The Annals of Statistics
1980, Vol. 8, No. 4, 870–882

OPTIMUM KERNEL ESTIMATORS OF THE MODE[1]

BY WILLIAM F. EDDY

Carnegie-Mellon University

Let $X_1, \cdots, X_n$ be independent observations with common density f. A kernel estimate of the mode is any value of t which maximizes the kernel estimate of the density f_n. Conditions are given restricting the density, the kernel, and the bandwidth under which this estimate of the mode has an asymptotic normal distribution. By imposing sufficient restrictions, the rate at which the mean squared error of the estimator converges to zero can be decreased from $n^{-\frac{4}{7}}$ to $n^{-1+\varepsilon}$ for any positive ε. Also, by bounding the support of the kernel it is shown that for any particular bandwidth sequence the asymptotic mean squared error is minimized by a certain truncated polynomial kernel.

1. Introduction. A mode of a probability density $f(t)$ is a value of t which maximizes f. Relatively little attention has been paid to estimating the mode perhaps because of the delicacy of the problem: any method for estimating the mode must estimate a density, either explicitly or implicitly, and this is itself a difficult problem. An excellent review of nonparametric density estimation methods may be found in Wegman (1972). Here, attention is focused on the class of kernel estimators introduced by Rosenblatt (1956).

Let $X_1, \cdots, X_n$ be independent observations with common (unknown) density f. Rosenblatt proposed estimating $f(t)$ by

$$f_n(t) = \frac{1}{na_n} \Sigma_{i=1}^n K\left(\frac{t - X_i}{a_n}\right) \tag{1.1}$$

where the kernel K is a bounded measurable function and the bandwidth a_n is a positive constant. It is desirable that

$$\lim_{|x| \to \infty} K(x) = 0$$

and that

$$\lim_{n \to \infty} a_n = 0$$

so that observations which are far from t will have little influence on $f_n(t)$; but if $\{a_n\}$ converges to zero too quickly $\{f_n(t)\}$ will not be a consistent estimator of $f(t)$.

Parzen (1962) proposed using the location of the maximum of the density estimate (1.1) to estimate the mode of f. More precisely, let

$$M(f) = \inf\{t | f(t) = \sup_s f(s)\}. \tag{1.2}$$

Received January 1977; revised October 1978.

[1]This work is based on a portion of the author's Ph.D. dissertation which was supported in part by the National Science Foundation under grant DCR-75-08374 to Yale University. Revision was supported under grant MCS-78-02422 to Carnegie-Mellon University.

AMS 1970 *subject classifications.* Primary 62F10, 62G05.

Key words and phrases. Location of the maximum, parabolic process, polynomial kernel.

Then the mode θ of a density f is $M(f)$ and Parzen's estimate is $\theta_n = M(f_n)$. Parzen (1962, Theorem 3A) gave conditions under which $\{\theta_n\}$ is a consistent estimator of θ. Nadaraya (1965) and Van Ryzin (1969) have derived stronger consistency results.

Parzen (1962, Theorem 5A) also gave conditions under which θ_n (appropriately normalized) has an asymptotic normal distribution. Samanta (1973) and Konakov (1974) have given multivariate versions of Parzen's results.

THEOREM 1.1 (Parzen). *Let $K(x)$ be a probability density with characteristic function $k(u)$ and let the density $f(t)$ have a characteristic function $\varphi(u)$. If, for some $r \geqslant 2$*

$$\lim_{u \to 0} \frac{1 - k(u)}{u^r} > 0, \tag{1.3}$$

and if, for some δ, $\frac{1}{2} < \delta < 1$,

$$\int u^{2+\delta} |\varphi(u)| du < \infty, \tag{1.4}$$

$$\int u^{2+\delta} |k(u)| du < \infty, \tag{1.5}$$

$$\lim_{n \to \infty} n a_n^{5+2\delta} = 0, \quad \text{and} \tag{1.6}$$

$$\lim_{n \to \infty} n a_n^6 = \infty, \tag{1.7}$$

then

$$(n a_n^3)^{\frac{1}{2}} (\theta_n - \theta) \to_D \mathfrak{N}\left(0, \frac{f(\theta)}{[f^{(2)}(\theta)]^2} V\right)$$

where $f^{(2)}(\theta)$ is the second derivative of the density at θ and $V = \int [K^{(1)}(x)]^2 \, dx$.

(All integrals here and elsewhere are taken over the whole real line unless specified otherwise. Also, for all i, $i = 1, 2, \cdots$, $g^{(i)}(x) = d^i g(x)/dx^i$.

Parzen's result has two serious drawbacks. First, condition (1.5) requires the kernel to have two uniformly continuous derivatives; the kernels satisfying the optimality property of Section 3 do not have even one continuous derivative. Second, (1.6) and (1.7) require that for some constant $d > 0$ and n large enough

$$n^{-\frac{1}{6}} < d a_n < n^{-1/(5+2\delta)}.$$

For any kernel the mean squared error usually converges to zero at the fastest rate when the asymptotic variance and the square of the asymptotic mean are of the same order. Since $\delta < 1$, the interesting case $a_n = (1/d) n^{-\frac{1}{7}}$ is excluded by Parzen's theorem; that is, $\{a_n\}$ must converge to zero so rapidly that the asymptotic mean is negligible compared to the asymptotic variance. As Chernoff (1964) has pointed out, the infimum of the mean squared error under Theorem 1.1 is $E(\theta_n - \theta)^2 = 0(n^{-\frac{4}{7}})$; but this limiting case is specifically excluded by (1.6). By imposing sufficient restrictions on K, it is possible to achieve not only $0(n^{-\frac{4}{7}})$, but in fact $0(n^{-1+\varepsilon})$ for any positive ε.

A theorem is given in Section 2 which overcomes these drawbacks by using the classical techniques of weak convergence (Billingsley (1968)). It will be shown that in a decreasing interval of t values near the mode the (appropriately normalized) kernel estimator $f_n(t)$ converges to a randomly located parabola in t. Since M, defined in (1.2), is continuous at the set of parabolas with fixed second derivative, the asymptotic distribution of the kernel estimator of the mode can be determined (c.f. Corollary 2.2).

For each sequence $\{a_n\}$ there is *no* kernel satisfying the conditions of Corollary 2.2 which minimizes the mean squared error of the asymptotic distribution. When the tail behavior of the kernel is restricted sufficiently it is possible to find an optimal kernel. If the kernel is restricted to be zero outside the interval $|x| \leqslant 1$ then for each rate at which $\{a_n\}$ converges to zero the calculus of variations yields an optimal kernel. Also it is noted that for each kernel there is an optimal rate for $\{a_n\}$ to approach zero. Unfortunately, for a particular optimal kernel the optimal rate is *not* the rate for which the kernel is optimal; this will be discussed in Section 3.

2. Asymptotic normality of θ_n.

Let $b = (na^3)^{-\frac{1}{2}}$ (the dependence of a and b on n will be suppressed henceforth) and define the random process

$$Z_n(t) = b^{-2}[f_n(\theta + bt) - f_n(\theta)],\ t \in [-T, T]$$

for some $T < \infty$. The essential point in the proof of asymptotic normality of θ_n is that the process Z_n converges weakly to a limit process Z and with probability one the sample functions of the limit process are parabolas with fixed second derivative satisfying $Z(0) = 0$. Hence the parabolas are determined by a single random variable. Specifically, the result is

THEOREM 2.1. *Let $p \geqslant 2$ be an integer. Let K be a bounded, absolutely continuous function with bounded derivative $K^{(1)}$. If*

$$\int K(x)\,dx = B_0 = 1, \tag{2.1}$$

$$\int x^i K(x)\,dx = B_i = 0, \qquad i = 1, \cdots, p-1, \tag{2.2}$$

$$\int x^p K(x)\,dx = B_p < \infty, \tag{2.3}$$

$$\int x^{p+1} K(x)\,dx < \infty, \tag{2.4}$$

$$\int [K^{(1)}(x)]^2\,dx = V < \infty, \tag{2.5}$$

$$\int x [K^{(1)}(x)]^2\,dx < \infty, \tag{2.6}$$

and if $\{a\}$ is a sequence of positive constants which satisfies

$$\lim_{n\to\infty} na^5 = \infty, \tag{2.7}$$

$$\lim_{n\to\infty} (na^{3+2p})^{\frac{1}{2}} = d < \infty, \tag{2.8}$$

and if the density f is bounded, has an absolutely continuous $(p+1)$st *derivative, and satisfies*

$$\sup_t |f^{(i)}(t)| < \infty, \qquad i = 1, \cdots, p+2 \tag{2.9}$$

then

$$Z_n(t) \to_W Z(t) = \frac{f^{(2)}(\theta)}{2} t^2 + (-1)^{p+1} \cdot \frac{f^{(p+1)}(\theta)}{p!} \cdot d \cdot B_p \cdot t + Y \cdot t$$

where Y *is a random variable having the normal distribution* $\mathfrak{N}(0, f(\theta) \cdot V)$.

The proof of the theorem shows first that EZ_n converges to a parabola, then that $Z_n - EZ_n$ converges in probability to a straight line, $\dot{Z}_n(1) - EZ_n(1)$ converges in distribution to a normal distribution, and finally that $\{Z_n\}$ is tight. The detailed proof is in Section 4.

The asymptotic normality of θ_n is an immediate consequence.

COROLLARY 2.2. *If the conditions of Theorem* 2.1 *are satisfied and if* $f^{(2)}(\theta) \neq 0$ *then*

$$(na^3)^{\frac{1}{2}}(\theta_n - \theta) \to_D \mathfrak{N}\left[(-1)^p \cdot \frac{d}{p!} \cdot \frac{f^{(p+1)}(\theta)}{f^{(2)}(\theta)} \cdot B_p, \frac{f(\theta)}{[f^{(2)}(\theta)]^2} \cdot V\right].$$

PROOF. Since

$$M(f_n) = \theta_n, M(Z_n) = \frac{\theta_n - \theta}{b} = (na^3)^{\frac{1}{2}}(\theta_n - \theta).$$

Also

$$M(Z) = (-1)^p \cdot \frac{d}{p!} \cdot \frac{f^{(p+1)}(\theta)}{f^{(2)}(\theta)} \cdot B_p - \frac{Y}{f^{(2)}(\theta)}$$

where $Y \sim_D \mathfrak{N}(0, f(\theta) \cdot V)$. So the corollary just states that $M(Z_n) \to_W M(Z)$. From Billingsley (Theorem 5.1), if $Z_n \to_W Z$ and if M is a measurable function such that its discontinuities have Z-measure zero then $M(Z_n) \to_W M(Z)$. Since M is a measurable function (see Section 5) and continuous at the set of parabolas with fixed second derivative (with probability one), the proof is complete.

At this point an examination of the assumptions of Theorem 2.1 and comparison (for $p = 2$) with those of Theorem 1.1 is in order. The moment conditions on K, (2.1)–(2.4), or the slightly weaker (1.3), are necessary so that $E\theta_n$ will be close to θ. They do not imply that $\theta_n - \theta = 0_p((na^3)^{-\frac{1}{2}})$ which would be necessary to insure that f_n has a unique maximum (with probability approaching one as $n \to \infty$). The smoothness conditions (2.5) and (2.6) are about the weakest possible (however, see Le Cam (1970), page 805 ff.) and are much weaker than (1.5); in the next section the importance of weakening (1.5) will become clear. Table 2.1 contains several

TABLE 2.1
Selected Kernels

$K(x)$	B_2	V
$\frac{1}{2}$, $\lvert x\rvert < 1$ 0, otherwise	$\frac{1}{3}$	0
$\frac{1}{(2\pi)^{\frac{1}{2}}}\exp(-x^2/2)$	1	$\frac{1}{2(2\pi)^{\frac{1}{2}}}$
$1-\lvert x\rvert$, $\lvert x\rvert \leq 1$ 0, otherwise	$\frac{1}{6}$	2
$\frac{15}{16}(1-x^2)^2$, $\lvert x\rvert < 1$ 0, otherwise	$\frac{1}{7}$	$\frac{15}{7}$
$\frac{3}{4}(1-x^2)$, $\lvert x\rvert < 1$ 0, otherwise	$\frac{1}{5}$	$\frac{3}{2}$
$\frac{15}{32}(1-x^2)(3-7x^2)$, $\lvert x\rvert < 1$ 0, otherwise	0	$\frac{75}{16}$

kernels together with their values of B_2 and V. All satisfy (2.5) and (2.6) and none except the second satisfy (1.5); of couse, the first kernel in Table 2.1 is not continuous. Condition (2.7) is needed so that $f^{(2)}(\theta)$ is consistently estimated; consistency of $f^{(2)}(t)$ uniformly in t requires (1.7). Evaluation of the asymptotic bias of θ_n is possible under (2.8) but not under (1.6); the accuracy of a normal approximation for moderate sample sizes would be severely affected. Under (2.8) the infinum of the mean-squared error is $E(\theta_n - \theta)^2 = 0(n^{-(2p/2p+3)})$; this rate is achieved when $d \neq 0$. Finally, (2.9) allows a Taylor expansion of f with $f^{(p+2)}$ in the remainder term and hence allows f_n to be locally parabolic; (1.4) is weaker.

3. Optimization of the estimator. Under the conditions of Corollary 2.2 a formal expansion of the mean-square error of the estimator θ_n is

(3.1)
$$E(\theta_n - \theta)^2 = \left[\frac{a^p \cdot B_p \cdot f^{(p+1)}(\theta)}{f^{(2)}(\theta)\cdot p!}\right]^2 + \frac{f(\theta)\cdot V}{na^3\cdot\{f^{(2)}(\theta)\}^2} + 0\left(\frac{1}{na} + a^{2p+2}\right).$$

If $a^{2p} = o(1/na^3)$ then the bias term is negligible. In this case it is desirable to choose the kernel so that the conditions of Corollary 2.2 are satisfied and V is as small as possible.

Let $p = 2$ for the moment and notice that K may be chosen so that the conditions of the corollary are satisfied and V is arbitrarily close to zero, for example:

$$K(x) = \frac{1}{(2\pi\sigma)^{\frac{1}{2}}}\exp(-x^2/2\sigma^2).$$

There is, however, no K satisfying the conditions with $V = 0$. If the tail behavior of the kernel is restricted in some fashion then an optimal kernel can be chosen from the restricted class. Although there are no compelling choices, the following restriction seems appropriate:

$$K(x) \equiv 0, |x| > 1. \tag{3.2}$$

Bartlett (1963), Epanechnikov (1969), and Johns and Van Ryzin (1972) also have used kernels which are zero outside bounded intervals for estimating densities.

The problem of choosing K satisfying the conditions to mimimize V is an *isoperimetric problem with constraints* (see, e.g., Gelfand and Fomin (1963), page 43, Theorem 1) in the calculus of variations. The solution is given by:

THEOREM 3.1. *If K satisfies both the conditions of Theorem* 2.1 *for $p = 2$ and* (3.2) *then $V = \int[K^{(1)}(x)]^2\,dx$ is minimized when*

$$K(x) = K_2(x) = (\tfrac{3}{4})(1 - x^2), \qquad |x| \leq 1$$
$$= 0, \qquad \textit{otherwise}.$$

PROOF. Euler's equation for this problem is (Gelfand and Fomin, page 15, Theorem 1)

$$K^{(1)}(x) + \lambda_1 x + \lambda_2 = 0$$

for some constants λ_1, λ_2. The only kernel satisfying the conditions and Euler's equation is the one given in the statement of the theorem.

This same kernel was found by Epanechnikov to minimize $\int E[f_n(t) - f(t)]^2\,dt$. This optimal kernel does not satisfy condition (1.5) of Theorem 1.1.

Recall that it was assumed that $na^7 \to 0$. Notice that if instead it is assumed that $K = K_2$ then (3.1) is minimized when

$$na^7 \to \frac{225f(\theta)}{4[f^{(3)}(\theta)]^2}.$$

However, when $na^7 \to d^2 > 0$ choosing a kernel with $B_2 = 0$ allows $\{a\}$ to converge to zero even more slowly, i.e., $d^2 = \infty$. Specifically, suppose $na^7 \to \infty$ and $na^{11} \to 0$, i.e., $p = 4$. Again K may be chosen to satisfy the conditions of Theorem 2.1 and to minimize V. The solution is given by:

THEOREM 3.2. *If K satisfies the conditions of theorem* 2.1 *for $p = 4$ and satisfies* (3.2) *then V is minimized when*

$$K(x) = K_4(x) = (\tfrac{15}{32})(3 - 10x^2 + 7x^4), \qquad |x| \leq 1$$
$$= 0, \qquad \textit{otherwise}.$$

Again notice that if it is assumed that $K = K_4$ then (3.1) is minimized when

$$na^{11} \to \frac{9f(\theta)\cdot V}{B_4^2[f^{(5)}(\theta)]^2}.$$

A curious phenomenon is occurring. If $\{a\}$ converges to zero at a certain rate then it is possible to find an optimum K for that rate. On the other hand if that optimum K is chosen then the optimum rate for $\{a\}$ to converge to zero is slower than the rate which gave rise to the kernel in the first place.

It is possible to continue in this fashion setting $B_i = 0$, $i \leqslant p - 1$ and letting $na^{3+2p} \to 0$. The optimal kernel is a polynomial of degree p. These polynomials have variation increasing in p; this is necessary to satisfy the bias constraints (2.2), but there is no kernel which satisfies the bias constraints for all p. Bartlett (1963) and Johns and Van Ryzin (1972) have also used kernels satisfying the constraints $\{B_i = 0\}$ to reduce the order of the bias in density estimates. For small and moderate sample sizes it seems risky to use high degree polynomial kernels although they are asymptotically better. Limited Monte-Carlo experiments have suggested that the kernel of K_4 of Theorem 3.2 does reduce the mean squared error of θ_n when compared with the kernel K_2 of Theorem 3.1 for sample sizes as small as $n = 20$ for a variety of densities.

4. Proof of Theorem 2.1. The first step is to show that $EZ_n(t)$ converges to a parabola. Since $X_1, \cdots, X_n$ are independent with common density f,

$$EZ_n(t) = \frac{1}{ab^2}\int\left[K\left(\frac{\theta + bt - x}{a}\right) - K\left(\frac{\theta - x}{a}\right)\right]f(x)\,dx.$$

Changing variables once for each term in square brackets yields

$$EZ_n(t) = \frac{1}{b^2}\int[f(\theta + ax + bt) - f(\theta + ax)]K(-x)\,dx.$$

Expanding f at $\theta + ax$ by Taylor's theorem, this becomes

$$EZ_n(t) = \frac{1}{b^2}\int\left[btf^{(1)}(\theta + ax) + \frac{b^2t^2}{2}f^{(2)}(\theta + ax) + \frac{b^3t^3}{6}f^{(3)}(\xi_1)\right]K(-x)\,dx \tag{4.1}$$

where ξ_1 lies between $\theta + ax$ and $\theta + ax + bt$.

Expanding $f^{(1)}$ at θ, the first term in (4.1) becomes

$$\frac{t}{b}\int\left[\Sigma_{i=1}^{p+1}\frac{(ax)^{i-1}f^{(i)}(\theta)}{(i-1)!} + \frac{(ax)^{p+1}f^{(p+2)}(\xi_2)}{(p+1)!}\right]K(-x)\,dx$$

where ξ_2 lies between θ and $\theta + ax$. Since $f^{(i)}(\theta) \cdot B_{i-1} = 0$, $1 \leqslant i \leqslant p$, this reduces to

$$\frac{t}{b}\left[\frac{a^pf^{(p+1)}(\theta)}{p!}\int x^pK(-x)\,dx + \frac{a^{p+1}}{(p+1)!}\int\int f^{(p+2)}(\xi_2)x^{p+1}K(-x)\,dx\right].$$

Because $a^p/b \to d$, the first of these two terms converges to

$$(-1)^{p+1} \cdot \frac{f^{(p+1)}(\theta)}{p!} \cdot B_p \cdot d \cdot t.$$

The second converges to zero since it is smaller in absolute value than

$$\frac{a^{p+1}t}{b(p+1)!} \cdot \sup_t |f^{(p+2)}(t)| \cdot \int |x^{p+1}K(x)|dx$$

and $a^{p+1}/b \to 0$.

Expanding $f^{(2)}$ at θ, the second term in (4.1) becomes

$$\frac{t^2}{2}\int \left[f^{(2)}(\theta) + axf^{(3)}(\xi_3)\right] K(-x)\, dx.$$

The first of these two terms is exactly

$$\frac{f^{(2)}(\theta)}{2} \cdot B_0 \cdot t^2$$

and the second is smaller in absolute value than

$$\frac{at^2}{2} \cdot \sup_t |f^{(3)}(t)| \cdot \int |xK(x)|dx$$

which converges to zero since $a \to 0$.

The third term in (4.1) is smaller in absolute value than

$$\frac{bt^3}{6} \cdot \sup_t |f^{(3)}(t)| \cdot \int |K(-x)|dx$$

which converges to zero since $b \to 0$. Thus

$$\lim_{n\to\infty} EZ_n(t) = \frac{f^{(2)}(\theta)}{2} \cdot B_0 \cdot t^2 + (-1)^{p+1} \cdot \frac{f^{(p+1)}(\theta)}{p!} \cdot B_p \cdot d \cdot t.$$

The second step in the proof of the theorem is to show that the deviations of Z_n from its expected value lie, with probability converging to one, on a straight line. It sufficies to show that $\mathrm{Var}[Z_n(t) - tZ_n(1)] \to 0$ as $n \to \infty$. Notice that $Z_n(t) - tZ_n(1)$ is the average of random variables identically distributed as

$$R_n(t) = \frac{1}{ab^2}\left\{ K\left(\frac{\theta + bt - X}{a}\right) - K\left(\frac{\theta - X}{a}\right) - t\left[K\left(\frac{\theta + b - X}{a}\right) - K\left(\frac{\theta - X}{a}\right)\right]\right\}.$$

So

$$\mathrm{Var}[Z_n(t) - tZ_n(1)] = \frac{1}{n}\mathrm{Var}[R_n(t)] \leqslant \frac{1}{n}\left[ER_n(t)^2\right]$$

$$= \frac{1}{na^2b^4}\int\Big\{K\Big(\frac{\theta + bt - x}{a}\Big) - K\Big(\frac{\theta - x}{a}\Big)$$

$$-t\Big[K\Big(\frac{\theta + b - x}{a}\Big) - K\Big(\frac{\theta - x}{a}\Big)\Big]\Big\}^2 f(x)\,dx$$

$$= \frac{1}{nab^4}\int\Big\{K\Big(\frac{bt}{a} - x\Big) - K(-x) - t\Big[K\Big(\frac{b}{a} - x\Big) - K(-x)\Big]\Big\}^2 f(\theta + ax)\,dx$$

$$(4.2) \qquad \leqslant \frac{f(\theta)t^2}{na^3b^2}\int\left\{\frac{K\Big(\frac{bt}{a} - x\Big) - K(-x)}{\frac{bt}{a}} - \frac{K\Big(\frac{b}{a} - x\Big) - K(-x)}{\frac{b}{a}}\right\}^2 dx.$$

Let $\delta = \frac{b}{a}$ and define

Then $q_\delta(x) = \frac{1}{\delta}$, $\leqslant X \leqslant \delta$, 0, otherwise

$$\frac{K(\delta t - x) - K(-x)}{\delta t} = \int_{-x}^{\delta t - x}\frac{K^{(1)}(u)}{\delta t}\,du$$

$$= \int_{-\infty}^{\infty}K^{(1)}(-x + y)q_{\delta t}(y)\,dy = \int_{-\infty}^{\infty}K^{(1)}(-x + yt)q_\delta(y)\,dy.$$

Thus

$$\left[\frac{K(\delta t - x) - K(-x)}{\delta t} - \frac{K(\delta - x) - K(-x)}{\delta}\right]^2$$

$$= \left[\int K^{(1)}(-x + yt)q_\delta(y)\,dy - \int K^{(1)}(-x + y)q_\delta(y)\,dy\right]^2$$

$$= \left\{\int\left[K^{(1)}(-x + yt) - K^{(1)}(-x + y)\right]q_\delta(y)\,dy\right\}^2$$

$$\leqslant \int\left[K^{(1)}(-x + yt) - K^{(1)}(-x + y)\right]^2 q_\delta(y)\,dy$$

by Jensen's inequality. Consequently the integral in (4.2) is smaller than

$$\int\left\{\int\left[K^{(1)}(-x + yt) - K^{(1)}(-x + y)\right]^2 q_\delta(y)\,dy\right\}dx$$

$$= \int\left\{\int\left[K^{(1)}(-x + yt) - K^{(1)}(-x + y)\right]^2 dx\right\}q_\delta(y)\,dy$$

(by Fubini's Theorem). By Theorem 13.24 of Hewitt and Stromberg (1965) the inner integral converges to zero as y converges to zero. Recall that $t \in [-T, T]$ is fixed and choose $\varepsilon(t) > 0$. There is a $\gamma(t)$ so that if $|y| < \gamma(t)$ then the inner integral is less than $\varepsilon(t)$. Choose $\delta < \gamma(t)$. Then the integral becomes

$$\frac{1}{\delta}\int_0^{\infty}\left\{\int_{-\infty}^{\infty}\left[K^{(1)}(-x + yt) - K^{(1)}(-x + y)\right]^2 dx\right\}dy < \frac{1}{\delta}\int_0^{\delta}\varepsilon(t)\,dy = \varepsilon(t).$$

Thus (4.2) converges to zero and

$$Z_n(t) - EZ_n(t) - t[Z_n(1) - EZ_n(1)] \to_P 0.$$

The next step in the proof of the theorem is to show that the deviation of the process at $t = 1$, $Z_n(1) - EZ_n(1)$, has an asymptotic normal distribution. $Z_n(1)$ is the average of random variables identically distributed as

$$U_n = \frac{1}{ab^2}\left[K\left(\frac{\theta + b - X}{a}\right) - K\left(\frac{\theta - X}{a}\right)\right].$$

So

$$\operatorname{Var}[Z_n(1)] = \frac{1}{n}\operatorname{Var}[U_n] = \frac{1}{n}E[U_n^2] - \frac{1}{n}[EU_n]^2.$$

Since $EU_n = EZ_n(1)$, as $n \to \infty$, $\frac{1}{n}[EU_n]^2$ converges to zero and

$$\frac{1}{n}E[U_n^2] = \frac{1}{na^2b^4}\int\left[K\left(\frac{\theta + b - x}{a}\right) - K\left(\frac{\theta - x}{a}\right)\right]^2 f(x)\,dx$$

$$= \frac{1}{nab^4}\int\left[K\left(\frac{b}{a} - x\right) - K(-x)\right]^2 f(\theta + ax)\,dx$$

$$(4.3) \qquad = \frac{1}{nab^4}\left\{f(\theta)\int\left[K\left(\frac{b}{a} - x\right) - K(-x))\right]^2 dx\right.$$

$$\left. + a\int x f^{(1)}(\xi_1)\left[K\left(\frac{b}{a} - x\right) - K(-x)\right]^2 dx\right\}$$

where ξ^1 lies between θ and $\theta + ax$.

The second term in (4.3) is smaller in absolute value than

$$\sup_t|f^{(1)}(t)| \cdot a \cdot \int|x|\left[K\left(\frac{b}{a} - x\right) - K(-x)\right]^2 dx.$$

This integral is smaller than

$$4\sup_x|K(x)|\left[\int|x| \cdot |K(x)|dx + \frac{b}{a}\int|K(x)|dx\right]$$

which is bounded. Since $a \to 0$ and $b \to 0$ the entire term may be neglected.

Letting $\delta = b/a$ and using an argument similar to the one used at (4.2)

$$\frac{K(\delta - x) - K(-x)}{\delta} = \int K^{(1)}(-x + y)q_\delta(y)\,dy$$

and

$$K^{(1)}(-x) = \int K^{(1)}(-x)q_\delta(y)\,dy,$$

so that

$$\int\left[\frac{K(\delta - x) - K(-x)}{\delta} - K^{(1)}(-x)\right]^2 dx$$

$$\leq \int\left\{\int[K^{(1)}(-x + y) - K^{(1)}(-x)]^2\,dx\right\}q_\delta(y)\,dy.$$

Choose $\varepsilon > 0$. Again there is a γ so that $|y| < \gamma$ implies the inner integral is less than ε. Choose $\delta < \gamma$. Then the whole expression is less than ε. Consequently the integral in the first term in (4.3) converges to

$$\int [K^{(1)}(-x)]^2 \, dx$$

and thus

$$\mathrm{Var}[Z_n(1)] \to f(\theta) \cdot V.$$

Since $Z_n(1)$ is the average of n random variables with the same distribution as U_n, Lindeberg's condition for asymptotic normality of $Z_n(1)$ (Billingsley (1968), Theorem 7.2) requires that

$$n \cdot \int_{\left|\frac{U_n - EU_n}{n}\right| > \varepsilon} \left[\frac{U_n - EU_n}{n}\right]^2 f(x) \, dx \to 0$$

as $n \to \infty$ for every $\varepsilon > 0$. Since K is bounded, say $\sup_x |K(x)| \leqslant D$,

$$|U_n| = \left|\frac{1}{b^2 a}\left[K\left(\frac{\theta + b - x}{a}\right) - K\left(\frac{\theta - x}{a}\right)\right]\right| \leqslant 2 \cdot D \cdot na^2.$$

Since $a \to 0$, for each $\varepsilon > 0$ there is an $n(\varepsilon)$ so that for all $n > n(\varepsilon)$, $|U_n - EU_n| \leqslant \varepsilon n$. Hence, for $n > n(\varepsilon)$

$$\frac{1}{n} \int_{|U_n - EU_n| > \varepsilon n} [U_n - EU_n]^2 f(x) \, dx = 0$$

and Lindeberg's condition is easily satisfied. Thus,

$$Z_n(1) - EZ_n(1) \to_D \mathfrak{N}(0, f(\theta) \cdot V).$$

At this point it has been shown that the finite-dimensional distributions of Z_n converge to those of Z. Since K is continuous (and hence Z_n is continuous), from Billingsley Theorem 8.1) it only remains to show that $\{Z_n\}$ is a tight sequence to complete the proof of the theorem. From Billingsley (Theorem 12.3), a sufficient condition that $\{Z_n\}$ be tight is that $\{Z_n(0)\}$ is tight and there exist $\gamma \geqslant 0$, $\alpha > 1$, and a continuous nondecreasing function H so that for all $s, t \in [-T, T]$ and $n \geqslant 1$

$$E\{|Z_n(s) - Z_n(t)|^\gamma\} \leqslant |H(s) - H(t)|^\alpha.$$

Now $Z_n(0) \equiv 0$ and hence is tight. So choose $\gamma = \alpha = 2$ and $H(s) = A \cdot s$ for some constant $A < \infty$ so that if

$$E\left\{\frac{Z_n(s) - Z_n(t)}{s - t}\right\}^2 \leqslant A^2$$

then $\{Z_n\}$ is tight. But

$$E\left\{\frac{Z_n(s) - Z_n(t)}{s - t}\right\}^2$$

$$= \frac{a^2}{b^2}\int \frac{1}{a}\left[\frac{K\left(\frac{\theta + bs - x}{a}\right) - K\left(\frac{\theta + bt - x}{a}\right)}{s - t}\right]^2 f(x)\, dx$$

$$= \int \left[\frac{K\left(\frac{b(s-t)}{a} - x\right) - K(-x)}{\frac{b(s-t)}{a}}\right]^2 f(\theta + bt + ax)\, dx$$

$$\leqslant f(\theta) \cdot \int \left[\frac{K(\delta - x) - K(-x)}{\delta}\right]^2 dx$$

letting $\delta = b(s - t)/a$. From the argument applied to the first term of (4.3), this integral is bounded (uniformly in s and t); hence $\{Z_n\}$ is tight. The proof of Theorem 2.1 is complete.

It should be noted that the proof was carried through for $t \in [-T, T]$. However, since T was arbitrary, by Theorem 5 of Whitt (1970), the proof is valid for all $t \in (-\infty, \infty)$.

5. Measurability of the Functional *M*. Because there is apparently no previous proof of the fact that M, defined in (1.2), is a measurable function on $C[0, 1] = C$, a proof is included here. Define, for each $c \in C$,

$$M_{\varepsilon n}(c) = \min_{0 \leqslant I \leqslant 2^n}\left\{\frac{I}{2^n}\middle| c\left(\frac{I}{2^n}\right) \geqslant \sup_{0 \leqslant t \leqslant 1} c(t) - \varepsilon\right\}$$

or $M_{\varepsilon n} = 1$ if the inequality is never satisfied. Since $M_{\varepsilon n}$ is a measurable function on C, if it can be shown that

$$\lim_{\varepsilon \to 0} \lim_{n \to \infty} M_{\varepsilon n}(c) = M(c)$$

for each $c \in C$, then M is a measurable function. Notice that for fixed ε, $M_{\varepsilon n}$ is a nonincreasing function of n. Also notice that continuity of c guarantees, for sufficiently small ε, the existence of an $n(\varepsilon)$ such that for all $n > n(\varepsilon)$, $M_{\varepsilon n}(c) \leqslant M(c)$. Now define

$$M_\varepsilon(c) = \inf_{0 \leqslant t \leqslant 1}\{t | c(t) \geqslant \sup_{0 \leqslant s \leqslant 1} c(s) - \varepsilon\}.$$

For each n, $M_{\varepsilon n}(c) \geqslant M_\varepsilon(c)$. Thus for ε small and n large $M_\varepsilon(c) \leqslant M_{\varepsilon n}(c) \leqslant M(c)$ and it only remains to show that

$$\lim_{\varepsilon \to 0} M_\varepsilon(c) = M(c).$$

Observe that M_ε is a nonincreasing function of ε. Since $M(c)$ is the smallest t which maximizes $c(t)$, for each $\delta > 0$ the maximum value of $c(t)$, $0 \leqslant t \leqslant M(c) - \delta$, is smaller than $c(M(c))$ (or $M(c) = 0 = M_\varepsilon(c)$). That is, there exists an $\varepsilon > 0$ so that $c(t) < c(M(c)) - \varepsilon$ for $0 \leqslant t \leqslant M(c) - \delta$. Thus, $M_\varepsilon(c) \leqslant M(c) - \delta$ so that $M_\varepsilon(c)$ increases to $M(c)$ as ε decreases to 0. Therefore M is measurable.

6. Acknowledgement. I am indebted to my dissertation advisor, J. A. Hartigan for his assistance and counsel in this research. I am grateful to D. Lambert for comments which led to improvements in presentation.

I would also like to thank an anonymous referee for his persistence and his patience. His persistence led to improved conditions for Theorem 2.1 and his patience permitted removal of several errors in the original manuscript.

REFERENCES

BARTLETT, M. S. (1963). Statistical estimation of density functions. *Sankhya A* **25** 245–254.

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

CHERNOFF, H. (1964). Estimation of the mode. *Ann. Inst. Statist. Math.* **16** 31–41.

EPANECHNIKOV, V. A. (1969). Nonparametric estimation of a multivariate probability density. *Theor. Probability Appl.* **14** 153–158.

GELFAND, I. M. and FOMIN, S. V. (1963). *Calculus of Variations*, Prentice-Hall, Englewood Cliffs.

HEWITT, E. and STROMBERG, K. (1965). *Real and Abstract Analysis*, Springer, New York.

JOHNS, M. V. and VAN RYZIN, J. (1972). Convergence rates for empirical Bayes two-action problems II. Continuous Case. *Ann. Math. Statist.* **43** 934–947.

KONAKOV, V. D. (1974). On the asymptotic normality of the mode of multidimensional distributions. *Theor. Probability Appl.* **19**, 794–799.

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

NADARAYA, E. A. (1965). On nonparametric estimates of density functions and regression curves. *Theor. Probability Appl.* **10** 186–190.

PARZEN, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* **33** 1065–1076.

ROSENBLATT, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* **27** 832–837.

SAMANTA, M. (1973). Nonparametric estimation of the mode of a multivariate density. *South African Statist. J.* **7** 109–117.

VAN RYZIN, J. (1969). On strong consistency of density estimates. *Ann. Math. Statist.* **40** 1765–1772.

WEGMAN, E. J. (1972). Nonparametric probability density estimation I. *Technometrics* **14** 533–546.

WHITT, W. (1970). Weak convergence of probability measures on the function space $C[0, \infty)$. *Ann. Math. Statist.* **41** 939–944.

DEPARTMENT OF STATISTICS
CARNEGIE-MELLON UNIVERSITY
SCHENLEY PARK
PITTSBURGH, PENNSYLVANIA 15213