1 + \frac{2c}{nr}\right). \tag{3.14}$$

It follows from (3.12) and Hölder's inequality that for $\varepsilon \geqq 0$, $\delta \geqq 0$, and $\delta + 2\varepsilon \leqq 1$

$$E \int_{-c}^{c} \frac{f_n^{1-\delta}(x; r)}{f^\varepsilon(x; r)} 1_{\{n^\gamma f(x;r)<1\}} dx \leqq \frac{2c}{n^{\gamma(1-\varepsilon-\delta)}}\left(1 + \frac{n^\gamma}{nr}\right). \tag{3.15}$$

Let s_n, $n \geqq 1$, be positive random variables such that $s_n/s = 1 + o_p(1)$ for some positive nonrandom number s. Suppose that $r_n = s_n t_n$ where t_n are positive constants that approach zero as $n \to \infty$. Set $q_n = st_n$. Suppose also that $c_n = s_n d_n$, where d_n are positive constants that approach $+\infty$ as $n \to \infty$.

PROPOSITION 3.1. *Suppose that $c_n/nr_n = O_p(1)$. Then for $0 \leqq \varepsilon \leqq 2$*

$$\int_{-c_n}^{c_n} \frac{f_n^{2-\varepsilon}(x; r_n)}{f(x; r_n)} 1_{\{n^2 f(x;r_n)\geqq 1\}} dx = O_p(1)n^{2\varepsilon} .$$

PROOF. This result follows from Proposition 2.1 and Equations (3.3) and (3.13).

PROPOSITION 3.2. *Suppose that $c_n/nr_n = O_p(1)$. Then for $\varepsilon \geqq 0$, $\delta \geqq 0$, and $\delta + 2\varepsilon < 1$*

$$\int_{-c_n}^{c_n} \frac{f_n^{1-\delta}(x; r_n)}{f^\varepsilon(x; r_n)} dx = O_p(1)n^{2(\delta+\varepsilon)} .$$

PROOF. This result follows from Proposition 2.1 and Equations (3.3) and (3.14).

PROPOSITION 3.3. *Suppose that $c_n/nr_n = O_p(1)$. Then for $\varepsilon \geqq 0$, $\delta \geqq 0$, and $\delta + 2\varepsilon < 1$*

$$\int_{-c_n}^{c_n} \frac{f_n^{1-\delta}(x; r_n)}{f^\varepsilon(x; r_n)} 1_{\{n^2 f(x;r_n)<1\}} dx = \frac{O_p(1)c_n n^{2(\varepsilon+\delta)}}{nr_n} .$$

PROOF. This result follows from Proposition 2.1 and Equations (3.3) and (3.15).

PROPOSITION 3.4. *For ν a nonnegative integer*

$$\int_{-\infty}^{\infty} (f_n^{(\nu)}(x; r_n) - f^{(\nu)}(x; r_n))^2 \, dx = \frac{O_p(1)}{nr_n^{2\nu+1}} .$$

PROOF. It follows from (2.3) and (3.2) that

$$(f_n^{(\nu)}(x, r_n) - f^{(\nu)}(x; r_n)) = (f_n^{(\nu)}(x; q_n) - f^{(\nu)}(x; q_n)) + \int_{q_n}^{r_n} r(f_n^{(\nu+2)}(x; r) - f^{(\nu+2)}(x; r)) \, ds .$$

We conclude from Schwarz's inequality that

$$(3.16) \qquad \begin{aligned} &(f_n^{(\nu)}(x; r_n) - f^{(\nu)}(x, r_n))^2 \\ &\quad \leqq 2(f_n^{(\nu)}(x; q_n) - f^{(\nu)}(x; q_n))^2 \\ &\quad\quad + 2\left|\int_{q_n}^{r_n} r^2 \, dr \int_{q_n}^{r_n} (f_n^{(\nu+2)}(x; r) - f^{(\nu+2)}(x; r))^2 \, dr\right| . \end{aligned}$$

The desired result now follows from (3.9).

PROPOSITION 3.5. *Suppose that $1/nr_n = O_p(1)$. Then for ν a nonnegative integer and $\varepsilon > 0$*

$$\int_{-c_n}^{c_n} \frac{(f_n^{(\nu)}(x; r_n) - f^{(\nu)}(x; r_n))^2}{f(x; r_n)} 1_{\{n^2 f(x; r_n) \geqq 1\}} \, dx = \frac{O_p(1) c_n n^{\varepsilon}}{nr_n^{2\nu+1}} .$$

PROOF. It follows from (3.10) that for $0 < \varepsilon < 1$

$$(3.17) \qquad \int_{-c_n}^{c_n} \left[\int_{-q_n(1-\varepsilon)}^{q_n(1+\varepsilon)} \frac{(f_n^{(\nu+2)}(x; r) - f^{(\nu+2)}(x; r))^2}{f(x; r)} \, dr \right] dx = \frac{O_p(1) c_n}{nr_n^{2\nu+4}} .$$

The desired result now follows from (3.10), (3.16), (3.17), and Proposition 2.1.

PROPOSITION 3.6. *Suppose that $c_n/nr_n = O_p(1)$. Then for ν a nonnegative integer and $\varepsilon \geqq 0$*

$$\int_{-c_n}^{c_n} \frac{(f_n^{(\nu)}(x; r_n) - f^{(\nu)}(x; r_n))^2}{f^{\varepsilon}(x; r_n)} \, dx = \frac{O_p(1) n^{2\varepsilon}}{nr_n^{2\nu+1}} .$$

PROOF. By Proposition 3.4 we can assume that $\varepsilon > 0$. This result then follows from (2.4), (3.8), (3.16), and Proposition 2.1.

4. Nonadaptive estimators. Let $g(x)$, $-\infty < x < \infty$, be a twice continuously differentiable function vanishing outside $[-1, 1]$ and such that $0 \leqq g \leqq 1$ and $g(0) = 1$. Let $\mathscr{I}$, $L(x; r)$, and $A(r, c)$ be as in the Introduction.

THEOREM 4.1.

$$\lim_{r \to 0, c \to \infty'} A(r, c) = \mathscr{I} .$$

PROOF. Since $A(r, c) \leqq \mathscr{I}$ (see Port and Stone [15]), we need only show that

$$(4.1) \qquad \lim_{c \to \infty, r \to 0} \liminf \int_{-c}^{c} L^2(x; r) f(x; r) \, dx \geqq \mathscr{I} .$$

According to Huber [7]

$$\mathscr{I} = \sup_{\psi} \frac{(E\psi'(X))^2}{E\psi^2(X)} ,$$

where the sup extends over all $\psi \in C_c^1$ such that $E\psi^2(X) > 0$. Suppose first that $\mathscr{I} < \infty$. Choose $\varepsilon > 0$. Then we can find a $\psi \in C_c^1$ such that $\int \psi^2(x)F(dx) > 0$ and

$$(\int \psi'(x)F(dx))^2 \geqq (1-\varepsilon)\mathscr{I} \int \psi^2(x)F(dx) .$$

Thus for r sufficiently small

$$(\int \psi'(x)f(x; r)\, dx)^2 \geqq (1-2\varepsilon)\mathscr{I} \int \psi^2(x)f(x; r)\, dx .$$

Choose $c > 0$ such that $\psi(x) = 0$ for $|x| \geqq c$. Then by Schwarz's inequality

$$\begin{aligned}(\int \psi'(x)f(x; r)\, dx)^2 &= (\int \psi(x)f'(x; r)\, dx)^2 \\ &\leqq \int \psi^2(x)f(x; r)\, dx \int_{-c}^{c} L^2(x; r)f(x; r)\, dx .\end{aligned}$$

Thus for r sufficiently small

$$\int_{-c}^{c} L^2(x; r)f(x; r)\, dx \geqq (1-2\varepsilon)\mathscr{I}$$

and hence (4.1) holds.

Suppose instead that $\mathscr{I} = \infty$. Choose $M > 0$. By arguing as before we can find a $c > 0$ such that for r sufficiently small

$$\int_{-c}^{c} L^2(x; r)f(x; r)\, dx \geqq \frac{M}{2} .$$

Thus (4.1) holds in this case also. This completes the proof of the theorem.

We now study the translation invariant nonrandomized estimator $\tilde{\theta}_n$ given by (see the motivation in the Introduction)

$$\tilde{\theta}_n = \bar{\theta}_n - \frac{1}{A(r_n, c_n)} \int L(x; r_n)g(x/c_n)(f_n(x + \bar{\theta}_n - \theta; r_n) - f(x; r_n))\, dx ,$$

where $\bar{\theta}_n$, r_n, and c_n are chosen as described in the Introduction, so that (1.3) and (1.12) hold.

THEOREM 4.2. *Suppose that*

$$\frac{c_n}{n^{1-\varepsilon} r_n^{6}} = O_p(1) \qquad \textit{for some} \quad \varepsilon > 0 . \tag{4.2}$$

Then $\mathscr{L}(n^{\frac{1}{2}}(\tilde{\theta}_n - \theta)) \to N(0, 1/\mathscr{I})$ *as* $n \to \infty$.

PROOF. It follows from Propositions 2.2 and 3.6 and Schwarz's inequality that

$$\int L(x; r_n)g(x/c_n)(f_n'(x; r_n) - f'(x; r_n))\, dx = o_p(1) . \tag{4.3}$$

It follows from (3.4), (3.5) and Propositions 2.2 and 3.2 that

$$\int |L(x; r_n)| g(x/c_n) \sup_{|t| \leqq |\bar{\theta}_n - \theta|} |f_n''(x + t; r_n)|\, dx = \frac{O_p(1)n^{\varepsilon}}{r_n^{3}} = o_p(1)n^{\frac{1}{2}} . \tag{4.4}$$

We conclude from (4.3) and (4.4) that

$$\tilde{\theta}_n - \theta = -\frac{1}{A(r_n, c_n)} \int L(x; r_n)g(x/c_n)(f_n(x; r_n) - f(x; r_n))\, dx + \frac{o_p(1)}{n^{\frac{1}{2}}} . \tag{4.5}$$

We can write

$$(4.6)\qquad \begin{aligned} &L(x; r_n)(f_n(x; r_n) - f(x; r_n)) \\ &\qquad = L(x; q_n)(f_n(x; q_n) - f(x; q_n)) \\ &\qquad\quad + \int_{q_n}^{r_n} \frac{\partial}{\partial r} [L(x; r)(f_n(x; r) - f(x; r))]\, dr , \end{aligned}$$

where q_n is as in Section 3. The indicated differentiation with respect to r can be obtained by using (2.2) and (3.2). We then substitute (4.6) into (4.5) and integrate first with respect to x and then with respect to r. By using Propositions 2.2 and 3.6 we conclude that for $\varepsilon > 0$

$$(4.7)\qquad \begin{aligned} &\int L(x; r_n)g(x/c_n)(f_n(x; r_n) - f(x; r_n))\, dx \\ &\qquad = \int L(x; q_n)g(x/c_n)(f_n(x; q_n) - f(x; q_n))\, dx \\ &\qquad\quad + \frac{O_p(1)}{n^{\frac{1}{2}}} \left(\frac{c_n}{r_n^{\,3} n^{1-\varepsilon}}\right)^{\frac{1}{2}} \left[n^{\frac{1}{2}} \left(1 - \frac{r_n}{q_n}\right)\right] . \end{aligned}$$

We conclude from (1.12), (4.2), (4.5), and (4.7) that

$$(4.8)\qquad \tilde{\theta}_n - \theta = -\frac{1}{A(r_n; c_n)} \int L(x; q_n)g(x/c_n)(f_n(x; q_n) - f(x; q_n))\, dx + \frac{o_p(1)}{n^{\frac{1}{2}}} .$$

By a similar argument we conclude from (4.2) and (4.8) that

$$(4.9)\qquad \tilde{\theta}_n - \theta = -\frac{1}{A(r_n, c_n)} \int L(x; q_n)g(x/b_n)(f_n(x; q_n) - f(x; q_n))\, dx + \frac{o_p(1)}{n^{\frac{1}{2}}}$$

where $b_n = sd_n$. Again we can use (4.2) to conclude that

$$(4.10)\qquad A(r_n, c_n) = A(q_n, b_n) + o_p(1) .$$

The random variable

$$\int L(x; q_n)g(x/b_n)(f_n(x; q_n) - f(x; q_n))\, dx$$

has mean zero and variance

$$\begin{aligned} &\frac{1}{n} \operatorname{Var} \int L(x; q_n)g(x/b_n)\varphi(x - X_i; q_n)\, dx \\ &\qquad \leqq \frac{1}{n} E \int L^2(x; q_n)g(x/b_n)\varphi(x - X_i; q_n)\, dx \\ &\qquad = \frac{1}{n} \int L^2(x; q_n)g(x/b_n)f(x; q_n)\, dx = \frac{1}{n} A(q_n, b_n) . \end{aligned}$$

Thus

$$(4.11)\qquad E\left(\frac{1}{A(q_n, b_n)} \int L(x; q_n)g(x/b_n)(f_n(x; q_n) - f(x; q_n))\, dx\right)^2 \leqq \frac{1}{nA(q_n, b_n)} .$$

It follows from (4.9)—(4.11) that

$$(4.12)\qquad \tilde{\theta}_n - \theta = -\frac{1}{A(q_n, b_n)} \int L(x; q_n)g(x/b_n)(f_n(x; q_n) - f(x; q_n))\, dx + \frac{o_p(1)}{n^{\frac{1}{2}}} .$$

We conclude from (4.11), (4.12), and Theorem 4.1 that the asymptotic second moment of $n^{\frac{1}{2}}(\tilde{\theta}_n - \theta)$ is at most $1/\mathscr{I}$. The conclusion of Theorem 4.2 now follows either from Theorem 4.1 of Hájek [6] or Proposition 2.1 of Stone [18].

5. Adaptive estimators. We assume in this section that F is symmetric about the origin, that $\bar{\theta}_n$ satisfies (1.3), and that s_n satisfies (1.12).

We estimate $f(x; r)$ based on the sample $\theta + X_1, \cdots, \theta + X_n$ by $\hat{f}_n(x, r)$, where

$$\hat{f}_n(x; r) = \frac{1}{2n} \sum_{j=1}^{n} [\varphi(x + \bar{\theta}_n - \theta - X_j; r) + \varphi(-x + \bar{\theta}_n - \theta - X_j; r)]$$
$$= \tfrac{1}{2}(f_n(x + \bar{\theta}_n - \theta; r) + f_n(-x + \bar{\theta}_n - \theta; r)) .$$

Clearly $\hat{f}_n(\cdot; r)$ is a symmetric probability density function. It follows as in the proof of Proposition 2.2 that for $\varepsilon > 0$

$$\hat{f}_n^{(\nu)}(x; r_n) = \frac{O_p(1)}{r_n^{\nu+\varepsilon}} (\hat{f}_n(x; r_n) + \hat{f}_n^{1-\varepsilon}(x; r_n)) . \tag{5.1}$$

PROPOSITION 5.1. *If* $1/nr_n^4 = O_p(1)$, *then*

$$\int_{-\infty}^{\infty} (\hat{f}_n^{(\nu)}(x; r_n) - f^{(\nu)}(x; r_n))^2 \, dx = \frac{O_p(1)}{nr_n^{2\nu+1}} .$$

The proof of this proposition is based on the next result.

LEMMA 5.1. *If f is twice continuously differentiable, then*

$$\int (f(x + t) - f(x - t) - 2f(x))^2 \, dx \leqq \frac{4t^4}{3} \int (f''(x))^2 \, dx .$$

PROOF OF LEMMA. We can assume that $t > 0$. Observe that

$$f(x + t) + f(x - t) - 2f(x) = \int_{-t}^{t} f''(x + s)(t - |s|) \, ds .$$

We conclude from Schwarz's inequality that

$$(f(x + t) + f(x - t) - 2f(x))^2 \leqq \frac{2t^3}{3} \int_{-t}^{t} (f''(x + s))^2 \, ds ,$$

from which the conclusion of the lemma follows easily.

PROOF OF PROPOSITION. It follows from (2.1) and Lemma 5.1 that

$$\int (f^{(\nu)}(x + \bar{\theta}_n - \theta; r_n) + f^{(\nu)}(x - \bar{\theta}_n + \theta; r_n) - 2f^{(\nu)}(x; r_n))^2 \, dx = \frac{O_p(1)}{n^2 r_n^{2\nu+5}} .$$

We now conclude from Proposition 3.4 and the symmetry of $f(\cdot; r)$ that

$$\int (\hat{f}_n^{(\nu)}(x; r_n) - f^{(\nu)}(x; r_n))^2 \, dx = O_p(1) \left(\frac{1}{n^2 r_n^{2\nu+5}} + \frac{1}{nr_n^{2\nu+1}} \right) ,$$

from which the conclusion of Proposition 5.1 follows.

Let $\hat{L}_n(x; r)$ and $\hat{A}_n(r, c)$ be as defined in the Introduction. We conclude from (5.1) that for $\varepsilon > 0$

$$\hat{L}_n(x; r_n) = \frac{O_p(1)}{r_n^{1+\varepsilon}} (1 + \hat{f}_n^{-\varepsilon}(x; r_n)) . \tag{5.2}$$

Similarly we conclude from Proposition 2.2 that for $\varepsilon > 0$

$$L(x; r_n) = \frac{O_p(1)}{r_n^{1+\varepsilon}} (1 + f_n^{-\varepsilon}(x; r_n)) . \tag{5.3}$$

We assume throughout the remainder of this section that g is symmetric about the origin.

THEOREM 5.1. *Suppose that*

$$\frac{c_n}{n^{1-\varepsilon} r_n^5} = O_p(1) \qquad \textit{for some} \quad \varepsilon > 0 . \tag{5.4}$$

Then $\hat{A}_n(r_n, c_n)/A_n(r_n, c_n) \to 1$ *in probability as* $n \to \infty$.

PROOF. Choose $\varepsilon > 0$. By (5.2)

$$\int \hat{L}_n^2(x; r_n) g(x/c_n) \hat{f}_n(x; r_n) 1_{\{n\hat{f}_n(x; r_n) < 1\}} \, dx = \frac{O_p(1) c_n}{n^{1-\varepsilon} r_n^2} . \tag{5.5}$$

By (5.3)

$$\int L_n^2(x; r_n) g(x/c_n) f(x; r_n) 1_{\{nf(x; r_n) < 1\}} \, dx = \frac{O_p(1) c_n}{n^{1-\varepsilon} r_n^2} . \tag{5.6}$$

Choose $\delta > 0$. By (5.2), Proposition 5.1, and Schwarz's inequality

$$\begin{aligned} \int \hat{L}_n^2(x; r_n) g(x/c_n) \hat{f}_n(x; r_n) 1_{\{n\hat{f}_n(x; r_n) \geqq 1, |\hat{f}_n(x; r_n) - f(x; r_n)| \geqq \delta \hat{f}_n(x; r_n)\}} \, dx \\ = o_p(1) \left(\frac{c_n}{n^{1-\varepsilon} r_n^5} \right)^{\frac{1}{2}} . \end{aligned} \tag{5.7}$$

Similarly

$$\begin{aligned} \int L^2(x; r_n) g(x/c_n) f(x; r_n) 1_{\{nf(x; r_n) \geqq 1, |\hat{f}_n(x; r_n) - f(x; r_n)| \geqq \delta f(x; r_n)\}} \, dx \\ = o_p(1) \left(\frac{c_n}{n^{1-\varepsilon} r_n^5} \right)^{\frac{1}{2}} . \end{aligned} \tag{5.8}$$

Moreover

$$\begin{aligned} \int \hat{L}_n^2(x; r_n) g(x/c_n) \hat{f}_n(x; r_n) 1_{\{n\hat{f}_n(x; r_n) \geqq 1, |\hat{f}_n'(x; r_n) - f'(x; r_n)| \geqq \delta |\hat{f}_n'(x; r_n)|\}} \, dx \\ = o_p(1) \left(\frac{c_n}{n^{1-\varepsilon} r_n^5} \right)^{\frac{1}{2}} \end{aligned} \tag{5.9}$$

and

$$\begin{aligned} \int L^2(x; r_n) g(x/c_n) f(x; r_n) 1_{\{nf(x; r_n) \geqq 1, |\hat{f}_n'(x; r_n) - f'(x; r_n)| \geqq \delta |f'(x; r_n)|\}} \, dx \\ = o_p(1) \left(\frac{c_n}{n^{1-\varepsilon} r_n^5} \right)^{\frac{1}{2}} . \end{aligned} \tag{5.10}$$

The conclusion of the theorem follows from (5.5)—(5.10).

Let $\hat{\theta}_n$ be the estimator defined by (1.9) or (1.15). We can rewrite this estimator as

$$\hat{\theta}_n = \bar{\theta}_n - \frac{1}{\hat{A}_n(r_n, c_n)} \int \hat{L}_n(x; r_n) g(x/c_n) (f_n(x + \bar{\theta}_n - \theta; r_n) - f(x; r_n)) \, dx .$$

THEOREM 5.2. *Suppose that* (1.13) *holds. Then* $n^{\frac{1}{2}}(\hat{\theta}_n - \tilde{\theta}_n) \to 0$ *in probability as* $n \to \infty$.

PROOF. We need to show that

$$\int \left(\frac{\hat{f}_n'(x; r_n)}{\hat{f}_n(x; r_n)} - \frac{f'(x; r_n)}{f(x; r_n)}\right) g(x/c_n)(f_n(x + \bar{\theta}_n - \theta; r_n) - f(x; r_n))\, dx = \frac{o_p(1)}{n^{\frac{1}{2}}}. \tag{5.11}$$

Choose $\varepsilon > 0$. Let I_1 denote the contribution to this integral for x's such that $n^2\hat{f}_n(x; r_n) < 1$ and $f(x; r_n) \geqq \hat{f}_n(x; r_n)$. Since $f_n(x + \bar{\theta}_n - \theta; r_n) \leqq 2\hat{f}_n(x; r_n)$, we conclude from (5.2), (5.3) and symmetry that

$$I_1 = \frac{O_p(1)c_n}{n^{2-\varepsilon}r_n}. \tag{5.12}$$

Let I_2 denote the contribution to the integral for x's such that $n^2f(x; r_n) < 1$ and $f(x; r_n) < \hat{f}_n(x; r_n)$. From (3.5), (5.2), (5.3), Proposition 3.3 and symmetry we see that

$$I_2 = \frac{O_p(1)c_n}{n^{1-\varepsilon}r_n}. \tag{5.13}$$

Let I_3 denote the contribution to the integral for x's such that $n^2f_n(x; r_n) \geqq 1$ and $n^2f(x; r_n) \geqq 1$. We can write $I_3 = I_4 + I_5$, where I_4 corresponds to the integrand

$$-\frac{f_n'(x; r_n)(\hat{f}_n(x; r_n) - f(x; r_n))}{\hat{f}_n(x; r_n)f(x; r_n)}(f_n(x + \bar{\theta}_n - \theta; r_n) - f(x; r_n))g(x/c_n)$$

and I_5 corresponds to the integrand

$$\frac{\hat{f}_n'(x; r_n) - f'(x; r_n)}{f(x; r_n)}(f_n(x + \bar{\theta}_n - \theta; r_n) - f(x; r_n))g(x/c_n).$$

It follows from (3.4), (3.5), and Propositions 3.1 and 3.5 that

$$\int_{-c_n}^{c_n} \frac{(f_n(x + \bar{\theta}_n - \theta; r_n) - f(x; r_n))^2}{f(x; r_n)} 1_{\{n^2f(x;r_n)\geqq 1\}}\, dx = \frac{O_p(1)c_n n^{\varepsilon/2}}{nr_n^2} \tag{5.14}$$

and

$$\int_{-c_n}^{c_n} \frac{(f_n'(x + \bar{\theta}_n - \theta; r_n) - f'(x; r_n))^2}{f(x; r_n)} 1_{\{n^2f(x;r_n)\geqq 1\}}\, dx = \frac{O_p(1)c_n n^{\varepsilon/2}}{nr_n^4}. \tag{5.15}$$

We conclude from (5.14) and (5.15) respectively that

$$\int_{-c_n}^{c_n} \frac{(\hat{f}_n(x; r_n) - f(x; r_n))^2}{f(x; r_n)} 1_{\{n^2f(x;r_n)\geqq 1\}}\, dx = \frac{O_p(1)c_n n^{\varepsilon/2}}{nr_n^2} \tag{5.16}$$

and

$$\int_{-c_n}^{c_n} \frac{(\hat{f}_n'(x; r_n) - f'(x; r_n))^2}{f(x; r_n)} 1_{\{n^2f(x;r_n)\geqq 1\}}\, dx = \frac{O_p(1)c_n n^{\varepsilon/2}}{nr_n^4}. \tag{5.17}$$

We conclude from (5.2), (5.14), (5.16) and Schwarz's inequality that

$$I_4 = \frac{O_p(1)c_n}{n^{1-\varepsilon}r_n^{\ 3}}. \tag{5.18}$$

We conclude from (5.14), (5.17), and Schwarz's inequality that

$$I_5 = \frac{O_p(1)c_n}{n^{1-\varepsilon}r_n^{\ 3}}. \tag{5.19}$$

The conclusion of the theorem follows from (5.12), (5.13), (5.18), and (5.19).

THEOREM 5.3. *Suppose that* (1.13) *holds. Then* $\mathscr{L}(n^{\frac{1}{2}}(\hat{\theta}_n - \theta)) \to N(0, 1/\mathscr{I})$ *as* $n \to \infty$.

PROOF. This result follows immediately from Theorems 4.2 and 5.2.

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: Survey and Advances.* Princeton Univ. Press.

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

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

[4] FABIAN, V. (1973). Asymptotically efficient stochastic approximation; the RM case. *Ann. Statist.* **1** 486–495.

[5] GOOD, I. J. and GASKINS, R. A. (1972). Global nonparametric estimation of probability densities. *Virginia J. Sci.* **23** 171–193.

[6] HÁJEK, J. (1972). Local asymptotic minimax and admissibility in estimation. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* **1** 175–194. Univ. of California Press.

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

[8] HUBER, P. J. (1969). *Théorie de l'inférence statistique robuste.* Les Presses de l'Université de Montréal.

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

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

[11] LE CAM, L. (1970). On the assumptions used to prove asymptotic normality of the maximum likelihood estimates. *Ann. Math. Statist.* **41** 802–828.

[12] MIKÉ, V. (1971). Efficiency-robust systematic linear estimators of location. *J. Amer. Statist. Assoc.* **66** 594–601.

[13] MIKÉ, V. (1973). Robust Pitman-type estimators of location. *Ann. Inst. Statist. Math.* **25** 65–86.

[14] PFANZAGL, J. (1974). Investigating the quantile of an unknown distribution. *Preprints in Statistics* **7** Univ. of Cologne.

[15] PORT, S. C. and STONE, C. J. (1974). Fisher information and the Pitman estimator of a location parameter. *Ann. Statist.* **2** 225–247.

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

[17] STEIN, C. (1956). Efficient nonparametric testing and estimation. *Proc. Third Berkeley Symp. Math. Statist. Prob.* **1** 187–196. Univ. of California Press.

[18] STONE, C. J. (1974). Asymptotic properties of estimators of a location parameter. *Ann. Statist.* **2** 1127–1137.

[19] SWITZER, P. (1972). Efficiency robustness of estimators. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* **1** 283-291. Univ. of California Press.

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

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

[22] WEISS, L. and WOLFOWITZ, J. (1970). Asymptotically efficient nonparametric estimators of location and scale parameters. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **16** 134-150.

[23] WOLFOWITZ, J. (1974). Asymptotically efficient nonparametric estimators of location and scale parameters. II. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **30** 117-128.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA
LOS ANGELES, CALIFORNIA 90024