The Philippine Statistician
1990, Vol. 39, Nos. 1-2 & 3-4
pages 71-81

ON ELIMINATING THE BIAS OF KERNEL DENSITY ESTIMATORS

by

Walfredo R. Javier
Jose Ramon Albert
DLSU, Math Department

Abstract: This paper shows that there exists a kernel function K and a trimming proportion α^* for the trimmed Jacknifed Kernel Density Estimators that will render the kernel density estimators unbiased. Practical ways are suggested for implementing the result of the paper.

Keywords and phrases: kernel density estimates, bias of estimates, trimmed Jacknifed kernel density estimator, statistical functional.

I. INTRODUCTION

An estimate of the probability density function f(x) of a distribution F is an important input in most data analysis procedures. In applications, the kernel density method of density estimation is extensively used. The estimates obtained, using non-negative kernels, however, turn out to be biased. The method of Jackknifing can be used to reduce the bias. Robust methods are preferable, in as much as in practice, data may contain outliers and also, the pseudo-values in jackknifing procedures are fat-tailed in distribution and are not independent. A simple alternative to make it robust was developed by Hinkley and Wang (1980). The method consists in taking the trimmed average of the pseudo-values.

II. KERNEL DENSITY ESTIMATORS

Given an iid sample Y_1, Y_2, ..., Y_n, it is often necessary to estimate the density f(x) at a point x of the distribution F which generated the random sample. Kernel density estimators, denoted by $\hat{f}_n(x)$, form a class of estimators used for this task; they are defined by

$$\hat{f}_n(x) = \frac{1}{n\ h_n} \sum_{i=1}^{n} K\left[\frac{x - Y_i}{h_n}\right] \qquad (2.1)$$

where n is the sample size;

h_n is the window or band width and is chosen in an optimal manner in the sense of integrated mean square error. It is a function of the distribution F; hence we can write $h_n = h_n(F)$; and

K(.) is the specified kernel function which satisfies:

(i) $\sup \mid k(x) \mid < \infty$; (2.2)

(ii) $\int k(x)\, dx = 1$; (2.3)

(iii) $\lim_{x \to \infty} \mid x\, k(x) \mid = 0$ (2.4)

(iv) $\exists\ r \in \mathbb{Z}^+ \ni$ (2.5)

$\int x^i\, k(x)\, dx = 0$ for $i = 1, 2, \ldots, r-1$;

$\int x^r\, k(x)\, dx \neq 0$; and

$\int \mid x^r\, k(x) \mid dx < \infty$.

Remark: The usual choices for K(.) are probability density functions like the normal distribution.

III. KERNEL DENSITY ESTIMATORS AS STATISTICAL FUNCTIONALS

Let F_n denote the empirical distribution function determined by the sample $Y_1, Y_2, \ldots, Y_n$ thus

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \delta\,(x - Y_i)$$

$$\text{where } \delta\,(t) = \begin{cases} 0 & \text{if } t < 0 \\ 1 & \text{if } t \geq 0 \end{cases}$$

We can rewrite (2.1) as

$$\hat{f}_n(x) = \int \frac{1}{h_n} K\left[\frac{x - y}{h_n} \right] dF_n(y) \tag{3.1}$$

This shows that $f_n(x)$ depends on the data $Y_1, Y_2, \ldots, Y_n$ only through the distribution function F_n; thus $f_n(x)$ is a functional of the form $T(F_n)$. Rustagi, Javier and Victoria (1989) derived its influence curve as

$$t_1(F_n;x,y) = \frac{1}{h_n} K\left[\frac{x - y}{h_n} \right] + \frac{1}{b} K\left[\frac{x - y}{b} \right] \tag{3.2}$$

where $b = \sup\{h_n\}$. At the underlying distribution F, the influence curve is

$$t_1(F;x,y) = \frac{1}{b} K\left[\frac{x - y}{b} \right] \tag{3.3}$$

IV. CHARACTERIZATION OF A TRIMMED JACKNIFE ESTIMATOR AS A STATISTICAL FUNCTIONAL

Following Hinkley and Wang (1980), let the statistical functional $T(F) \equiv f(x)$ be defined by $T(F_n) = \hat{f}_n(x)$. Denote by $F_{n,-j}$ the empirical distribution based on the observations $Y_1, Y_2, \ldots, Y_n$ with the jth observation Y_j removed. Then the jth pseudovalue $P_{n,-j}$ is defined by

$$P_{n-j} \equiv n\,T(F_n) - (n-1)\,T(F_{n,-j})$$
$$= T(F_n) + (n-1)\,[\,T(F_n) - T(F_{n,-j})\,] \qquad (4.1)$$

for $j = 1, 2, \ldots, n$.

Let these pseudo-values be arranged as ordered statistics

$$P_{(n,-1)} \le P_{(n,-2)} \le \ldots \le P_{(n,-n)}$$

For a given trimming proportion α (say $\alpha = 0.05$), we remove the first α smallest and the α largest pseudo-values and average the remaining $n(1-2\alpha)$ pseudo-values. This average is the α-trimmed jacknifed estimate for $T(F) = f(x)$, given by

$$T_{n,\alpha} \equiv \frac{1}{n\,(1-2\alpha)} \sum_{j=r_\alpha+1}^{n-r_\alpha} P_{(n,-j)} \qquad (4.2)$$

where $r_\alpha = [\![\, n\alpha \,]\!]$.

Using a Taylor's expansion for $T(F_n)$ about the parent distribution F, Hinley and Wang (1980) evaluated the expression inside the square bracket in (4.1), giving a characterization of $T_{n,\alpha}$ in terms of its first-order Von Mises derivative:

$$T_{n,\alpha} = t^{\alpha}(F) + \frac{1}{n}\sum_{j=1}^{n} t_1^{\alpha}(F;x_j) + O_p(1/n) \qquad (4.3)$$

where

$$t^{\alpha}(F) \equiv \frac{1}{1-2\alpha}\int_{\alpha}^{1-\alpha} L^{-1}(\mu)\,d\mu \; ; \qquad (4.4)$$

$$L(z) \equiv \Pr\Big[\, T(F) + t_1(F;Y) \le z \,\Big] \; ; \qquad (4.5)$$

$$t_1^{\alpha}(F;x) \equiv \frac{d}{dt}\,[\, T(F + t(\delta_x - F)) - T(F)\,]\Big|_{t=0} \qquad (4.6)$$

The function defined by (4.6) is the first-order Von Mises derivative, also called the influence curve, of the statistical functional T(F), (Hampel, 1974) where δ_x is the degenerate distribution function with mass concentrated at x.

$$t_1^{\alpha}(F;x) \equiv [t_1(F;x)]_{\alpha}^{1-\alpha} + E[t_2(F;x;Y)] \tag{4.7}$$

$t_2(F;x,Y)$ is the second order Von Mises derivative; it is zero for kerneal density estimators and finally $[t_1(F;x)]_{\alpha}^{1-\alpha}$ is defined by

$$(1 - 2\alpha)[t_1(F;x)]_{\alpha}^{1-\alpha} \equiv \begin{cases} L^{-1}(1-\alpha) & \text{if } t_1(F;x) > L^{1}(1-\alpha) \\ t_1(F;x) & \text{if } L^{-1}(\alpha) \le t_1(F;x) \le L^{-1}(1-\alpha) \\ L^{-1}(\alpha) & \text{if } t_1(F;x) < L^{-1}(\alpha) \end{cases} \tag{4.8}$$

V. THE BIAS FUNCTION

Using equations (4.3) - (4.8) with T(F) and $t_1(F;x)$ replaced by f(y) and $\left[\frac{1}{b} K\left(\frac{x-y}{b} \right) - m(x) \right]$ respectively with

$$m(x) \equiv E\left[\frac{1}{b} K\left(\frac{x-Y}{b} \right) \right] = \int_{-\infty}^{\infty} \frac{1}{b} K\left(\frac{x-y}{b} \right) dF(y) \tag{5.1}$$

then the bias of the trimmed-Jackknifed kernel density estimator $T_{n,\alpha}$ may be written as :

$$B(\alpha;x) \equiv \frac{1}{1-2\alpha} \left\{ \int_{y_\alpha}^{y_{1-\alpha}} \left[\frac{1}{b} K\left(\frac{x-y}{b} \right) - m(x) \right] dF(y) + \right.$$

$$\int_{L^{-1}(\alpha)}^{L^{-1}(1-\alpha)} \left[\frac{1}{b} K\left(\frac{x-y}{b} \right) - m(x) \right] dF(y) +$$

$$L^{-1}(1-\alpha) \;\; L\left(L^{-1}(1-\alpha) + f(x) \right) +$$

$$L^{-1}(\alpha) \quad L\left[L^{-1}(\alpha) + f(x) \right] \Big\} + O_p(1/n) \tag{5.2}$$

where y_α is the α^{th} percentile of F.

The bias function $B(\alpha;x)$ is a continuous function of α, where $0 \le \alpha \le 1/2$.

VI. BEHAVIOR OF BIAS FUNCTION NEAR $\alpha = 1/2$.

We shall now investigate how $B(\alpha,x)$ behaves near $\alpha = 1/2$. For this purpose, we shall compute the limit $\lim_{\alpha \to 1/2^-} B(\alpha;x)$.

Assuming L to be a symmetric distribution, then we can rewrite (4.10) as

$$B(\alpha;x) = \frac{y_{1-\alpha} - y_\alpha}{1 - 2\alpha} \left[\frac{1}{b} K\left(\frac{x - y^*}{b} \right) - m(x) \right] f(y^*) +$$

$$\frac{L^{-1}(1-\alpha) - L^{-1}(\alpha)}{1 - 2\alpha} \left[\frac{1}{b} K\left(\frac{x - y^{**}}{b} \right) - m(x) \right] f(y^{**}) +$$

$$\frac{L^{-1}(1-\alpha)}{1 - 2\alpha} \left[L\left[L^{-1}(1-\alpha) + f(x) \right] - L\left[L^{-1}(\alpha) + f(x) \right] \right] +$$

$$O_p(1/n) , \qquad \begin{array}{c} y_\alpha \le y^* \le y_{1-\alpha} \\ L^{-1}(\alpha) \le y^{**} \le L^{-1}(1-\alpha) \end{array} \tag{6.1}$$

using the mean-value theorem of integral calculus. Also, assuming that the unknown density function $f(y)$ is symmetric, then by L'Hospital's rule we have

$$\lim_{\alpha \to 1/2^-} \frac{y_{1-\alpha} - y_\alpha}{1 - 2\alpha} = \lim_{\alpha \to 1/2^-} \frac{\frac{d}{dx}\left(F^{-1}(1-\alpha) - F^{-1}(\alpha) \right)}{-2} =$$

$$= \lim_{\alpha \to 1/2^-} \frac{1}{2} \left(\frac{1}{f(y_{1-\alpha})} + \frac{1}{f(y_\alpha)} \right) = \frac{1}{f(0)};$$

Similarly,

$$\lim_{\alpha \to 1/2^-} \frac{L^{-1}(1-\alpha) - L^{-1}(\alpha)}{1 - 2\alpha} = \frac{1}{L'(0)}$$

where L′(0) is the density at z = 0 of the random variable Z defined by

$$Z = f(x) + \frac{1}{b} K \left(\frac{x - Y}{b} \right) - m(x)$$

Also by L′Hospital′s rule, we have

$$\lim_{\alpha \to 1/2^-} \frac{L^{-1}(1-\alpha)}{1 - 2\alpha} \left[L\left(L^{-1}(1-\alpha) + f(x) \right) - L\left(L^{-1}(\alpha) + f(x) \right) \right] = 0$$

Thus,

$$\lim_{\alpha \to 1/2^-} B(\alpha;x) = \frac{1}{f(0)} \left[\frac{1}{b} K\left(\frac{x}{b} \right) - m(x) \right] f(0) +$$

$$\frac{1}{L'(0)} \left[\frac{1}{b} K\left(\frac{x}{b} \right) - m(x) \right] f(0)$$

$$= \left[\frac{1}{b} K\left(\frac{x}{b} \right) - m(x) \right] \left[1 + \frac{f(0)}{L'(0)} \right] \qquad (6.2)$$

From the last expression, we see that the Bias will be positive or negative depending on whether $\frac{1}{b} K \left(\frac{x}{b}\right) - m(x)$ is positive or negative.

We notice that by appropriately choosing the Kernel function K through variance reduction or inflation, we can control the sign of the bias (see Section IX for practical issues). Hence, we are led to the following result.

<u>Theorem</u> 1 The sign of the bias near $\alpha = 1/2$ is dependent only upon the choice of the Kernel function K.

VII. BIAS WHEN α = 0.

When $\alpha = 0$, the estimates given by $T_{n,\alpha}$ reduces to the jacknifed density estimators considered by Rustagi and Dynin (1985). They gave the expression for bias

$$\text{bias}(\hat{f}_j) = \frac{h_{n-1}^{r} h_n^{r} (h_n - h_{n-1})}{(h_{n-1}^{r} - h_n^{r})\ (r+2)!} f^{(r+2)}(x) \int K(-z)\, z^{r+2}\, dz \qquad (7.1)$$

where r is the number defined by (2.5) in choosing a Kernel and in practice, is taken to be 2.

VIII. IMPLICATIONS

The sign of the bias at $\alpha = 0$ is independent of the choice of the Kernel function k(.); it only depends upon the 4th (when r=2) order derivative of the density function f(x); i.e. the sign of the bias at $\alpha = 0$ is constant for any choice of the kernel function K(.). With an appropriate choice of the kernel K(.), the sign of the bias near $\alpha = 1/2$ can be made opposite to the sign of the bias at zero. Since the bias function $B(\alpha;x)$ is continuous (and in fact, differentiable) for an appropriate choice of the Kernel function K(.), there exists a proportion α^* at which bias is zero. We have thus proven the following result:

<u>Theorem 2.</u> For a given point x, there exists a Kernel function K(.) and a trimming proportion α^* for which the trimmed Jacknifed Kernel density estimator is unbiased.

IX. PRACTICAL ISSUES AND PROPOSED SOLUTIONS

Three practical issues immediately call for attention. They are :

(9i) The expression for the bias at $\alpha = 0$ is in terms of the unknown pdf f(.). Since the estimates for the derivative $f_n^{(r+2)}(x)$ is the $(r+2)^{th}$ derivative (Roscom, 1990) of the Kernel estimate (2.1), we can estimate (7.1) using $\hat{f}_n^{(r+2)}(x)$. This gives a practical way of determining the sign of the bias at $\alpha = 0$.

(9ii) How to choose the appropriate kernel function K(.). We may start with a particular kernel function K(.).

We may choose a rectangular kernel function K(.):

If we choose K(.) to be the symmetric rectangular density at (-a,a), then by choosing the endpoint a with $|\frac{x}{b}| > a$. K(x/b) have a value of zero; consequently, $\frac{1}{b} K\left(\frac{x}{b}\right)$ - m(x) would be negative. On the other hand, if K(x/b) is made to be positive specifically made to have a value 1/(2a), then

$$m(x) = \int_{x-ab}^{x+ab} \frac{1}{b} \frac{1}{(2a)} \, dF(y) \le \frac{1}{b\ (2a)} = \frac{1}{b} K\left(\frac{x}{b}\right)$$

and thus $\frac{1}{b} K\left(\frac{x}{b}\right)$ - m(x) would be non-negative.

Or we may choose a normal kernel:

If we choose K(.) as the standard normal, then our procedure goes as follows:

If $B\left(\hat{f}_j(x)\right) > 0$ then we want

$$\frac{1}{b} \varphi\left[\frac{x}{b}\right] < \frac{1}{2\sqrt{2\Pi}} \tag{9.1}$$

If (9.1) is satisfied, then we adopt K(.) = φ(.), the standard normal kernel; otherwise choose a variance reducing factor β such that

$$\frac{1}{b} \varphi\left[\beta \frac{x}{b}\right] < \frac{1}{2\sqrt{2\Pi}} \tag{9.2}$$

$$\beta > \frac{b}{x}\sqrt{-\ln(b/2)} \qquad (9.3)$$

If $B\left[\hat{f}_j(x)\right] < 0$ then we choose a variance-inflating factor β:

$$\beta < \frac{b}{x}\sqrt{-\ln(b/2)} \qquad (9.4)$$

(9iii) How to find α^*.

The bias function $B(\alpha;x)$ is a continuous (in fact diffferentiable) function of α. A computer search for α^* can be initiated using standard numerical analysis methods. This search can be done alongside the empirical verification that indeed bias is eliminated through the calculation of both the Mean Squared Error and variance of $\hat{f}_n(x)$. These two quantities being equal would be an indication that $\hat{f}_n(x)$ is unbiased for $f(x)$. As a first approximation to α^*, we may take the first term in the Newton-Raphson algorithm : $\alpha_1 = -\frac{\hat{B}(0)}{\hat{B}'(0)}$.

X. ACKNOWLEDGMENTS

To Prof. Jagdish Rustagi under whose direction the first named author has been introduced to the area of kernel density estimation; to Dr. Jose Victoria for some discussion on thhe project; to NRCP for providing support to the first author; to De La Salle University, for providing the ambiance for researh.

Refereuces:

(1) Hampel, F. R. (1974). The Influence Curve and Its Role in Robust Estimation, JASA Vol. No. 69, No. 346, 383-393.

(2) Hinkley, D. and Wang, H. (1980). A Trimmed Jackknife, J. Roy.Statist. Soc B, 42, 347-356.

(3) Javier, W R. (1989). Influence Curves for Kernel Density Estimators, paper presented at the Research and Graduate Seminar, Statistical Center, University of the Philippines, August 25, 1989.

(4) Rustagi, J. S. and Dynin, S.(1983). Jackknifing Kernel Density and Hazard Function Estimates, Technical Report No. 280, The Ohio State University, Department of Statistcs, Columbus , Ohio, 1-27.

(5) Rustagi, J. S., Javier. W. R., and Victoria, J. S.(1989). Trimmed Jackknife Kernel Density Estimate for the Probabilty Density Function. Invited paper : International Conference on Recent Developments in Statistical Data Analysis and Inference, Aug., 1989, Univ. of Neuchatel, Switzerland.

(5) Tapia, R. A. and Thomson, J. R. (1978). Non-parametric Density Estimation, Johhn's Hopkins University Press, Baltimore, Maryland.

(6) Silverman, B. W. (1986). Density Estimation for Statistics and Data Analysis, Chapman and Hall, New York.

(7) Von Mises (1947) On the Asymptotic Distribution of Differentiable Statistical Functions, Ann. Math. Statist, 18, 309-348.