ON ESTIMATION OF A PROBABILITY DENSITY FUNCTION AND MODE

BY
EMANUEL PARZEN

TECHNICAL REPORT NO. 40
AUGUST 21, 1961

PREPARED UNDER CONTRACT Nonr-225(21)
(NR-042-993)
FOR
OFFICE OF NAVAL RESEARCH

APPLIED MATHEMATICS AND STATISTICS LABORATORIES
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
ON ESTIMATION OF A PROBABILITY DENSITY FUNCTION AND MODE

0. Introduction

Given a sequence of independent identically distributed random variables $X_1, X_2, \ldots, X_n, \ldots$ 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 results in this paper were obtained over a period of years. Since I have not had time in more than a year to return to this problem, I am writing up the incomplete results I have obtained so far in the hope they may be of interest nevertheless.
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 \( X_1, X_2, \ldots, X_n \) be independent random variables identically distributed as a random variable \( X \) whose distribution function \( F(x) = P[X \leq x] \) satisfies the relation

\[
(1.1) \quad F(x) = \int_{-\infty}^{x} f(x') \, dx'
\]

for some 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

\[
(1.2) \quad F_n(x) = \frac{1}{n} \{ \text{no. of observations } \leq x \text{ among } X_1, \ldots, X_n \}
\]

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

\[
(1.3) \quad E[F_n(x)] = F(x)
\]

\[
(1.4) \quad \text{Var}[F_n(x)] = \frac{1}{n} F(x) (1 - F(x))
\]

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

\[
(1.5) \quad f_n(x) = \frac{F_n(x+h) - F_n(x-h)}{2h}
\]
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 \( \infty \). 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

\[
(1.6) \quad K(y) = \begin{cases} 
\frac{1}{2}, & |y| \leq 1 \\
0, & |y| > 1
\end{cases}
\]

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

\[
(1.7) \quad f_n(x) = \int_{-\infty}^{\infty} \frac{1}{nh} K\left(\frac{x-y}{h}\right) d\bar{F}_n(y) = \frac{1}{nh} \sum_{j=1}^{n} K\left(\frac{x-X_j}{h}\right)
\]

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 choosen.
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

\[
\lim_{n \to \infty} h(n) = 0
\]

then

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

Now

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

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

\[
\sup_{-\infty < y < \infty} |K(y)| < \infty
\]

(1.11)

\[
\int_{-\infty}^{\infty} |K(y)| \, dy < \infty
\]

(1.12)

\[
\lim_{y \to \infty} |yK(y)| = 0
\]

(1.13)

Let \( g(y) \) satisfy

\[
\int_{-\infty}^{\infty} |g(y)| \, dy < \infty
\]

(1.14)
Let \((h(n))\) be a sequence of positive constants satisfying (1.8).

Define

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

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

\[
(1.16) \quad \lim_{n \to \infty} g_n(x) = g(x) \int_{-\infty}^{\infty} K(y) \, dy
\]

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, \(|y| \leq \delta\) and \(|y| > \delta\). Then

\[
|g_n(x) - g(x) \int_{-\infty}^{\infty} K(y) \, dy| \leq \max_{|y| \leq \delta} |g(x-y) - g(x)| \int_{-\infty}^{\infty} |K(z)| \, dz
\]

\[
+ \int_{|y| \geq \delta} \frac{|g(x-y)|}{y} \frac{1}{h(n)} K\left(\frac{y}{h(n)}\right) \, dy
\]

\[
+ |g(x)| \int_{|y| \geq \delta} \frac{1}{h(n)} K\left(\frac{y}{h(n)}\right) \, dy
\]

\[
\leq \max_{|y| \leq \delta} |g(x-y) - g(x)| \int_{-\infty}^{\infty} |K(z)| \, dz
\]

- 6 -
\[ + \frac{1}{\delta} \sup_{|z| \geq \delta/h(n)} |z K(z)| \int_{-\infty}^{\infty} |g(y)| \, dy \]

\[ + |g(x)| \int_{|z| \geq \delta/h(n)} |K(z)| \, dz \]

which tends to 0 as one lets \( n \) tend to \( \infty \), and then lets \( \delta \) 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 \( h \) satisfy (1.8) and if the function \( K(y) \) satisfies (1.11) – (1.13) and in addition satisfies

\[ (1.17) \quad \int_{-\infty}^{\infty} K(y) \, dy = 1 \]

For ease of exposition, a 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 I.
\begin{table}
\centering
\begin{tabular}{|c|c|c|}
\hline
$K(y)$ & $k(u) = \int_{-\infty}^{\infty} e^{iuy}K(y)dy$ & $\int_{-\infty}^{\infty} K^2(y)dy = \frac{1}{2\pi} \int_{-\infty}^{\infty} k^2(u)du$
\hline
$\frac{1}{2}, \ |y| \leq 1$ & $\frac{\sin u}{u}$ & $\frac{1}{2}$
\hline
$0, \ |y| \geq 1$ & & 
\hline
$1 - |y|, \ |y| \leq 1$ & $\frac{(\sin(u/2))^2}{u^2}$ & $\frac{2}{3}$
\hline
$0, \ |y| \geq 1$ & & 
\hline
$(4/3) - 8y^2 + 8|y|^3, \ |y| \leq \frac{1}{2}$ & $\frac{\sin(u/4)^4}{u^4}$ & $0.96$
\hline
$\frac{8}{3}(1-|y|)^3, \ \frac{1}{2} \leq |y| \leq 1$ & & 
\hline
$0, \ |y| > 1$ & & 
\hline
$\frac{1}{\sqrt{2\pi}} e^{-(1/2)y^2}$ & $e^{-(1/2)u^2}$ & $\frac{1}{2\sqrt{\pi}}$
\hline
$\frac{1}{2} e^{-|y|}$ & $\frac{1}{1+u^2}$ & $\frac{1}{2}$
\hline
$\frac{1}{\pi} \frac{1}{1+y^2}$ & $e^{-|u|}$ & $\frac{1}{\pi}$
\hline
$\frac{1}{2\pi} \left(\frac{\sin(y/2)}{y/2}\right)^2$ & $1-|u|, \ |u| \leq 1$ & $\frac{1}{3\pi}$
\hline
$0, \ |u| \geq 1$ & & 
\hline
\end{tabular}
\caption{}
\end{table}
2. **Consistency and asymptotic normality.**

The variance of the estimate \( f_n(x) \) is given by

\[
(2.1) \quad \text{Var}[f_n(x)] = \frac{1}{n} \text{Var} \left[ \frac{1}{h} K \left( \frac{x-X}{h} \right) \right]
\]

Now by Theorem 1A

\[
(2.2) \quad h \mathbb{E} \left[ \left( \frac{1}{h} K \left( \frac{x-X}{h} \right) \right)^2 \right] = \frac{1}{h} \int_{-\infty}^{\infty} k^2 \left( \frac{x-y}{h} \right) f(y) \, dy \to f(x) \int_{-\infty}^{\infty} k^2(y) \, dy
\]

since (1.11) and (1.12) imply that

\[
(2.3) \quad \int_{-\infty}^{\infty} k^2(y) \, dy < \infty
\]

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

\[
(2.4) \quad \lim_{n \to \infty} nh \text{Var}[f_n(x)] = f(x) \int_{-\infty}^{\infty} k^2(y) \, dy
\]

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

\[
(2.5) \quad \mathbb{E} |f_n(x) - f(x)|^2 \to 0 \quad \text{as} \quad n \to \infty
\]
The mean square error may be written

\begin{equation}
E[f_n(x) - f(x)]^2 = \sigma^2 [f_n(x)] + b[f_n(x)]
\end{equation}

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

\begin{equation}
b[f_n(x)] = E[f_n(x)] - f(x)
\end{equation}

is the bias of the estimate. Consequently if in addition to satisfying

\begin{equation}
(1.8) \quad h = h(n)
\end{equation}

the constants \( h = h(n) \) are required to satisfy the condition

\begin{equation}
(2.8) \quad \lim_{n \to \infty} nh(n) = \infty
\end{equation}

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 [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 a sum,

\begin{equation}
f_n(x) = \sum_{k=1}^{n} V_{nk}, \quad \text{where} \quad V_{nk} = \frac{1}{h(n)} K \left( \frac{x - X_k}{h(n)} \right)
\end{equation}
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 \( c \),

\[
\lim_{n \to \infty} P \left[ \frac{f_n(x) - E[f_n(x)]}{\sigma[f_n(x)]} \leq c \right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{c} e^{-(1/2)y^2} dy
\]

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

\[
(2.11) \quad n P \left[ \frac{|V_n - E[V_n]|}{\sigma[V_n]} \geq \epsilon \sqrt{n} \right] \to 0 \text{ as } n \to \infty.
\]

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

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

Now

\[
(2.13) \quad E|V_n|^{2+\delta} = \int_{-\infty}^{\infty} |\frac{1}{h} K(\frac{x-x_i}{h})|^{2+\delta} f(y) dy
\]

\[
\sim \frac{1}{h^{1+\delta}} \int_{-\infty}^{\infty} f(x) |K(y)|^{2+\delta} dy
\]
while by (1.2)

\[(2.14)\quad \sigma^2[n] \sim \frac{1}{n} f(x) \int_{-\infty}^{\infty} k^2(y) \, dy \quad .\]

Now the quantity in (2.12) can be written

\[(2.15)\quad \frac{h^{1+\delta} \operatorname{E}[V_n - \operatorname{E}[V_n]]^{2+\delta}}{(nh)^{(5/2)}} \frac{1}{n^{1+\delta/2}} \frac{\sigma^{2+\delta}[V_n]}{\sigma^{2+\delta}[V_n]} \quad .\]

which [in view of (2.13), (2.14), and (2.8), and the \(c_r\)-inequality of Loeve [3, p. 155] tends to 0 as \(n\) tends to \(\infty\) if it is assumed that for some \(\delta > 0\)

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

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

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

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 I except the first.

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

\begin{equation}
(3.2) \quad \varphi_n(u) = \int_{-\infty}^{\infty} e^{iux} \, d\varphi_n(x) = \frac{1}{n} \sum_{k=1}^{n} e^{iux_k}.
\end{equation}

It is easily verified that one may write

\begin{equation}
(3.3) \quad f_n(x) = \frac{1}{nh} \sum_{k=1}^{n} K\left(\frac{x-x_k}{h}\right) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iux} k(hu) \varphi_n(u) \, du.
\end{equation}

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 $\theta_n$ such that

- 13 -
(3.4) \[ f_n(\theta_n) = \max_{-\infty < x < \infty} f_n(x) \, . \]

We call \( \theta_n \) 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 \( \theta \) defined by

(3.5) \[ f(\theta) = \max_{-\infty < x < \infty} f(x) \, . \]

We assume that \( \theta \) 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) \[ \lim_{n \to \infty} nh^2 = \infty \]

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

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

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

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

Proof: To show (3.6), it suffices to show that
\[ (3.9) \quad \lim_{n \to \infty} \frac{1}{2} \mathbb{E} \left[ \sup_{-\infty < x < \infty} |f_n(x) - f(x)|^2 \right] = 0. \]

To prove (3.9), it suffices to show that
\[ (3.10) \quad \mathbb{E} \left[ \sup_{-\infty < x < \infty} |f_n(x) - \mathbb{E}[f_n(x)]|^2 \right] \to 0, \]
as \( n \to \infty \), since by Theorem 1A [modified to take account of the uniform continuity of \( f(x) \)] it follows that
\[ (3.11) \quad \lim_{n \to \infty} \sup_{-\infty < x < \infty} |\mathbb{E}[f_n(x)] - f(x)| = 0. \]

Now
\[ (3.12) \quad \sup_{-\infty < x < \infty} |f_n(x) - \mathbb{E}[f_n(x)]| \leq \frac{1}{2\pi} \int_{-\infty}^{\infty} |k(hu)| \left| \varphi_n(u) - \mathbb{E}[\varphi_n(u)] \right| du. \]

Therefore, by Minkowski's inequality, the quantity in (3.10) is no greater than
\[ (3.13) \quad \frac{1}{2\pi} \int_{-\infty}^{\infty} |k(hu)| \sigma[\varphi_n(u)] du \leq \frac{1}{\sqrt{\nu h}} \int_{-\infty}^{\infty} |k(u)| du \]
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 \( \theta \), it has the following property: for every \( \epsilon > 0 \) there exists an \( \eta > 0 \) such that, for every point \( x \), \( |\theta - x| \geq \epsilon \) implies \( |f(\theta) - f(x)| \geq \eta \).

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

\[
(3.14) \quad |f(\theta) - f(x_n)| < \frac{1}{n} \quad \text{and} \quad |\theta - x_n| \geq \epsilon .
\]

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 \( \theta \).

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

\[
(3.15) \quad f(\theta_n) \to f(\theta) \quad \text{in probability as} \quad n \to \infty
\]

Now

\[
|f_n(\theta_n) - f(\theta)| \leq |f(\theta_n) - f_n(\theta_n)| + |f_n(\theta_n) - f(\theta)|
\]

\[
(3.16) \quad \leq 2 \sup_{x} |f_n(x) - f(x)|
\]

since

\[
|f_n(\theta_n) - f(\theta)| = |\sup_{x} f_n(x) - \sup_{x} f(x)| \leq \sup_{x} |f_n(x) - f(x)| .
\]

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

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

$$
= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iux} k(u) \varphi(u) \, du .
$$

Consequently

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

Now let $r$ be a positive number such that

$$
k_r = \lim_{u \to 0} \frac{1 - k(u)}{|u|^r}
$$

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 $h \to 0$,
\[
\frac{b[r_n(x)]}{h^r} = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iux} \frac{k(hu)-1}{|hu|^r} \varphi(u) \, du = k_r \, f(r)(x)
\]

in which

\[
f^{(r)}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iux} |\varphi(u)|^r \varphi(u) \, du
\]

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

\[
f^{(2)}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iux} u^2 \varphi(u) \, du
\]

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

\[
\int_{-\infty}^{\infty} y K(y) \, dy = 0, \, \int_{-\infty}^{\infty} y^2 |K(y)| \, dy < \infty.
\]

Then, as \( u \to 0 \),

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

\[
= (-1/2) \int_{-\infty}^{\infty} y^2 K(y) \, dy - \frac{1}{2} \int_{-\infty}^{\infty} y^2 (e^{-iu\theta y}-1) K(y) \, dy
\]

\[
\to (-1/2) \int_{-\infty}^{\infty} y^2 K(y) \, dy = k_2.
\]
Thus if the integrals in (4.6) and (4.7) all converge absolutely the bias of the estimate \( f_n(x) \) satisfies

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

One may also obtain (4.10) directly since

\[
(4.11) \quad \frac{b[f_n(x)]}{h^2} = \int_{-\infty}^{\infty} K(w) \left( \frac{f(x-wh) - f(x)}{h^2} \right) \, dw.
\]

**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 \( k_r \):

\[
(4.12) \quad E[f_n(x) - f(x)]^2 = \frac{f(x)}{nh} \int_{-\infty}^{\infty} k^2(y) \, dy + h^{2r} |k_r f^{(r)}(x)|^2.
\]

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 ha:** Let \( A, B, \alpha, \) and \( \beta \) be given positive numbers. Then

\[
(4.13) \quad \min_{x > 0} \frac{A \frac{\alpha}{\beta} + B x^{-\beta}}{A^{1+\alpha/\beta}} = A(1 + \alpha/\beta) \left( \frac{B}{\alpha A} \right)^{\alpha/(\alpha+\beta)}
\]

\[
= (\alpha+\beta) \left\{ \left( \frac{A}{\beta} \right)^{\beta/\alpha} \left( \frac{B}{A} \right)^{1/(\alpha+\beta)} \right\}
\]

- 19 -
and the value of $x$ at which the minimum is achieved is

$$x_{\text{min}} = \left( \frac{\beta B}{\alpha A} \right)^{1/(\alpha + \beta)}.$$

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

$$h = \left( \frac{f(x) \int \infty_{-\infty} k^2(y) \, dy}{n2r |k_r f(r)(x)|^2} \right)^{1/(2r+1)}$$

then the mean square error

$$E |f_n(x) - f(x)|^2 = (2r+1) \left( \frac{f(x)}{n2r} \int \infty_{-\infty} k^2(y) \, dy \right)^{2r/(1+2r)} |k_r f(r)(x)|^{2/(1+2r)}$$

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

tends to a finite positive limit as $n$ tends to $\infty$. 

- 20 -
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 \( \theta_n \) is asymptotically normal.

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

\[ f'(\theta) = 0, \quad f''(\theta) < 0 \quad (5.1) \]

Similarly if the estimated probability density function \( f_n(x) \) is chosen to be twice differentiable, then

\[ f_n'(\theta_n) = 0, \quad f_n''(\theta_n) < 0 \quad (5.2) \]

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

\[ 0 = f''(\theta_n) = f_n'(\theta) + (\theta_n - \theta) f_n''(\theta_\ast) \quad (5.3) \]

for some random variable \( \theta_\ast \) between \( \theta_n \) and \( \theta \).

From (5.3) one may write

\[ \theta_n - \theta = \frac{f_n'(\theta)}{f_n''(\theta_\ast)} \quad (5.4) \]

if the denominator does not vanish. Using (5.4) as a basis, we now state conditions under which the estimated mode \( \theta_n \) is asymptotically normal.
Theorem 5A  Asymptotic normality of the sample mode. Suppose that the transform $k(u)$ satisfies

\begin{equation}
\int_{-\infty}^{\infty} u^2 |k(u)| \, du < \infty
\end{equation}

and that $h$ is a function of $n$ satisfying

\begin{equation}
\lim_{n \to \infty} nh^6 = \infty
\end{equation}

Suppose further that the characteristic function $\varphi(u)$ satisfies

\begin{equation}
\int_{-\infty}^{\infty} u^2 |\varphi(u)| \, du < \infty
\end{equation}

Then as $n \to \infty$

\begin{equation}
\mathbb{E}\left[ \sup_{-\infty < x < \infty} |f_n''(x) - f''(x)|^2 \right] \to 0
\end{equation}

\begin{equation}
f_n''(\theta_n^*) \to f''(\theta) \quad \text{in probability}
\end{equation}

\begin{equation}
\sqrt{nh^3} f_n'(\theta) \to \mathcal{N}(0, f(\theta)J) \quad \text{in distribution}
\end{equation}

\begin{equation}
\sqrt{nh^3} (\theta_n - \theta) \to \mathcal{N}(0, \frac{f(\theta)}{[f''(\theta)]^2} J) \quad \text{in distribution}
\end{equation}

where we define

\begin{equation}
J = \int_{-\infty}^{\infty} K^2(y) \, dy = \frac{1}{2\pi} \int_{-\infty}^{\infty} u^2 k^2(u) \, du
\end{equation}
Proof: That (5.8) holds may be inferred from the following facts:

\[ |f''_n(x) - E[f''_n(x)]| \leq \frac{1}{2\pi} \int_{-\infty}^{\infty} |k(hu)| u^2 |\varphi_n(u) - E[\varphi_n(u)]| \, du \]

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

\[ \leq \frac{1}{\sqrt{nh^3}} \int_{-\infty}^{\infty} |k(v)|v^2 \, dv \]

\[ |E[f''_n(x)] - f(x)| \leq \frac{1}{2\pi} \int_{-\infty}^{\infty} |1-k(hu)||u^2 |\varphi(u)| |u| \, du \]

That (5.9) holds follows from (5.8) and the fact that \( \theta^*_n \) tends to \( \theta \), since it is between \( \theta_n \) and \( \theta \), and \( \theta_n \) tends to \( \theta \).

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

\[ f'_n(\theta) = \sum_{k=1}^{n} V_{nk}' \]

\[ V_{nk} = \frac{1}{h^2} K' \left( \frac{\theta - X_k}{h} \right) \]

\[ V_{nk} \text{ independent and identically distributed as } V_n = \frac{1}{h^2} K' \left( \frac{\theta - X}{h} \right) \]
\[ h^{2m-1} E|V_n|^m \rightarrow f(\theta) \int_{-\infty}^{\infty} |K'(y)|^m \, dy , \]

\[ \frac{E|V_n - E[V_n]|^{2+8}}{n^{(8/2)\sigma^{2+8}[V_n]}} \rightarrow 0 , \]

\[ \frac{f_n'(\theta) - E[f_n'(\theta)]}{\sigma[f_n'(\theta)]} \rightarrow N(0,1) \quad \text{in distribution} , \]

\[ E[f_n'(\theta)] \rightarrow f'(\theta) = 0 , \]

\[ nh^3 \text{var}[f_n'(\theta)] = \frac{1}{h^2} \int_{-\infty}^{\infty} K'(\frac{\theta - Y}{h}) f(y) \, dy \]

\[ \rightarrow f(\theta) \int_{-\infty}^{\infty} K'(y) \, dy , \]

\[ \sqrt{nh^3} f_n'(\theta) \rightarrow N(0, f(\theta) J) \]

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


Armed Services Technical Information Agency
Arlington Hall Station
Arlington 12, Virginia

Head, Logistics and Mathematical Statistics Branch
Code 436
Office of Naval Research
Washington 25, D. C.

Commanding Officer
Office of Naval Research Branch Office
346 Broadway
New York 13, N. Y.
Attn: J. Laderman

Commanding Officer
Office of Naval Research Branch Office
1030 East Green Street
Pasadena 1, California
Attn: Dr. A. R. Laufer

Commanding Officer
U. S. Army Signal Electronic Research
P. O. Box 205 - Major Adornetto
Mountain View, California

Commanding Officer
U. S. Navy Powder Factory
Attn: F. Freshman, R and D
Indianhead, Maryland

Director, Naval Research Lab.
Attn: Tech. Infor. Officer
Washington 25, D. C.

Director
National Security Agency
Attn: REMP-1
Fort George G. Meade, Maryland

Dept. of Mathematical Statistics
University of North Carolina
Chapel Hill, North Carolina

Department of Statistics
University of California
Berkeley 4, California

Department of Statistics
Michigan State College
East Lansing, Michigan

Library
Institute for Defense Analyses
Communications Research Division
Von Neumann Hall
Princeton, New Jersey

Office of the Ass't Naval Attaché for Research
American Embassy
Navy No. 100, Fleet P. O.
New York, N. Y.

Scientific Section
Office of Naval Research
1000 Geary Street
San Francisco 9, California

Statistical Engineering Lab.
National Bureau of Standards
Washington 25, D. C.

Statistical Laboratory
University of Washington
Seattle 5, Washington

Superintendent
U. S. Navy Postgraduate School
Attn: Library
Monterey, California

Technical Information Officer
Naval Research Laboratory
Washington 25, D. C.

U. S. Naval Avionics Facility
Attn: Library
Indianapolis 18, Indiana

Contract Nonr-225 (21)
August 1961
(165)
<table>
<thead>
<tr>
<th>Name</th>
<th>Address</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dr. Stephen G. Allen</td>
<td>Stanford Research Institute</td>
<td>Menlo Park, California</td>
</tr>
<tr>
<td>Professor T. W. Anderson</td>
<td>Dept. of Mathematical Statistics</td>
<td>New York 27, New York</td>
</tr>
<tr>
<td>Professor Fred C. Andrews</td>
<td>Department of Mathematics</td>
<td>Eugene, Oregon</td>
</tr>
<tr>
<td>Professor Robert Bechhofer</td>
<td>Dept. of Industrial and Eng. Admin.</td>
<td>Sibley School of Mechanical Eng.</td>
</tr>
<tr>
<td>Mr. J. Berkowitz</td>
<td>Quality Control Supervisor</td>
<td>Commack, L.I., New York</td>
</tr>
<tr>
<td>Dr. Julian H. Bigelow</td>
<td>School of Mathematics</td>
<td>Princeton, New Jersey</td>
</tr>
<tr>
<td>Professor Z. W. Birnbaum</td>
<td>Laboratory of Statistical Research</td>
<td>University of Washington</td>
</tr>
<tr>
<td>Professor David Blackwell</td>
<td>Department of Mathematical Sciences</td>
<td>University of California</td>
</tr>
<tr>
<td>Dr. Julius R. Blum</td>
<td>Sandia Corporation</td>
<td>Albuquerque, New Mexico</td>
</tr>
<tr>
<td>Dr. Paul Blunk</td>
<td>Box 532</td>
<td>Fair Oaks, California</td>
</tr>
<tr>
<td>Dr. Charles Boll</td>
<td>Aerospace Corp. F-2813</td>
<td>Los Angeles 45, California</td>
</tr>
<tr>
<td>Professor G. E. P. Box</td>
<td>Department of Statistics</td>
<td>Madison, Wisconsin</td>
</tr>
<tr>
<td>Professor Ralph A. Bradley</td>
<td>Department of Statistics</td>
<td>Tallahassee, Florida</td>
</tr>
<tr>
<td>Professor W. G. Cochran</td>
<td>Department of Statistics</td>
<td>Cambridge 38, Massachusetts</td>
</tr>
<tr>
<td>Professor Lee Cronbach</td>
<td>Bureau of Education Research</td>
<td>Champaign, Illinois</td>
</tr>
<tr>
<td>Professor Cyrus Derman</td>
<td>Dept. of Industrial Engineering</td>
<td>New York 27, New York</td>
</tr>
<tr>
<td>Dr. Joseph Daly</td>
<td>Bureau of the Census</td>
<td>Washington 25, D. C.</td>
</tr>
<tr>
<td>Dr. Francis Dresp</td>
<td>Stanford Research Institute</td>
<td>Menlo Park, California</td>
</tr>
<tr>
<td>Mr. John A. Dutton</td>
<td>Department of Meterology</td>
<td>Madison, Wisconsin</td>
</tr>
<tr>
<td>Professor Meyer Dwass</td>
<td>Department of Mathematics</td>
<td>Evanston, Illinois</td>
</tr>
</tbody>
</table>

Contract Nonr-225 (21)
August 1961
(165)
Professor W. T. Federer
Cornell University
Department of Plant Breeding
Biometrics Unit
Ithaca, New York

Dr. Merrill M. Flood
Mental Health Research Institute
205 North Forest Avenue
Ann Arbor, Michigan

Professor H. P. Goode
Dept. of Industrial and
Engineering Administration
Cornell University
Ithaca, New York

Professor E. J. Gumbel
Industrial Engr. Department
409 Engineering Building
Columbia University
New York 27, New York

Professor Donald Guthris, Jr.
Stanford Research Institute
333 Ravenswood Avenue
Menlo Park, California

Professor Bernard Harris
Department of Mathematics
The University of Nebraska
Lincoln 8, Nebraska

Dr. Theodore E. Harris
The RAND Corporation
1700 Main Street
Santa Monica, California

Professor W. Hirsch
Institute of Mathematical Sciences
New York University
New York 3, New York

Professor Harold Hotelling
Associate Director
Institute of Statistics
University of North Carolina
Chapel, Hill, North Carolina

Professor Stanley Isaacson
4715 Pleasant Street
Des Moines, Iowa

Professor Oscar Kempthorne
Statistics Laboratory
Iowa State College
Ames, Iowa

Professor Solomon Kullback
Department of Statistics
George Washington University
Washington 7, D. C.

Dr. Eugene Lukacs
Department of Mathematics
Catholic University
Washington 15, D. C.

Professor G. W. McElrath
Department of Mechanical Engineering
University of Minnesota
Minneapolis 14, Minnesota

Professor Paul Meier
Dept. of Statistics
University of Chicago
Chicago, Illinois

Professor James M. Moore
Dept. of Industrial Engineering
Northeastern University
Boston 15, Massachusetts

D. E. Newnham
Chief, Industrial Eng. Division
Comptroller
HQAM: San Bernardino Air Materiel
Area
Norton Air Force Base, California

Professor J. Neyman
Department of Statistics
University of California
Berkeley 4, California

Mr. Monroe Norden
Engineering Statistics Group
College of Engineering
New York University
New York 5, New York

Contract Nonr-225 (21)
August 1961
(165)
<table>
<thead>
<tr>
<th>Name</th>
<th>Institution</th>
<th>Address</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mr. W. J. Peterson</td>
<td>Dept. 65-24 Lockheed Missile Systems Div.</td>
<td>Sunnyvale, California</td>
</tr>
<tr>
<td>Dr. Richard Post</td>
<td>Department of Mathematics</td>
<td>San Jose State College</td>
</tr>
<tr>
<td></td>
<td>San Jose, California</td>
<td></td>
</tr>
<tr>
<td>Professor Ronald Pyke</td>
<td>Department of Mathematics</td>
<td>University of Washington</td>
</tr>
<tr>
<td></td>
<td>Seattle 5, Washington</td>
<td></td>
</tr>
<tr>
<td>Professor Stanley Reiter</td>
<td>Department of Economics</td>
<td>Purdue University</td>
</tr>
<tr>
<td></td>
<td>Lafayette, Indiana</td>
<td></td>
</tr>
<tr>
<td>Professor George J. Resnikoff</td>
<td>Department of Industrial Engineering</td>
<td>Illinois Institute of Technology</td>
</tr>
<tr>
<td></td>
<td>Chicago 16, Illinois</td>
<td></td>
</tr>
<tr>
<td>Professor Herbert Robbins</td>
<td>Mathematical Statistics Dept.</td>
<td>Columbia University</td>
</tr>
<tr>
<td></td>
<td>New York 27, New York</td>
<td></td>
</tr>
<tr>
<td>Professor Murray Rosenblatt</td>
<td>Department of Mathematics</td>
<td>Brown University</td>
</tr>
<tr>
<td></td>
<td>Providence 12, Rhode Island</td>
<td></td>
</tr>
<tr>
<td>Professor Herman Rubin</td>
<td>Department of Statistics</td>
<td>Michigan State University</td>
</tr>
<tr>
<td></td>
<td>East Lansing, Michigan</td>
<td></td>
</tr>
<tr>
<td>Professor I. R. Savage</td>
<td>Department of Statistics</td>
<td>University of Minnesota</td>
</tr>
<tr>
<td></td>
<td>Minneapolis, Minnesota</td>
<td></td>
</tr>
<tr>
<td>Professor Seymour Sherman</td>
<td>Department of Mathematics</td>
<td>Wayne State University</td>
</tr>
<tr>
<td></td>
<td>Detroit 2, Michigan</td>
<td></td>
</tr>
<tr>
<td>Professor W. L. Smith</td>
<td>Statistics Department</td>
<td>University of North Carolina</td>
</tr>
<tr>
<td></td>
<td>Chapel Hill, North Carolina</td>
<td></td>
</tr>
<tr>
<td>Dr. Milton Sobel</td>
<td>Statistics Department</td>
<td>University of Minnesota</td>
</tr>
<tr>
<td></td>
<td>Minneapolis, Minnesota</td>
<td></td>
</tr>
<tr>
<td>Professor Frank Spitzer</td>
<td>Department of Mathematics</td>
<td>University of Minnesota</td>
</tr>
<tr>
<td></td>
<td>Minneapolis, Minnesota</td>
<td></td>
</tr>
<tr>
<td>Professor Gerald L. Thompson</td>
<td>Department of Mathematics</td>
<td>Carnegie Institute of Technology</td>
</tr>
<tr>
<td></td>
<td>Pittsburgh 13, Pennsylvania</td>
<td></td>
</tr>
<tr>
<td>Professor Donald Truax</td>
<td>Department of Mathematics</td>
<td>University of Oregon</td>
</tr>
<tr>
<td></td>
<td>Eugene, Oregon</td>
<td></td>
</tr>
<tr>
<td>Dr. Geoffrey Watson</td>
<td>The Research Triangle Institute</td>
<td>P. O. Box 490</td>
</tr>
<tr>
<td></td>
<td>Durham, North Carolina</td>
<td></td>
</tr>
<tr>
<td>Mr. Harry Weingarten</td>
<td>Special Project Office</td>
<td>SP2016</td>
</tr>
<tr>
<td></td>
<td>Navy Department</td>
<td>Washington 25, D. C.</td>
</tr>
<tr>
<td>Professor Irving Weiss</td>
<td>The Mitre Corporation</td>
<td>Bedford, Massachusetts</td>
</tr>
<tr>
<td>Capt. Burton L. Weller</td>
<td>AFPR Office</td>
<td>Martin Aircraft Corp.</td>
</tr>
<tr>
<td></td>
<td>Denver, Colorado</td>
<td></td>
</tr>
<tr>
<td>Professor Oscar Wesler</td>
<td>Department of Mathematics</td>
<td>University of Michigan</td>
</tr>
<tr>
<td></td>
<td>Ann Arbor, Michigan</td>
<td></td>
</tr>
</tbody>
</table>