IMPROVED DENSITY ESTIMATION

by

John P. Sommers

Technical Report No. 114
Department of Statistics Series

February 28, 1972

Research sponsored by the Office of Naval Research[1]
and U.S. Department of Labor[2]

DEPARTMENT OF STATISTICS
&
DEPARTMENT OF ECONOMICS

SOUTHERN METHODIST UNIVERSITY

[1]Department of Statistics Office of Naval Research Contract N00014-68-A-0515, Project NR 042-260.

[2]Department of Economics Institutional Grant No. 31-46-70-06 from the Manpower Administration, U.S. Department of Labor, under the authority of Title 1 of the Manpower Development and Training Act of 1962, as amended. Researchers undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment. Therefore, points of view or opinions stated in this document do not necessarily represent the official position or policy of the Department of Labor. The researchers are solely responsible for the factual accuracy of all material developed in the report.

sufficient background for the results we shall present later.

Let $x_1, x_2, \ldots x_n$ be a random sample distributed as the random variable X whose distribution function $F(x) = P[X \leq x]$ is absolutely continuous. If f is the probability density function for F such that

$$(1.1) \qquad F(x) = \int_{-\infty}^{x} f(t)dt$$

then the natural estimate of F(x) would be

$$F_n(x) = \frac{1}{n} [\text{number of observations} \leq x \text{ among } x_1, \ldots x_n]$$

Given this estimator Parzen [1962] reasoned that perhaps a good estimator of f(y) would be

$$(1.2) \qquad f_n(y) = \frac{F_n(y + h) - F_n(y - h)}{2h} \quad \text{for } h > 0.$$

If $h \to 0$ as $n \to \infty$, then $F_n(y)$ would for large n be a good approximation to the derivative of F evaluated at y, which is the desired density at y.

Parzen [1962] then suggested that (1.2) can be written as a weighted average over the sample distribution function:

$$f_n(y) = \int_{-\infty}^{\infty} \frac{1}{h} K\left(\frac{x-y}{h}\right) d\, F_n(x)$$

$$= \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x_i-y}{h}\right),$$

where

$$K(x) = \begin{cases} 1/2, & \text{for } -1 < x < 1 \\ 0, & \text{otherwise.} \end{cases}$$

This then led to a variety of estimators of $f(y)$ by using a variety of functions $K(x)$ and different values of h, where $K(x)$ and h are properly chosen.

Parzen [1962] restricted $K(x)$ and proved the following result.

Theorem 1A: Suppose $K(z)$ is a Borel function satisfying the following conditions:

$$\sup_{-\infty < z < \infty} | K(z) | < \infty,$$

$$\int_{-\infty}^{\infty} | K(z) | \, dz < \infty,$$

and

$$\lim_{z \to \infty} | zK(z) | = 0.$$

Also let $g(y)$ satisfy

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

Let $h(n)$ be a sequence of real numbers such that $\lim_{n \to \infty} h(n) = 0$ and furthermore let

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

then at every point of continuity of g

$$\lim_{n \to \infty} g_n(y) = g(y) \int_{-\infty}^{\infty} K(z) \, dz.$$

Further if $\int_{-\infty}^{\infty} K(z)dz = 1$ and if g is a probability density function, then $g_n(y)$ is asymptotically unbiased for g(y).

Parzen [1962] then showed further that

$$\lim_{n\to\infty} nh(n) \text{ Var } [g_n(y)] = g(y) \int_{-\infty}^{\infty} K^2(z)dz$$

and thus if $\lim_{n\to\infty} nh(n) = \infty$, $g_n(y)$ is a mean square consistent estimator of g(y).

Parzen [1962] also showed that under the proper conditions

$$\lim_{n\to\infty} \frac{E_x[g_n(y) - g(y)]}{h^2(n)} = \frac{1}{2} g^{(2)}(y) \int_{-\infty}^{\infty} z^2 K(z)dz,$$

which provides an approximation for the bias of these estimators.

Since 1962 many authors have examined this type of estimator further. Cacoullis [1966] and Epanechnikov [1969] extended the estimators to the case of multivariate density functions and developed similar asymptotic results concerning bias, variance, and consistency. Epanechnikov [1969] also developed a relative global mean square error μ^2 defined as follows:

$$\mu^2 = \frac{1}{Q} \int_{-\infty}^{\infty} E\ [f_n(x) - f(x)]^2\ dx,$$

where

$$Q = \int_{-\infty}^{\infty} f^2(x)dx.$$

He then developed the following asymptotic approximation for μ^2 by use of previous results:

$$\mu^2 \cong \frac{1}{Q}\left(\frac{L}{nh(n)} + \frac{h^4(n)M}{4}\right),$$

where

$$L = \int_{-\infty}^{\infty} K^2(z)\,dz$$

and

$$M = \int_{-\infty}^{\infty} \left[f^{(2)}(x)\right]^2 dx.$$

This approximation was then minimized with respect to h(n) to yield a minimum asymptotic relative global error. Further by restricting L to be equal to one he obtained an optimal function K(z) under this criterion.

Besides these extensions of the original estimators authors have worked on the properties of the estimators as first proposed. Murthy [1965] relaxed some of Parzen's [1962] assumptions for consistency. Nadaraya [1965] showed that if f is uniformly continuous then for a large class of kernels, $f_n(x)$ converges uniformly on the real line to f(x) with probability one. Schuster [1971] showed these conditions to be necessary and also stated conditions under which $f_n^{(r)}(x)$ converges to $f^{(r)}(x)$ uniformly with probability one.

Van-Ryzin [1969] showed the kernel-type estimators to be strongly consistent under conditions similar to those of Nadaraya [1965]. Bhattacharya [1967] also proved some theorems on consistency and convergence of the derivatives of $f_n(x)$.

Finally Pickands III [1969] and Woodroofe [1970] have presented two different methods of choosing sequences of estimators which are in some sense optimal. Pickands III [1969] defined I_n^2 and J_n^2 as follows:

$$I_n^2 = 2\pi E\left[\int_{-\infty}^{\infty} | f_n(x) - f(x) |^2 dx\right]$$

and $J_n^2 = \min I_n^2$, for all possible estimators $f_n(x)$.

He then presents a method of picking a sequence of estimators $f_n(x)$ so that if $\{I_n^2\}$ is the corresponding set of I_n^2 values then $\lim_{n\to\infty} \frac{J_n^2}{I_n^2} = 1$. Woodroofe, on the other hand, presents a method of estimating an optimal value of h(n), for a particular kernel by use of prior results. He further showed that asymptotically $E\,[f_n(x) - f(x)]^2$ is of the same order whether the optimal sequence of h(n)'s are used or the estimated optimal h(n)'s are used. We shall make further use of these results in Chapter IV.

CHAPTER II

THE JACKKNIFE

1. Introduction

The jackknife is the name given to an estimator developed by Quenouille [1956] and later used and named by Tukey [1958]. The name has survived because this technique is a useful tool for the statistician in much the same way as the jackknife is a "rough and ready" tool for man.

As it was originally developed by Quenouille [1956] the jackknife was a method of taking an estimator with bias of a certain form and by some alteration producing a new estimator, which had some of the bias of the original estimator eliminated. Schucany, Gray, and Owen [1971] have generalized Quenouille's [1956] methods. We shall describe these methods more fully.

2. The Generalized Jackknife

The generalized jackknife of Schucany, Gray, and Owen [1971] is defined as follows: let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two different estimators of a parameter θ, then if R is a real number not equal to one, define

$$G(\hat{\theta}_1, \hat{\theta}_2) = \frac{\hat{\theta}_1 - R\hat{\theta}_2}{1-R}$$

Further, if

$$E\,[\hat{\theta}_k] \;=\; \theta + \sum_{i=1}^{\infty} b_{k,i}(n,\theta)\;, \quad k = 1, 2, \tag{2.1}$$

where $\qquad b_{2,1}(n,\theta) \;\neq\; 0$ and

$$R(n) = \frac{b_{1,1}(n,\theta)}{b_{2,1}(n,\theta)} \;\neq\; 1, \tag{2.2}$$

then

$$E\,[G(\hat{\theta}_1,\hat{\theta}_2)] \;=\; \theta + \sum_{i=2}^{\infty} \left(\frac{b_{1,i}(n,\theta) - Rb_{2,i}(n,\theta)}{1-R}\right) \tag{2.3}$$

We thus see that by proper choice of R we have eliminated the first term of the expansion. This will generally reduce the bias since often this first term will dominate such a bias expansion.

3. The Method of Quenouille

One can easily see that in order to use the generalized jackknife one must have two estimators. The original method developed by Quenouille [1956] included a method of developing a second estimator from a first estimator.

Assume $\hat{\theta}$ is an estimator based on a random sample $x_1, \ldots x_n$. Assume further that $n = p.q$ where p and q are integers. We can then break the random sample into p subsets of size q. We then form p different subsamples of size n-q by eliminating the various subsets of size q from the original sample. With each of these subsamples of size n-q we form an estimator $\hat{\theta}^i$, formed in the same manner as $\hat{\theta}$ except based on a sample of size n-q. We then let

$$J_i(\hat{\theta}) = p\hat{\theta} - (p-1)\hat{\theta}^i \qquad i = 1, \ldots, p,$$

and

$$J(\hat{\theta}) = \sum_{i=1}^{p} \frac{J_i(\hat{\theta})}{p}$$

$$= p\hat{\theta} - (p-1)\,\bar{\hat{\theta}}^i ,$$

where $\qquad \bar{\hat{\theta}}^i = \frac{1}{p}\sum_{i=1}^{p} \hat{\theta}^i .$

The estimator $J(\hat{\theta})$ is what was originally called the jackknife by Tukey [1958]. We can see that if we let

$$R(n) = \frac{p-1}{p}$$

then

$$J(\hat{\theta}) = \frac{\hat{\theta} - R(n)\bar{\hat{\theta}}^i}{1-R(n)} .$$

Thus we can see that $J(\hat{\theta})$ is a special case of the generalized jackknife where the second estimator is $\bar{\hat{\theta}}^i$ and $R(n)$ is given above.

It is possible that $J(\hat{\theta})$ may not reduce the bias at all. However if

$$E[\hat{\theta}] = \theta + \sum_{i=1}^{\infty} b_i n^{-i}$$

for all n then

$$E[\hat{\theta}^i] = \theta + \sum_{i=1}^{\infty} b_i (n-q)^{-i} = E[\bar{\hat{\theta}}^i].$$

the existence of two estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ such that

$$E\left[\hat{\theta}_j\right] = \theta + \sum_{i=1}^{\infty} b_{j,i}(n,\theta) \qquad j = 1, 2,$$

and we thus eliminated the first term of the bias by use of the estimator

$$\frac{\hat{\theta}_1 - R(n)\hat{\theta}_2}{1-R(n)}, \text{ where}$$

$$R(n) = \frac{b_{1,1}(n,\theta)}{b_{2,1}(n,\theta)} .$$

It must be noted that R(n) still depends on θ as written. Actually, however, the case that is hoped for is that

$$b_{j,i}(n,\theta) = f_{i,j}(n) \cdot b_i(\theta) .$$

In this case then if $\hat{\theta}_2$ is formed by the method of Quenouille [1956] then $b_{2i}(n,\theta) = b_{1,i}(n-1,\theta) = f_{1,i}(n-1)b_i(\theta)$. Thus

$$R(n) = \frac{f_{1,1}(n)}{f_{1,1}(n-1)}$$

an expression which does not contain θ. In Quenouille's [1956] case we merely see a special form of $f_{i,j}(n)$.

Another possibility for use of the generalized jackknife is when one has two estimators where $\hat{\theta}_2$ is not formed by the method of Quenouille [1956] but still

$$E\left[\hat{\theta}_j\right] = \theta + \sum_{i=1}^{\infty} f_{i,j}(n)b_i(\theta), \; j = 1, 2.$$

In this case, because we have the same function of θ in the ith term of each bias expansion, then

$$R(n) = \frac{f_{1,1}(n)}{f_{2,1}(n)}, \text{ since}$$

the term $b_1(\theta)$ cancells. This possibility presents itself many times in density estimation because of the variety of possible choices of the kernel K and the spreading coefficient h. The relevant bias expansions are taken up in Chapter III.

5. Approximate Confidence Intervals

Another useful result which accompanies the generalized jackknife is a method of forming approximate confidence intervals for the parameter θ. These results are seen in Gray and Schucany [1972]. Here they have proved the following theorem.

Theorem: Let $x_1, x_2, \ldots, x_n$ be a random sample from a distribution with mean μ and finite variance σ^2. Let the parameter of interest be a function of the population mean, i.e.,

$$\theta = f(\mu).$$

Also let $\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$

and $\hat{\theta} = f(\bar{x}_n)$ (the corresponding function of the sample mean),

where f is a real-valued function, defined on the real line, which possesses a bounded second derivative in a neighborhood of μ. Further suppose

$$\alpha = \lim_{n\to\infty} \frac{R(n)}{(1-R(n))(n-1)} \neq \pm\infty, 0.$$

Then the random variable

$$\frac{\alpha\,(G(\hat{\theta}) - \theta)\sqrt{n}}{\sqrt{\sum_{i=1}^{n} (G_i(\hat{\theta}) - G(\hat{\theta}))^2}} \rightarrow N(0,1) \text{ in distribution}$$

as $n \rightarrow \infty$, where

$$G_i(\hat{\theta}) = \frac{\hat{\theta} - R(n)\hat{\theta}^{i}}{1-R(n)},$$

$$G(\hat{\theta}) = \sum_{i=1}^{n} \frac{G_i(\hat{\theta})}{n}, \text{ and}$$

the $\hat{\theta}^i$ are formed as described in section 3.

From this theorem it may be seen that for a large class of parameters θ we may develop approximate confidence intervals. It should also be noted that this result may be extended to the case where $\bar{x}_n$ is replaced by any U-statistic.

CHAPTER III

THEORETICAL CONSIDERATION

1. Bias Expansion Rigorously Established and Examined

In order to use the generalized jackknife we have seen that a series expansion for the bias was advantageous. Parzen [1962] has shown that as $n \to \infty$

$$\frac{E\,[f_n(y) - f(y)]}{h^2(n)} \to \int \frac{x^2 K(x)dx}{2} \; f^{(2)}(y) \,, \tag{3.1}$$

provided that $f^{(2)}(y)$ exists and that $K(x)$ is of exponential order 2 which is true if

$$\int xK(x)dx = 0 \quad \text{and} \quad \int x^2 \, K(x)dx < \infty.$$

We see from this result that if one multiplies (3.1) by $h^2(n)$ then the following approximation for the bias of $f_n(y)$ results:

$$E_x\,[f_n(y) - f(y)] \cong h^2(n) \int \frac{x^2K(x)dx}{2} \cdot f^{(2)}(y).$$

This asymptotic result immediately suggests a possible first term of a bias expansion and thus a place for the jackknife. Consequently we shall now develop a series expansion for the bias in order that we might use the jackknife.

Before developing our series we shall define some notation for our estimators of $f(y)$ and examine the form of their expected values. Define

$$f_n(y,K,h) = \sum_{i=1}^{n} \frac{K\left(\frac{x_i - y}{h}\right)}{nh} .$$

If $x_1 \ldots x_n$ is a set of independent, identically distributed random variables with probability density funtion f then

$$(3.2) \qquad E\,[f_n(y,K,h)] = \int_{-\infty}^{\infty} \frac{K\left(\frac{x-y}{h}\right)}{h} f(x)\,dx.$$

We shall now investigate an integral similar to that in (3.2).

<u>Theorem IA</u>: Let the following conditions hold:

1) $f^{(m+1)}(x)$ is continuous for $x \varepsilon [a,b]$;

2) $\int_{-\infty}^{\infty} z^t K(z)\,dz$ exists for $t = 0, 1, \ldots, m + 1$, where m is an integer greater than zero in 1) and 2);

3) $0 < h < \infty$;

4) $y \varepsilon \ (a,b)$;

then

$$(3.3) \qquad \int_a^b \frac{K\left(\frac{x-y}{h}\right)}{h} f(x)\,dx =$$

$$\sum_{p=o}^{m} \left[\frac{h^p f^{(p)}(y)}{p!} \int_{\frac{a-y}{h}}^{\frac{b-y}{h}} z^p K(z)\,dz \right] + \frac{h^{m+1}}{(m+1)!} \int_{\frac{a-y}{h}}^{\frac{b-y}{h}} z^{m+1} K(z) f^{(m+1)}(z_o)\,dz,$$

where $a \leq z_o \leq b$ and z_o is a function of z.

<u>Proof</u>: Since $f^{(m+1)}(x)$ exists throughout $[a,b]$ we may let $z = \frac{x-y}{h}$ in the integral on the left of (3.3) to obtain

$$\int_a^b \frac{K\left(\frac{x-y}{h}\right)}{h} f(x)dx = \int_{\frac{a-y}{h}}^{\frac{b-y}{h}} K(z)\ f(y + hz)\ dz.$$

We then expand $f(y+zh)$ about y by use of Taylor's theorem and the desired result follows.

This theorem leads to several corollaries which we shall tailor to our statistical needs.

Corollary I: If $h = h(n)$ is a function of n and $\lim_{n\to\infty} h(n) = 0$ and the other conditions of Theorem IA are true, then

$$\lim_{n\to\infty} \int_a^b \frac{K\left(\frac{x-y}{h(n)}\right)}{h(n)} f(x)dx = f(y) \int_{-\infty}^{\infty} K(z)dz.$$

Proof: For each $1 \le t \le m$

$$\lim_{n\to\infty} f^{(t)}(y) \frac{h^t(n)}{t!} \int_{\frac{a-y}{h(n)}}^{\frac{b-y}{h(n)}} z^t\ K(z)\ dz = 0 \text{ since },$$

$$\int_{-\infty}^{\infty} z^t K(z)dz$$ exists and $f^{(t)}(y)$ is bounded.

That is, each term of the sum in (3.3), except the $p = 0$ term, goes to zero.

Likewise $\left| \frac{h^{m+1}(n)}{(m+1)!} \int_{\frac{a-y}{h(n)}}^{\frac{b-y}{h(n)}} z^{m+1} K(z) f^{(m+1)}(z_o) \, dz \right| \leq$

$$M \frac{h^{m+1}(n)}{(m+1)!} \int_{\frac{a-y}{h(n)}}^{\frac{b-y}{h(n)}} \left| z^{m+1} K(z) \right| dz$$

where $\left| f^{(m+1)}(z_o) \right| \leq M$ for $z_o \varepsilon [a,b]$.

Thus, again, since $\int_{-\infty}^{\infty} z^{m+1} K(z) \, dz$ exists,

$$\lim_{n \to \infty} \frac{h^{m+1}(n)}{(m+1)!} \int_{\frac{a-y}{h(n)}}^{\frac{b-y}{h(n)}} z^{m+1} K(z) f^{(m+1)}(z_o) dz = 0.$$

This leaves the limit of the term

$$\int_{\frac{a-y}{h(n)}}^{\frac{b-y}{h(n)}} K(z) \, dz \cdot f(y) \quad \text{to be determined.}$$

But since $a < y < b$

$$\lim_{n \to \infty} \int_{\frac{a-y}{h(n)}}^{\frac{b-y}{h(n)}} K(z) \, dz = \int_{-\infty}^{\infty} K(z) \, dz$$

from which the desired result follows.

Corollary II:

If $f(x)$ is a probability density function and

$\int_{-\infty}^{\infty} K(z) \, dz = 1$ and the conditions of Theorem IA are

true then

$$E_x\left[\frac{K\left(\frac{x-y}{h(n)}\right)}{h(n)} - f(y)\right]$$

$$= \sum_{t=1}^{m} \frac{h^t(n)}{t!} f^{(t)}(y) \int_{\frac{a-y}{h(n)}}^{\frac{b-y}{h(n)}} z^t K(z)\,dz + \frac{h^{m+1}(n)}{(m+1)!} \int_{\frac{a-y}{h(n)}}^{\frac{b-y}{h(n)}} z^{m+1} K(z) f^{(m+1)}(z_o)\,dz$$

$$+ \int_{\frac{b-y}{h(n)}}^{\infty} K(z)\ [f(y + h(n)z) - f(y)]\ dz$$

$$+ \int_{-\infty}^{\frac{a-y}{h(n)}} K(z)\ [f(y + h(n)z) - f(y)]\ dz.$$

Proof: Since

$$E_x\left[\frac{K\left(\frac{x-y}{h(n)}\right)}{h(n)} - f(y)\right] =$$

$$\int_{-\infty}^{a}\left[\frac{K\left(\frac{x-y}{h(n)}\right)}{h(n)} - f(y)\right] f(x)\,dx + \int_{a}^{b}\left[\frac{K\left(\frac{x-y}{h(n)}\right)}{h(n)} - f(y)\right] f(x)\,dx$$

$$+ \int_{b}^{\infty}\left[\frac{K\left(\frac{x-y}{h(n)}\right)}{h(n)} - f(y)\right] f(x)\,dx\ ,$$

we may make the variable change $z = \frac{x-y}{h(n)}$ and apply Theorem IA to obtain:

$$E_x\left[\frac{K\left(\frac{x-y}{h(n)}\right)}{h(n)} - f(y)\right] =$$

$$\int_{-\infty}^{\frac{a-y}{h(n)}} K(z)f(y+h(n)z)dz + \int_{\frac{b-y}{h(n)}}^{\infty} K(z)f(y+h(n)z)dz$$

$$+ f(y)\int_{\frac{a-y}{h(n)}}^{\frac{b-y}{h(n)}} K(z)dz - f(y) + \sum_{t=1}^{m} \frac{h^t(n)}{t!} f^{(t)}(y)\int_{\frac{a-y}{h(n)}}^{\frac{b-y}{h(n)}} z^t K(z)dz$$

$$+ \frac{h^{m+1}(n)}{(m+1)!}\int_{\frac{a-y}{h(n)}}^{\frac{b-y}{h(n)}} z^{m+1} K(z)f^{(m+1)}(z_o)dz.$$

Adding and subtracting $f(y)\left[\int_{\frac{b-y}{h(n)}}^{\infty} K(z)dz + \int_{-\infty}^{\frac{a-y}{h(n)}} K(z)dz\right]$

give the desired result since $\int_{-\infty}^{\infty} K(z)dz = 1.$

We now have an expression for the bias of our estimators and thus can apply the jackknife. However, first we shall note several other important relationships which shall be used to simplify our task.

Consider a function $g(x) \geq 0$ such that for all x such that $|x| \geq p > 0$, $|x^t g(x)| \leq M$ where t is an integer greater than one and $M > 0$. Further assume

$$\int_p^{\infty} g(x)\,dx \quad \text{and} \quad \int_{-\infty}^{-p} g(x)\,dx \quad \text{exist.}$$

Then $$\left| \int_p^{\infty} g(x)\,dx \right| \leq \int_p^{\infty} \frac{M}{x^t}\,dx = \frac{M}{p^{t-1}(t-1)}$$

and $$\left| \int_{-\infty}^{-p} g(x)\,dx \right| \leq \frac{M}{p^{t-1}(t-1)} \quad .$$

Using this result we shall now prove a theorem which alters the series derived in Corollary II in such a way that the resulting series lends itself more easily to the jackknife.

Theorem IB: If f(x) is a probabilty density function such that for some positive integer t $f^{(2t+1)}(x)$ is continuous for $x \varepsilon [a,b]$ and if $y \varepsilon (a,b)$ and if K(x) is a function such that:

1) $\int_{-\infty}^{\infty} K(x)\,dx = 1$;

2) $K(x) = K(-x)$ for all x ;

3) $\int_{-\infty}^{\infty} x^p K(x)\,dx < \infty$ for $p = 1, \ldots, 2t + 1$;

4) $| x^{2t+2} K(x) | < M$ for all x ;

then for $h > 0$;

$$(3.4) \quad E_x\left[\frac{K\left(\frac{x-y}{h}\right)}{h} - f(y) \right] = \sum_{p=1}^{t} \frac{f^{(2p)}(y)\, h^{2p}}{(2p)!} \int_{-\infty}^{\infty} z^{2p} K(z)\,dz + 0\left(h^{2t+1}\right) .$$

Proof: From Corollary II we get

$$(3.5) \qquad E_x\left[\frac{K\left(\frac{x-y}{h}\right)}{h} - f(y)\right] = \sum_{p=1}^{2t} \frac{h^p}{p!} f^{(p)}(y) \int_{\frac{a-y}{h}}^{\frac{b-y}{h}} z^p K(z)dz$$

$$+ \frac{h^{2t+1}}{(2t+1)!} \int_{\frac{a-y}{h}}^{\frac{b-y}{h}} z^{2t+1} K(z) \cdot f^{(2t+1)}(z_o)\, dz$$

$$+ \int_{\frac{b-y}{h}}^{\infty} K(z)\,[f(y+hz) - f(y)]\, dz$$

$$+ \int_{-\infty}^{\frac{a-y}{h}} K(z)\,[f(y+hz) - f(y)]\, dz\ .$$

We now add and subtract

$$\sum_{p=1}^{2t} h^p \frac{f^{(p)}(y)}{p!}\left[\int_{\frac{b-y}{h}}^{\infty} z^p K(z)dz + \int_{-\infty}^{\frac{a-y}{h}} z^p K(z)dz\right]$$

to get

$$(3.6) \quad E_x\left[\frac{K\left(\frac{x-y}{h}\right)}{h} - f(y)\right] = \sum_{p=1}^{2t} \frac{h^p}{p!} f^{(p)}(y) \int_{-\infty}^{\infty} z^p K(z)dz$$

$$+ \int_{\frac{b-y}{h}}^{\infty} K(z)\,[f(y+hz) - f(y)]\, dz + \int_{-\infty}^{\frac{a-y}{h}} K(z)[f(y+hz)-f(y)]dz$$

$$+ \frac{h^{2t+1}}{(2t+1)!} \int_{\frac{a-y}{h}}^{\frac{b-y}{h}} z^{2t+1} K(z) f^{(2t+1)}(z_o)\,dz$$

$$- \sum_{p=1}^{2t} \frac{h^p}{p!} f^{(p)}(y) \left[\int_{\frac{b-y}{h}}^{\infty} z^p K(z)\,dz + \int_{-\infty}^{\frac{a-y}{h}} z^p K(z)\,dz \right] .$$

Now we can develop bounds for several of the terms in this expression, beginning with

$$(3.7) \qquad \left| \int_{\frac{b-y}{h}}^{\infty} K(z)\,[f(y+zh) - f(y)]\,dz \right| \leq$$

$$\int_{\frac{b-y}{h}}^{\infty} K(z) f(y+zh)\,dz + \int_{\frac{b-y}{h}}^{\infty} K(z) f(y)\,dz$$

$$\leq \sup_{z > \frac{b-y}{h}} K(z) \int_{\frac{b-y}{h}}^{\infty} f(y+zh)\,dz + f(y) \cdot \int_{\frac{b-y}{h}}^{\infty} K(z)\,dz .$$

But $\quad | z^{2t+2} K(z) | \leq M$ and thus

$$K(z) \leq \frac{M}{z^{2t+2}} \qquad \text{for all} \quad z > 0 .$$

Further for $z \geq \frac{b-y}{h}$

$$\frac{M}{z^{2t+2}} \leq \left(\frac{h}{(b-y)} \right)^{2t+2} M \quad \text{thus}$$

$$\sup_{z>\frac{b-y}{h}} K(z) \le \frac{h^{2t+2}}{(b-y)^{2t+2}} \cdot M .$$

Likewise $\displaystyle\int_{\frac{b-y}{h}}^{\infty} K(z)dz \le M \int_{\frac{b-y}{h}}^{\infty} \frac{1}{z^{2t+2}}\, dz = \frac{M}{(2t+1)} \frac{h^{2t+1}}{(b-y)^{2t+1}}$.

Inserting these results in (3.7) we see that

$$\left| \int_{\frac{b-y}{h}}^{\infty} K(z)[f(y+hz) - f(y)]dz \right| \le \frac{Mh^{2t+2}}{(b-y)^{2t+2}} \int_{\frac{b-y}{h}}^{\infty} f(y+zh)dz$$

$$+ \frac{Mh^{2t+1}}{(b-y)^{2t+1}(2t+1)} .$$

Thus $\displaystyle\int_{\frac{b-y}{h}}^{\infty} K(z)[f(y+hz) - f(y)]dz$

is $O(h^{2t+1})$. By a similar argument

$$\int_{-\infty}^{\frac{a-y}{h}} K(z)[f(y+hz) - f(y)]dz \text{ is } O(h^{2t+1}) .$$

Consider now from (3.6) the terms

$$- \sum_{p=1}^{2t} \frac{h^p}{p!} f^{(p)}(y) \left[\int_{\frac{b-y}{h}}^{\infty} z^p K(z)dz + \int_{-\infty}^{\frac{a-y}{h}} z^p K(z)dz \right].$$

For $z \neq 0$ $\left| z^{2t+2}K(z) \right| \le M$ and thus for $1 \leq p \leq 2t$ $\left| z^p K(z) \right| \leq \frac{M}{\left| z^{2t+2-p} \right|}$ and therefore the above sum is bounded in the following manner.

$$\left| -\sum_{p=1}^{2t} \frac{h^p}{p!} f^{(p)}(y) \left[\int_{\frac{b-y}{h}}^{\infty} z^p K(z)dz + \int_{-\infty}^{\frac{a-y}{h}} z^p K(z)dz \right] \right|$$

$$< \left| \sum_{p=1}^{2t} \frac{h^p}{p!} f^{(p)}(y) \cdot M \left[\frac{h^{2t+1-p}}{(2t+1-p)} \left[\frac{1}{(b-y)^{2t+1-p}} + \frac{1}{|a-y|^{2t+1-p}} \right] \right] \right|$$

$$= M\, h^{2t+1} \left| \sum_{p=1}^{2t} \frac{f^{(p)}(y)}{p!} \cdot \left[\frac{1}{(2t+1-p)} \left[\frac{1}{(b-y)^{2t+1-p}} + \frac{1}{|a-y|^{2t+1-p}} \right] \right] \right|$$

which is $0(h^{2t+1})$.

With these results (3.6) becomes

$$E_x \left[\frac{K\left(\frac{x-y}{h}\right)}{h} - f(y) \right] = \sum_{p=1}^{2t} \frac{h^p}{p!} f^{(p)}(y) \int_{-\infty}^{\infty} z^p K(z)dz + 0(h^{2t+1}) .$$

But since $K(z) = K(-z)$, for p odd $\int_{-\infty}^{\infty} z^p K(z)dz = 0$. Hence we finally obtain

$$E_x \left[\frac{K\left(\frac{x-y}{h}\right)}{h} - f(y) \right] = \sum_{p=1}^{t} \frac{h^{2p}}{(2p)!} f^{(2p)}(y) \int_{-\infty}^{\infty} z^{2p} K(z)dz + 0(h^{2t+1}).$$

2. Employing the Generalized Jackknife

a) Bias of the Standard Estimator

We have seen that for a general set of densities $f(x)$ and symmetric $K(z)$, we may write

$$E_x\left[\frac{K\left(\frac{x-y}{h}\right)}{h} - f(y)\right] = \sum_{p=1}^{t} \frac{h^{2p}}{(2p)!} f^{(2p)}(y) \int_{-\infty}^{\infty} z^{2p}K(z)\,dz + 0(h^{2t+1}).$$

Now with an expansion for our bias we may use the jackknife to eliminate terms of the bias expansion. To do this one must have two different estimators. In this problem due to the variety of functions $K(z)$ and the infinite number of h values we have to choose from we can develop an uncountable number of estimators from which to choose. We shall show how to select these estimators and combine them and present theorems concerning the asymptotic properties of their jackknife combination which should assist the reader in his choice of estimators.

b) Methods of Forming Estimators and the New Bias

As was mentioned we may choose different estimators by varying $K(z)$, h or both. Consider $K_1(z)$ and $K_2(z)$ both symmetric functions of the type described previously. If $x_1 \ldots x_n$ is a random sample of size n from a probability distribution with probability density f which satisfies the restrictions of Theorem IB, we may estimate f at the point y by

$$f_n(y,K_i,h_i) = \sum_{p=1}^{n} \frac{K_i\left(\frac{x_p-y}{h_i}\right)}{nh_i}$$

where $i = 1, 2$ and $h_i > 0$.

If we define $I(K,p) = \int_{-\infty}^{\infty} z^p K(z)dz$ then

$$E_x\left[f_n(y,K_i,h_i) - f(y)\right] = \sum_{p=1}^{t} \frac{f^{(2p)}(y)}{(2p)!} h_i^{2p} I(K_i,2p) + 0(h^{2t+1}).$$

If we let $R = \dfrac{h_1^2 I(K_1,2)}{h_2^2 I(K_2,2)}$ provided $R \neq 1$ as suggested for the generalized jackknife by Schucany, Gray, and Owen [1971] then we get a new estimator (to be denoted G when there is no possibility of confusion) by combining $f_n(y,K_1h_1)$ and $f_n(y,K_2,h_2)$ according to

$$G(f_n(y,K_1,h_1)\ ,\ f_n(y,K_2,h_2)\ ,\ R) = \frac{f_n(y,K_1,h_1) - R\ f_n(y,K_2,h_2)}{1 - R}\ ,$$

which has the following bias expansion, obtained by combining the individual bias expansions:

$$E_x\ [G(f_n(y,K_1h_1)\ ,\ f_n(y,K_2,h_2),\ R) - f(y)] =$$

$$\frac{1}{1-R}\sum_{p=1}^{t} \frac{\left[h_1^{2p}I(K_1,2p) - Rh_2^{2p}I(K_2,2p)\right]}{(2p)!} f^{(2p)}(y) + 0(h_1^{2t+1}) + 0(h_2^{2t+1})$$

But $\quad R = \dfrac{h_1^2 I(K_1,2)}{h_2^2\ I(K_2,2)}$, thus

$h_1^2 I(K_1,2) - h_2^2 RI(K_2,2) = 0$ and therefore we have eliminated the term of the bias expansions which involved $f^{(2)}(y)$, h_1^2 and h_2^2 leaving

$$\frac{1}{1-R} \quad \frac{h_1^4 I(K_1,4) - Rh_2^4 I(K_2,4)}{4!}\ f^{(4)}(y) \text{ as the leading term.}$$

Actually the process of jackknifing these kernel type estimators in this fashion to eliminate bias is a general method of producing the functions u(x) suggested by Bartlett [1963]. Bartlett [1963] following the work of Parzen [1962] defined u(x) in the same manner as K(x) except for two differences. Bartlett's u(x) could be negative and he imposed the requirement that $u(x) = 0$ for $|x| \geq h$. With this in mind Bartlett then noted that the first term of Parzen's bias could be made zero simply by choosing u(x) such that $\int_{-\infty}^{\infty} x^2 u(x)dx = 0$. Furthermore, if higher even-ordered "moments" of u(x) were zero, then more terms of the bias would be eliminated.

If we employ two kernels with finite range and take $h_1/h_2 = c$ (a constant) then we may let

$$u(x) = \frac{K_1(x) - \frac{I(K_1,2)}{I(K_2,2)} c^3 K_2(cx)}{1 - \frac{I(K_1,2)}{I(K_2,2)} c^2} .$$

This yields a function which fits Bartlett's criteria. We see that the methods proposed in this dissertation give a general procedure for constructing such u(x) and further remove the restriction that u(x) be zero outside a finite interval. Furthermore we shall establish the properties of these estimators in more detail than Bartlett did.

At this point one begins to see where to look to make a judgement concerning the choice of the two required estimators.

For instance if the values of h_1 and h_2 were small then the above term would dominate the bias expansion and thus we would like to find kernels $K_1(z)$ and $K_2(z)$ which make the term

$$\frac{1}{1-R} \quad \frac{h_1^4 \, I(K_1,4) - Rh_2^4 \, I(K_2,4)}{4!} \quad \text{small.}$$

Two such kernels might be

$K_1(z) = 1/2$, $z \varepsilon [-1,1]$

$K_1(z) = 0$, otherwise or

$K_2(z) = 1 - |z|$, $z \varepsilon [-1,1]$

$K_2(z) = 0$, otherwise.

The respective coefficients for the series expansion are

$$I\,(K_1,2t) = \frac{1}{(2t+1)} \quad \text{and}$$

$$I\,(K_2,2t) = \frac{1}{(t+1)(2t+1)} \quad .$$

Contrast these values with

$$K_3 \quad (z) = \frac{e^{-|z|}}{2} \;, \; -\infty < z < \infty \;, \quad \text{for which } I(K_3,2t) = (2t)!.$$

We can see that if bias were our only consideration we would always choose truncated functions for K(z). However, as we shall show there are other properties which recommend kernels such as $K_3(z)$ above.

We have given the most general form of our estimator where K_1 and K_2 are different and h_1 and h_2 are different. However, this need not be so. Only one of the two sets need be made up of different elements. For instance if h_1 and h_2 are the same and K_1 and K_2 are different then

$$R = \frac{I(K_1,2)}{I(K_2,2)}$$ and the

bias expansion for the estimator is

$$E_x\left[G(f_n(y,K_1,h)\ ,\ f_n(y,K_2,h)\ ,\ R) - f(y)\right] =$$

$$\sum_{p=1}^{t} \frac{f^{(2p)}(y)h^{2p}(I(K_1,2p) - RI(K_2,2p))}{(2p)!\ (1-R)} + 0(h^{2t+1})\ .$$

The first non-zero term is then

$$\frac{h^4 f^{(4)}(y)}{4!}\ \frac{I(K_1,4) - RI(K_2,4)}{(1-R)}\ .$$

If, on the other hand, we chose to use different h's and the same K then

$$R = \frac{h_1^2}{h_2^2}$$ and the bias expansion would become

$$E_x\ [G(f_n(y,K_1,h_1)\ ,\ f_n(y,K_1,h_2)\ ,\ R) - f(y)] =$$

$$\sum_{p=1}^{t} \frac{f^{(2p)}(y)}{(2p)!}\ \frac{\left(h_1^{2p}-Rh_2^{2p}\right)I(K_1,2p)}{1-R} + 0(h_1^{2t+1}) + 0(h_2^{2t+1})\ .$$

The first non-zero term is then

$$f^{(4)}(y)\quad I\ \frac{(K_1,4)}{4!}\ \frac{(h_1^4-Rh_2^4)}{1-R} = -h_1^2h_2^2\ f^{(4)}(y)\quad I\ \frac{(K_1,4)}{4!}.$$

A special form of this case relates to the original jackknife of Quenouille [1956] where R was taken as $\frac{n-1}{n}$. If we let $h_1 = \frac{c}{n^p}$ and $h_2 = c/(n-1)^p$ where $p > 0$ then $R = \left(\frac{n-1}{n}\right)^{2p}$ which has a limit of 1 as n increases without bound. This can cause difficulties in our theorems concerning asymptotic properties of this estimator because of the 1-R term in the denominator which approaches zero and thus causes a need for special theorems which we shall see later.

A final method of forming our estimators should be noted. If we select $K_1(z) \neq K_2(z)$, then we may choose h_1 and h_2 so that the second term of the bias expansion of G is also zero, that is, we can find values of h_1 and h_2 for which

$$h_1^4\, I(K_1,4) - \frac{h_1^2}{h_2^2}\frac{I(K_1,2)}{I(K_2,2)}\, h_2^4\, I(K_2,4) = 0.$$

Solving we get

$$\frac{h_1^2}{h_2^2} = \frac{I(K_2,4)\ \ I(K_1,2)}{I(K_1,4)\ \ I(K_2,2)}\text{, which gives}$$

$$R = \frac{I^2(K_1,2)}{I^2(K_2,2)}\ \frac{I(K_2,4)}{I(K_1,4)}\ .$$

If $R \neq 1$ the first two terms of the bias expansion are zero. Examples of this, as well as the other methods of forming estimators, may be found in the next chapter.

c) Asymptotic Properties

i) Case I, $R(n) = c$, a constant for all n.

The asymptotic unbiasedness of our estimator follows quite easily from Theorem 1A of Parzen [1962].

Theorem IIA: Suppose $K_i(z)$ $(i = 1,2)$ is a Borel function satisfying

1) $\sup 0 < K_i(z) < \infty$ for $-\infty < z < \infty$.

2) $\int_{-\infty}^{\infty} K_i(z)dz = 1$, and

3) $\lim_{z\to\infty} |zK_i(z)| = 0.$

Further suppose

4) $f(y)$ is a probability density function,

5) $h_1(n)$, $h_2(n)$ are sequences of positive constants such that $\lim_{n\to\infty} h_i(n)=0$ $i=1,2$ and $I(K_i,2)=\int_{-\infty}^{\infty} x^2 K_i(x)dx$ exists $i = 1,2$, so that

$$\frac{h_1^2(n)}{h_2^2(n)} \frac{I(K_1,2)}{I(K_2,2)} = R \neq 1 \text{ for all n,}$$

then

$$\lim_{n\to\infty} E_x \left[\frac{\left(\sum_{i=1}^{n} \frac{K_1\left(\frac{x_i-y}{h_1(n)}\right)}{nh_1(n)} - R \sum_{i=1}^{n} \frac{K_2\left(\frac{x_i-y}{h_2(n)}\right)}{nh_2(n)} \right)}{1-R} \right] = f(y)$$

at each point of continuity of $f(y)$.

Proof: Follows immediately since by Corollary 1A from Parzen [1962]

$$\lim_{n\to\infty} E_x \left[\sum_{i=1}^{n} \frac{K_j\left(\frac{x_i-y}{h_j(n)}\right)}{h_j(n)} \right] = f(y) \quad \text{for } j = 1, 2.$$

It then follows that the limit of our combined estimator is

$$\frac{f(y) - Rf(y)}{1-R} = f(y).$$

Thus the estimator is asymptotically unbiased. Generally the estimators are also consistent.

Theorem IIB: Under the conditions of Theorem IIA the estimator $G(f_n(y,K_1,h_1(n)) , f_n(y,K_2,h_2(n)) , R)$ is mean square consistent for $f(y)$ provided $nh_1(n) \to \infty$ as $n \to \infty$. Furthermore,

$$\lim_{n\to\infty} nh_1(n) \text{ Var } [G(f_n(y,K_1,h_1(n) , f_n(y,K_2,h_2(n)), R)]$$

$$= f(y) \int_{-\infty}^{\infty} \left(\frac{K_1(z) - R \cdot cK_2(c \cdot z)}{1-R} \right)^2 dz ,$$

where $c = \frac{h_1(n)}{h_2(n)}$, which is a constant for all n under the condition 5) of Theorem IIA.

Proof: First note if $\hat{\theta}$ is a random variable and θ a constant such that

$E[\hat{\theta} - \theta] = b$ then, if it exists, the mean square error (MSE) is given by

$$E[(\hat{\theta} - \theta)^2] = \text{Var } \hat{\theta} + b^2 .$$

Also if $x_1, x_2 \ldots x_n$ are independent identically distributed random variables then

$$\text{Var} \left[\sum_{i=1}^{n} \frac{x_i}{n} \right] = \frac{E[x^2]}{n} - \frac{(E[x])^2}{n} .$$

We can now get an expression for the MSE of $G(f_n(y,K_1,h_1(n)),\ f_n(y,K_2,h_2(n)),R)$ by employing

$$E_x\left[\left(G\left(f_n(y,K_1,h_1(n)),\ f_n(y,K_2,h_2(n)),R\right) - f(y)\right)^2\right] =$$

$$\mathrm{Var}\left[G(f_n(y,K_1,h_1(n)),\ f_n(y,K_2,h_2(n)),R\right]$$

$+ \quad b^2(n)$, where $b(n)$ is the bias of the estimator.

Then $E_x\left[\left(G(f_n(y,K_1,h_1(n)),\ f_n(y,K_2,h_2(n)),R) - f(y)\right)^2\right]$

$$= \frac{1}{(1-R)^2 n} E_x\left[\left(\frac{K_1\left(\frac{x-y}{h_1(n)}\right)}{h_1(n)} - R\,\frac{K_2\left(\frac{x-y}{h_2(n)}\right)}{h_2(n)}\right)^2\right]$$

$$- \quad \frac{1}{n}\left(E_x\left[\frac{\frac{K_1\left(\frac{x-y}{h_1(n)}\right)}{h_1(n)} - R\,\frac{K_2\left(\frac{x-y}{h_2(n)}\right)}{h_2(n)}}{(1-R)}\right]\right)^2 + b^2(n)\ .$$

Now since we are interested in the limit of this quantity as $n\to\infty$ let us first note that $\lim_{n\to\infty} b(n) = 0$ and

$$\lim_{n\to\infty} \frac{1}{n}\left(E_x\left[\frac{\frac{K_1\left(\frac{x-y}{h_1(n)}\right)}{h_1(n)} - R\,\frac{K_2\left(\frac{x-y}{h_2(n)}\right)}{h_2(n)}}{1-R}\right]\right)^2 = \lim_{n\to\infty} \frac{f^2(y)}{n} = 0\ .$$

Thus we need only consider the first term. However, since $\frac{h_1(n)}{h_2(n)} = c$, the first term becomes

$$\frac{1}{nh_1(n)(1-R)^2} \quad E_x\left[\frac{\left(K_1\left(\frac{x-y}{h_1(n)}\right) - RcK_2\left(c\,\frac{x-y}{h_1(n)}\right)\right)^2}{h_1(n)}\right].$$

This function fits the conditions of Theorem IA of Parzen [1962] since each of the individual functions $K_1(z)$ and $K_2(z)$ fit these conditions. The only one that might not be obvious is whether

$$\lim_{x\to\infty} \left| y(K_1(x-y) - RcK_2(c(x-y)))^2 \right| = 0 .$$

However, this is

$$\lim_{x\to\infty} \left| y(K_1(x-y) - RcK_2(c(x-y)) \right| \left| K_1(x-y) - RcK_2(c(x-y)) \right|$$

both parts of which go to zero.

Therefore $$\lim_{n\to\infty} \int_{-\infty}^{\infty} \left(\frac{K_1\left(\frac{x-y}{h_1(n)}\right) - RcK_2\left(c\left(\frac{x-y}{h_2(n)}\right)\right)}{h_1(n)} \right)^2 f(x)\,dx =$$

$$f(y)\int_{-\infty}^{\infty} [K_1(z) - RcK_2(cz)]^2\, dz = P .$$

Thus $$\lim_{n\to\infty} E_x \left[\left(G(f_n(y,K_1,h_1(n)), f_n(y,K_2,h_2(n)),R) - f(y) \right)^2 \right] =$$

$$\lim_{n\to\infty} \frac{P}{nh_1(n)(1-R)^2} .$$

Since K_1 and K_2 are bounded and integrable

$$\int_{-\infty}^{\infty} K_1^2\,(z)\,dz, \int_{-\infty}^{\infty} K_2^2\,(cz)\,dz, \int_{-\infty}^{\infty} K_1(z)K_2(cz)\,dz$$

are all finite; and hence P is finite.

Consequently $\lim_{n\to\infty} \frac{P}{nh_1(n)(1-R)^2} = 0$, since $\lim_{n\to\infty} nh_1(n) = \infty$.

We have shown that the MSE tends to zero and hence the estimator is mean square consistent. Further

$$\lim_{n\to\infty} nh_1(n) \text{ Var } [G(f_n(y,K_1,h_1(n)),f_n(y,K_2,h_2(n)),R)] =$$

$$\lim_{n\to\infty} \left[\frac{P}{(1-R)}2 - h_1(n) \; E[G(f_n(y,K_1,h_1(n)),f_n(y,K_2,h_2(n)),R)]^2\right]$$

$$= \frac{P}{(1-R)^2} \quad .$$

Thus the proof is complete .

ii) Case II, $\lim_{n\to\infty} R(n) = 1$

We shall now consider the asymptotic properties of the estimator when $K_1(z) = K_2(z)$ and $h_1(n) = \frac{c}{n^p}$, $h_2(n) = h_1(n-1)$ with $p > 0$, $c > 0$. Such an estimator $f_n(y,K,h_2(n))$ could be obtained by averaging $f_{n-1}(y,K,h_1(n-1))$ over all subsamples of size n-1 as described by Quenouille [1956]. For this estimator $R(n) = \left(\frac{n-1}{n}\right)^{2p}$ and $\lim_{n\to\infty} R(n) = 1$, which could cause problems since the estimator is

$$\sum_{i=1}^{n}\left(\frac{\frac{K}{h_1(n)}\left(\frac{x_i-y}{h_1(n)}\right)-R(n)\frac{K}{h_2(n)}\left(\frac{x_i-y}{h_2(n)}\right)}{n(1-R(n))}\right)$$

and has a limit of zero in the denominator. For this reason and because of the similarity to the original jackknife we shall investigate this estimator in more detail.

We shall now prove several theorems which show that a large class of these estimators are both asymptotically unbiased and mean-square consistent.

Theorem III: Let $K(z)$ be a function satisfying conditions 1,2,3, and 4 of Theorem IB and let $f(x)$ be a probability density function which satisfies the conditions of the density in Theorem IB. If $(2t + 1)p - 1 > 0$, $y \varepsilon (a,b)$ and $t > 2$ (there are $2t + 1$ continuous derivatives of f)

then

$$\lim_{n\to\infty} E_x \left[\frac{\dfrac{K\left(\frac{x-y}{h_1(n)}\right)}{h_1(n)} - R(n) \dfrac{K\left(\frac{x-y}{h_2(n)}\right)}{h_2(n)}}{1-R(n)} \right] = f(y) ,$$

where $h_1(n) = c/n^p$, $h_2(n) = h_1(n-1)$, $R(n) = \dfrac{h_1^2(n)}{h_2^2(n)}$ for all n and $c > 0$ and $p > 0$.

Proof: By Theorem IB

$$E_x \left[\frac{K\left(\frac{x-y}{h}\right)}{h} \right] = \sum_{q=0}^{t} \frac{f^{(2q)}(y) \cdot h^{2q} \int_{-\infty}^{\infty} z^{2q} K(z)dz}{(2q)!} + 0(h^{2t+1}) .$$

Then

$$(3.8) \quad E_x \left[\frac{\dfrac{K\left(\frac{x-y}{h_1(n)}\right)}{h_1(n)} - R(n) \dfrac{K\left(\frac{x-y}{h_2(n)}\right)}{h_2(n)}}{1-R(n)} \right]$$

$$= f(y) + \sum_{q=2}^{t} h_2^2(n) f^{(2q)}(y) \frac{(h_1^{2q}(n) - h_1^2(n) h_2^{2q-2}(n)) \, I(K,2q)}{h_2^2(n) - h_1^2(n)}$$

$$+ \frac{0\left(h_1^{2t+1}(n)\right) + 0\left(h_2^{2t+1}(n)\right)}{1 - (1-1/n)^{2p}} .$$

To evaluate the limit note that for an integer $q > 2$

$$(3.9) \qquad \frac{h_2^2(n)\left(h_1^{2q}(n) - h_1^2(n)h_2^{2q-2}(n)\right)}{h_2^2(n) - h_1^2(n)},$$

$$= - h_1^2(n)h_2^2(n)\left[h_1^{2q-4}(n) + h_1^{2q-6}(n)h_2^2(n)+\ldots+h_1^2(n)h_2^{2q-6}(n) + h_2^{2q-4}(n)\right],$$

which goes to zero a $n \to \infty$.

If $q = 2$ the expression on the left of (3.9) is $-h_1^2(n)h_2^2(n)$ which also goes to zero as $n \to \infty$.

Next consider the limit of the remainder by examining

$$\frac{h_1^{2t+1}(n)}{1-R(n)} = \frac{c\,\frac{1}{n}(2t+1)p}{1 - (1-1/n)^{2p}} .$$

We may use L'Hospital's Rule to determine this limit. Differentiating with respect to n we get

$$\frac{\dfrac{-c(2t+1)p}{n^{(2t+1)p+1}}}{-(1-1/n)^{2p-1}\dfrac{1}{n^2}} = \frac{c(2t+1)p}{(1-1/n)^{2p-1}} \cdot \frac{1}{n^{(2t+1)p-1}},$$

which goes to zero as n increases without bound since $(2t+1)p-1 > 0$.

Thus $\lim_{n\to\infty} \frac{0\left(h_i^{2t+1}\right)}{1-R(n)} = 0$, for $i = 1, 2$.

We have shown that all terms in (3.8) go to zero except f(y) and thus

$$\lim_{n\to\infty} E_x \left[\frac{\dfrac{K\left(\frac{x-y}{h_1(n)}\right)}{h_1(n)} - R(n)\,\dfrac{K\left(\frac{x-y}{h_2(n)}\right)}{h_2(n)}}{1-R(n)} \right] = f(y) .$$

Therefore the estimator

$$G(f_n(y)) = \sum_{i=1}^{n} \left[\frac{\dfrac{K\left(\frac{x_i-y}{h_1(n)}\right)}{h_1(n)} - R(n)\,\dfrac{K\left(\frac{x_i-y}{h_2(n)}\right)}{h_2(n)}}{n(1-R(n))} \right]$$

is asymptotically unbiased.

We shall now consider the mean-square consistency of this type of estimator, considering only estimators which are asymptotically unbiased. For estimators with this property we need only concern ourselves with

$$(3.10)\quad \frac{E_x}{n} \left[\frac{\left(\dfrac{K\left(\frac{x-y}{h_1(n)}\right)}{h_1(n)} - R(n)\,\dfrac{K\left(\frac{x-y}{h_2(n)}\right)}{h_2(n)} \right)^2}{(1-R(n))^2} \right]$$

$$= \frac{1}{n} \int_{-\infty}^{\infty} \left(\frac{\dfrac{K\left(\frac{x-y}{h_1(n)}\right)}{h_1(n)} - R(n)\,\dfrac{K\left(\frac{x-y}{h_2(n)}\right)}{h_2(n)}}{1-R(n)} \right)^2 f(x)\,dx .$$

If $|f(x)| \leq M$ for all x then the expression in (3.10) is less than or equal to

$$\frac{M}{n}\int_{-\infty}^{\infty}\left(\frac{\frac{K\left(\frac{x-y}{h_1(n)}\right)}{h_1(n)} - R(n)\,\frac{K\left(\frac{x-y}{h_2(n)}\right)}{h_2(n)}}{1-R(n)}\right)^2 dx \,.$$

Letting $z = \frac{x-y}{h_1(n)}$ we get

$$\frac{M}{nh_1(n)}\int_{-\infty}^{\infty}\left(\frac{K(z)-R^{3/2}(n)\;K(R^{1/2}(n)z)}{1-R(n)}\right)^2 dz$$ for an upper bound. Then if $\lim_{n\to\infty} nh_1(n) = \infty$ and if

$$\int_{-\infty}^{\infty}\left(\frac{K(z)-R^{3/2}(n)\;K(R^{1/2}(n)z)}{1-R(n)}\right)^2 dz$$ is bounded then the estimator is mean-square consistent. We thus have the following theorem.

Theorem IV: Assume f(x) is a bounded probability density function and

$$\int_{-\infty}^{\infty}\left(\frac{K(z)-R^{3/2}(n)\;K(R^{1/2}(n)z)}{1-R(n)}\right)^2 dz$$ exists and is bounded for all n. Further assume $R(n) = \frac{h^2(n)}{h^2(n-1)}$ where $\lim_{n\to\infty} h(n) = 0$ and $\lim_{n\to\infty} nh(n) = \infty$, and $\lim_{n\to\infty} R(n) = 1$.

Further assume $$\lim_{n\to\infty} E_x\left[\frac{\frac{K\left(\frac{x-y}{h(n)}\right)}{h(h)} - R(n)\,\frac{K\left(\frac{x-y}{h(n-1)}\right)}{h(n-1)}}{1-R(n)}\right] = f(y)$$

then $\lim_{n\to\infty} E_x\left[\left(G\left(f_n(y,K,h(n)),f_n(y,K,h(n-1)),R(n)\right) - f(y)\right)^2\right] = 0$.

We have a theorem now which shows us a general class of functions K(z) which yield mean-square consistent estimators. We shall now limit this class to a smaller set for which we shall develop more properties.

Theorem V: Let f(x) be a bounded probability density function such that $f'(x)$ is continuous for $x \varepsilon [a,b]$ and let $y \varepsilon (a,b)$.

Then if 1) $R(n) = \frac{h^2(n)}{h^2(n-1)}$ where $h(n) = \frac{c}{n^p}$, $c > 0$, $p > 0$

for all n

2) $\lim_{n\to\infty} nh(n) = \infty$;

3) $$\lim_{n\to\infty} E_x\left[\frac{\frac{K\left(\frac{x-y}{h(n)}\right)}{h(n)} - R(n) \cdot \frac{K\left(\frac{x-y}{h(n-1)}\right)}{h(n-1)}}{1-R(n)}\right] = f(y) ;$$

4) $$\lim_{n\to\infty}\left(\frac{K(z) - R^{3/2}(n)K(R^{1/2}(n)z)}{1-R(n)}\right)^2 = g(z) \text{ for all } z ;$$

5) $$\left(\frac{K(z) - R^{3/2}(n)K(R^{1/2}(n)z)}{1-R(n)}\right)^2 \leq M \text{ for all } n \text{ and all } z;$$

6) $$p(n) = \int_{-\infty}^{\infty}\left(\frac{K(z) - R^{3/2}(n)K(R^{1/2}(n)z)}{1-R(n)}\right)^2 dz \text{ exists and is finite}$$

for all n and $\lim_{n\to\infty} p(n) = \int_{-\infty}^{\infty} g(z)dz < \infty$;

7) $$\left| \int_c^d z\left(\frac{K(z) - R^{3/2}(n)K(R^{1/2}(n)z)}{1-R(n)}\right)^2 dz \right| < A < \infty$$

for all $c < d$ and for all n;

then the estimator $\left[G\left(f_n(y,K,h(n)), f_n(y,K,h(n-1)), R(n)\right)\right]$ is mean-square consistent for $f(y)$ and

$$\lim_{n\to\infty} nh(n)\operatorname{Var}\left[G(f_n(y,K,h(n)), f_n(y,K,h(n-1)), R(n)\right] = f(y)\int_{-\infty}^{\infty} g(z)dz.$$

<u>Proof:</u> Consider $\frac{1}{n}\, E_x\left[\left(\frac{\frac{K\left(\frac{x-y}{h(n)}\right)}{h(n)} - R(n)\,\frac{K\left(\frac{x-y}{h(n-1)}\right)}{h(n-1)}}{1-R(n)}\right)^2\right]$

$$= \frac{1}{n}\int_{-\infty}^{\infty}\left(\frac{\frac{K\left(\frac{x-y}{h(n)}\right)}{h(n)} - R(n)\,\frac{K\left(\frac{x-y}{h(n-1)}\right)}{h(n-1)}}{1-R(n)}\right)^2 f(x)dx\ .$$

Letting $z = \frac{x-y}{h(n)}$ the integral becomes

$$(3.11)\quad \frac{1}{nh(n)}\left[\int_{\frac{a-y}{h(n)}}^{\frac{b-y}{h(n)}}\left(\frac{K(z) - R^{3/2}(n)K(R^{1/2}(n)z)}{1-R(n)}\right)^2 f(y+h(n)z)dz\right.$$

$$+\int_{\frac{b-y}{h(n)}}^{\infty}\left(\frac{K(z) - R^{3/2}(n)K(R^{1/2}(n)z)}{1-R(n)}\right)^2 f(y+h(n)z)dz$$

$$\left.+\int_{-\infty}^{\frac{a-y}{h(n)}}\left(\frac{K(z) - R^{3/2}(n)K(R^{1/2}(n)z)}{1-R(n)}\right)^2 f(y+h(n)z)\,dz\right].$$

Expanding f about y in the first term of (3.11) yields

$$(3.12)\quad \frac{1}{nh(n)}\left[\int_{\frac{a-y}{h(n)}}^{\frac{b-y}{h(n)}}\left(\frac{K(z)-R^{3/2}(n)K(R^{1/2}(n)z)}{1-R(n)}\right)^2 dz \cdot f(y)\right.$$

$$\left.+\int_{\frac{a-y}{h(n)}}^{\frac{b-y}{h(n)}} h(n)z\left(\frac{K(z)-R^{3/2}(n)K(R^{1/2}(n)z)}{1-R(n)}\right)^2 f'(z_o)\,dz\right],$$

where z_o is a function of z and h(n), and $z_o \varepsilon [a,b]$.

The second integral in (3.12) is bounded by

$$h(n)\, A \cdot \left|\sup_{z\varepsilon(a,b)} f'(z)\right| \quad \text{which goes to zero since } h(n)$$

goes to zero. Now there exists an n_o such that

$$\int_{\frac{b-y}{h(n_o)}}^{\infty} g(z)dz < \varepsilon \quad \text{and} \int_{-\infty}^{\frac{a-y}{h(n_o)}} g(z)dz < \varepsilon$$

$$\text{and} \left|\int_{\frac{a-y}{h(n_o)}}^{\frac{b-y}{h(n_o)}}\left(\frac{K(z)-R^{3/2}(n)K(R^{1/2}(n)z)}{1-R(n)}\right)^2 - \int_{\frac{a-y}{h(n_o)}}^{\frac{b-y}{h(n_o)}} g(z)dz\right| < \varepsilon ,$$

the last expression following from the Lebesgue Dominated Convergence Theorem.

$$\text{Letting} \quad B_n(z) = \left(\frac{K(z)-R^{3/2}(n)K(R^{1/2}(n)z)}{1-R(n)}\right)^2$$

then there exists $n_1 \geq n_o$ such that for all $n > n$.

$$\left| \int_{-\infty}^{\infty} B_n(z)\,dz - \int_{-\infty}^{\infty} g(z)\,dz \right| < \varepsilon .$$

Then for all $n \geq n_1$

$$\left| \int_{\frac{b-y}{h(n_o)}}^{\infty} B_n(z)\,dz + \int_{-\infty}^{\frac{a-y}{h(n_o)}} B_n(z)\,dz \right| < \varepsilon + \int_{\frac{b-y}{h(n_o)}}^{\infty} g(z)\,dz$$

$$+ \int_{-\infty}^{\frac{a-y}{h(n_o)}} g(z)\,dz$$

$$+ \left| \int_{\frac{a-y}{h(n_o)}}^{\frac{b-y}{h(n_o)}} \Big(B_n(z) - g(z)\Big)\,dz \right|$$

$$< 4\,\varepsilon$$

But since $B_n(z) \geq 0$ then for all $n \geq n_1$

$$\int_{\frac{b-y}{h(n)}}^{\infty} B_n(z)\,dz + \int_{-\infty}^{\frac{a-y}{h(n)}} B_n(z)\,dz < 4\,\varepsilon .$$

Thus the sum of the two integrals goes to zero for large n.

We see then that for $n \geq n_1$

$$\left| \int_{\frac{a-y}{h(n)}}^{\frac{b-y}{h(n)}} B_n(z)\,dz - \int_{-\infty}^{\infty} g(z)\,dz \right| < \varepsilon + \int_{\frac{b-y}{h(n)}}^{\infty} B_n(z)\,dz + \int_{-\infty}^{\frac{a-y}{h(n)}} B_n(z)\,dz$$

$$< 5\,\varepsilon .$$

Thus we have that for all "large n" greater than $N \geq n_1$

$$(3.13)\quad \left| \frac{1}{nh(h)} \left[f(y) \int_{\frac{a-y}{h(n)}}^{\frac{b-y}{h(n)}} B_n(z)dz - f(y) \int_{-\infty}^{\infty} g(z)dz \right.\right.$$

$$+ \int_{\frac{b-y}{h(n)}}^{\infty} B_n(z)f(y+h(n)z)dz + \int_{-\infty}^{\frac{a-y}{h(n)}} B_n(z)f(y+h(n)z)dz$$

$$\left.\left. + \; h(n) \int_{\frac{a-y}{h(n)}}^{\frac{b-y}{h(n)}} zB_n(z) \circ f'(z_o)dz \right] \right|$$

$$\leq \frac{1}{nh(n)} \left[f(y) \quad 5\varepsilon + \varepsilon \sup_{x\varepsilon(a,b)} f(x) + \varepsilon \sup_{x\varepsilon(a,b)} f(x) \right.$$

$$\left. + \; h(n) \quad A \quad \sup_{z \,\varepsilon\, [a,b]} f'(z) \right] .$$

Multiplying both sides of (3.13) by $nh_1(n)$ we see that for $n \geq N$,

$$(3.14)\quad \left| f(y) \int_{\frac{a-y}{h(n)}}^{\frac{b-y}{h(n)}} B_n(z)dz + \int_{\frac{b-y}{h(n)}}^{\infty} B_n(z)f(y+h(n)z)dz \right.$$

$$+ \int_{-\infty}^{\frac{a-y}{h(n)}} B_n(z)f(y+h(n)z)dz + \int_{\frac{a-y}{h(n)}}^{\frac{b-y}{h(n)}} h(n) \quad zB_n(z)f(z_o)dz$$

$$\left. - \int_{-\infty}^{\infty} g(z)dz \right|$$

$$= \left| nh(n)\, E\left[\left(G(f_n(y,K,h(n)), f_n(y,K,h(n-1)), R(n)\right)^2\right] - f(y)\int_{-\infty}^{\infty} g(z)dz \right|$$

$$\leq f(y)\, 5\varepsilon + 2\varepsilon \sup_{x\varepsilon(a,b)} f(x) + h(n)\, A \sup_{z\varepsilon[a,b]} f'(z) .$$

But ε is arbitrary and $\lim_{n\to\infty} h(n) = 0$,

$$\text{thus} \quad \lim_{n\to\infty} nh(n)\, E\left[\left(G(f_n(y,K,h(n)), f_n(y,K,h(n-1)), R(n))\right)^2\right]$$

$$= f(y)\int_{-\infty}^{\infty} g(z)dz .$$

Further since $nh(n)\left(\frac{E[(G(f(y,K,h(n)), f(y,K,h(n-1)), R(n))]}{n}\right)^2$

goes to zero then

$$\lim_{n\to\infty} nh(n) \operatorname{Var}\,[G(f_n(y,K,h(n)), f_n(y,K,h(n-1)), R(n))]$$

$$= f(y)\int_{-\infty}^{\infty} g(z)dz .$$

Also

$$(3.15) \quad \lim_{n\to\infty}\left[E_x\,[\,G(f_n(y,K,h(n)), f_n(y,K,h(n-1)), R(n)) - f(y)]^2\right]$$

$$= \lim_{n\to\infty}\left[\frac{f(y)\int_{-\infty}^{\infty} g(z)dz + b^2(n)}{nh(n)}\right],$$

where $b(n) = E_x\,[G(f_n(y,K,h(n)), f_n(y,K,h(n-1)), R(n)) - f(y)]$.

Therefore, the expression in (3.15) then has a limit of zero since $nh(n)$ goes to infinity and by assumption 3) $b(n) \to 0$ and thus the theorem is complete.

We shall now consider a set of functions $K(z)$ which satisfy the conditions of Theorem V. Many densities with infinite support belong to this set. Among these are the normal and the t.

<u>Theorem VI:</u> Let $K(z)$ be a function such that

1) $K(z) > 0$ for all z ;

2) $\int_{-\infty}^{\infty} K(z)dz$ exists ;

3) $K'(z)$ exists and is continuous for all z ;

4) $\int_{-\infty}^{\infty} |z||K'(z)|dz$ exists ;

5) $|K(z)|$ and $|K'(z)|$ are bounded for all z ;

6) there exists a $z_a > 0$ such that

   $|K'(z_1)| > |K'(z_2)|$ for $z_2 > z_1 > z_a$ and for all $z_2 < z_1 < -z_a$;

7) $\int_c^d z\left(\frac{K(z) - R^{3/2}(n)K(R^{1/2}(n)z)}{1-R(n)}\right)^2 dz < p < \infty$ for all $c<d$ and

   for all n ;

then if $1 > R(n) > 0$ for all n and $\lim_{n\to\infty} R(n) = 1$,

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} \left[\frac{K(z) - R^{3/2}(n)K(R^{1/2}(n)z)}{1-R(n)}\right]^2 dz$$

$$= \int_{-\infty}^{\infty} \left(3/2\ K(z) + 1/2\ zK'(z)\right)^2 dz \quad \text{and}$$

$$\lim_{n\to\infty} \left(\frac{K(z) - R^{3/2}(n)K(R^{1/2}(n)Z)}{1-R(n)}\right)^2 = \left(\frac{3}{2}K(z) + \frac{1}{2}zK'(z)\right)^2$$

for all z .

<u>Proof</u>: By Taylor's Theorem letting $R^{1/2}(n)z = z - \left((1-R^{1/2}(n))z\right)$

$$(3.16) \int_{-\infty}^{\infty} \left[\frac{K(z) - R^{3/2}(n)K(R^{1/2}(n)z)}{1-R(n)}\right]^2 dz =$$

$$\int_{-\infty}^{\infty} \left(\frac{K(z)\left(1-R^{3/2}(n)\right) + \left(1-R^{1/2}(n)\right) zK'(z_o(z))}{1 - R(n)}\right)^2 dz =$$

$$\int_{-\infty}^{\infty} \left(\left[\frac{1 + R^{1/2}(n) + R(n)}{1+R^{1/2}(n)}\right]K(z) + \frac{1}{1+R^{1/2}(n)} z\cdot K'(z_o(z))\right)^2 dz$$

where $z_o(z)$ is a function of z such that

$$z_o(z) \in (R^{1/2}(n)z, z) \text{ for } z \geq 0 \text{ and } z_o(z) \in (z, R^{1/2}(n)z) \text{ if } z < 0.$$

Since $|K'(z)| \leq M$ and $|K(z)| \leq M$ for all z then for $-2z_a \leq z \leq 2z_a$

for n large enough so R(n) is close enough to 1 ,

$$\left(\left(\frac{1+R^{1/2}(n)+R(n)}{1+R^{1/2}(n)}\right)K(z) + \frac{zK'(z_o(z))}{1+R^{1/2}(n)}\right)^2 \leq \left(3M + 2z_aM\right)^2 .$$

Further for $|z| \geq |2z_a|$ and n large enough so $R^{1/2}(n) > 1/2$

$$\left[\left(\frac{1+R^{1/2}(n)+R(n)}{1+R^{1/2}(n)}\right)K(z) + z\,\frac{K'(z_o(z))}{1+R^{1/2}(n)}\right]^2 \le \left(3K(z)+\left|zK'\left(\tfrac{z}{2}\right)\right|\right)^2 .$$

But since $K(z)$ and $|zK'(z)|$ are integrable so is

$$\left(3K(z) + |zK'(z/2)|\right)^2 = B(z).$$

Thus if $Q(z) = (3M + 2z_a M)^2$ for $|z| \le 2|z_a|$ and $Q(z) = B(z)$ for $|z| > 2|z_a|$, then we see that for n sufficiently large the function

$$\left(\left[\frac{1+R^{1/2}(n) + R(n)}{1+R^{1/2}(n)}\right]K(z) + \frac{1}{1+R^{1/2}(n)} \quad K'(z_o(z))\right)^2$$

is bounded by an integrable function.

Further since $\lim_{n\to\infty} R^{1/2}(n) = 1$ and $R^{1/2}(n)z \le z_o(z) \le z$ for $z \ge 0$ and

$$R^{1/2}(n)\ z > z_o(z) > z \quad \text{for } z < 0 \quad \text{then}$$

$$\lim_{n\to\infty} R^{1/2}(n)\ z = z \quad \text{and} \quad \lim_{n\to\infty} zK'(z_o(z)) = zK'(z) ,$$

since $K'(z)$ is continuous. Thus

$$\lim_{n\to\infty}\left[\left(\frac{1+R^{1/2}(n)+R(n)}{1+R^{1/2}(n)}\right)K(z) + \frac{zK'(z_o(z))}{1+R^{1/2}(n)}\right]^2 = \left(\frac{3}{2}\,K(z)+\frac{zK'(z)}{2}\right)^2 .$$

Then since for all n the integral in (3.16) exists then the integral in (3.16) converges to

$$\int_{-\infty}^{\infty} \left(\frac{3}{2} K(z) + \frac{zK'(z)}{2}\right)^2 dz \quad \text{by}$$

the Lebesgue Dominated Convergence Theorem and the theorem is complete.

We now see that if a kernel K(z) meets the requirements of Theorem VI and if

$$\lim_{n\to\infty} E_x \left[\frac{\dfrac{K\left(\frac{x-y}{h(n)}\right)}{h(n)} - R(n)\,\dfrac{K\left(\frac{x-y}{h(n-1)}\right)}{h(n-1)}}{1-R(n)}\right] = f(y)$$

then we would have a class of estimators which fit Theorem V. We would thus have a class of asymptotically unbiased estimators which are mean-square consistent and for which we have an idea of the asymptotic variance.

Another set of functions which fit Theorems III and V are those K(z) such that

$$(3.17)\quad K(z) = \begin{cases} c\left(1 - |z|^p\right)^s & \text{for } -1 \le z \le 1,\ p > 0 \\ & \quad s \ge 1\,, \\ 0\,, & \text{otherwise, where } c \text{ is chosen so that} \end{cases}$$

$$\int_{-\infty}^{\infty} K(z)dz = 1\ .$$

These functions easily satisfy Theorem III. Further

$$\int_{-\infty}^{\infty} \left[\frac{K(z) - R^{3/2}(n)K(R^{1/2}(n)z)}{1-R(n)}\right]^2 dz \quad \text{becomes}$$

$$(3.18)\quad 2\int_0^1 c^2\left[\frac{\left(1-|z|^p\right)^s - R^{3/2}(n)\left(1-|R^{1/2}(n)z|^p\right)^s}{(1-R(n))^2}\right]dz$$

$$+\ 2\int_1^{1/R^{1/2}(n)} R^3(n)\left(\frac{1-|R^{1/2}(n)z|^p}{(1-R(n))^2}\right)^{2s} dz\ .$$

But by L'Hospital's rule for all z such that $0 \leq z \leq 1$ the integrand in the first integral in (3.18) converges to

$$(3.19)\qquad 2c^2\left(3/_2\ \left(1-|z|^p\right)^s + \frac{sp}{2}\left(1-|z|^p\right)^{s-1} z^p\right)^2 .$$

Then by the Lebesgue Dominated Convergence Theorem since the first integrand in (3.18) converges to a bounded continuous integrable function on a finite domain then the first integral in (3.18) converges to the integral over [0,1] of (3.19).

Further for $1/R^{1/2}(n) \geq z \geq 1$ the integrand in the second term of (3.18) is bounded by

$$(3.20)\qquad R^3(n)\left(\left(\frac{1-R^{p/2}(n)}{1-R(n)}\right)^s\right)^2 \quad \text{for all } n$$

provided $0 < R(n) < 1$. However (3.20) is bounded by $\left(p/_2 + 1\right)^s$ and thus

$$0 \leq \lim_{n\to\infty} \int_1^{1/R^{1/2}(n)} R^3(n)\left(\frac{\left(1-R^{p/2}(n)z^p\right)^s}{1-R(n)}\right)^2 dz$$

$$\leq \lim_{n\to\infty} \int_1^{1/R^{1/2}(n)} \left(\frac{p}{2} + 1\right)^s dz = 0 .$$

Therefore

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} \left[\frac{K(z) - R^{3/2}(n)K(R^{1/2}(n)z)}{1-R(n)}\right]^2 dz$$

$$= 2\int_0^1 c^2\left[\frac{3}{2}(1-z^p)^s + \frac{ps}{2} z^p (1-s^p)^{s-1}\right]^2 dz .$$

We thus have a set of functions K(z) with finite range which give a set of estimators by this method which are asymptotically unbiased, mean-square consistent, and for which we have an asymptotic value for the variance.

d. Higher Order Jackknife

We mentioned that the generalized jackknife introduced by Schucany, Gray, and Owen [1971] could be used to eliminate more than one term of the bias expansion. For this reason we present the following theorem similar to Theorems IIA and IIB for a special group of these higher order jackknife estimators.

Theorem VII: Suppose $K_i(z)$ is a Borel function for i = 1, ...p, such that for each i = 1, ...p, $K_i(z)$ satisfies the following:

1) $0 < \sup_{-\infty < z < \infty} K_i(z) < 1$;

2) $\int_{-\infty}^{\infty} K_i(z)dz = 1$;

3) $\lim_{|z|\to\infty} |zK_i(z)| = 0$;

4) $I(K_i,2s) = \int_{-\infty}^{\infty} x^{2s} K_i(x)\,dx < \infty \quad s = 1,\ldots,p.$

Suppose then that

5) $f(y)$ is a probability density function:

6) $h_1(n), \ldots h_p(n)$ are positive constants such that $\lim_{n\to\infty} h_i(n) = 0$

and $\dfrac{h_i(n)}{h_1(n)} = c_i$ for all n ;

7) $\lim_{n\to\infty} nh_1(n) = \infty$;

8)
$$D(n) = \begin{vmatrix} 1 & \cdots & 1 \\ I(K_1,2)h_1^2(n) & \cdots & I(K_p,2)h_p^2(n) \\ \vdots & \cdots & \vdots \\ I(K_1,2p-2)h_1^{2p-2}(n) & \cdots & I(K_p,2p-2)h_p^{2p-2}(n) \end{vmatrix} \neq 0, \text{ for all n.}$$

$$\text{Then } \lim_{n\to\infty} E_x \sum_{i=1}^{n} \frac{1}{n} \frac{\begin{vmatrix} \dfrac{K_1\left(\dfrac{x_i-y}{h_1(n)}\right)}{h_1(n)} & \cdots & \dfrac{K_p\left(\dfrac{x_i-y}{h_p(n)}\right)}{h_p(n)} \\ I(K_1,2)h_1^2(n) & & I(K_p,2)h_p^2(n) \\ \vdots & & \vdots \\ I(K_1,2p-2)h_1^{2p-2}(n) & & I(K_p,2p-2)h_p^{2p-2}(n) \end{vmatrix}}{D(n)} \quad (3.21)$$

$= f(y)$ and

$$\lim_{n\to\infty} nh_1(n)\ \mathrm{Var}\left[\frac{\sum_{i=1}^{n}}{n}\frac{\begin{vmatrix} \frac{K_1\left(\frac{x_i-y}{h_1(n)}\right)}{h_1(n)} & \cdots & \frac{K_p\left(\frac{x_i-y}{h_p(n)}\right)}{h_p(n)} \\ I(K_1,2)h_1^2(n) & \cdots & I(K_p,2)h_p^2(n) \\ \vdots & & \vdots \\ I(K_1,2p-2)h_1^{2p-2}(n) & & I(K_p,2p-2)h_p^{2p-2}(n) \end{vmatrix}}{D(n)}\right]$$

$$= f(y)\frac{\displaystyle\int_{-\infty}^{\infty}\begin{vmatrix} K_1(z)c_2^{-1}K_2(c_2^{-1}z) & & c_p^{-1}K_p(c_p^{-1}x) \\ I(K_1,2)\ c_2^2 I(K_2,2) & \cdots & c_p^2 I(K_p,2) \\ \vdots & & \vdots \\ I(K_1,2p-2)\ c_2^{2p-2}I(K_2,2p-2) & & c_p^{2p-2}I(K_p,2p-2) \end{vmatrix}^2 dz}{\begin{vmatrix} 1 & & 1 \\ I(K_1 2)\ \ c_2^2 I(K_2,2) & \cdots & c_p^2 I(K_p,2) \\ \vdots & & \vdots \\ I(K_1,2p-2)\ c_2^{2p-2}I(K_2,2p-2) & & c_p^{2p-2}I(K_p,2p-2) \end{vmatrix}^2}.$$

Proof: By factoring $h_1^{2p}(n)$ out of the 2nd to pth row of the top and bottom of (3.21) and then cancelling, (3.21) becomes

(3.22)

$$E_x \frac{\begin{vmatrix} \frac{K_1\left(\frac{x-y}{h_1(n)}\right)}{h_1(n)} & \frac{c_2^{-1}K_2\left(c_2^{-1}\left(\frac{x-y}{h_1(n)}\right)\right)}{h_1(n)} & \cdots & \frac{c_p^{-1}K_p\left(c_p^{-1}\left(\frac{x-y}{h_1(n)}\right)\right)}{h_1(n)} \\ I(K_1,2) & c_2^2 I(K_2,2) & \cdots & c_p^2 I(K_p,2) \\ \vdots & & & \vdots \\ I(K_1,2p-2) & c_2^{2p-2} I(K_2,2p-2) & & c_p^{2p-2} I(K_p,2p-2) \end{vmatrix}}{\begin{vmatrix} 1 & 1 & & 1 \\ I(K,2) & c_2^2 I(K_2,2) & \cdots & c_p^2 I(K_p,2) \\ \vdots & & & \vdots \\ I(K,2p-2) & c_2^{2p-2} I(K_2,2p-2) & & c_p^{2p-2} I(K_p,2p-2) \end{vmatrix}} .$$

We see that (3.22) is nothing more than

$$(3.23) \quad E\sum_{i=1}^{p} a_i \; K\left(\frac{x-y}{h_i(n)}\right), \text{ where} \sum_{i=1}^{p} a_i = 1 .$$

Now $\lim_{n\to\infty} E_x\left[\frac{K\left(\frac{x-y}{h_i(n)}\right)}{h_i(n)}\right] = f(y)$ by Corollary IA of Parzen [1962]

and thus the desired result follows.

To prove the second part of the theorem we consider the function in (3.22) and apply the same procedures as were applied in Theorem IIB and the desired results follow.

e) Asymptotic Mean-Square Error

We have seen in Parzen [1962] and Epanechnikov [1969] that under quite general conditions the mean square error for the kernel-type estimator of $f(y)$ was approximated as follows

$$MSE(y) \doteq \frac{c(y)}{nh(n)} f(y) + \left[\frac{f^{(2)}(y)}{2} \quad I(K,2)\right]^2 h_1^4(n)$$

for large n .

Using this result both authors arrived at a value of $h(n)$ which in some way minimizes MSE (y), Parzen [1962] on a pointwise basis and Epanechnikov on what he terms a global basis. Both, however, arrive at the best value of $h(n)$ in the form $\frac{c}{n^{1/5}}$, where c is a positive constant.

This yields MSE $(y) = \frac{c_1(y)}{n^{4/5}}$ where $c_1(y)$ is a positive constant depending on y, the point at which the density is to be estimated.

Our estimator under quite general conditions has as the term of the bias expansion with lowest power of $h(n)$, a term whose exponent is $2t$ where $t \geq 2$. Since we also have a variance approximation of the form $\frac{c_2(y)}{nh(n)}$ we then arrive at an asymptotic mean square-error of the following form for our estimators

$$MSE\ (G,y) \doteq \frac{c_2(y)}{nh(n)} + c_3(y)h^{4t}(n),$$ where $c_2(y)$ and $c_3(y)$ are positive constants depending on y. If we minimize our mean-square error by either Parzen's or Epanechnikov's method we arrive at

$$h(n) = \frac{c_4(y)}{n^{\frac{1}{4t+1}}}, \quad t \geq 2 \text{ as the minimizing value.}$$

Then

$$MSE\ (G,y) = \frac{c_5(y)}{n^{4t/4t+1}}, \quad t \geq 2 .$$

We can see that taking the ratio of the two mean-square errors we get

$$\frac{MSE\ (G,y)}{MSE\ (y)} = \frac{c_6(y)}{n^{4(t-1)/5(4t+1)}}, \quad t \geq 2 .$$

Since this has a limit of zero as n increases without bound we see that the new estimator is asymptotically more efficient than the standard when both have been optimized in the above fashion.

CHAPTER IV

EXAMPLES AND PRACTICAL CONSIDERATIONS

1. Approximations

We have seen from the end of the previous chapter that the jackknifed kernel estimators are asymptotically better than the ordinary kernel type estimators when an optimal h is chosen. This is a good property; however, in general practice we will have a specific finite value of n and will not have the knowledge of the density to be estimated needed to find these optimal values. For this reason we need to make the best possible use of the approximations for variance and bias. We shall therefore list these approximations for the original type estimator and the jackknifed estimators so that we may refer to them as the need arise.

If K(x) is a symmetric kernel as described in Theorem IB then

$$(4.1) \qquad E_x\ [f_n(y,K,h) - f(y)] \cong \frac{h^2 f^{(2)}(y)}{2}\ I(K,2).$$

and

$$(4.2) \qquad \mathrm{Var}\ [f_n(y,K,h)] \cong \frac{f(y)}{nh} \int_{-\infty}^{\infty} K^2(x)\,dx.$$

These are the approximations for the standard estimators.

If $K_1(x)$ and $K_2(x)$ are symmetric kernels which satisfy Theorems IB and IIA then

(4.3) $$E_x\left[G\left(f_n(y, K_1,h_1)\ ,\ f_n(y,K_2,h_2)\ ,\ R\right)- f(y)\right]$$

$$\cong \left[\frac{h_1^4\ (I(K_1,4) - Rc^4 I(K_2,4))}{4!\ (1-R)}\right] f^{(4)}(y)\ ,$$

where $R = \dfrac{I(K_1,2)}{I(K_2,2)}\ \dfrac{h_1^2}{h_2^2} \neq 1$ and

$$\frac{h_1}{h_2} = c, \text{ where } c \text{ is a positive constant.}$$

If h_1 and h_2 are chosen so that

$$c^2 = \frac{h_1^2}{h_2^2} = \frac{I(K_2,4)}{I(K_1,4)} \cdot \frac{I(K_1,2)}{I(K_2,2)}\ , \text{ then the bias}$$

is approximated by

(4.4) $$E_x\ [G(f_n(y,K_1,h_1),f_n(y,K_2,h_2),R) - f(y)]$$

$$\cong h_1^6(I(K_1,6) - Rc^6 I(K_2,6))\ f^{(6)}(y)\ .$$

In either of the above cases

(4.5) $$\text{Var}\ [G(f_n(y,K_1,h_1),f_n(y,K_2,h_2),R)]$$

$$\cong \frac{f(y)}{nh_1}\int_{-\infty}^{\infty}\left(\frac{K_1(x) - cRK_2(cx)}{1-R}\right)^2 dx\ .$$

Note we do not mention the case where R is a function of n simply because we assume that only one value of n is available and thus even if R is taken to be close to one we can still assume for purposes of approximation that R would remain constant for higher values of n.

2. Choosing the Value of h

We shall now discuss the problem of choosing h values. First we shall find approximate optimal values of h by use of the given approximations for the mean-square error of two estimators. We shall then compare the approximate values with true expected values and make some observations. Finally we shall discuss finding the optimal h with respect to MSE in practice.

The example we shall give will be estimating a normal density at two points. We shall use two different estimators.

Let $\quad K(z) = \dfrac{e^{-z^2/2}}{\sqrt{2\pi}}$ then

$$\int_{-\infty}^{\infty} K^2(z)\,dz = \frac{1}{2\sqrt{\pi}}, \quad I(K,2) = 1, \text{ and } I(K,4) = 3.$$

We shall need these values for our approximations as seen from (4.1) - (4.5).

Our first estimator will be the standard type using K(z),namely

$$f_n(y,K,h) = \sum_{i=1}^{n} \frac{K\left(\frac{x_i-y}{h}\right)}{nh} = \sum_{i=1}^{n} \frac{e^{-\frac{1}{2}\left(\frac{x_i-y}{h}\right)^2}}{\sqrt{2\pi}\;nh} .$$

For this estimator the approximate bias is obtained from (4.1) as $\frac{h^2 f^2(y)}{2}$ and the approximate variance from (4.2) is

$\frac{f(y)}{2nh\sqrt{\pi}}$, where f is the density to be estimated at the point y.

The other estimator we shall consider is

$$G(f_n(y,K,h_1),f_n(y,K,h_2),.99) = \sum_{i=1}^{n}\left[\frac{K\left(\frac{x_i-y}{h_1}\right)}{(.01)nh_1} - .99\frac{K\left(\frac{x_i-y}{h_2}\right)}{(.01)nh_2}\right]$$

$$= \sum_{i=1}^{n}\left[\frac{e^{-\frac{1}{2}\left(\frac{x_i-y}{h_1}\right)^2}}{(.01)h_1 n\sqrt{2\pi}} - .99\frac{e^{-\frac{1}{2}\left(\frac{x_i-y}{h_2}\right)^2}}{(.01)nh_2\sqrt{2\pi}}\right],$$

where $\frac{h_1^2}{h_2^2} = .99 = R$.

For this estimator, which we shall refer to as G, the approximate bias obtained from (4.3) is $\frac{h_1^4}{7.92} f^{(4)}(y)$ and the approximate variance from (4.5) is $\frac{.47484}{nh_1} f(y)$.

$$\text{Thus}\quad \text{MSE }[f_n(y,K,h)] \cong \frac{f(y)}{2\sqrt{\pi}nh} + \frac{h^4}{4}[f^{(2)}(y)]^2$$

and

$$\text{MSE }[G] \cong \frac{.47484}{nh_1} f(y) + h_1^8\left[\frac{f^{(4)}(y)}{7.92}\right]^2 .$$

If f is the normal density function the above become

$$(4.6)\quad MSE[f_n(y,K,h)] \cong \frac{e^{-x^2/2}}{\sqrt{2\pi}} \cdot \frac{1}{nh2\sqrt{\pi}} + \frac{h^4}{4}\left(\frac{(y^2-1)e^{-y^2/2}}{\sqrt{2\pi}}\right)^2$$

and

$$(4.7)\quad MSE[G] \cong \frac{.47484e^{-y^2/2}}{\sqrt{2\pi}nh_1} + h_1^8\left(\frac{(y^4-6y^2+3)e^{-y^2/2}}{7.92\sqrt{2\pi}}\right)^2 .$$

The actual expected values when f is the normal density are

$$(4.8)\quad E_x[f_n(y,K,h)] = f(y)\frac{e^{\frac{1}{2}\frac{y^2h^2}{(1+h^2)}}}{(1+h^2)^{1/2}} ,$$

$$(4.9)\quad E_x[f_n^2(y,K,h)] = \frac{f(y)}{\sqrt{2\pi}} \cdot \frac{e^{\frac{y^2h^2}{2(2+h^2)}}}{(2+h^2)^{1/2}} ,$$

$$(4.10)\quad E_x[G] = \frac{f(y)}{.01}\left[\frac{e^{\frac{1}{2}\frac{y^2h_1^2}{1+h_1^2}}}{(1+h_1^2)^{1/2}} - .99\frac{e^{\frac{1}{2}\frac{y^2h_2^2}{1+h_2^2}}}{(1+h_2^2)^{1/2}}\right] , \text{ and}$$

$$(4.11)\quad E_x[G^2] = \frac{f(y)}{(.0001)nh_1\sqrt{2\pi}}\left[\frac{e^{\frac{1}{2}\frac{x^2h_1^2}{2+h_1^2}}}{(2+h_1^2)^{1/2}} + \frac{(.99)^{2.5}\, e^{\frac{1}{2}\frac{x^2h_2^2}{2+h_2^2}}}{(2+h_2^2)^{1/2}} - \frac{2(.99)^{1.5}}{(1.99+h_1^2)^{1/2}}\, e^{\frac{1}{2}\frac{x^2h_1^2}{(1.99+h_1^2)}}\right].$$

After finding an approximate optimal h with respect to MSE by using (4.6) and (4.7) we may then evaluate the approximate mean-square error as well as the true expected values. An example of these is listed on the following page, where in the last column

$$Q_1(y) = n^{-1}E_x[h^{-2}K^2(h^{-1}(x-y))] \text{ and } Q_2(y) = n^{-1}E_x[(h_1^{-1}K(h_1^{-1}(x-y)) - Rh_2^{-1}K(h_2^{-1}(x-y)))^2] .$$

The last column is thus the value the approximate MSE really approximates as seen from the derivation in Chapter III.

With this example perhaps we should make some observations about the estimators we have. The first is that the optimal values of h_1 are larger than those of h for the same value of n. As may be seen in (4.6) and (4.7), for $h_1 = h$ the variance of G tends to be larger than that of f_n. This seems to be a general rule; anything which reduces bias tends to increase variance. If a kernel has a relatively low variance it tends to yield higher bias.

Values for $f_n(0,K,h), y=0, f(0)=.3989423$

n	Optimal h	$E[f_n(0,K,h)]$	Bias	Approximate MSE	MSE	$(Bias)^2+Q_1(y)$
25	.46873	.361229	.037714	.011524	.0053156	.010538
50	.40806	.369373	.029569	.0066191	.0034452	.006174
100	.35523	.375928	.023014	.0038017	.0021891	.003602

Values for G, y=0, f(0) = .3989423

n	Optimal h_1	E[G]	Bias	Approximate MSE	MSE	$(Bias)^2+Q_2(y)$
25	.70214	.380269	.018673	.012141	.0046944	.010479
50	.65010	.384031	.014911	.0065564	.0027897	.0057393
100	.60191	.387161	.011781	.0035407	.0016428	.0031417

Values for $f_n(2,K,h), y=2, f(2)=.0533910$

n	Optimal h	$E[f_n(2,K,h)]$	Bias	Approximate MSE	MSE	$(Bias)^2+Q_1(y)$
25	.45061	.068989	.014998	.0016224	.0015835	.0017739
50	.39228	.065626	.011635	.00093184	.00091244	.00099858
100	.34150	.062962	.0089717	.00053520	.00052489	.00056453

Values for G, y = 2, f(2) = .0533910

n	Optimal h_1	E[G]	Bias	Approximate MSE	MSE	$(Bias)^2+Q_2(y)$.
25	.72638	.062875	.0088845	.0015883	.0016188	.0017769
50	.67253	.060767	.0067763	.00085771	.00086860	.00094245
100	.62268	.059113	.0051221	.00046319	.00046577	.00050072

For instance, if $K(z)$ is the standard normal density $\int_{-\infty}^{\infty} K^2(z)dz = \frac{1}{2\sqrt{\pi}}$ and $I(K,z) = 1$. On the other hand if $K_1(z) = \begin{cases} 1/2 & -1 < z < 1 \\ 0 & \text{otherwise} \end{cases}$

then $\int_{-\infty}^{\infty} K_1^2(z)dz = .5$ but $I(K_1,2) = 1/3$.

The kernel $K(z)$ for the same value of h tends to have higher approximate bias but lower approximate variance as seen from (4.1) and (4.2).

If one makes h smaller for the same kernel then again from (4.1) and (4.2) one raises variance and lowers bias. If one makes h larger the opposite holds true. Finally if one jackknifes and keeps h fixed the variance of G tends to exceed the variance of the kernels which made up the jackknife, however, the bias goes down. This is seen in (4.6) and (4.7).

Actually these observations would tend to lead one to believe that there would be some method of balancing these effects, and there is, once the specific estimator has been chosen. This balance lies in the optimal h value. One can see this in the example where the optimal h_1 value is greater than the optimal h. This leads to an MSE for the jackknife which is about the same or smaller than that of the standard kernel estimator. At the same time even with a higher h_1 value the bias of the jackknife is less than that of the standard type estimator. This is all as it should be, since the jackknife is supposed to reduce bias and with optimal h_1 is asymptotically better in the sense of MSE. However, how is the optimal h_1 to be chosen since it depends upon the density to be estimated? The answer is furnished by Woodroofe [1970] and is simple enough. Woodroofe [1970] notes that the optimal value of h_1 is approximately $an^{-\frac{1}{(2r+1)}}$, where r is the power of h_1 in the first non-zero term of the bias expansion and

$$a^{2r+1} = \frac{8r}{(r!)^2} \cdot \frac{d^2(f^{(r)}(y))^2}{ef(y)}, \quad \text{where}$$

d and e are positive constants. For a standard jackknife r = 4 and

$$d = \frac{I(K_1,4) - Rc^4\, I(K_2,4)}{1-R} \quad \text{and}$$

$$e = \int_{-\infty}^{\infty}\left(\frac{K_1(x) - RcK_2(cx)}{1-R}\right)^2 dx \quad \text{as shown in (4.3) and (4.5).}$$

In a case similar to that shown in (4.4) r = 6 and

$$d = \frac{I(K_1,(6) - Rc^6 I(K_2,6)}{1-R} \quad .$$

Woodroofe [1970] then gives an estimate for a^{2r+1} which thus gives an estimate for the optimal h_1, which we shall call 0_1 .

$$(4.12) \quad 0_1^{-(2r+1)} = n\left[\frac{8r}{(r!)^2}\left(\frac{(dG^{(r)}(f_n(y,K_3,h_3),f_n(y,K_4,h_4),R_1)^2+b_n)}{e|G(f_n(y,K_1,h_3),f_n(y,K_2,h_4)R)|+b_n}\right)\right],$$

where $0 < b_n$, $\lim_{n\to\infty} b_n = 0$, and nb_n is bounded away from zero for all n. Furthermore $K_3(z)$ and $K_4(z)$ are kernels with bounded, continuous rth derivatives such that $I(K_3,r) < \infty$ and $I(K_4,r) < \infty$ and $|K_3(z)| + |zK_3'(z)|$ and $|K_4(z)| + |zK_4'(z)|$ are both dominated by a function L which is bounded, real-valued, symmetric and absolutely integrable on the real line.

Thus Woodroofe [1970] simply uses the same type of estimator to estimate the optimal h_1 and then uses this estimated optimal h_1, along with the same data again to estimate the density. If $G(0_1)$ is the estimator using the estimated optimal h_1 and $G(h_1)$ is the estimator using the optimal h_1 then Woodroofe [1970] proved that under the conditions of (4.12)

$$MSE[G(0_1)] \cong MSE[G(h_1)], \text{ as } n \to \infty ,$$

provided the density f is bounded on the real line and has r continuous derivatives in a neighborhood about the point y where f is to be estimated.

This method is an asymptotically optimal one. However, since we do not know the properties of this estimator for a fixed n, we may not get a good value of h at all. On the other hand, it is generally better to use all available information rather than to make guesses. Thus until further studies can be made by Monte Carlo or exact methods, this procedure must be recommended.

One can see that whether or not the above method is used, an h must be chosen at sometime and one would like to choose the best possible value. To pick such an h one perhaps needs to do some guessing, for although we have a reasonable approximation for the MSE, it is a function of the unknown density as can be seen from (4.1)-(4.5). To have any idea of the size of the bias in comparison to the variance or the bias of another estimator we would need to know the comparative sizes of $f(y)$, $f^{(2)}(y)$, $f^{(4)}(y)$ and in some cases $f^{(6)}(y)$. Perhaps the best solution is to assume $|f^{(4)}(y)| > |f^{(2)}(y)| > f(y)$ as is true for many points of the normal density for instance. This would lead one to the worst of all possible situations for bias. However, consider

$$(4.13) \qquad \frac{a\ f(y)}{nh} + h^4\ b\ (f^{(2)}(y))^2 ,$$

the approximate MSE for the standard estimator arrived at from (4.1) and (4.2) and consider

$$(4.14) \qquad \frac{c\ f(y)}{nh_1} + h^8\ d\ (f^{(4)}(y))^2 ,$$

the approximate MSE for one of the new estimators as seen from (4.3) and (4.5). Generally in (4.13) and (4.14) a somewhat larger value of $f^{(2)}(y)$ and $f^{(4)}(y)$ does not affect the MSE very much because the bias squared is usually small compared to the variance portion of these approximations. Furthermore even if $|f^{(4)}(y)| > |f^{(2)}(y)|$ the bias of the new estimator is still smaller because d in (4.14) is generally much less than b in (4.13) and h_1^8 is likewise less than h^4. This is illustrated in (4.6) and (4.7).

What does this all finally mean? It means that one should worry about variance and probably the bias will take care of itself provided h is kept from being extremely large. Thus one should pick h_1 in (4.14) so that $\frac{c}{nh_1}$ in (4.14) is somewhat less than $\frac{a}{nh}$ in (4.13). Thus the variance of the jackknife should be smaller than that of the original type estimator chosen. At the same time the biases will be of the same general magnitude provided h_1 is not taken to be unreasonably large. For instance if $a = 2$ and $c = 3$ one might take $h_1 = 2h$. This would allow for less variance without much effect on bias. One should not get carried away and take something like $h_1 = 10h$ because this would probably have an unreasonable effect on bias. One should always remember the rule that a decrease in variance gives an increase in bias and vice versa. So one should not attack one without thinking of the other.

To illustrate these points the following are examples of expected values when the above principles were put into practice. A standard normal density is to be estimated. The first estimator considered is

$$f_n(y,K_1,h) = \sum_{i=1}^{n} \frac{K_1\left(\frac{x_i-y}{h}\right)}{nh} ,$$

a standard type estimator, where

$$K_1(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}} \quad \text{and} \quad h = .1 .$$

For this estimator $\int_{-\infty}^{\infty} K_1^2(x)\,dx = \frac{1}{2\sqrt{\pi}} \cong .28209$, which would be the value of a in equation (4.12) .

The second estimator considered is a generalized jackknife which we will denote G_1

$$G_1 = \sum_{i=1}^{n} \left[\frac{\dfrac{K_1\left(\dfrac{x_i-y}{h_1}\right)}{h_1} - .99\,\dfrac{K_1\left(\dfrac{x_i-y}{h_2}\right)}{h_2}}{.01n} \right]$$

where $\dfrac{h_1^2}{h_2^2} = .99$ and $h_1 = .2$.

$$\text{Here } \int_{-\infty}^{\infty} \left(\frac{K_1(z) - R^{3/2} K_1(R^{1/2} z)}{1-R} \right)^2 dz \cong .4748 \quad ,$$

which is the value of c in equation (4.14) .

We see that $\dfrac{.28209}{.1} > \dfrac{.4748}{.2}$ as was desired but h_1 is not extreme.

We then examine another generalized jackknife which we shall denote as G_2.

$$G_2 = \sum_{i=1}^{n} \left[\frac{\dfrac{K_1\left(\dfrac{x_i-y}{h_3}\right)}{h_3} - .6K_2\,\dfrac{\left(\dfrac{x_i-y}{h_4}\right)}{h_4}}{.4n} \right], \text{ where}$$

$h_4 = .75 \qquad \dfrac{h_3^2}{h_4^2} = .2$ and thus

$h_3 = .3354101966$, and

$$K_2(x) = \begin{cases} 1/2 , & -1 < x < 1 \\ 0 , & \text{otherwise} \end{cases} .$$

Here $$\int_{-\infty}^{\infty} \left(\frac{K_1(z) - .6\sqrt{.2}\, K_2\left(\sqrt{.2}z\right)}{.4} \right)^2 dz \cong .63239 .$$

Again $\frac{.28209}{.1} > \frac{.63239}{.3354101966}$ but

h_3 is not extreme.

The following example gives expected values for these estimators. Listed are the expected values of the estimators and n times the expected values of the estimator squared which, since the expected values and biases of the estimators are "similar", gives an approximation to the relative difference in the various MSE's considered.

y	$f(y)$	$E[f_n(y,K_1,h)]$	$nE[f_n^2(y,K_1,h)]$	$E[G_1]$	$nE[G_1^2]$	$E[G_2]$	$nE[G_2^2]$
0	.3989423	.3969624	1.12259	.3987159	.941947	.3988387	.771126
.5	.3520653	.3507519	.991301	.3519586	.832409	.3520334	.646480
1	.2419707	.2419648	.682583	.2420586	.574471	.2420352	.455036
1.5	.1295176	.1303183	.366498	.1296525	.309583	.1295713	.252019
2	.0539910	.0547974	.153466	.0540469	.130320	.0539856	.110359
2.5	.0175283	.0179894	.050096	.0175156	.042840	.0175043	.038254
3	.0044318	.0046108	.012753	.0044062	.010999	.0044217	.010516
4	.0001338	.0001441	.000392	.0001291	.000345	.0001364	.000396

From these examples one can see in practice what was discussed. Perhaps one should note that these particular estimators deteriorate in relative bias and some in variance as y moves towards the tails of the density where one would expect to have more troubles making estimates. This is especially true for the normal and similar densities where the derivatives in the tails become larger relative to $f(y)$.

3. Truncated Densities

We shall now consider one other instructive example. Consider

$$f(y) = e^{-y} \quad , \quad y > 0$$
$$f(y) = 0 \quad , \quad \text{otherwise} .$$

This density has a discontinuity at the origin. Suppose we desire to estimate the density at y with $f_n(y,K_2,h)$ where $K_2(z)$ is the same as described in the previous example. Then

$$E\left[f_n(y,K_2,h)\right] = f(y)\left[\frac{e^{h}-e^{-h}}{2h}\right] , \text{ for } y \geq h ;$$

$$E\left[f_n(y,K_2,h)\right] = \frac{1-e^{-h-y}}{2h} = e^{-y}\left[\frac{e^{y}-e^{-h}}{2h}\right], \text{ for } |y| < h;$$

and $$E\left[f_n(y,K_2,h)\right] = 0, \quad \text{for } y \leq -h .$$

Now $\frac{e^{h}-e^{-h}}{2h}$ approaches 1 as h approaches 0 and thus for small h and

$y \geq h$ the expected value of $f_n(y)$ is the value of the density times a constant close to one. In this case everything is relatively good. However, for $-h < y < h$ we see that problems occur. This is because, as was shown in Theorem IB the bias expansion depends on the nearest discontinuity. Here where it is close enough so that in the bias expansion the term

$$f(y) \int_{\frac{a-y}{h}}^{\frac{b-y}{h}} K(z)dz$$

yields $mf(y)$ where $m < 1$, then there can be a considerable amount of bias if m is small as in this case. Fortunately, if $K(z)$ has a finite support, this problem exists only at points "close" to the discontinuity since if $K(z) > 0$ only for $-\varepsilon < z < \varepsilon$ then when $\frac{b-y}{h}$ and $\left|\frac{a-y}{h}\right|$ are both greater than ε then

$$\int_{\frac{a-y}{h}}^{\frac{b-y}{h}} K(z)dz = 1.$$ If $K(z)$ had an infinite support

then

$$\int_{\frac{a-y}{h}}^{\frac{b-y}{h}} K(z)dz$$ would never equal one although it would approach one as $h \to 0$. Such is the case of $K(z) = \frac{e^{-z^2/2}}{\sqrt{2\pi}}$. In this instance if $y > 0$, $\int_{\frac{a-y}{h}}^{\frac{b-y}{h}} K(z)dz = \int_{\frac{-y}{h}}^{\infty} \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz$, which

is never quite equal to one for $h > 0$. Likewise if $y < 0$ the expected value will never equal zero like the density. We will always have problems with that part of the expected value given by

$$\int_{-y/h}^{\infty} \frac{e^{-z^2/2}}{\sqrt{2\pi}}\, dz,$$ which goes to zero as Theorem IB promises.

However, for all $y < 0$ we will always have a positive expected value, although the further away from zero y is, the smaller the expected value becomes. For these reasons it is suggested that if one has a truncated density one should use a "finite" kernel. Of course if one knows ε is the truncation point, then he should simply modify the estimator so that it is zero beyond the truncation point. One should at this time note that for a truncated density, if one takes a larger h, as suggested, then as may be seen from this example, one will have a bad estimate for a slightly longer interval since this region of bad estimation depends on the size of h.

We thus see that we have an estimator which is asymptotically better than the previous type estimator. We see further that if the proper care is taken a general method is available for choosing the estimator we desire. This estimator should thus consistently lead to better estimates of $f(y)$.

CHAPTER V

USES OF DENSITY ESTIMATION IN MANPOWER RESEARCH

There are many ways that one is able to use density estimation in manpower research. One such instance would be in the question of income distribution in the United States. By taking a relatively large random sample of a thousand or more and assuming a continuous infinite population which we may do because of the large size of the population, an extremely good estimate of all but the very outmost fringes of the probability density function of this population is obtainable. The sample could easily be obtained by some random selection from the information gathered at the various Internal Revenue Centers throughout the United States.

Once such a sample is taken and the density estimate is obtained, then the researcher has the picture which says a thousand words. From the density function the researcher can obtain estimates of probabilities between any points on the curve. Normally he would have to rework his sample if he wanted to compare groups with new sets of endpoints. He can also get a quick idea from a glance what the nature of the spread of the distribution is. He can tell if there are abnormal amounts of rich or poor, if there is a uniform distribution over a wide range or if there is a large grouping about some average.

One could also sample from smaller groups within the United States, for instance, by state, by section of the country, by race, or by sex. From such samples one could obtain density estimates for the various groups. From these density estimates one could obtain the previously discussed information. One could also perform all types of comparative studies. From such comparative curves one could tell where the various percentiles lie in relationship to one another and how the probabilities or the curve overlap. One could also see whether differences in average incomes are caused by abnormally high numbers of rich or poor on one curve or perhaps lack of many rich or poor. Perhaps on the other hand one population is very much like the other in shape, but shifted up or back.

One can do all this with a good estimate curve which can be derived and plotted easily without a great deal of computational difficulty, whereas to keep recycling data in the normal fashion to obtain as much information might take much more time and effort.

Another possible use of the density estimator could be in reduction of bulk data. Rather than try to sort through a great amount of governmental data, a researcher might instead take a random sample of the data and estimate the probability density of the data. In this manner the researcher could get a good view of the overall picture of his data without digging through the entire bulk of the data.

CHAPTER VI

ADDITIONAL TOPICS

1. Confidence Intervals

We saw in Chapter II that under certain conditions

$$\alpha\sqrt{n(n-1)}\ \frac{(G(\hat{\theta}) - \theta)}{\sqrt{\sum_{i=1}^{n} (G_i(\hat{\theta})-G(\hat{\theta}))^2}}$$

converges in distribution to a standard normal as n increases without bound. Unfortunately in our approach to the problem there is not always a set of pseudo-values similar to $G_i(\hat{\theta})$ so this theorem cannot always be used. However, for the estimator

$G(f_n(y,K_1,h_1(n)),f_n(y,K_2,h_2(n)),R)$, where

$$R = \frac{h_1^2(n)}{h_2^2(n)} \cdot \frac{I(K_1,2)}{I(K_2,2)} \qquad \text{and}$$

$$\frac{h_1(n)}{h_2(n)} = c, \text{ a constant for all } n ,$$

we obtain a result similar to that of Parzen [1962]. This result is

$$\frac{G(f_n(y,K_1,h_1(n)),f_n(y,K_2,h_2(n)),R)-E[G(f_n(y,K_1,h_1(n)),f_n(y,K_2,h_2(n),R)]}{\sqrt{Var[G(f_n(y,K_1,h_1(n)),f_n(y,K_2,h_2(n)),R]}}$$

converges in distribution to a standard normal density.

It is perhaps better, however, to consider this problem as if $h_1(n)$ were fixed for all n. Then because the estimator

$$G = G(f_n(y,K_1,h_1(n)),f_n(y,K_2,h_2(n)),R)$$

$$= \frac{1}{1-R} \sum_{i=1}^{n} \left[\frac{K_1\left(\frac{x_i-y}{h_1(n)}\right)}{nh_1(n)} - \frac{RK_2\left(\frac{x_i-y}{h_2(n)}\right)}{nh_2(n)} \right]$$

is an average over a sum of independent identically distributed random variables with finite variance, we may use the central limit theorem and the convergence in probability of

$$S_k^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left[\left(\frac{\frac{K_1\left(\frac{x_i-y}{h_1(n)}\right)}{h_1(n)} - \frac{RK_2\left(\frac{x_i-y}{h_2(n)}\right)}{h_2(n)}}{1-R} \right) - G \right]^2$$

to $\quad \text{Var}\left[\frac{\frac{K_1\left(\frac{x-y}{h_1(n)}\right)}{h_1(n)} - \frac{RK_2\left(\frac{x-y}{h_2(n)}\right)}{h_2(n)}}{1-R} \right]$

to assert that

(5.1) $\quad \sqrt{n}\,\frac{(G - E(G))}{S_k}$ converges in distribution to the standard normal density as n increases without bound.

From this we see that we may obtain an approximate confidence interval for E[G]. This expected value is close to f(y) and thus we have an approximate confidence interval for f(y). We must note that to substitute f(y) for E[G] in (5.1) can cause some problems. For small h_1 $E_x\,[G-f(y)] \cong ch_1^4$ where c is a positive constant. Furthermore the confidence interval is of the form

$$G - \frac{dS_k}{n} \le E[G] \le G + \frac{dS_k}{n}, \quad \text{where}$$

d is a positive constant. The variable S_k converges in probability to

$$\sqrt{\operatorname{Var}\left[\frac{\dfrac{K_1\left(\frac{x-y}{h_1(n)}\right)}{h_1(n)} - \dfrac{RK_2\left(\frac{x-y}{h_2(n)}\right)}{h_2(n)}}{1-R}\right]}, \text{ which for small } h_1(n)$$

is approximately $\sqrt{\dfrac{f(y)b}{h_1(n)} - f^2(y)}$, where b is a positive constant.

Hence we see that the confidence interval becomes approximately

$$G - d\sqrt{\frac{bf(y)}{nh_1(n)} - \frac{f^2(y)}{n}} \le f(y) + ch_1^4(n)f^{(4)}(y)$$

$$\le G + d\sqrt{\frac{bf(y)}{nh_1(n)} - \frac{f^2(y)}{n}} .$$

We then note that our original symmetric interval for the expected value of our estimator would yield an asymmetric confidence interval for f(y) if we knew $f^{(4)}(y)$. But, since we cannot obtain this information, we cannot use the new interval. For this reason one might simply use the original confidence interval replacing E[G] with f(y).

The effects of such a move depend on the size of the correction factor $ch_1^4(n)f^{(4)}(y)$ relative to the length of the entire interval. To make such a comparison we must compare the correction factor $ch_1^4(n)f^{(4)}(y)$ to the length of the interval $2d\sqrt{\frac{bf(y)}{nh_1(n)} - \frac{f^2(y)}{n}}$. Doing this and assuming $f(y) \neq 0$ and $h_1(n) = t/n^p$ $\quad 0 < p < 1$, then for large enough n the following ratio and inequality is obtained:

$$\left| \frac{ch_1^4(n)f^{(4)}(y)}{2d\sqrt{\frac{bf(y)}{nh_1(n)} - \frac{f^2(y)}{n}}} \right| \leq \left| \frac{ch_1^4(n)f^{(4)}(y)}{d\sqrt{\frac{bf(y)}{nh_1(n)}}} \right| \tag{5.2}$$

which is a constant times $n^{-\frac{1}{2}(9p-1)}$.

In this case if $p > 1/9$, (5.2) goes to zero as n increases without bound. Hence in such a case for large n the correction factor should not affect the interval very much and therefore the use of the original interval as a confidence interval for f(y) should be fairly good.

We should note that there are other statistics which could be investigated for their merits in forming confidence intervals. For instance since Var [G] for small h_1 is approximately $\frac{bf(y)}{nh_1}$

then we might substitute this factor for S_k^2 and $f(y)$ for $E[G]$ in (5.1) to obtain

$$(5.3) \qquad \frac{G - f(y)}{\sqrt{\frac{bf(y)}{nh_1}}}$$

as an approximate normal statistic to be used in finding confidence intervals for $f(y)$.

We could also use another estimate for the variance of G, namely

$$(5.4) \qquad S_T^2 = \frac{T}{n-1} \sum_{i=1}^{n} \left(\frac{K\left(\frac{x_i-y}{h_1}\right)}{h_1} - \sum_{i=1}^{n} \frac{K\left(\frac{x_i-y}{h_1}\right)}{nh_1} \right)^2 ,$$

where K is another kernel. This estimator converges in probability to $T\mathrm{Var}\left[\frac{1}{h_1}K\left(\frac{x-y}{h_1}\right)\right]$ which for small h_1 is approximately $T\left(\frac{df(y)}{h_1}\right)$.

Since $\mathrm{Var}\left[\frac{1}{h_1}K\left(\frac{x-y}{h_1}\right)\right]$ is approximately $\frac{df(y)}{h_1}$ for small h_1, then if we choose $T = \frac{c}{d}$,

$$\sqrt{n}\,\frac{(G - f(y))}{S_T}$$

could also be used as an approximate standard normal.

These methods, however, tend to require more and more approximations in forming the final statistic to be used and thus should not be used without further investigation of the effects of these approximations.

2. Application to Empirical Bayes Theory

Another use of these estimators arises in empirical Bayes' theory. Rutherford and Krutchkoff [1969] and Clemmer and Krutchkoff [1968] have shown the need for estimates of $\frac{f'(y)}{f(y)}$, where $f(y)$ is a probability density function, arises when empirical Bayes' techniques are applied. Nadaraya [1965] and later Schuster [1969] have shown that for a large class of kernels K and a large class of densities $f(y)$, $f_n^{(t)}(y,K,h)$ converges with probability one to $f^{(t)}(y)$, where t is a positive integer. Furthermore Maritz [1969] gave conditions under which the simple derivative $f_n'(y,K,h)$ converges in mean-square to $f'(y)$. With this result then for a large class of kernels $\frac{f_n'(y,K,h)}{f_n(y,K,h)}$ converges in probability to $\frac{f'(y)}{f(y)}$.

Thus we have a logical estimator for $\frac{f'(y)}{f(y)}$.

Furthermore if K_1 and K_2 are kernels which fit the conditions of either Schuster [1969] or Maritz [1969] and if $\frac{h_1(n)}{h_2(n)} = c$ for all n

and $$R = \frac{I(K_1,2)}{I(K_2,2)} \frac{h_1^2(n)}{h_2^2(n)} \neq 1$$

then $G(f_n(y,K_1,h_1(n)),f_n(y,K_2,h_2(n)),R)$ has the properties of the original simpler type estimator $f_n(y,K,h)$. Thus

$$(5.5) \qquad \frac{G'(f_n(y,K_1,h_1(n)),f_n(y,K_2,h_2(n)),R)}{G(f_n(y,K_1,h_1(n)),f_n(y,K_2,h_2(n)),R)}$$

could also be used to estimate $\frac{f'(y)}{f(y)}$. Furthermore it might be better since $G(f_n(y,K_1,h_1(n)),f_n(y,K_2,h_2(n)),R)$ is asymptotically better than $f_n(y,K,h)$. However, this has yet to be determined.

One should also note that some kernels such as $K(x) = \frac{\sin^2 x}{x^2}$, which are used quite often to form $f_n(y,K,h)$, do not fit well into the generalized jackknife because $I(K,2)$ does not exist. For this reason and for the reason mentioned in the preceeding paragraph it is recommended that further studies be conducted into this new ratio estimator in general and also the particular estimators that arise from the use of specific kernels. Furthermore, work also should be done to determine whether estimators such as $\frac{f_n'(y,K,h)}{f_n(y,K,h)}$ could be jackknifed directly instead of using the estimator in (5.5)

CHAPTER VII

CONCLUSION

Since, without some knowledge of the form of the density, unbiased estimators of a probability density function f are rare, a method of reducing the bias of a non-parametric estimate of a density function is to be desired. Taking note of a suggestion by Bartlett [1963], estimators of the form introduced by Rosenblatt [1956] and Parzen [1962] have been combined according to the generalized jackknife method introduced by Schucany, Gray, and Owen [1971] to produce an estimator which does reduce this bias considerably.

The various forms of these estimators have been described along with some of their asymptotic properties. The estimators have been compared with the standard type estimator introduced by Parzen [1962] and have been shown to be better asymptotically when using mean-square error as a comparison. Also, examples of small sample values have been presented along with rules for constructing the estimator properly.

It is also shown that the new estimator will often reduce both bias and variance. However, the general rule is that a large reduction of bias is accompanied by an increase in variance and the user has some control of how much of each because of his wide choice of estimators and his control of the h value. Because of these multiple choices, the user sometimes must decide between estimators which have little difference in mean-square error, but differ greatly in variance and bias. Here the user may choose for himself from among the various options.

Further work also needs to be done with these estimators. The many combinations of kernels must be examined for their properties. The use of these estimators in empirical Bayes' techniques should be investigated. The estimators need to be compared with the other non-parametric methods mentioned in Chapter I and also might be compared with some parametric method when information about the general form of the density is known. Lastly large Monte Carlo studies could shed some light on the utility of these new point and interval estimators.

LIST OF REFERENCES

[1] Bartlett, M.S. (1963). "Statistical estimation of density functions." Sankhyā, 25A, pp. 245-254.

[2] Bovena, L., Kendall, D.G. and Stepanov, I. (1971). "Spline transformations: three new diagnostic aids for the statistical data-analyst." Journal of the Royal Statistical Society, Series B, 33, pp. 1-71.

[3] Bhattacharya, P.K. (1967). "Estimation of a probability density function and its derivatives." Sankhyā, 29A, pp. 373-383.

[4] Cacoullis, T. (1966). "Estimation of a multivariate density." Annals of the Institute of Statistical Mathematics, 18, pp. 179-190.

[5] Clemmer, B.A. and Krutchkoff, R.G. (1968). "Use of empirical Bayes estimators in a linear regression model." Biometrika, 55, pp. 525-534.

[6] Epanechnikov, V.A. (1969). "Non-parametric estimation of a multivariate probability density." Theory of Probability and its Applications, 10, pp. 186-190.

[7] Good, I.J. and Gaskins, R.A. (1971). "Non-parametric roughness penalities for probability densities." Biometrika, 58, pp. 255-277.

[8] Ghurye, S.G. and Olkin, I. (1969). "Unbiased estimation of some multivariate probability densities." Annals of Mathematical Statistics, 40, pp. 1261-1271.

[9] Gray, H.L. and Schucany, W.R. (1972). The Generalized Jackknife Statistic. Marcel Decker, New York.

[10] Loftsgaarden, D.O. and Quesenberry, C.P. (1965). "A non-parametric estimate of a multivariate density function." Annals of Mathematical Statistics, 36, pp. 1049-1051.

[11] Martz Jr., H.F. (1969). "A note on estimating the derivative of a density function with empirical Bayes applications." Proceedings of the Symposium on Empirical Bayes Estimation and Computing in Statistics. Texas Tech University Mathematics Series No. 6, pp. 141-149.

[12] Mood, A.M. and Graybill, F.A. (1963). Introduction to the Theory of Statistics. 2nd Edition, McGraw-Hill, New York.

[13] Moore, D.S. and Henrichon, E.G. (1969). "Uniform consistency of some estimates of a density function." Annals of Mathematical Statistics, 40, pp. 1499-1502.

[14] Murthy, V.K. (1965). "Estimation of a probability density." Annals of Mathematical Statistics, 36, pp. 1027-1031.

[15] Nadaraya, E.A. (1965). "On non-parametric estimates of density functions and regression curves." Theory of Probability and its Applications, 10, pp. 186-190.

[16] Olmsted, John M.A. (1966). Advanced Calculus. Appleton-Century-Crofts, New York.

[17] Parzen, E. (1962). "On estimation of a probability density function and its mode." Annals of Mathematical Statistics, 33, pp. 1065-1076.

[18] Pickands III, J. (1969). "Efficient estimation of a probability density function." Annals of Mathematical Statistics, 40, pp. 854-864.

[19] Quenouille, M. (1965). "Notes on bias in estimation." Biometrika, 43, pp. 353-360.

[20] Rosenblatt, M. (1956). "Remarks on some non-parametric estimates of a density function." Annals of Mathematical Statistics, 27, pp. 832-837.

[21] Rutherford, J.R. and Krutchkoff, R.G. (1969). "Some empirical Bayes techniques in point estimation." Biometrika, 56, pp. 133-137.

[22] Schucany, W.R., Gray, H.L., and Owen, D.B. (1971). "On bias reduction in estimation." Journal of the American Statistical Association, 66, pp. 524-533.

[23] Schuster, E.F. (1969). "Estimation of a probability density function and its derivatives." Annals of Mathematical Statistics, 40, pp. 1187-1195.

[24] Seheult, A.H. and Quesenberry, C.P. (1971). "On unbiased estimation of density functions." Annals of Mathematical Statistics, 42, pp. 1434-1438.

[25] Specht, D.F. (1971). "Series estimation of a probability density function." Technometrics, 13, pp. 409-424.

[26] Taylor, A.E. (1965). General Theory of Functions and Integration. Blaisdell, New York.

[27] Tukey, J.W. (1958). "Bias and confidence in not quite large samples." (abstract) Annals of Mathematical Statistics, 29, p. 614.

[28] Van Ryzin, J. (1969). "On strong consistency of density estimates." Annals of Mathematical Statistics, 40, pp. 1765-1772.

[29] Watson, G.S. and Leadbetter, M.R. (1963). "On the estimation of a probability density function I." Annals of Mathematical Statistics, 34, pp. 480-491.

[30] Whittle, P. (1958). "On the smoothing of probability density functions." Journal of the Royal Statistical Society, Series B, 20, pp. 334-343.

[31] Woodroofe, M. (1970). "On choosing a delta-sequence." Annals of Mathematical Statistics, 41, pp. 1665-1671.

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)* SOUTHERN METHODIST UNIVERSITY	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b. GROUP UNCLASSIFIED

3. REPORT TITLE

Improved Density Estimation

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

Technical Report

5. AUTHOR(S) *(First name, middle initial, last name)*

John P. Sommers

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
February 28, 1972	90	31

8a. CONTRACT OR GRANT NO. N00014-68-A-0515 b. PROJECT NO. NR 042-260 c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) 114 9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited. Reproduction in whole or in part is permitted for any purpose of the United States Government.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	OFFICE OF NAVAL RESEARCH

13. ABSTRACT

Non-parametric estimation of a continuous probability density function almost always leads to a biased estimator. The purpose of this paper is to attack the problem of bias reduction. The problem is approached by using combinations of estimators of the form studied by Parzen [1962]. Combining more than one of these estimators by the jackknife method of Schucany, Gray, and Owen [1971], new estimators are formed which generally have a substantial decrease in bias.

This paper studies the properties of these new estimators in detail. Approximations are derived for their variance and bias. General classes of these new estimators are shown to be asymptotically unbiased and mean square consistent. Furthermore, the estimators are shown to be asymptotically better than the original estimators using mean square error as a criterion.

Examples and applications are given and small sample properties studied. Rules are developed so that one can choose estimators of this new form which have less bias and less variance than a comparable original type estimator. A rule is also discussed which will improve the estimation of densities with truncation points.

Finally approximate confidence intervals for the value of a density at a point are derived and studied. Possible uses in empirical Bayes' estimation of $f'(y)/f(y)$ are recommended for further study along with several other topics.

Parzen, E. (1962). "On estimation of a probability density function & its mode." Annals of Math. Stat., 33, pp. 1065-1076.

Schucany, Gray & Owen (1971). "On Bias reduction in estimation." JASA, 66, pp. 524-533