Bias Correction and Higher Order Kernel Functions

Tien-Chung Hu[1]
Department of Mathematics
National Tsing-Hua University
Hsinchu, Taiwan

Jianqing Fan
Department of Statistics
University of North Carolina
Chapel Hill, N.C. 27514

Abstract

Kernel density estimates are frequently used, based on a second order kernel. Thus, the bias inherent to the estimates has an order of $O(h_n^2)$. In this note, a method of correcting the bias in the kernel density estimates is provided, which reduces the bias to a smaller order. Effectively, this method produces a higher order kernel based on a second order kernel. For a kernel function K, the functions

$$W_k(x) = \sum_{l=0}^{k-1} \binom{k}{l+1} x^l K^{(l)}(x)/l!$$

and

$$\frac{1}{\int_{-\infty}^{\infty} K^{(k-1)}(x)/x dx} K^{(k-1)}(x)/x$$

are kernels of order k, under some mild conditions.

[1]Completed while visiting Department of Statistics, University of North Carolina, Chapel Hill.
Abbreviated title. Bias Correction.
AMS 1980 subject classification. Primary 62G05. Secondary 62G20.
Key words and phrases. Bias correction, higher order kernel, kernel density estimate, nonparametrics.

1 Introduction

Consider data that can be thought of as a random sample from a distribution having an unknown density. It is common practice to summarize the data with some kinds of statistics. Unless the form of the density is known, it is also very helpful to examine graphical representations and overall structures of the data. Kernel density estimates provide a useful tool for these purpose. See Silverman (1986), Eubank (1988), Müller (1988), Härdle (1990) and Wahba (1990) for many examples of this, and good introductions to the general subject area.

Great efforts have been made to select a bandwidth for a kernel density estimate based on a second order kernel, because such an estimate is easily explainable. A large amount of recent progress has been obtained on data based smoothing parameter selection, see Rice (1984), Marron (1988), Hall *et al.* (1990), Jones, Marron and Park (1990), Chiu (1990), Fan and Marron (1990) and among others. Most of these bandwidth selectors have extremely fast rates of convergence to their theoretical optimal. However, since the second order kernel is used in the density estimate, the bias inherent to the estimate is always of order $n^{-2/5}$, no matter how good an automatic bandwidth selector is. This amount of bias may sometimes obscure the interesting features such the number of modes and height of the underlying density at the modes. In such cases, bias correction to the kernel density estimate is desirable. The discussion on this issue forms the core of this paper.

For a set of random sample $X_1, \cdots, X_n$, a kernel density estimator is defined by (See Rosenblatt (1956))

$$\hat{f}_n(x) = \frac{1}{nh_n} \sum_1^n K\left(\frac{x - X_j}{h_n}\right), \tag{1.1}$$

where K is a kernel function and h_n is a bandwidth. We will concentrate on how the bias of (1.1) can be estimated for a nonrandom bandwidth h_n.

A method of correcting bias is given in section 2. Effectively, we give a method for constructing a higher order kernel based on a second order kernel. This provides a new insight to the effect of an higher order kernel.

There are several methods of constructing a higher order kernel. Schucany and Sommers (1977) propose a method based on the generalized jackknife to higher order kernels. A useful class of higher order kernel based on Gaussian density can be found in Wand and Schucany (1990). Berlinet (1990) using the idea of reproducing kernel in a Hilbert space to construct a class of higher order kernel and discuss its consequences. The optimalities of higher order kernels are discussed in Müller (1984), Gasser *et al.* (1985), and Granovsky and Müller (1990).

Mathematically, most of methods above are directly targeted at finding a function K_r satisfying

$$\int_{-\infty}^{\infty} K_r(z)dz = 1, \int_{-\infty}^{\infty} z^q K_r(z)dz = 0, q = 1, \cdots, r-1, \text{ and } \int_{-\infty}^{\infty} |z^r K_r(z)|dz < \infty.$$

We take a different approach from the pioneering work by correcting bias directly. As a result of bias correction, a class of higher order kernel is constructed.

Section 2 gives a precise formulation, and discussion, of the main results. Proofs are in section 3.

2 Main Results

Let's illustrate how the bias of the kernel density can be corrected. Mathematical justifications are given in section 3. Observe that the kernel density estimator (1.1) is an unbiased estimator of $\int_{-\infty}^{\infty} f(x - h_n y)K(y)dy$:

$$E\hat{f}_n(x) = \int_{-\infty}^{\infty} f(x - h_n y)K(y)dy. \tag{2.1}$$

Taking derivatives j times with respective to h_n yields an unbiased estimate of the functional (viewing x as a fixed point)

$$\theta_j(x) \equiv \int_{-\infty}^{\infty} f^{(j)}(x - h_n y)(-y)^j K(y)dy, \tag{2.2}$$

and the unbiased estimate of θ_j is given by

$$\hat{\theta}_j(x) = \frac{\partial^j}{\partial h_n^j}\hat{f}_n(x), \tag{2.3}$$

where $\hat{f}_n$ is the kernel density estimate defined by (1.1).

Let's assume that the unknown density has k bounded continuous derivatives. Now, the Taylor expansion of $f(x)$ yields

$$\begin{aligned} f(x) &= f(x - h_n y + h_n y) \\ &= \sum_{j=0}^{k-1} \frac{1}{j!} f^{(j)}(x - h_n y)(h_n y)^j + O(h_n^k) \end{aligned} \tag{2.4}$$

Multiplying $K(y)$ and then integrating both sides of (2.4) with respect to y, we have

$$f(x) = \sum_{j=0}^{k-1} \frac{(-h_n)^j}{j!} \theta_j(x) + O(h_n^k),$$

where the fact $\int_{-\infty}^{\infty} K(y)dy = 1$ is used. Thus, one can use $\hat{\theta}_j(x)$ to correct the bias

$$\theta_0(x) - f(x) = -\sum_{j=1}^{k-1} \frac{(-h_n)^j}{j!} \theta_j(x) + O(h_n^k)$$

of kernel density estimate (1.1). In other words, a bias-corrected estimate is defined as

$$\hat{f}_b(x) = \sum_{j=0}^{k-1} \frac{(-h_n)^j}{j!} \hat{\theta}_j(x). \tag{2.5}$$

Let's give a simpler formula for the bias-corrected estimate (2.5).

Lemma 2.1. *If $K(\cdot)$ has bounded k^{th} derivative, then*

$$\frac{(-h_n)^j}{j!} \hat{\theta}_j(x) = \frac{1}{nh_n} \sum_{i=1}^{n} \sum_{l=0}^{j} \binom{j}{l} K_l\left(\frac{x - X_i}{h_n}\right) / l!,$$

where $K_l(z) = z^l K^{(l)}(z)$, and $\hat{\theta}_j$ was defined by (2.3).

By Lemma 2.1, the bias-corrected estimate (2.5) can be written as

$$\hat{f}_b(x) = \frac{1}{nh_n} \sum_{1}^{n} W_k\left(\frac{x - X_i}{h_n}\right) \tag{2.6}$$

with

$$\begin{aligned} W_k(x) &= \sum_{j=0}^{k-1} \sum_{l=0}^{j} \binom{j}{l} x^l K^{(l)}(x)/l! \\ &= \sum_{l=0}^{k-1} \binom{k}{l+1} x^l K^{(l)}(x)/l!, \end{aligned} \tag{2.7}$$

where the identity that

$$\sum_{j=l}^{k-1} \binom{j}{l} = \binom{k}{l+1}$$

was used. Thus, effectively the efforts of bias correction of kernel density estimate produce another kernel function $W_k(\cdot)$ defined by (2.7). As intuitively expected, $W_k(x)$ is a k^{th} order kernel, which is justified by

Theorem 1. *If the kernel function $K(\cdot)$ satisfies*

$$\int_{-\infty}^{\infty} K(y)dy = 1 \quad and \quad \int_{-\infty}^{\infty} |y^{2k}K^{(k)}(y)|dy < \infty,$$

then the function $W_k(\cdot)$ is a k^{th} order kernel:

$$\int_{-\infty}^{\infty} W_k(x)dx = 1, \quad \int_{-\infty}^{\infty} x^s W_k(x)dx = 0, for\ s = 1, \cdots k-1,$$

and

$$\int_{-\infty}^{\infty} x^k W_k(x)dx = (-1)^{k-1} \int_{-\infty}^{\infty} x^k K(x)dx. \tag{2.8}$$

Since $W_k(x)$ is a k^{th} order kernel, it follows that

Theorem 2. Let K satisfy the condition of Theorem 1 and let $f(\cdot)$ have k^{th} bounded continuous derivative. Then,

$$E\hat{f}_b(x) = f(x) - \frac{f^{(k)}(x)}{k!} \int_{-\infty}^{\infty} x^k K(x)dx h_n^k(1 + o(1)).$$

Thus, the bias-corrected estimate does have the order of bias as expected. Since W_k is a k^{th} order kernel, a similar conclusion holds for the Mean Integrated Square Error (MISE).

Remark 1. When K is symmetric, the kernel function (2.7) is also symmetric. In such a case, if $k = 2r - 1$ is an odd integer, then W_{2r-1} is also a kernel of order $2r$ which can easily justified by (2.8), and satisfies

$$\int_{-\infty}^{\infty} x^{2r} W_{2r-1}(x)dx = (2r-1) \int_{-\infty}^{\infty} x^{2r} K(x)dx.$$

Example 1. Let's take a standard normal density

$$K(x) = \phi(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$$

as a kernel function. Then,

$$\phi^{(l)}(x) = (-1)^l H_l(x)\phi(x),$$

where $H_l(x)$ is the Hermite polynomial of order l. Thus, by (2.7),

$$W_k(x) = \sum_{l=0}^{k-1} \binom{k}{l+1} (-x)^l H_l(x)\phi(x)/l! \tag{2.9}$$

is a k^{th} order kernel with

$$\int_{-\infty}^{\infty} x^k W_k(x)dx = \int_{-\infty}^{\infty} x^k \phi(x)dx = (-1)^{k-1} \begin{cases} 0, & \text{if } k = 2r-1 \\ -(2r-1)(2r-3)\cdots 1 & \text{if } k = 2r. \end{cases}.$$

Note that also that $W_{2r-1}(x)$ is a kernel of order $2r$ with

$$\int_{-\infty}^{\infty} x^{2r} W_{2r-1}(x)dx = (2r-1)\int_{-\infty}^{\infty} x^{2r}\phi(x)dx = (2r-1)^2(2r-3)\cdots 1.$$

These kernel functions are different from the kernel functions derived by Wand and Schucany (1990). The following table list the first few kernel functions (2.9), which is computed by a computer program.

Table 1: Gaussian-based kernels of order 2—7

k	$W_k(x)$	$\int_{-\infty}^{\infty} W_k^2(x)dx$
2	$(-x^2+2)\phi(x)$	0.7758
3	$(x^4-7x^2+6)\phi(x)/2$	1.4149
4	$(-x^6+15x^4-48x^2+24)\phi(x)/6$	2.2336
5	$(x^8-26x^6+183x^4-360x^2+120)\phi(x)/24$	3.3145
6	$(-x^{10}+40x^8-495x^6+2190x^4-3000x^2+720)\phi(x)/120$	4.8096
7	$(x^{12}-57x^{10}+1095x^8-8625x^6+27090x^4-27720x^2+5040)\phi(x)/720$	6.9908

It appears that the higher kernels produced by (2.9) are quite complicated, which make them less useful. However, a simple method is possible. Observe that for $l \geq 1$

$$\int_{-\infty}^{\infty} K^{(l)}(x)dx = 0.$$

By integration by parts, we obtain

$$\int_{-\infty}^{\infty} x^l K^{(k-1)}(x)dx = \begin{cases} 0, & \text{if } l = 0, \cdots, k-2 \\ (-1)^{(k-1)}(k-1)!, & \text{if } l = k-1 \end{cases} . \tag{2.10}$$

Thus, if $\int_{-\infty}^{\infty} |K^{(k-1)}(x)/x|dx < \infty$, then by (2.10)

$$K_{k-1}(x) = \frac{1}{\int_{-\infty}^{\infty} K^{(k-1)}(x)/x dx} K^{(k-1)}(x)/x \tag{2.11}$$

is a kernel of order k.

Theorem 3. *Let $K(x)$ be a kernel function satisfying*

$$\int_{-\infty}^{\infty} |K^{(k-1)}(x)/x|dx < \infty,$$

and

$$x^l K^{(l-1)}(x) \to 0, \text{ as } |x| \to \infty, \text{for } l = 1, \cdots, k-1.$$

Then, K_k defined by (2.11) is a kernel function satisfying

$$\int_{-\infty}^{\infty} x^l K_k(x)dx = \begin{cases} 0, & \text{if } l = 1, \cdots, k-1 \\ \frac{(-1)^{(k-1)}(k-1)!}{\int_{-\infty}^{\infty} K^{(k-1)}(x)/x dx}, & \text{if } l = k \end{cases} .$$

Remark 2. When $K(\cdot)$ is a even function, then $K^{(2r-1)}(0) = 0$. Thus,

$$\lim_{x \to 0} K^{(2r-1)}(x)/x = K^{(2r)}(0),$$

if $K^{(2r)}(0)$ exits. In other words, the function $K^{(2r-1)}(x)/x$ is well defined at point $x = 0$. Consequently, if $\int_{-\infty}^{\infty} |K^{(2r-1)}(x)|dx < \infty$, then $\int_{-\infty}^{\infty} |K^{(2r-1)}(x)|/x dx < \infty$ and $K_{2r-1}(x)$ is a kernel of order $2r$, if other conditions of Theorem 3 is satisfied.

Example 1. (*continued*) If $K(x) = \phi(x)$, then

$$\begin{aligned} K_{2r-1}(x) &= \phi^{(2r-1)}(x)/[x \int_{-\infty}^{\infty} \phi^{(2r-1)}(x)/x dx] \\ &= \frac{(-1)^r \phi^{(2r-1)}(x)}{2^{r-1}(r-1)!x}. \end{aligned}$$

is a kernel function of order 2r. The result is found in Wand and Schucany (1990). See also Wand and Schucany (1990) for the kernel functions $K_{2r-1}(x)$, $r = 1, \cdots, 5$.

Example 2. Let $K(x) = \frac{1}{\pi}\frac{1}{1+x^2}$ be the standard Cauchy density. Then,

$$(1+x^2)K^{(2r-1)}(x) + 2(2r-1)xK^{(2r-2)}(x) + (2r-1)(2r-2)K^{(2r-3)}(x) = 0.$$

The recursive formula is used to compute higher order kernels. The following table gives the higher order kernel function resulting from (2.11). The renormalization constants and $\int_{-\infty}^{\infty} K_{2r-1}^2(x)dx$ are computed by using numerical integration.

Table 2: Cauchy density based kernels of order 2—8

order 2r	$K_{2r-1}(x)$	$\int_{-\infty}^{\infty} K_{2r-1}^2(x)dx$
2	$\frac{2}{\pi(1+x^2)^2}$	0.4
4	$\frac{16(x^2-1)}{\pi(1+x^2)^4}$	0.3581
6	$\frac{90.5415(3x^4-10x^2+3)}{(1+x^2)^6}$	0.4023
8	$\frac{130.380(x^6-7x^4+7x^2-1)}{(1+x^2)^8}$	0.4464

Example 3. Let $K_n(x) = c_n(1-x^2)_+^n$ be a kernel function, where c_n is a normalization constant. Then, by (2.11)

$$K_{n,2r-1} = C_{n,r}^{-1}\sum_{j=r}^{n}(-1)^j \binom{n}{j} \frac{(2j)!}{(2j-2r+1)!} x^{2j-2r} 1_{[|x|\le 1]}, \text{ for } r = 1, \cdots, [n/2],$$

where $C_{n,r} = 2\sum_{j=r}^{n}(-1)^j \binom{n}{j} \frac{(2j)!}{(2j-2r+1)!(2j-2r+1)}$. The following Table gives the result of $K_{8,2r-1}(\mathrm{x})$ for $x \in [-1, 1]$.

Table 3: Polynomial based kernels of order 4—8

order 2r	$K_{8,2r-1}(x)$	$\int_{-\infty}^{\infty} K_{2r-1}^2(x)dx$
4	$\frac{1}{0.4547}(1-x^2)_+^5(1-5x^5)$	1.8190
6	$\frac{1}{1.1821}(1-x^2)_+^3(3-26x^2+39x^4)$	2.2435
8	$\frac{1}{13.00}(1-x^2)_+(35-385x^2+1001x^4-715x^6)$	2.5333

If one is interested in finding a fourth order kernel, a simpler one would be $K_{4,3}(x) = \frac{15}{32}(1-x^2)_+(3-7x^2)$ with $\int_{-\infty}^{\infty} K_{4,3}^2(x) = \frac{5}{4}$.

3 Proofs

3.1 Proof of Lemma 2.1

Since differentiation is a linear operator, we need only to show for the case $n = 1$. We use the induction to prove the result. Note that Lemma 2.1 holds for $j = 0$. Assume that Lemma 2.1 holds for $j = m$. Then,

$$\begin{aligned}
\hat{\theta}_{m+1}(x) &= \frac{\partial}{\partial h_n}\hat{\theta}_m(x) \\
&= \frac{(-1)^{m+1}m!}{h_n^{m+2}}\sum_{l=0}^{m}\begin{pmatrix} m \\ l \end{pmatrix}\left[K_{l+1}(\frac{x-X_1}{h_n}) + (l+m+1)K_l(\frac{x-X_1}{h_n})\right]/l! \\
&= \frac{(-1)^{m+1}m!}{h_n^{m+2}}\left[\frac{m+1}{(m+1)!}K_{m+1}(\frac{x-X_1}{h_n}) + (m+1)K_0(\frac{x-X_1}{h_n})\right. \\
&\quad \left.\sum_{l=0}^{m-1} a_{m,l}K_{l+1}(\frac{x-X_1}{h_n})\right],
\end{aligned}$$

where

$$\begin{aligned}
a_{m,l} &= \begin{pmatrix} m \\ l \end{pmatrix}/l! + (m+l+2)\begin{pmatrix} m \\ l+1 \end{pmatrix}/(l+1)! \\
&= (m+1)\begin{pmatrix} m+1 \\ l+1 \end{pmatrix}/(l+1)!.
\end{aligned}$$

Combining the last two displays yields that

$$\hat{\theta}_{m+1}(x) = \frac{(-1)^{m+1}(m+1)!}{h_n^{m+2}}\sum_{l=0}^{m+1}\begin{pmatrix} m+1 \\ l \end{pmatrix}K_l(\frac{x-X_1}{h_n})/l!.$$

Thus, Lemma 2.1 holds for $l = m+1$.

3.2 Proof of Theorem 1

Let's give two simple Lemmas, which will be used in the proof of Theorem 1.

Lemma 3.1. $\sum_{i=\max(0,k-s)}^{\min(r,k)}\begin{pmatrix} r \\ i \end{pmatrix}\begin{pmatrix} s \\ k-i \end{pmatrix} = \begin{pmatrix} r+s \\ k \end{pmatrix}.$

Proof. Think of products consisting of r good and s bad products. Choosing k products is equivalent to selecting i good products and $k-i$ bad products, for all possible i.

Lemma 3.2. Under the condition of Theorem 1,

$$\int_{-\infty}^{\infty} x^{s+j} K^{(j)}(x)/j! = (-1)^j \binom{s+j}{s} \tau_s,$$

where $\tau_s = \int_{-\infty}^{\infty} x^s K(x) dx$.

Proof. Integration by parts j times yields the results.

Proof of Theorem 1. By Lemma 3.2 and the definition of W_k, we have for $1 \leq s \leq k$

$$\begin{aligned}
&\int_{-\infty}^{\infty} x^s W_k(x) dx \\
= &\int_{-\infty}^{\infty} x^s \sum_{l=0}^{k-1} \binom{k}{l+1} x^l K^{(l)}(x)/l! dx \\
= &\tau_s \sum_{l=1}^{k} \binom{k}{l} \binom{l+s-1}{s} (-1)^{l-1} \qquad (3.1)
\end{aligned}$$

By Lemma 3.1, the summation in (3.1) can be written as

$$\begin{aligned}
&\sum_{l=1}^{k} (-1)^{l-1} \binom{k}{l} \sum_{i=1}^{\min(l,s)} \binom{l}{i} \binom{s-1}{i-1} \\
= &\sum_{i=1}^{s} \sum_{l=i}^{k} (-1)^{l-1} \binom{k}{l} \binom{l}{i} \binom{s-1}{i-1} \\
= &\sum_{i=1}^{s} \sum_{l=i}^{k} (-1)^{l-1} \binom{k}{i} \binom{k-i}{k-l} \binom{s-1}{i-1} \\
= &\sum_{i=1}^{s} \left[\sum_{l=0}^{k-i} (-1)^{l+i-1} \binom{k-i}{l} \right] \binom{k}{i} \binom{s-1}{i-1} \qquad (3.2)
\end{aligned}$$

Note that

$$\sum_{l=0}^{k-i} (-1)^l \binom{k-i}{l} = \begin{cases} 0, & \text{if } i < k \\ 1, & \text{if } i = k \end{cases}$$

Thus, by (3.1) and (3.2)

$$\int_{-\infty}^{\infty} x^s W_k(x)dx = \begin{cases} 0, & \text{if } s < k \\ (-1)^k \tau_k, & \text{if } s = k \end{cases}$$

Similarly, by (3.1) we have

$$\int_{-\infty}^{\infty} W_k(x)dx = \sum_{l=1}^{k}(-1)^k \binom{k}{l} = 1.$$

This completes the proof.

ACKNOWLEDGEMENTS

We would also like to express our sincere thanks to Professor J.S. Marron for many helpful discussions.

References

[1] Berlinet, A. (1990) Reproducing kernels and finite order kernels. *Manuscript.*

[2] Chiu, S.T. (1990). Bandwidth selection for kernel density estimation, *Ann. Statist.*, to appear.

[3] Eubank, R. L. (1988). *Spline Smoothing and Nonparametric Regression.* Dekker, New York.

[4] Fan, J. and Marron, J.S. (1990). Best possible constant for bandwidth selection. *Institute of Mimeo Series #2041*, University of North Carolina, Chapel Hill.

[5] Gasser, T., Müller, H.G., and Mammitzsch, V. (1985). Kernels for nonparametric curve estimation. *J. Roy. Statist. Soc. Ser. B*, **47**, 238-252.

[6] Härdle, W. (1990). *Applied Nonparametric Regression.* Cambridge University Press, Boston.

[7] Hall, P., Sheather, S. J., Jones, M.C. and Marron, J.S. (1990). On optimal data-based bandwidth selection in kernel density estimation. *Biometrika*, to appear.

[8] Jones, M. C., Marron, J. S. and Park, B. U. (1990). A simple root n bandwidth selector. *Annals of Statistics*, to appear.

[9] Marron, J. S. (1988). Automatic smoothing parameter selection: a survey. *Empirical Economics*, **13**, 187-208.

[10] Müller, H.G. (1984). Smooth optimum kernel estimators of densities, regression curves and modes. *Ann. Statist.*, **12**, 766-774.

[11] Müller, H.G. (1988). *Nonparametric Analysis of Longitudinal Data.* Springer Verlag, Berlin.

[12] Granovsky, B. and Müller, H.-G. (1990). Optimizing kernel methods for the nonparametric estimation of functions and characteristic points: a unifying variational principle. *Manuscript.*

[13] Rice, J. (1984). Bandwidth choice for nonparametric regression. *Ann. Statist.*, **12**, 1215-1230.

[14] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function, *Ann. Math. Statist.*, **42**, 1815-1842.

[15] Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, London.

[16] Wahba, G. (1990). *Spline Models for Observational Data.* SIAM, Philadelphia.

[17] Wand, M. P. and Schucany, W.R. (1990). Gaussian-based kernels. *Canadian J. Statist.*, **18**, 197-204.

[18] Schucany, W. R. and Sommers, J. P. (1977). Improvement of kernel type density estimators. *J. Ameri. Statist. Assoc.*, **72**, 420-423.