ON ESTIMATION OF A PROBABILITY DENSITY FUNCTION AND MODE[1]

By Emanuel Parzen

Stanford University

0. Introduction. Given a sequence of independent identically distributed random variables $X_1, X_2, \cdots, X_n, \cdots$ with common probability density function $f(x)$, how can one estimate $f(x)$?

The problem of estimation of a probability density function $f(x)$ is interesting for many reasons. As one possible application, we mention the problem of estimating the hazard, or conditional rate of failure, function $f(x)/\{1 - F(x)\}$, where $F(x)$ is the distribution function corresponding to $f(x)$. In this paper we discuss the problem of estimation of a probability density function and the problem of determining the mode of a probability density function. Despite the obvious importance of these problems, we are aware of only two previous papers on the subject of estimation of the probability density function (Rosenblatt [5] and Whittle [6]).

In this paper we show how one may construct a family of estimates of $f(x)$, *and of the mode,* which are consistent and asymptotically normal. We shall see that there are a multitude of possible estimates. We do not examine here the question of which estimate to use.

The problem of estimating a probability density function is in some respects similar to the problem of estimating the spectral density function of a stationary time series; the methods employed here are inspired by the methods used in the treatment of the latter problem (see Parzen [4] for references). The problem of estimating the mode of a probability density function is somewhat similar to the problem of maximum likelihood estimation of a parameter; the methods employed here are inspired by the methods used in the treatment of the latter problem (see Le Cam [2] for references).

1. A class of estimates of the probability density function. Let $X_1, X_2, \cdots, X_n$ be independent random variables identically distributed as a random variable X whose distribution function $F(x) = P[X \leqq x]$ is absolutely continuous,

$$F(x) = \int_{-\infty}^{x} f(x')\,dx', \tag{1.1}$$

with probability density function $f(x)$.

As an estimate of the value $F(x)$ of the distribution function at a given point x, it is natural to take the sample distribution function

Received August 29, 1961.

[1] This work was supported in part by Office of Naval Research Contract Nonr-225(21) at Stanford University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

$$(1.2) \quad F_n(x) = (1/n)\{\text{no. of observations } \leqq x \text{ among } X_1, \cdots, X_n\}$$

which is essentially a binomially distributed random variable whose mean and variance are respectively given by

$$(1.3) \qquad E[F_n(x)] = F(x),$$

$$(1.4) \qquad \text{Var}\,[F_n(x)] = \{(1/n)F(x)\}\{1 - F(x)\}.$$

Various possible estimates of the probability density functions suggest themselves but none of them appear to be naturally superior. For example, as an estimate of $f(x)$ one might take

$$(1.5) \qquad f_n(x) = \{F_n(x + h) - F_n(x - h)\}/2h$$

where h is a suitably chosen positive number. However, how should one choose h? It is clear that h should be chosen as a function of n which tends to 0 as n tends to ∞. But how fast should h tend to zero? In order to answer this question we will have to study the statistical properties of the estimate defined by (1.5). In particular we must study how the mean and variance of $f_n(x)$ depends on h.

It turns out that to study the estimate defined by (1.5) one may as well study a very general class of estimates to be defined by (1.7) below. Let $K(y)$ be the function defined by

$$(1.6) \qquad \begin{aligned} K(y) &= \tfrac{1}{2}, & |y| &\leqq 1, \\ &= 0, & |y| &> 1. \end{aligned}$$

Then the estimate given in (1.5) can essentially be written as a weighted average over the sample distribution function:

$$(1.7) \qquad f_n(x) = \int_{-\infty}^{\infty} \frac{1}{h} K\left(\frac{x - y}{h}\right) dF_n(y) = \frac{1}{nh} \sum_{j=1}^{n} K\left(\frac{x - X_j}{h}\right).$$

By writing (1.5) in the form (1.7) we are immediately made aware of a multitude of possible estimates $f_n(x)$ for the probability density function $f(x)$. Instead of the function $K(y)$ defined by (1.6) we could choose other functions $K(y)$. We are thus led to the problem of studying the statistical properties of estimates of the form of (1.7) where h and $K(y)$ are suitably chosen.

We first examine what are the conditions under which estimates of the form of (1.7) are asymptotically unbiased in the sense that if $h = h(n)$ is chosen as a function of n such that

$$(1.8) \qquad \lim_{n\to\infty} h(n) = 0,$$

then

$$(1.9) \qquad \lim_{n\to\infty} E[f_n(x)] = f(x).$$

Now

$$(1.10) \quad E[f_n(x)] = E\left[\frac{1}{h(n)} K\left(\frac{x - X}{h(n)}\right)\right] = \int_{-\infty}^{\infty} \frac{1}{h(n)} K\left(\frac{x - y}{h(n)}\right) f(y)\, dy.$$

In order for (1.9) to hold, the last expression in (1.10) must tend to $f(x)$. Conditions under which this happens are given by the following theorem.

THEOREM 1A. *Suppose $K(y)$ is a Borel function satisfying the conditions*

$$\sup_{-\infty<y<\infty} |K(y)| < \infty, \tag{1.11}$$

$$\int_{-\infty}^{\infty} |K(y)|\, dy < \infty, \tag{1.12}$$

$$\lim_{y\to\infty} |yK(y)| = 0. \tag{1.13}$$

Let $g(y)$ satisfy

$$\int_{-\infty}^{\infty} |g(y)|\, dy < \infty. \tag{1.14}$$

Let $\{h(n)\}$ be a sequence of positive constants satisfying (1.8). *Define*

$$g_n(x) = \frac{1}{h(n)} \int_{-\infty}^{\infty} K\left(\frac{y}{h(n)}\right) g(x-y)\, dy. \tag{1.15}$$

Then at every point x of continuity of $g(\cdot)$,

$$\lim_{n\to\infty} g_n(x) = g(x) \int_{-\infty}^{\infty} K(y)\, dy. \tag{1.16}$$

REMARK. This theorem may essentially be found in Bochner [1]. Because it plays such a central role in this paper, we give the proof here since it is brief.

PROOF. Note first that

$$g_n(x) - g(x) \int_{-\infty}^{\infty} K(y)\, dy = \int_{-\infty}^{\infty} \{g(x-y) - g(x)\} \frac{1}{h(n)} K\left(\frac{y}{h(n)}\right) dy.$$

Let $\delta > 0$, and split the region of integration into two regions, $|y| \leqq \delta$ and $|y| > \delta$. Then

$$\begin{aligned}
\left| g_n(x) - g(x) \int_{-\infty}^{\infty} K(y)\, dy \right| &\leqq \max_{|y|\leqq\delta} |g(x-y) - g(x)| \int_{|z|\leqq\delta/h(n)} |K(z)|\, dz \\
&\quad + \int_{|y|\geqq\delta} \frac{|g(x-y)|}{y} \frac{y}{h(n)} K\left(\frac{y}{h(n)}\right) dy + |g(x)| \int_{|y|\geqq\delta} \frac{1}{h(n)} K\left(\frac{y}{h(n)}\right) dy \\
&\leqq \max_{|y|\leqq\delta} |g(x-y) - g(x)| \int_{-\infty}^{\infty} |K(x)|\, dz + \frac{1}{\delta} \sup_{|z|\geqq\delta/h(n)} |zK(z)| \int_{-\infty}^{\infty} |g(y)|\, dy \\
&\quad + |g(x)| \int_{|z|\geqq\delta/h(n)} |K(z)|\, dz,
\end{aligned}$$

which tends to 0 as one lets n tend to ∞, and then lets δ tend to 0.

COROLLARY 1A. *The estimates defined by* (1.7) *are asymptotically unbiased at all points x at which the probability density function is continuous if the constants h*

TABLE 1

$K(y)$	$k(u) = \int_{-\infty}^{\infty} e^{iuy} K(y)\, dy$	$\int_{-\infty}^{\infty} K^2(y)\, dy = \frac{1}{2\pi} \int_{-\infty}^{\infty} k^2(u)\, du$
$\begin{cases} \frac{1}{2}, & \lvert y \rvert \leqq 1 \\ 0, & \lvert y \rvert \geqq 1 \end{cases}$	$(\sin u)/u$	$\frac{1}{2}$
$\begin{cases} 1 - \lvert y \rvert, & \lvert y \rvert \leqq 1 \\ 0, & \lvert y \rvert \geqq 1 \end{cases}$	$\left\{ \frac{\sin (u/2)}{u/2} \right\}^2$	$\frac{2}{3}$
$\begin{cases} (\frac{4}{3}) - 8y^2 + 8 \lvert y \rvert^3, & \lvert y \rvert < \frac{1}{2} \\ \frac{8}{3}(1 - \lvert y \rvert)^3, & \frac{1}{2} \leqq \lvert y \rvert \leqq 1 \\ 0, & \lvert y \rvert > 1 \end{cases}$	$\left\{ \frac{\sin (u/4)}{u/4} \right\}^4$	0.96
$(2\pi)^{-\frac{1}{2}} e^{-(\frac{1}{2}) y^2}$	$e^{-(\frac{1}{2}) u^2}$	$[2(\pi^{\frac{1}{2}})]^{-1}$
$\frac{1}{2} e^{-\lvert y \rvert}$	$(1 + u^2)^{-1}$	$\frac{1}{2}$
$(1/\pi)(1 + y^2)^{-1}$	$e^{-\lvert u \rvert}$	$(1/\pi)$
$\frac{1}{2\pi} \left(\frac{\sin (y/2)}{y/2} \right)^2$	$\begin{cases} 1 - \lvert u \rvert, & \lvert u \rvert \leqq 1 \\ 0, & \lvert u \rvert \geqq 1 \end{cases}$	$1/(3\pi)$

satisfy (1.8) *and if the function* $K(y)$ *satisfies* (1.11)-(1.13) *and in addition satisfies*

$$\int_{-\infty}^{\infty} K(y)\, dy = 1. \tag{1.17}$$

For ease of exposition, an even function $K(y)$ satisfying (1.11)-(1.13) and (1.17) will be called a *weighting function.* Some examples of weighting functions are given in Table 1.

2. Consistency and asymptotic normality. The variance of the estimate $f_n(x)$ is given by

$$\operatorname{Var}[f_n(x)] = n^{-1} \operatorname{Var}[h^{-1} K((x - X)/h)]. \tag{2.1}$$

Now by Theorem 1A

$$hE\left[\left\{ \frac{1}{h} K\left(\frac{x - X}{h} \right) \right\}^2 \right] = \frac{1}{h} \int_{-\infty}^{\infty} K^2\left(\frac{x - y}{h} \right) f(y)\, dy \to f(x) \int_{-\infty}^{\infty} K^2(y)\, dy, \tag{2.2}$$

since (1.11) and (1.12) imply that

$$\int_{-\infty}^{\infty} K^2(y)\, dy < \infty. \tag{2.3}$$

In view of (2.1), (2.2), and (1.8) we have proved the following theorem.

THEOREM 2A. *Limits for variance. The estimates defined by* (1.7) *have variances satisfying*

$$\lim_{n\to\infty} nh \operatorname{Var}[f_n(x)] = f(x) \int_{-\infty}^{\infty} K^2(y)\, dy \tag{2.4}$$

at all points x of continuity of $f(x)$, *if the constants h satisfy* (1.8).

From Theorem 2A one can state conditions under which the estimate $f_n(x)$ is consistent in quadratic mean in the sense that

$$E\,|f_n(x) - f(x)|^2 \to 0 \quad \text{as} \quad n \to \infty. \tag{2.5}$$

The mean square error may be written

$$E\,|f_n(x) - f(x)|^2 = \sigma^2[f_n(x)] + b^2[f_n(x)] \tag{2.6}$$

in which $\sigma^2[f_n(x)] = \operatorname{Var}[f_n(x)]$ is the variance and

$$b[f_n(x)] = E[f_n(x)] - f(x) \tag{2.7}$$

is the bias of the estimate. Consequently if in addition to satisfying (1.8) the constants $h = h(n)$ are required to satisfy the condition

$$\lim_{n\to\infty} nh(n) = \infty, \tag{2.8}$$

it then follows that $f_n(x)$ is a consistent estimate of $f(x)$.

In the remainder of this paper we shall always be considering estimates of the form of (1.7) *in which* $K(y)$ *is a weighting function* [that is, an even function satisfying (1.11)-(1.13) and (1.17)] *and h is a sequence satisfying* (2.8).

Since the estimate $f_n(x)$ may be written as an average,

$$f_n(x) = n^{-1}\sum_{k=1}^{n} V_{nk}, \qquad V_{nk} = \{1/h(n)\}K\{(x - X_k)/h(n)\}, \tag{2.9}$$

of independent random variables identically distributed as a random variable $V_n = K(\{(x - X)/h(n)\})/h(n)$, it is easy to state conditions under which the sequence $f_n(x)$ is asymptotically normal, in the sense that, for every real number c,

$$\lim_{n\to\infty} P\left[\frac{f_n(x) - E[f_n(x)]}{\sigma[f_n(x)]} \leqq c\right] = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{c} e^{-(\frac{1}{2})y^2}\, dy = \Phi(c). \tag{2.10}$$

From Loève [3], p. 316, it follows that a necessary and sufficient condition for (2.10) to hold is that, for every $\epsilon > 0$,

$$n\, P\left[\left|\frac{V_n - E[V_n]}{\sigma[V_n]}\right| \geqq \epsilon n^{\frac{1}{2}}\right] \to 0, \quad \text{as} \quad n \to \infty. \tag{2.11}$$

A sufficient condition for (2.11) to hold is that, for some $\delta > 0$,

$$\frac{E|V_n - E[V_n]|^{2+\delta}}{n^{(\delta/2)}\sigma^{2+\delta}[V_n]} \to 0 \quad \text{as} \quad n \to \infty. \tag{2.12}$$

Now

$$(2.13)\quad E|V_n|^{2+\delta} = \int_{-\infty}^{\infty} \left|\frac{1}{h} K\left(\frac{x-y}{h}\right)\right|^{2+\delta} f(y)\, dy \sim \frac{1}{h^{1+\delta}} f(x) \int_{-\infty}^{\infty} |K(y)|^{2+\delta}\, dy,$$

while by (1.2),

$$(2.14)\quad \sigma^2[V_n] \sim (1/h) f(x) \int_{-\infty}^{\infty} K^2(y)\, dy.$$

Now the quantity in (2.12) can be written

$$(2.15)\quad \frac{h^{1+\delta} E|V_n - E[V_n]|^{2+\delta}}{(nh)^{(\delta/2)} h^{1+(\delta/2)} \sigma^{2+\delta}[V_n]},$$

which [in view of (2.13), (2.14), and (2.8)] tends to 0 as n tends to ∞ since, for every $\delta > 0$,

$$(2.16)\quad \int_{-\infty}^{\infty} |K(y)|^{2+\delta}\, dy < \infty.$$

We have thus shown that the sequence of estimates $\{f_n(x)\}$ are asymptotically normal, as well as consistent.

Some idea of the closeness of the normal approximation can be obtained from the Berry-Esseen bound (see Loève [3], p. 288): for a suitable constant C,

$$\sup_{-\infty<a<\infty} \left| P\left[\frac{f_n(x) - E[f_n(x)]}{\sigma[f_n(x)]} \leqq a\right] - \Phi(a)\right| \leqq C \frac{E|V_n|^3}{n^{\frac{1}{2}}\sigma^3[V_n]}$$

$$\sim \frac{1}{\{nhf(x)\}^{\frac{1}{2}}} \left[\left\{\int_{-\infty}^{\infty} |K(y)|^3\, dy\right\} \Big/ \left\{\int_{-\infty}^{\infty} K^2(y)\, dy\right\}^{\frac{3}{2}}\right].$$

3. Uniform consistency and estimation of the mode. In this section we determine conditions under which the estimated probability density function $f_n(x)$ tends uniformly (in probability) to the true probability density function, in the sense that (3.7) holds. Using this fact, we are able to obtain consistent estimates of the mode.

It is convenient to introduce the Fourier transform

$$(3.1)\quad k(u) = \int_{-\infty}^{\infty} e^{-iuy} K(y)\, dy$$

of the weighting function $K(y)$. We *assume* hereafter that $k(u)$ is absolutely integrable; note that this assumption holds for all the functions in Table 1 except the first.

We may then express estimates of the form of (1.7) as weighted averages over the sample characteristic function

$$(3.2)\quad \varphi_n(u) = \int_{-\infty}^{\infty} e^{iux}\, dF_n(x) = n^{-1} \sum_{k=1}^{n} e^{iuX_k}.$$

It is easily verified that one may write (since $k(u)$ is even)

$$(3.3) \quad f_n(x) = (nh)^{-1} \sum_{k=1}^{n} K\{(x - X_k)/h\} = (2\pi)^{-1} \int_{-\infty}^{\infty} e^{-iux} k(hu)\varphi_n(u)\, du.$$

It is clear that $f_n(x)$ is continuous and tends to 0 as x tends to $\pm\infty$. Consequently, there is a random variable θ_n such that

$$(3.4) \qquad f_n(\theta_n) = \max_{-\infty<x<\infty} f_n(x).$$

We call θ_n the *sample* mode.

We next assume that the true probability density function $f(x)$ is uniformly continuous in x (this is the case if it has an absolutely integrable characteristic function). It follows that $f(x)$ possesses a mode θ defined by

$$(3.5) \qquad f(\theta) = \max_{-\infty<x<\infty} f(x).$$

We *assume that θ is unique.*

THEOREM 3A. *Consistency of the sample mode as an estimate of the mode. If h is a function of n satisfying*

$$(3.6) \qquad \lim_{n\to\infty} nh^2 = \infty,$$

and if the probability density $f(x)$ is uniformly continuous, then for every $\epsilon > 0$

$$(3.7) \qquad P[\sup_{-\infty<x<\infty} f_n(x) - f(x)| < \epsilon] \to 1, \qquad \text{as } n \to \infty.$$

If $\{\theta_n\}$ are the sample modes, and if the population mode θ is unique, then for every $\epsilon > 0$

$$(3.8) \qquad P[|\theta_n - \theta| < \epsilon] \to 1, \qquad \text{as } n \to \infty.$$

PROOF. To show (3.7), it suffices to show that

$$(3.9) \qquad \lim_{n\to\infty} E^{\frac{1}{2}}[\sup_{-\infty<x<\infty} |f_n(x) - f(x)|^2] = 0.$$

To prove (3.9), it suffices to show that

$$(3.10) \qquad E^{\frac{1}{2}}[\sup_{-\infty<x<\infty} |f_n(x) - E[f_n(x)]|^2] \to 0,$$

as $n \to \infty$, since by Theorem 1A [modified to take account of the uniform continuity of $f(x)$] it follows that

$$(3.11) \qquad \lim_{n\to\infty} \sup_{-\infty<x<\infty} |E[f_n(x)] - f(x)| = 0.$$

Now

$$(3.12) \quad \sup_{-\infty<x<\infty} |f_n(x) - E[f_n(x)]| \leqq (2\pi)^{-1} \int_{-\infty}^{\infty} |k(hu)||\varphi_n(u) - E[\varphi_n(u)]|\, du.$$

Therefore, by Minkowski's inequality, the quantity in (3.10) is no greater than

$$(3.13) \qquad (2\pi)^{-1} \int_{-\infty}^{\infty} |k(hu)|\sigma[\varphi_n(u)]\, du \leqq (n^{\frac{1}{2}}h)^{-1} \int_{-\infty}^{\infty} |k(u)|\, du$$

which tends to 0. The proof of (3.9) is complete.

To prove (3.8), we first show that because $f(x)$ is a uniformly continuous probability density function with a unique mode θ, it has the following property: for every $\epsilon > 0$ there exists an $\eta > 0$ such that, for every point x, $|\theta - x| \geqq \epsilon$ implies $|f(\theta) - f(x)| \geqq \eta$.

If the assertion were false, then there would exist an $\epsilon > 0$ and a sequence $\{x_n\}$ such that

$$|f(\theta) - f(x_n)| < n^{-1} \quad \text{and} \quad |\theta - x_n| \geqq \epsilon. \tag{3.14}$$

Now (3.14), and the fact that $f(x) \to 0$ as $x \to \pm\infty$, implies that there exists a point $\theta' \neq \theta$ such that $f(\theta') = f(\theta)$, which contradicts the assumption that $f(x)$ has a unique mode θ.

From this assertion it follows that to prove $\theta_n \to \theta$ in probability, it suffices to prove that

$$f(\theta_n) \to f(\theta) \quad \text{in probability as} \quad n \to \infty. \tag{3.15}$$

Now

$$\begin{aligned} |f(\theta_n) - f(\theta)| &\leqq |f(\theta_n) - f_n(\theta_n)| + |f_n(\theta_n) - f(\theta)| \\ &\leqq 2 \sup_x |f_n(x) - f(x)|, \end{aligned} \tag{3.16}$$

since

$$|f_n(\theta_n) - f(\theta)| = |\sup_x f_n(x) - \sup_x f(x)| \leqq \sup_x |f_n(x) - f(x)|.$$

From (3.16) and (3.7), one obtains (3.15).

4. Limits for bias and mean square error. The properties of estimates $f_n(x)$ of the probability density function of the form of (3.3) depend on the constant h and the weighting function $K(y)$. In order to gain further insight into this dependence, in this section we see how the bias and mean square error of estimates of the form of (3.3) depend on h and $k(u)$.

Evaluation of bias. From (3.3) it follows that

$$\begin{aligned} E[f_n(x)] &= \int_{-\infty}^{\infty} h^{-1} K\{(x - y)/h\} f(y)\, dy \\ &= (2\pi)^{-1} \int_{-\infty}^{\infty} e^{-iux} k(hu) \varphi(u)\, du. \end{aligned} \tag{4.1}$$

Consequently

$$b[f_n(x)] = (2\pi)^{-1} \int_{-\infty}^{\infty} e^{-iux} \{k(hu) - 1\} \varphi(u)\, du. \tag{4.2}$$

Now let r be a positive number such that

$$k_r = \lim_{u \to 0} \{[1 - k(u)]/|u|^r\} \tag{4.3}$$

is finite. If there exists a value of r such that k_r is non-zero, it is called the characteristic exponent of the transform $k(u)$, and k_r is called the characteristic coefficient. If (4.3) holds then, as $h \to 0$,

$$(4.4) \qquad \frac{b[f_n(x)]}{h^r} = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-iux}\,\frac{k(hu)-1}{|hu|^r}\,|u|^r\varphi(u)\,du \to k_r f^{(r)}(x)$$

in which

$$(4.5) \qquad f^{(r)}(x) = -(2\pi)^{-1}\int_{-\infty}^{\infty} e^{-iux}|u|^r\varphi(u)\,du$$

where it is assumed that the integral in (4.5) converges absolutely.

To gain more insight into (4.4) let us examine in more detail the important case of $r = 2$. Then

$$(4.6) \qquad f^{(2)}(x) = -(2\pi)^{-1}\int_{-\infty}^{\infty} e^{-iux}u^2\varphi(u)\,du$$

is the second derivative of $f(x)$. The transform $k(u)$ has characteristic exponent $r = 2$ if the following conditions are satisfied:

$$(4.7) \qquad \int_{-\infty}^{\infty} yK(y)\,dy = 0, \qquad \int_{-\infty}^{\infty} y^2|K(y)|\,dy < \infty.$$

Then, as $u \to 0$,

$$[1 - k(u)]/u^2 = \int_{-\infty}^{\infty} \{[1 - e^{-iuy}]/u^2\}\,K(y)\,dy$$

$$(4.8) \qquad = (-1/2)\int_{-\infty}^{\infty} y^2K(y)\,dy - \int_{-\infty}^{\infty} \frac{e^{-iuy} - 1 + iuy - \frac{1}{2}(u^2y^2)}{u^2}\,K(y)\,dy$$

$$\to (-1/2)\int_{-\infty}^{\infty} y^2K(y)\,dy = k_2\,.$$

Thus if the integrals in (4.6) and (4.7) all converge absolutely the bias of the estimate $f_n(x)$ satisfies

$$(4.10) \qquad \{b[f_n\,(x)]/h^2\} \to -\tfrac{1}{2}f''(x)\int_{-\infty}^{\infty} y^2K(y)\,dy.$$

One may also obtain (4.10) directly since

$$(4.11) \qquad b[f_n(x)]/h^2 = \int_{-\infty}^{\infty} K(w)\{[f(x - wh) - f(x)]/h^2\}\,dw.$$

Evaluation of mean square error: We may now write an approximate expression for the mean square error of the estimate $f_n(x)$ assuming that the transform $k(u)$ of the function $K(y)$ used to form $f_n(x)$ has characteristic exponent r and characteristic coefficient k_r :

$$(4.12) \qquad E\,|f_n\,(x) - f(x)|^2 \sim [f(x)/nh]\int_{-\infty}^{\infty} K^2(y)\,dy + h^{2r}|k_r f^{(r)}(x)|^2$$

Let us find the value of h which minimizes the mean square error for a fixed value of n. The following lemma is easily verified.

LEMMA 4a. *Let A, B, α, and β be given positive numbers. Then*

$$\begin{aligned} \min\nolimits_{x>0} Ax^{\alpha} + Bx^{-\beta} &= A(1 + \alpha/\beta)(\beta B/\alpha A)^{\alpha/(\alpha+\beta)} \\ &= (\alpha + \beta)\{(A/\beta)^{\beta}(B/\alpha)^{\alpha}\}^{1/(\alpha+\beta)} \end{aligned} \tag{4.13}$$

and the value of x at which the minimum is achieved is

$$x_{\min} = (\beta B/\alpha A)^{1/(\alpha+\beta)}. \tag{4.14}$$

Consequently if one chooses h, as a function of n, by

$$h = \{f(x) \int_{-\infty}^{\infty} K^2(y)\, dy\}/\{n2r\, |k_r f^{(r)}(x)|^2\}^{1/(2r+1)}, \tag{4.15}$$

then the mean square error

$$\begin{aligned} E\, |f_n(x) - f(x)|^2 \sim (2r+1)\{[f(x)/n2r] \\ \cdot \int_{-\infty}^{\infty} K^2(y)\, dy\}^{2r/(1+2r)} |k_r f^{(r)}(x)|^{2/(1+2r)} \end{aligned} \tag{4.16}$$

tends to 0 as $n^{-2r/(1+2r)}$. In particular if $r = 2$, the estimates $f_n(x)$ have order of consistency $n^{\frac{4}{5}}$ in the sense that $n^{\frac{4}{5}}E\, |f_n(x) - f(x)|^2$ tends to a finite positive limit as n tends to ∞.

5. Asymptotic normality of the sample mode. In this section we state conditions on the constants $h(n)$ and the kernel $k(u)$ such that the estimated mode θ_n is asymptotically normal.

Consider a probability density function $f(x)$ with a unique mode at θ. If $f(x)$ has a continuous second derivative, then

$$f'(\theta) = 0, \qquad f''(\theta) < 0. \tag{5.1}$$

Similarly if the estimated probability density function $f_n(x)$ is chosen to be twice differentiable (that is, the weighting function $K(y)$ is chosen to be twice differentiable), then

$$f_n'(\theta_n) = 0, \qquad f_n''(\theta_n) < 0, \tag{5.2}$$

if θ_n is the mode of $f_n(x)$. By Taylor's theorem,

$$0 = f_n'(\theta_n) = f_n'(\theta) + (\theta_n - \theta) f_n''(\theta_n^*) \tag{5.3}$$

for some random variable θ_n^* between θ_n and θ. From (5.3) one may write

$$\theta_n - \theta = -f_n'(\theta)/f_n''(\theta_n^*) \tag{5.4}$$

if the denominator does not vanish. Using (5.4) as a basis, we now state conditions under which the estimated mode θ_n is asymptotically normal.

THEOREM 5A. *Asymptotic normality of the sample mode. Suppose that there exists*

δ, $0 < \delta < 1$, such that the transform $k(u)$ has characteristic exponent $r \geqq 2$ and satisfies

$$\int_{-\infty}^{\infty} u^{2+\delta}|k(u)|\, du < \infty. \tag{5.5}$$

that h is a function of n satisfying

$$\lim_{n\to\infty} nh^6 = \infty, \qquad \lim_{n\to\infty} nh^{5+2\delta} = 0, \tag{5.6}$$

and that the characteristic function $\varphi(u)$ satisfies

$$\int_{-\infty}^{\infty} u^{2+\delta}|\varphi(u)|\, du < \infty. \tag{5.7}$$

Then as $n \to \infty$,

$$E[\sup_{-\infty<x<\infty} |f_n''(x) - f''(x)|^2] \to 0 \tag{5.8}$$

$$f_n''(\theta_n^*) \to f''(\theta) \qquad \textit{in probability} \tag{5.9}$$

$$(nh^3)^{\frac{1}{2}} f_n'(\theta) \to N(0, f(\theta)J) \qquad \textit{in distribution} \tag{5.10}$$

$$(nh^3)^{\frac{1}{2}}(\theta_n - \theta) \to N(0, \{f(\theta)/[f''(\theta)]^2\}J) \qquad \textit{in distribution,} \tag{5.11}$$

where we define

$$J = \int_{-\infty}^{\infty} K'^2(y)\, dy = (2\pi)^{-1} \int_{-\infty}^{\infty} u^2 k^2(u)\, du. \tag{5.12}$$

Proof. That (5.8) holds may be inferred from the following facts.

$$|f_n''(x) - E[f_n''(x)]| \leqq (2\pi)^{-1} \int_{-\infty}^{\infty} |k(hu)| u^2 |\varphi_n(u) - E[\varphi_n(u)]|\, du,$$

$$E^{1/2}[\sup_{-\infty<x<\infty} |f_n''(x) - E[f_n''(x)]|^2] \leqq \int_{-\infty}^{\infty} |k(hu)| u^2 \sigma[\varphi_n(u)]\, du$$

$$\leqq (n^{\frac{1}{2}} h^3)^{-1} \int_{-\infty}^{\infty} |k(v)| v^2\, dv,$$

$$|E[f_n''(x)] - f''(x)| \leqq (2\pi)^{-1} \int_{-\infty}^{\infty} |1 - k(hu)| u^2 |\varphi(u)|\, du.$$

That (5.9) holds follows from (5.8) and the fact that θ_n^* tends to θ, since it is between θ_n and θ, and θ_n tends to θ.

That (5.10) holds may be inferred from the following facts:

$$f_n'(\theta) = n^{-1} \sum_{k=1}^{n} V_{nk}, \qquad V_{nk} \equiv \frac{1}{h^2} K'\{(\theta - X_k)/h\},$$

V_{nk} independent and identically distributed as $V_n = (h^2)^{-1} K'\{(\theta - X)/h\}$,

$$h^{2m-1}E|V_n|^m \to f(\theta) \int_{-\infty}^{\infty} |K'(y)|^m \, dy,$$

$$\{E|V_n - E[V_n]|^{2+\delta}/n^{(\delta/2)}\sigma^{2+\delta}[V_n]\} \to 0,$$

$$\frac{f'_n(\theta) - E[f'_n(\theta)]}{\sigma[f'_n(\theta)]} \to N(0,1) \qquad \text{in distribution,}$$

$$(nh^3)^{\frac{1}{2}}E[f'_n(\theta)] = (nh^3)^{\frac{1}{2}}(-i/2\pi) \int_{-\infty}^{\infty} e^{-iu\theta}\{k(hu) - 1\}u\varphi(u) \, du \to 0,$$

$$nh^3\operatorname{Var}[f'_n(\theta)] = h^{-1} \int_{-\infty}^{\infty} K'^2\{(\theta - y)/h\}f(y) \, dy - nh^3E^2[f'_n(\theta)]$$

$$\to f(\theta) \int_{-\infty}^{\infty} K'^2(y) \, dy,$$

$$(nh^3)^{\frac{1}{2}}f'_n(\theta) \to N(0, f(\theta)J).$$

Finally (5.11) follows by standard large sample theory from (5.8)-(5.10).

REFERENCES

[1] Bochner, S. (1955). *Harmonic Analysis and the Theory of Probability*. Univ. of California Press.

[2] Le Cam, L. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes' estimates. **1** 277–330. *Univ. of California Publ. in Statist.* Univ. of California Press.

[3] Loève, M. (1960). *Probability Theory*, 2nd Ed. Van Nostrand, Princeton.

[4] Parzen, E. (1961). Mathematical considerations in the estimation of spectra. *Technometrics* **3** 167–190.

[5] Rosenblatt, M. (1956). Remarks on some non-parametric estimates of a density function. *Annals Math. Statist.* **27** 832–837.

[6] Whittle, P. (1958). On the smoothing of probability density functions. *J. Roy. Statist. Soc., Ser. B* **20** 334–343.