Z. Wahrscheinlichkeitstheorie verw. Gebiete
59, 279–290 (1982)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete

The Asymptotic Distributions of Kernel Estimators of the Mode

William F. Eddy*

Department of Statistics, Carnegie-Mellon University, Pittsburgh, PA 15213, USA

Summary. In a decreasing sequence of intervals centered on the true mode the normalized kernel estimate of the density converges weakly to a non-stationary Gaussian random process. The expected value of this process is a parabola through the origin. The covariance function of this process depends on the smoothness of the kernel. When the kernel is mean-square differentiable the location of the maximum of this process has a normal distribution. When the kernel is discontinuous the location of the maximum has a distribution related to a solution of the heat equation.

1. Introduction

A mode of the probability density $f(t)$ is a value of t which maximizes f. To make this precise define the functional

$$M(g)=\inf\{t \mid g(t)=\sup_s g(s)\}. \tag{1.1}$$

The mode θ of a density f is $\theta=M(f)$. Let $X_1, \ldots, X_n$ be independent observations with common unknown density f. The kernel estimate of $f(t)$ (Rosenblatt (1956)) is

$$f_n(t)=\frac{1}{n a_n} \sum_{i=1}^{n} K\left(\frac{t-X_i}{a_n}\right) \tag{1.2}$$

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

The kernel estimate of the mode proposed by Parzen (1962) is $\theta_n=M(f_n)$. Parzen gives conditions under which $\{\theta_n\}$ is a consistent estimator of θ and

* Research supported in part by the National Science Foundation under grant MCS-78-02422 and MCS-80-05115 to Carnegie-Mellon University

0044-3719/82/0059/0279/$02.40

also gives conditions under which θ_n has an asymptotic normal distribution. He shows that

$$(na_n^3)^{\frac{1}{2}}(\theta_n-\theta)\xrightarrow{W}\mathcal{N}(0,f(\theta)\cdot V/[f^{(2)}(\theta)]^2)$$

where $f^{(2)}(\theta)$ is the second derivative of f evaluated at the mode and $V=\int[K^{(1)}(x)]^2dx$ (all integrals unless specified otherwise are taken over the whole real line); the symbol $\xrightarrow{W}$ stands for weak convergence (see, e.g., Billingsley (1968)).

Chernoff (1964) defines the naive estimator of the mode θ_n^* as "the center of that interval of length $2a_n$ which contains the most observations". This is nearly the kernel estimate θ_n with the kernel

$$K(x)=\tfrac{1}{2},\ |x|\leqq 1$$
$$=0,\ \text{otherwise}. \tag{1.3}$$

The distinction between θ_n^* and θ_n is due to the nonuniqueness of the location of the maximum of f_n for this kernel. Chernoff chooses the midpoint of an interval where f_n is maximized; here, for simplicity, the left-hand end point of the leftmost interval is chosen (1.1). This difference does not affect the local asymptotic behavior.

Since the kernel (1.3) is not continuous Parzen's regularity conditions are not satisfied. Chernoff (Sect. 5, Theorem 2) shows that $\{\theta_n^*\}$ is a consistent estimator of θ. Wegman (1971) gives a stronger consistency result. Chernoff derives the asymptotic distribution of

$$\left\{\frac{na_n^2[f^{(2)}(\theta)]^2}{2\cdot f(\theta)}\right\}^{\frac{1}{3}}(\theta_n^*-\theta).$$

He shows that the asymptotic distribution is the same as the distribution of the location of the maximum of the process $W(t)-t^2$ where $W(t)$ is a two-sided standard Brownian motion (i.e., $W(t)$, $t\geqq 0$ and $W(-t)$, $t\leqq 0$ are each an independent standard Brownian motion).

Recently, Eddy (1980), using the techniques of weak convergence in the space of continuous functions, derived a stronger form of Parzen's distributional result under generally weaker conditions. Eddy shows that an appropriately normalized version of the kernel estimate of the density converges weakly to a randomly located parabola through the origin with fixed second derivative. This result, together with the continuous mapping theorem (Billingsley (1968), Theorem 5.1), allows a proof that the location of the maximum of the normalized kernel estimator converges weakly to the Gaussian distribution

$$\mathcal{N}\left((-1)^p\cdot\frac{d\cdot f^{(p+1)}(\theta)\cdot B_p}{f^{(2)}(\theta)\cdot p!},\ \frac{f(\theta)\cdot V}{[f^{(2)}(\theta)]^2}\right) \tag{1.4}$$

where $B_p=\int x^pK(x)\,dx$ and $d^2=\lim_{n\to\infty}na_n^{3+2p}$. The non-zero mean of the asymptotic distribution occurs because $\{a_n\}$ may converge to zero at a slower rate under Eddy's conditions than it may under Parzen's conditions.

The objective of this work is to derive a more general form of the result in Eddy (1980). As special cases the result will include the work of Parzen (1962) and Chernoff (1964). The essential idea is that the density estimator f_n in a decreasing interval near the mode, under very general conditions, converges weakly to a particular Gaussian process. The mean function of the process is a parabola through the origin which depends on the moments of the kernel and on the unknown density and its derivatives at the mode. The covariance function depends on the cross product-moment of the divided differences of the kernel.

The asymptotic distribution of the kernel estimate of the mode is the same as the distribution of the location of the maximum of this process. When the kernel is mean-square differentiable a functional limit theorem shows that the location of the maximum has a normal distribution; when the kernel is discontinuous an argument given by Chernoff (1964) relates the distribution of the location of the maximum to the heat equation.

2. Weak Convergence of the Density Estimator

Let b_n be a positive constant and define the random process

$$Z_n(t)=b_n^{-2}[f_n(\theta+b_n t)-f_n(\theta)],\ t\in[-T,T] \tag{2.1}$$

for some $T<\infty$. The process Z_n is a normalized version of the density estimate in an interval centered at the mode. Also define the Gaussian random process

$$Z(t)=Y_\alpha(t)+(-1)^{p+1}\frac{d}{c\cdot p!}\cdot f^{(p+1)}(\theta)\cdot B_p\cdot t$$
$$+\tfrac{1}{2}f^{(2)}(\theta)\cdot t^2,\ t\in(-\infty,\infty) \tag{2.2}$$

where $c<\infty$ is a positive constant, $d<\infty$ is a non-negative constant, $p\geqq 2$ is a fixed integer, $B_p=\int x^p K(x)\,dx$, and Y_α is a mean-zero Gaussian random process with covariance function

$$R(s,t)=\frac{f(\theta)}{c^{2+\alpha}}\cdot V_\alpha(s/t)\cdot s\cdot t^{1-\alpha},\ 0\leqq\alpha\leqq 1,\ |s|\leqq|t|, \tag{2.3}$$

where

$$V_\alpha(\gamma)=\lim_{\delta\to 0}\delta^\alpha\int\left[\frac{K(\gamma\delta-x)-K(-x)}{\gamma\delta}\right]\left[\frac{K(\delta-x)-K(-x)}{\delta}\right]dx.$$

The purpose of this section is to prove Theorem 2.1 which gives conditions under which $\{Z_n\}$ converges weakly to Z.

When the kernel used to estimate f_n is the uniform kernel (1.3) the conditions of Theorem 2.1 will require $\alpha=1$. In this case the covariance function (2.3) will reduce to

$$R(s,t)=\frac{f(\theta)}{2c^3}\cdot|s|\cdot\delta(s,t)$$

$$\text{where}\quad \delta(s,t)=1,\ s\cdot t\geqq 0,$$
$$\qquad\qquad =0,\ s\cdot t<0.$$

This is the covariance function of the two-sided Brownian motion defined above. For this special case, Chernoff (1964, Sect. 3) outlines a proof of the theorem. When the kernel has a bounded square-integrable derivative the conditions will allow $\alpha=0$. In this case the covariance function reduces to

$$R(s,t)=\frac{f(\theta)}{c^2}\cdot V\cdot s\cdot t$$

where $V=\int[K^{(1)}(x)]^2\,dx$. The Gaussian process $Y_0(t)$ with this covariance function can be represented as $t\cdot Y$ where Y is a single Gaussian random variable. This process has also occurred in other contexts (see, e.g. Antille (1976)). Theorem 2.1 is proved in this case by Eddy (1980, Theorem 2.1).

In order that the theorem may include the uniform kernel it is necessary to consider convergence of a sequence of processes whose sample functions may be discontinuous. For this reason assume henceforth that the kernel K is continuous from the right and has limits from the left; this is equivalent to the assumption that the sample functions of the process Z_n lie (w.p.1) in the function space $\mathscr{D}=\mathscr{D}[-T,T]$. For the properties of the equivalent space $\mathscr{D}[0,1]$ refer to Billingsley (1968, Chap. 3).

Theorem 2.1. *Let $p\geqq 2$ be a fixed integer. Suppose that K is a bounded measurable function which is continuous from the right, has limits from the left, has a finite number of discontinuities, and satisfies*

$$\begin{aligned}&\int K(x)=B_0=1,\\ &\int x^i K(x)\,dx=B_i=0,\qquad 1\leqq i\leqq p-1,\\ &\int x^i K(x)\,dx=B_i<\infty,\qquad i=p,p+1,\end{aligned}$$

and suppose there is an α, $0\leqq\alpha\leqq 1$, so that

$$0<\lim_{\delta\to 0}\delta^\alpha\int\left[\frac{K(\gamma\delta-x)-K(-x)}{\gamma\delta}\right]\left[\frac{K(\delta-x)-K(-x)}{\delta}\right]dx=V_\alpha(\gamma)$$

for every γ, $|\gamma|\leqq 1$, and

$$\lim_{\delta\to 0}\delta^\alpha\int\left|x\left[\frac{K(\gamma\delta-x)-K(-x)}{\gamma\delta}\right]\left[\frac{K(\delta-x)-K(-x)}{\delta}\right]\right|dx<\infty.$$

Let $\{a_n\}$ be a sequence of positive constants which satisfies

$$\lim_{n\to\infty} n a_n^5=\infty,$$

$$\lim_{n\to\infty}[n a_n^{3+2p+\alpha(p-1)}]^{\frac{1}{2+\alpha}}=d<\infty,$$

for d a non-negative constant, and let $\{b_n\}$ be a sequence of positive constants which satisfies

$$\lim_{n\to\infty}\frac{a_n^p}{b_n}=\frac{d}{c},$$

for c a positive constant. If the density f has an absolutely continuous $(p+1)^{st}$ *derivative and satisfies*

$$\sup_t |f^{(p+2)}(t)| < \infty,$$

then

$$Z_n(t) \xrightarrow{W} Z(t),\ t \in (-\infty, \infty).$$

Proof. (The dependence of $a=a_n$ and $b=b_n$ on n will be suppressed henceforth.)

The first step is to show that $EZ_n(t)$ converges to a parabola. Following a step in the proof of Theorem 2.1 in Eddy (1980) and noting that here $\lim_{n\to\infty} \frac{a^p}{b} = \frac{d}{c}$ it is immediate that

$$\lim_{n\to\infty} \sup_t \left| EZ_n(t) - (-1)^{p+1} \frac{f^{(p+1)}(\theta)}{p!} \cdot B_p \cdot \frac{d}{c} \cdot t - \frac{f^{(2)}(\theta)}{2} \cdot t^2 \right| = 0.$$

The second step in the proof of the theorem is to determine the asymptotic behavior of the covariance function. Since $Z_n(t)$ is an average of independent random variables identically distributed as

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

$$\begin{aligned}\operatorname{Cov}[Z_n(s), Z_n(t)] &= \frac{1}{n}\operatorname{Cov}[U_n(s), U_n(t)] \\ &= \frac{1}{n} E[U_n(s)\, U_n(t)] - \frac{1}{n} EU_n(s)\, EU_n(t).\end{aligned}$$

Because $EU_n(t) = EZ_n(t)$ and $\lim_{n\to\infty} EZ_n(t) < \infty$, $\lim_{n\to\infty} \frac{1}{n} EU_n(s)\, EU_n(t) = 0$. Now

$$\begin{aligned}&\frac{1}{n} E[U_n(s)\, U_n(t)] \\ &= \frac{1}{na^2b^4} \int \left[K\left(\frac{\theta+bs-x}{a}\right) - K\left(\frac{\theta-x}{a}\right)\right]\left[K\left(\frac{\theta+bt-x}{a}\right) - K\left(\frac{\theta-x}{a}\right)\right] f(x)\,dx \\ &= \frac{1}{nab^4} \int \left[K\left(\frac{bs}{a}-x\right) - K(-x)\right]\left[K\left(\frac{bt}{a}-x\right) - K(-x)\right] f(\theta+ax)\,dx \\ &= \frac{f(\theta)\cdot s\cdot t^{1-\alpha}}{na^{3-\alpha}b^{2+\alpha}} \left(\frac{bt}{a}\right)^{\alpha} \int \left[\frac{K\left(\frac{bs}{a}-x\right) - K(-x)}{\frac{bs}{a}}\right]\left[\frac{K\left(\frac{bt}{a}-x\right) - K(-x)}{\frac{bt}{a}}\right] dx \\ &\quad + \frac{a\cdot s\cdot t^{1-\alpha}}{na^{3-\alpha}b^{2+\alpha}} \left(\frac{bt}{a}\right)^{\alpha} \int f^{(1)}(\xi)\, x \left[\frac{K\left(\frac{bs}{a}-X\right) - K(-x)}{\frac{bs}{a}}\right]\left[\frac{K\left(\frac{bt}{a}-x\right) - K(-x)}{\frac{bt}{a}}\right] dx\end{aligned} \tag{2.4}$$

where ξ lies between θ and $\theta+ax$. Setting $\delta=\frac{bt}{a}$ and $\gamma=\frac{s}{t}$, the second term in (2.4) is smaller in absolute value than

$$\frac{a\cdot s\cdot t^{1-\alpha}}{na^{3-\alpha}b^{2+\alpha}}\cdot\sup_t |f^{(1)}(t)| \cdot \delta^\alpha \int \left| x\left[\frac{K(\gamma\delta-x)-K(-x)}{\gamma\delta}\right]\left[\frac{K(\delta-x)-K(-x)}{\delta}\right]\right| dx$$

which converges to zero by assumption. The first term in (2.4) converges to

$$\frac{f(\theta)}{c^{2+\alpha}}\cdot s\cdot t^{1-\alpha}\cdot V_\alpha(s/t) \tag{2.5}$$

as required.

The next step in the proof is to show that the finite-dimensional distributions of the process Z_n converge to those of a Gaussian process; for each set $(t_1,\dots,t_r)$ it must be shown that the set

$$\{Z_n(t_1)-EZ_n(t_1),\dots,Z_n(t_r)-EZ_n(t_r)\}$$

has an asymptotic joint Gaussian distribution. Since, for each fixed t, $Z_n(t)$ is the average of n independent random variables with the same distribution as

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

application of the multivariate central limit theorem for triangular arrays will complete the proof.

Let $V_n=\sum_{j=1}^r c_j[U_n(t_j)-EU_n(t_j)]$ and let $C=\sum_{j=1}^r |c_j|$ for any constants $c_1,\dots,c_r$. Lindeberg's condition for asymptotic normality of $Z_n^*=\sum_{j=1}^r c_j Z_n(t_j)$ (Billingsley, Theorem 7.2) is that

$$\frac{n}{s_n^2}\int_{\Omega_n(\varepsilon)}\left[\frac{V_n}{n}\right]^2 f(x)\,dx\to 0 \tag{2.6}$$

as $n\to\infty$ for every $\varepsilon>0$ where

$$\Omega_n(\varepsilon)=\left\{\left|\frac{V_n}{n}\right|\geqq \varepsilon s_n\right\} \tag{2.7}$$

and $s_n^2=\frac{1}{n}\operatorname{Var}(V_n)$. From the proof of (2.5), s_n^2 has a finite limit, say σ^2. Since K is bounded, $\sup_x |K(x)|\leqq A$,

$$|V_n|=\left|\frac{1}{b^2 a}\sum_{j=1}^{r} c_j\left[K\left(\frac{\theta+bt_j-X}{a}\right)-K\left(\frac{\theta-X}{a}\right)\right]\right|$$
$$\leqq\frac{4AC}{b^2 a}=n\cdot\frac{4AC}{nb^2 a}.$$

As $n\to\infty$ $nb^2a\to\infty$. Thus there is an $n(\varepsilon)$ so that if $n>n(\varepsilon)$, $|V_n|\leqq\varepsilon\cdot n$. Consequently, there is a $\delta(\varepsilon)$ so that for $n>n(\varepsilon)(\sigma+\delta(\varepsilon))$ the region (2.7) is vacuous and (2.6) is trivially satisfied. This is true for any constants $c_1,\dots,c_r$ and hence $\{Z_n(t_1)-EZ_n(t_1),\dots, Z_n(t_r)-EZ_n(t)\}$ have an asymptotic joint Gaussian distribution.

To complete the proof of the theorem all that remains is to show that $\{Z_n-EZ_n\}$ is a tight sequence (Billingsley, Theorem 15.6). From Billingsley (Theorem 15.7) a sufficient condition that $\{Z_n\}$ be tight is that there exist $\gamma\geqq 0$, $\beta>\frac{1}{2}$ and a continuous nondecreasing function H so that for $t_1\leqq s\leqq t_2$ and $n\geqq 1$,

$$E\{|Z_n(s)-Z_n(t_1)-E[Z_n(s)-Z_n(t_1)]|^\gamma \cdot|Z_n(t_2)-Z_n(s)-E[Z_n(t_2)-Z_n(s)]|^\gamma\}\leqq[H(t_2)-H(t_1)]^{2\beta}. \tag{2.8}$$

There are two cases: $0\leqq\alpha<1$ and $\alpha=1$.

First consider $0\leqq\alpha<1$ and choose $\gamma=1$. For convenience suppose $t_1\geqq 0$. By the Cauchy-Schwarz inequality the left-hand side of (2.8) is smaller than

$$\{\operatorname{Var}[Z_n(s)-Z_n(t_1)]\cdot\operatorname{Var}[Z_n(t_2)-Z_n(s)]\}^{\frac{1}{2}}. \tag{2.9}$$

Now

$$\operatorname{Var}[Z_n(s)-Z_n(t)]=\frac{1}{n}\operatorname{Var}[U_n(s)-U_n(t)]$$
$$\leqq\frac{1}{n}E[U_n(s)-U_n(t)]^2$$
$$=\frac{1}{na^2b^4}E\left[K\left(\frac{\theta+bs-X}{a}\right)-K\left(\frac{\theta+bt-X}{a}\right)\right]^2$$
$$=\frac{1}{na^2b^4}\int\left[K\left(\frac{\theta+bs-x}{a}\right)-K\left(\frac{\theta+bt-x}{a}\right)\right]^2 f(x)\,dx$$
$$=\frac{1}{nab^4}\int\left[K\left(\frac{b(s-t)}{a}-x\right)-K(-x)\right]^2 f(\theta+bt+ax)\,dx$$
$$\leqq\frac{(s-t)^{2-\alpha}}{na^{3-\alpha}b^{2+\alpha}}f(\theta)\cdot\delta^\alpha\int\left[\frac{K(\delta-x)-K(-x)}{\delta}\right]^2 dx$$

letting $\delta=b(s-t)/a$.

This last expression is smaller than

$$C(s-t)^{2-\alpha}$$

for some positive constant C. Thus (2.9) is smaller than

$$[C(s-t_1)^{2-\alpha}\cdot C(t_2-s)^{2-\alpha}]^{\frac{1}{2}}\leqq C(t_2-t_1)^{2-\alpha}.$$

Letting $\beta=(2-\alpha)/2$ and $H(s)=C^{\frac{1}{2-\alpha}}\cdot s$ (2.8) is satisfied for $0\leqq\alpha<1$.

The case $\alpha=1$ will be handled in a different way. First, note that if $\{A_n\}$ and $\{B_n\}$ are tight sequences then $\{A_n+B_n\}$ is a tight sequence. To see this, recall there are compact sets K_A and K_B so that

$$\Pr\{A_n\notin K_A\}<\varepsilon, \qquad \Pr\{B_n\notin K_B\}<\varepsilon.$$

and define

$$S=\{A_n+B_n \mid A_n\in K_A, B_n\in K_B\}.$$

The set S is compact and

$$\begin{aligned}\Pr\{A_n+B_n\in S\} &\geqq \Pr\{A_n\in K_A, B_n\in K_B\}\\ &=1-\Pr\{A_n\notin K_A \text{ or } B_n\notin K_B\}\\ &\geqq 1-[\Pr\{A_n\notin K_A\}+\Pr\{B_n\notin K_B\}]\geqq 1-2\varepsilon.\end{aligned}$$

Thus $\{A_n+B_n\}$ is tight.

Second, note that since K has a finite number of discontinuities Z_n can be written as

$$Z_n=C_n+D_n$$

where C_n is continuous and D_n is a step function. Tightness of $\{C_n\}$ has already been demonstrated. Since D_n is a linear combination of indicators of intervals, by repeated use of the argument above, tightness of $\{D_n\}$ (and hence $\{Z_n\}$) will follow from tightness when K is the uniform kernel (1.3).

Let

$$Y_1=\sum_i I\{X_i\in(\theta+bt_1+a, \theta+bs+a]\},$$

$$Y_2=\sum_i I\{X_i\in(\theta+bt_1-a, \theta+bs-a]\},$$

$$Y_3=\sum_i I\{X_i\in(\theta+bs+a, \theta+bt_2+a]\},$$

$$Y_4=\sum_i I\{X_i\in(\theta+bs-a, \theta+bt_2-a]\}.$$

and

$$p_1=F(\theta+bs+a)-F(\theta+bt_1+a)=b(s-t_1)f(\theta+a)+O(b^2),$$

$$p_2=F(\theta+bs-a)-F(\theta+bt_1-a)=b(s-t_1)f(\theta-a)+O(b^2),$$

$$p_3=F(\theta+bt_2+a)-F(\theta+bs+a)=b(t_2-s)f(\theta+a)+O(b^2),$$

$$p_4=F(\theta+bt_2-a)-F(\theta+bs-a)=b(t_2-s)f(\theta-a)+O(b^2).$$

Choose $\gamma=2$, $\beta=1$, and $H(t)=Ct$ in (2.8). Since Y_1 is independent of Y_2 and Y_3 is independent of Y_4 for $n \geqq n_0$ it is enough to show that

$$\frac{1}{(2nab^2)^4} E[(Y_i-EY_i)^2(Y_j-EY_j)^2] \leqq C^2(t_2-t_1)^2$$

for $n \geqq n_0$ and $i=1,2$ and $j=3,4$.
Condition on $Y_j=k$:

$$E[(Y_i-EY_i)^2(Y_j-EY_j)^2]=E\{(k-np_j)^2 E[(Y_i-EY_i)^2 \mid Y_j=k]\}.$$

When $Y_j=k$, Y_i is binomial with parameters $n-k$ and $p_i/(1-p_j)$ and thus has expectation

$$E_{ik}=\frac{(n-k)\,p_i}{1-p_j}.$$

So

$$E[(Y_i-EY_i)^2 \mid Y_j=k] \leqq 2E[(Y_i-E_{ik})^2 \mid Y_j=k]+2(E_{ik}-np_i)^2.$$

But

$$E[(Y_i-E_{ik})^2 \mid Y_j=k]=2(n-k)\frac{p_i}{1-p_j}\left(1-\frac{p_i}{1-p_j}\right) \leqq 2n\frac{p_i}{1-p_j}$$

and thus

$$E[(Y_i-EY_i)^2 \mid Y_j=k]\cdot(k-np_j)^2 \leqq \frac{2np_i}{1-p_j}(k-np_j)^2+\frac{2p_i^2}{(1-p_j)^2}(k-np_j)^4.$$

Since the fourth central moment of a binomial random variable is $O(n^2)$, integration yields

$$E[(Y_i-EY_i)^2(Y_j-EY_j)^2] \leqq \frac{2p_i}{1-p_j}\, n^2\, p_j(1-p_j)+C\frac{2p_i^2}{(1-p_j)^2}\, n^2\, p_j.$$

Hence

$$\begin{aligned}\frac{1}{(2nab^2)^4} E[(Y_i-EY_i)^2(Y_j-EY_j)^2] &\leqq \frac{Cp_i\,p_j}{n^2a^4b^8}\\ &\leqq C(s-t_1)(t_2-s) \leqq C^2(t_2-t_1)^2,\end{aligned}$$

as required. When $\alpha=1$, $\{Z_n-EZ_n\}$ is tight.

Thus it has been shown that $Z_n \xrightarrow{W} Z$ for $t\in[-T,T]$. By virtue of Theorem 3 of Lindvall (1973) the proof extends immediately from $t\in[-T,T]$ to $t\in(-\infty,\infty)$.

3. Asymptotic Distribution of the Mode Estimator

When $\alpha=0$ the continuous mapping theorem (Billingsley, Theorem 5.2.(i)) leads to the asymptotic distribution (1.4) for $M(Z_n)$ (Eddy, 1980)). This method of proof was possible because the sample functions of Z have an analytic representation.

When K is the uniform kernel (1.3) (Chernoff (1964), Theorem 1, Sect. 4) derived the distribution of $M(Z)$; actually, he derived the distribution of $M(Z^*)$ where $Z^*(t)=W(t)-t^2$ and $W(t)$ is a two-sided standard Brownian motion. Except for location and scale changes this distribution is the same as the distribution of $M(Z)$. He showed that the probability density of $M(Z^*)$ can be written as

$$f(t)=\tfrac{1}{2}U_x(t^2,t)\,U_x(t^2,-t) \tag{3.1}$$

where

$$U(x,y)=\Pr\{Z(t)>t^2, t>y \mid Z(y)=x\}$$

is a solution of the heat equation

$$\tfrac{1}{2}U_{xx}=-U_y$$

subject to the boundary conditions

(i) $U(x,y)=1,\ x\geqq y^2$

(ii) $\lim_{x\to-\infty} U(x,y)=0$

and U_x is the partial derivative of $U(x,y)$ with respect to x. This result was possible because the process $Y_1(t)$ has independent increments.

Since every discontinuous kernel requires $\alpha=1$, Chernoff's argument yields the same distribution (3.1), up to location and scale changes, for the location of the maximum; that is, the asymptotic distribution of the kernel estimator of the mode for any kernel satisfying the conditions of Theorem 2.1 with $\alpha=1$ has density (3.1). Prakasa Rao (1969) obtained (3.1) as the asymptotic distribution of the maximum likelihood estimator of a unimodal density at a fixed point.

When $\alpha \neq 0, 1$ little is known about the distribution of $M(Z)$. The sample functions of Z are continuous with probability one. This follows immediately from a corollary of Cramer and Leadbetter (1967, p. 65). Measurability of $M(\cdot)$ on the space of continuous functions was established in Eddy (1980). Proving continuity of $M(\cdot)$ seems quite difficult. It is possible however to compute the expected value of $M(Z)$. The process Y_α is symmetric around zero in the sense that for each t and y

$$\Pr\{Y_\alpha(t)>y\}=\Pr\{Y_\alpha(-t)>y\}.$$

Thus, if $M(Z)$ has an expected value then

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

The tail behavior of the density of $M(Z)$ for $\alpha>0$ can be determined from the following argument. Define

$$\tilde{Y}_\alpha(t)=Y_\alpha(t)/\sqrt{R(t,t)}$$

and

$$\tilde{Z}_\alpha(t)=\begin{cases} t^{\frac{2-\alpha}{2}}\,\tilde{Y}_\alpha(t)-t^2 & t\geqq 0 \\ -|t|^{\frac{2-\alpha}{2}}\,\tilde{Y}_\alpha(t)-t^2 & t<0 \end{cases}$$

so that $M(\tilde{Z}_\alpha)$ has the same distribution as $M(Z)$ (up to location and scale changes). Notice that the covariance function of $\tilde{Y}_\alpha(t)$ is

$$(s/t)^{\alpha/2} \qquad 0\leqq s\leqq t.$$

Consequently there is a standard Gaussian random variable X satisfying

$$\lim_{t\to\infty}|\tilde{Y}_\alpha(t)-X|=0 \text{ w.p.l.}$$

and thus

$$\lim_{t\to\infty}|\tilde{Z}_\alpha(t)-(t^{\frac{2-\alpha}{2}}X-t^2)|=0 \qquad \text{w.p.l.}$$

Therefore, if $M(\tilde{Z}_\alpha)\gg 0$ it will be approximately equal to $M(t^{\frac{2-\alpha}{2}}X-t^2)$. But this last random variable is simply

$$\left(\frac{2-\alpha}{4}X\right)^{\frac{2}{2+\alpha}}$$

which has the density function

$$f(t)=\frac{2}{\sqrt{2\pi}}\left(\frac{2+\alpha}{2-\alpha}\right)|t|^{\frac{\alpha}{2}}\exp\left[\frac{-8}{(2-\alpha)^2}|t|^{2+\alpha}\right].$$

This density has the same tail behavior as the density of $M(Z)$.

4. Concluding Remarks

If both the kernel K and the density f are symmetric then there is no asymptotic bias effect. In this case other methods may be more appropriate (see, for example, Stone (1975)). It should be noted however that if

$$\rho(x)\propto K\left(\frac{x}{a}\right)$$

then T_n, the value of t which minimizes

$$g_n(t)=-\sum_{i=1}^{n}\rho(t-x_i),$$

is equal to θ_n. The estimate T_n is Huber's (1964) M-estimator of θ. For any fixed sample size Huber's M-estimate and Parzen's kernel estimate are identical. With some effort Theorem 2.1 can be modified to cover this case.

The case when K is symmetric and f is not is also interesting. The variance of θ_n is

$$O((na^{3-\alpha})^{-\frac{2}{2+\alpha}})$$

and the bias is $O(a^p)$. Since K is symmetric, its odd moments are zero and hence p is even, say $p=2q$. The optimal rate for $\{a\}$ to converge to zero under these conditions is

$$a \approx n^{-\frac{1}{3+4q+\alpha(2q-1)}}$$

and this leads to a mean-square error of order

$$n^{-\frac{4q}{3+4q+\alpha(2q-1)}}.$$

When K is a symmetric probability density $q=1$. If K also has a bounded square integrable derivative then $\alpha=0$, $a \approx n^{-1/7}$ and the mean-square error is of order $n^{-4/7}$. If K is the uniform kernel (1.3) then $q=1$, $\alpha=1$, the optimal rate for $\{a\}$ is $a \approx n^{-1/8}$ and the mean-square error is of order $n^{-1/2}$. These two special cases were noted by Chernoff (1964).

The conditions of Theorem 2.1 do not guarantee that the global maximum of f_n is near θ. Additional conditions on the tail of f would be required.

Acknowledgement. The possibility of achieving the unification of these results using weak convergence was first suggested to the author by Leo Breiman. I am indebted to André Antille for the tightness proof when $\alpha=1$.

Bibliography

Antille, A.: Asymptotic linearity of Wilcoxon signed-rank statistics. Ann. Statist. **4**, 175–186 (1976)

Billingsley, P.: Convergence of Probability Measures. New York: John Wiley 1968

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

Cramer, H., Leadbetter, M.R.: Stationary and Related Stochastic Processes. New York: John Wiley 1967

Eddy, W.F.: Optimum kernel estimators of the mode. Ann. Statist. **8**, 870–882 (1980)

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

Lindvall, T.: Weak convergence of probability measures and random functions in the function space $D[0, \infty]$. J. Appl. Probab. **10**, 109–121 (1973)

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

Prakasa Rao, B.L.S.: Estimation of a unimodal density. Sankhya, **A**, **31**, 23–36 (1969)

Rosenblatt, M.: Remarks on some nonparametric estimates of a density function. Ann. Math. Statist. **27**, 823–837 (1956)

Stone, C.J.: Adaptive maximum likelihood estimators of a location parameter. Ann. Statist. **3**, 267–284 (1973)

Wegman, E.J.: A note on the estimation of the mode. Ann. Math. Statist. **42**, 1909–1915 (1971)

Received August 18, 1980; in revised form September 11, 1981