BIAS REDUCTION OF KERNEL DENSITY ESTIMATES

Jagdish S. Rustagi *

The Ohio State University and University of the Philippines

and

Svetlana Dynin **

The Ohio State University

Abstract

Among the many estimates available for the probability density function, the kernel type estimates are quite popular. Jackknife technique is used to reduce bias of these estimates. It is shown that the jackknife estimate has the same properties as the original estimates for certain well-behaved kernels. A Berry-Essen type central limit theorem is also given for these estimates.

1. Introduction

Jackknifing techniques are increasingly being applied to data analysis for bias reduction. They are used in many statistical contexts, such as robust estimation and density estimation. Often the asymptotic properties of jackknifed estimates turn out to be the same as the original estimates. Also, jackknifing is related to Efron's bootstrap technique which has found applications in many statistical settings. Kernel type estimates of the probability density functions have been studied by several authors. Bias reduction in kernel density estimate has been studied through combination of estimates by Schucany and Sommers (1977) using different density estimates in such a case may become negative.

Using jackknifing technique on the kernel density estimates we can reduce the bias of the estimates. In this paper, we define pseudovalues for kernel density estimates and study the jackknife properties of their jackknife estimates. It is shown that the jackknifed estimates have the same asymptotic properties as the original estimates for certain well-behaved kernels. A Berry-Esseen type central limit theorem is also given for the jackknifed estimates.

Application of jackknife technique has been made to the hazard function estimates for bias reduction in a paper by the authors (1989). The technique has also

---

Keywords and Phrases: nonparametric density estimates, jackknifing for bias reduction, asymptotic properties.

* Supported in part by Contract No. N00014-82-K-0291 (NR 042-403) by Office of Naval Research. Reproduction in whole or in part is permitted for any purpose of the United States Government.

** Based on part of doctoral dissertation submitted to the State University of New York, Stonybrook.

been applied to data on tumors in breast cancer, Rustagi and Dynin (1989).

2. Pseudovalues

Let $X_1, X_2, \ldots, X_n$ be a random sample from a population with cumulative distribution function $F(x)$ and probability density function $f(x)$. Let $K(x)$ be a given kernel function with the following properties:

(i) $\sup |K(x)| < \infty$

(ii) $\int K(x)\,dx = 1$

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

(iv) $\int_{-\infty}^{\infty} x^i K(x)\,dx = 0, i = 1,2,.., r-1,$

$\int_{-\infty}^{\infty} x^r K(x)dx \neq 0,$

and $\int_{-\infty}^{\infty} |x^r K(x)|\,dx < \infty$

Let $F_n(x)$ be the empirical distribution function based on the random sample, and let $\{h_n\}$ be a sequence of constants. Then the kernel density estimates of $f(x)$, are given by Rosenblatt (1956) and Parzen (1962)

$$f_{nh_n}(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right) = \frac{1}{h_n}\int_{-\infty}^{\infty} K\left(\frac{x-y}{h_n}\right) dF_n(y) \qquad (2.1)$$

Note that

$$E[f_n h_n(x)] = \frac{1}{h_n}\int K\left(\frac{x-y}{h_n}\right) dF(y) \text{ with}$$

$E[f_n h_n(x)] \to f(x)$ as $n \to \infty$ and $h_n \to 0$. That is the kernel estimate is asymptotically unbiased.

The variance of $f_n h_n(x)$ also approaches zero if in addition we assume that $n h_n \to \infty$, see for example, Tapia and Thompson (1978).

Let $F_{n,i}(x)$ be the empirical distribution function of the random sample $X_1, \ldots, X_n$ with the observation X_i removed,

Supposed we denote by

$$f^i_{n-1h_{n-1}}(x) = \frac{1}{h_{n-1}}\int K\left(\frac{x-y}{h_{n-1}}\right) dF_{n,i}(y) \qquad (2.2)$$

where h_{n-1} are constants based on n-1 observations. The notation h_{n-1} does not mean that it is the previous value to h_n, rather h_n, h_{n-1} are functions of n.

We define the pseudovalues as follows:

$$f_s^i(x) = \frac{h_n^{-r}}{h_n^{-r} - h_{n-1}^{-r}} f_{nh_n}(x) - \frac{h_{n-1}^{-r}}{h_n^{-r} - h_{n-1}^{-r}} f^i_{n-1h_{n-1}}(x) \qquad (2.3)$$

The jackknifed estimated of the probability density function is then defined by f_J given by,

$$f_J(x) = \frac{1}{n} \sum f_s^i(x) = \gamma f_n h_n(x) + (1-\gamma) f_{n-1h_{n-1}}(x) \qquad (2.4)$$

where $\gamma = \dfrac{h_n^{-r}}{h_n^{-r} - h_{n-1}^{-r}}$,

and $f_{n-1}h_{n-1}(x)$ is the average of the quantity defined in (2.2)

That is, the jackknife estimate is convex linear function of the classical estimate based on n observations and an average of estimates based on n-1 observations. The generalized jackknife estimate given by Schucany and Sommers (1977) for the density is linear combination of the density estimates based on two different kernels, K_1 and K_2. If we assume, $K_1, = K_2 = K$, their estimate (3.2) p. 421) reduces to the above estimate with appropriate adjustment of the bandwith, h_n.

3. Properties of Jackknifed Estimates of the Density

Bias reduction in the estimates of the density by using not necessarily positive kernels, has been demonstrated by several authors. For sufficiently smooth probability density functions, it is always possible to reduce the bias by choosing an appropriate kernel K. Among the class of non-negative K's A(iv) can be achieved only for $r=2$, giving $n^{4/5}$ as the best possible error rate. For better results, one has to also include those k's for which K(y) is negative, leading to a negative estimate of probability density function for some n and h_n and at some point x.

From now on, we shall assume that kernel K statisfies the following additional properties:

(v) The r^{th} derivative of density function satisfies a Lipschitz condition,

$$| f^{(r)}(x) - f^{(r)}(y) | < c | x-y |^{\alpha}, 0 \le \alpha \le 1$$

for all x and y,

(vi) $\int | x^{r+\alpha} K(x) | dx < \infty$, and

(vii) $\{h_n\}$ is a sequence of constants such that

$$\frac{h_n}{h_{n-1}} = 1 + o(1), h_n \to 0, nh_n \to \infty.$$

The following result gives the form of bias of the estimate (2.1).

Theorem 3.1 Under the conditions (i) - (vi),

$$\text{(i) Bias } [f_{nh_n}(x)] = h_n^r f^{(\delta)}(x) \int_{-\infty}^{\infty} z^r K(-z) dz / r! + 0(h_n^{r+\alpha}) \qquad (3.1)$$

$$\text{(ii) Bias } (f_J) = 0\,(h_{n-1}^{r-1}(h_n - h_{n-1}) h_n^{r-1} (h_{n-1}^r - h_n^r)^{-1} [\max(h_n, h_{n-1})]^{1+\alpha} \qquad (3.2)$$

Further, if (vii) holds, Bias $(f_J) = 0(h_n^{r+\alpha})$.

Proof: (3.1) is well known. For (3.2), we note that

$$E(f_{n-1h_{n-1}}{}^{i}(x)] = f(x) + h_{n-1}^{r-1} \int K(-z) \int \frac{(z-u)^{r-2}}{(r-2)!} f^{(r-1)}(x + h_{n-1}u) du dz$$

by using Taylor's expansion. Hence

$$\text{Bias } f_J(x) = \int K(-z) \int_0^z \frac{(z-u)^{r-2}}{(r-2)!} [\alpha h_n^{r-1} f^{(r-1)}(x + h_n u) - (1-\alpha) h_{n-1}^{r-1} f^{(r-1)}(x + h_{n-1}u)] du dz$$

$$= \int K(-z) \int_0^z \frac{(z-\mu)^{r-2}}{(r-2)!} \alpha h_n^{r-1} [f^{(r-1)}(x + h_n u) - f^{(r-1)}(x + h_{n-1}u)]$$

$$+ [\alpha h_n^{r-1} - (1-\alpha) h_{n-1}^{r-1}] f^{(r-1)}(x + h_{n-1}u)\} dz du$$

$$= A + B, \text{ say}$$

$$\text{Now } A = \int_{-\infty}^{\infty} K(-z) \int_0^z \frac{(z-\mu)^{r-2}}{(r-2)!} (\alpha h_n^{r-1}) \int_{x+h_{n-1}u}^{x+h_n u} f^r(v) \, dv du dz$$

and we may write

$$\int_{x+h_{n-1}u}^{x+h_n u} f^{(r)}(v) dv = \int_{x+h_{n-1}u}^{x+h_n u} \{[f^{(r)}(v) - f^{(r)}(x)] + f(r)(x)\} dv$$

$$= C + D, \text{ say}$$

We have here,

$$|C| \le \frac{1}{r!}(h_{n-1}-h_n)\int_{-\infty}^{\infty} |K(-z)|\, |z^r|\, \sup |f^{(r)}(v)-f^{(r)}(x)|\, dz$$

where supremum is taken for

$|v-x| \le \max(h_n, h_{n-1}|z|)$, so that

$$|C| = [\frac{h_{n-1}-h_n}{h_{n-1}^r-h_n^r} h_n^{r-1}h_{n-1}^{r-1}(\max(h_n, h_{n-1}))^{1+\alpha}]$$

$= 0(h_n^{r+\alpha})$ if (vii) is satisfied.

Similarly for B and D. Hence the theorem follows.

The following theorem provides the connection between the bias of jackknife estimate and that of the kernel estimate and is stated here without proof.

Theorem 3.2 Under the conditions (i) - (iv), and

$\int ||x|^{r+1}|K(x)|\,dx < \infty$ for any Lebesgue point of $f^{(r+1)}$.

$$\text{Bias}(f_J) = \frac{h_{n-1}rh_n^r(h_n-h_{n-1})}{(h_{n-1}^r-h_n^r)(r+1)!} f(x)^{(r+1)}\int K(-z)z^{r+1}dz.$$

Further, if (vii) is satisfied and K is differentiable, then

Bias $(f_J) \sim$ Bias (f_{nh})

with kernel $K_o = [zK'(z) + (r+1)k(z)]r^{-1}$, provided that K_o is square integrable.

Collorary: If K is symmetric on the interval (-a, a) (r even), $\int |x|^{r+2}K(x)\,dx < \infty$ and x is Lebesgue point for $f^{(r+2)}(x)$, then

$$\text{bias}(f_J) = \frac{h_{n-1}^r h_n^r(h_n-h_{n-1})}{(h_{n-1}^r-h_n^r)(r+2)!} f^{(r+2)}(x)\, K(-z)\, z^{r+2}$$

Now we obtain expression for the variance of the jackknife estimate,

Variance of $f_{J(x)}$.

Let σ_J^2 denote the variance of $f_J(x)$. Then

$$\sigma_J^2 = n^{-1}(h_n^{-r}-h_{n-1}^{-r})^{-2}\,\text{Var}\{h_n^{-r-1}K(\frac{x-Y}{h_n}) - h_{n-1}^{-r-1}K(\frac{x-Y}{h_{n-1}})\}$$

= A - B, say

where

$$A = n^{-1}(h_n^{-r}-h_{n-1}^{-r})^{-2}\int [h_n^{-r-1}K(\frac{x-Y}{h_n}) - h_{n-1}^{-r-1}K(\frac{x-Y}{h_{n-1}})]^2 f(y)dy$$

$$B = n^{-1}(h_n^{-r}-h_n^{-1-r})^{-2}\{\int [h_n^{-r-1}K(\frac{x-Y}{h_n}) - h_{n-1}^{-r-1}K(\frac{x-Y}{h_{n-1}})] f(y)\,dy\}^2$$

Notice that with $z = (x-y)h_n^{-1}$, we have

$$A = n^{-1}(h_n^{-r}-h_{n-1}^{-r})^{-2}h_n^{-2r-1}\int_{-\infty}^{\infty}[K(z) - (\frac{h_n}{h_{n-1}})^{r+1}K(z\frac{h_n}{h_{n-1}})]^2 f(x-zh_n)dz$$

$$= (nh_n)^{-1}(h_n^{-r}-h_{n-1}^{-r})^{-2}h_n^{-2r}(1-\frac{h_n}{h_{n-1}})^2 \int_{-\infty}^{\infty}\{[\frac{K(z)-K(z\frac{h_n}{h_{n-1}})}{1-\frac{h_n}{h_{n-1}}}] - \frac{K(z\frac{h_n}{h_{n-1}})}{1-\frac{h_n}{h_{n-1}}}[(\frac{h_n}{h_{n-1}})^{r+1}-1]\}^2 f(x-zh_n)dz.$$

In the limit when $\frac{h_n}{h_{n-1}} \to 1$, $h\,h_n \to \infty$, $h_n \to 0$ as

$n \to \infty$, we have

$$\lim_{n \to \infty} (n\, h_n A) = \frac{f(x)}{r^2} \int_{-\infty}^{\infty} \{z K'(z) + K(z)(r+1)\}^2 dz$$

so that $A = \frac{f(x)}{n\, h_n r^2} \int_{-\infty}^{\infty} (zK'(z) + K(z)(r+1)^2 dz + 0\left(\frac{1}{n\, h_n}\right)$,

where we used the conditions that $\int [zk'(z)]^2 dz < \infty$ and

$$\int_m^{\infty} K2(z) = 0(m^{-2})$$

Using results of theorem (3.2) we get

$$B = \frac{1}{n}[f(x)] + 0(h_n^{r+\alpha}).$$

Therefore, as $n \to \infty$, we have

$$\sigma_J^2 = \frac{f(x)}{nh_n^r} \int_{-\infty}^{\infty} \frac{\{zk'(z)+(r+1)K(z)\}^2 dz}{r} + 0\left(\frac{1}{n\, h_n}\right)$$

Notice that $\sigma_J^2 > 0$ since $[zK'(z) + (r+1)K(z)] > 0$ for all integrable functions $K'(z)$. If $zK'(z) + (r+1)K(z) = 0$, then $K(z) = z^{-(r+1)}$ which is not integrable.

If $\frac{h_n}{h_{n-1}} \to \beta$, $n \to \infty$, $\beta \neq 1$, then $\sigma_J^2 = \frac{f(x)}{nh_n(1-\beta r)^2}$

$$\int [K(z) - \beta^{r+1} K(z^{\beta})]^2 dz.$$

We see that if $\lim_{n \to \infty} h_n/h_{n-1}$ exists and does not depend on x, then MSE f_J is approximately equal to MSE f_{nhn} with kernel $k_{o\beta}$ with

$$K_{o\beta} = \frac{K(z) - \beta^{r+1} K(\beta z)}{1 - \beta^r}. \quad (3.3)$$

Observe that if $h_n/h_{n-1} \to \beta = 1$, then $K_{o\beta} \to K_o$.

Theorem 3.3. Let $f_{nhn} K_{o\beta}$ be a kernel density estimate with kernel $K_{o\beta}$ as defined in (3.3). Then

$$\min_{hn} \min_{hn-1} MSEf_J = \min_{hn} \min_{\beta} MSEf_J. \quad (3.4)$$

When $n \to \infty$, (3.4) equals $\min_{\beta} \min_{hn}$ MSE f_{nhn}, K_{ob} approximately.

Remarks:

(i) For the kernel $K = 1$, $0 \leq x \leq 1$, the minimizing parameter β is given by some root of

$$\beta^3 - 4\beta^2 + \beta - 1 = 0.$$

For the kernel $K = e^{-x}$, $x \geq 0$, the minimizing parameter $\beta = 1$

(ii) If h_n is larger than the optimal h_n for minimizing MSE of f_{nhn}, $K_{o\beta}$, then we can find another sequence of constants h'_n such that

$$MSE(f_J)/MSE\, f_{nhn} K_{o\beta} \to 0$$

as $n \to \infty$. Note that f_J is computed with the help of h_n and h'_n.

(iii) Note that the jackknife estimate is not asymptotically a kernel estimate except when

$$\lim_{n \to \infty} h_n/h_{n-1}$$

exists.

Proof: If $\int K(-z) z^{r+1} dz \neq 0$ and $f^{r+1}(x) \neq 0$, then

$$\min_{hn} \min_{\beta} MSEf_J = \min_{hn} \min_{\beta} \left[\frac{f(x)}{nh_n} \int K_{o\beta}^2(z) dz + \right.$$

$$h_n^{2r+1} \frac{f^{(r+1)}(x)}{(r+1)!} \int [K(-z)z^{r+1}dz)^2)]$$

$$= n^{-\frac{2r+2}{2r+3}} (2r+3)(2r+2)^{-\frac{2r}{2r+3}}$$

$$\min_{\beta} \left[\frac{1-\beta}{b(1-br)}\right]^{2/(2r+3)} \left[f(x) \int K_{o\beta}^2 dz\right]^{(2r+2)/(2r+3)}$$

$$\left[\frac{f^{(r+1)}(x)}{(r+1)!} \int K(-z)\, z^{r+1}dz\right]^{2/(2r+3)}.$$

Minimization of $\int K_{o\beta}^2(z)dz$ with respect to β occurs for $\beta = 0$. The right hand side becomes

$$\min_{\beta} \min_{h_n} MSE\ f_{nh_n} K_{o\beta}$$

Hence the variance of f_J is minimized for $\beta = 0$.

4. Berry-Esseen bounds for Jackknife Estimates

The jackknife estimate of the probability density function is the sum of n independent but not necessarily identically distributed random variables. Using the function k_1 defined below the estimate is given by

$$f_J(x) = \frac{1}{n} \sum_{i=1}^{n} K_1(x-X_i) \qquad (4.1)$$

where

$$K_1(z) = (h_n^{-r} - h_{n-1}^{-r})^{-1}[h_n^{-r}K(\frac{z}{h_n}) - h_{n-1}^{-r-1} K(\frac{z}{h_{n-1}})].$$

Assuming that

(i) K_1 is square integrable, and

(ii) $(2+\delta)$-th power of K_1 is integrable for some > 0 we can find Berry-Essen bound for g(y) where

$$g(y) = P\{\frac{f_J(x) - E[f_J(x)]}{\sqrt{Var\, f_J(x)}} \leq y\} - \Phi(y),$$

$\Phi(y)$ is the distribution function of the standard normal random variable. We use the following notation:

$$\sigma^2 = n^{-2} \sum_{i=1}^{n} E\{K_1(x-X_i) - E[K_1(x-X_i)]^2$$

$$\mu_{2+\delta} = n^{-2-\delta} \sum_{i=1}^{n} E \mid K_1(x-Xi) - E[K_1(x-X_i)] \mid^{2+\delta}$$

The Berry-Esseen bound is given by

$$\sup_{-\infty\, y\, \infty} \mid g(y) \mid \leq c_o \frac{\mu_{2+d}}{\sigma^{2+\delta}} \qquad (4.2)$$

where c_o is the universal constant, for reference, see Loeve (1955). This result provides also the asymptotic normality for the jackknife estimate if

$\frac{\mu_{2+\delta}}{\sigma^{2+\delta}} \to 0$ as $n \to \infty$. The main result is stated in the following theorem.

Theorem 4.1

Let assumptions (i) - (vii) be satisfied. Further, some $\delta > 0$, assume that

(a) $\int \mid z \mid^{2+\delta} K'(z)^{2+\delta} dz < \infty,$

(b) $\int_m^{\infty} \mid K^{2+\delta} dz = 0(\frac{1}{m^{2+\delta}})$ as $m \to \infty$,

(c) $\int_m^{\infty} K^2(z)dz = 0(\frac{1}{m^2})$ as $m \to \infty$,

(d) f(x) is continuous at x,

(e) $h_n \to 0$, $nh_n \to \infty$, and $\frac{h_{n+1}}{h_n} \to 1$ as $n \to \infty$

Then there exists a universal constant $c\delta$ such that

$$\sup_{-\infty < y < \infty} \left| \left\{ \frac{[f_J(x)-E[f_J(x))]}{\sqrt{\mathrm{Var}\, f_J(x)}} \leq y \right\} \Phi - (y) \right| \leq$$

$$c\delta \frac{2^{2+\delta} \int | zK'(z) + (r+1) K(z) |^{2+\delta} dz}{\{nh_n f(x)\}^{\delta/2} \int [zK'(z) + (r+1)(K(z)^2)dz]^{\delta/2+1}}$$

The proof of the theorem follows from that of Loeve (1955) p. 55.

References

Dynin, S. (1985) Nonparametric probability density estimator: Use of negative kernels and jackknifing to reduce bias, Doctoral dissertation submitted to State University of New York, Stonybrook, New York, pp 1-103.

Rustagi, J. S. and Dynin, S. (1989). Statistical Analysis of Tumors in Breast Cancer, *Statistics in Medicine* (T. Krishnan, Editor). Himalya Publishing House, Bombay p. 52-63.

Rustagi, J. S. and Dynin, S. (1989) Jackknifing Hazard Function estimates, *Proceedings of the Conference in Recent Developments in Statistics and their applications*, (Editors, John Klein and Jae C. Lee). Korea University, Seoul.