ADDITIVE KERNEL ESTIMATOR FOR THE PROBABILITY DENSITY FUNCTION

By

Sami F. Al-Kayed

Supervisor

Dr. Omar M. Eidous

Program: Statistics

July 11, 2011

ADDITIVE KERNEL ESTIMATOR FOR THE PROBABILITY DENSITY FUNCTION

By

Sami F. Al-Kayed

B. Sc. Actuarial Sciences, University of Jordan, 2007

Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science in the Department of Statistics, Yarmouk University, Irbid, Jordan.

Approved by:

Dr. Omar M. Eidous..(Chairman)

Associate Professor of Statistics, Yarmouk University.

Dr. Moh'd Alodat..(Member)

Associate Professor of Statistics, Yarmouk University.

Dr. Abdel-Rzzaq Mugdadi..(Member)

Associate Professor of Statistics, Jordan University of Science and Technology.

DECLARATION

I hereby declare that the work in this thesis is my own has been in accordance with generally accepted scientific bases.

July 11, 2011 SAMI. F. AL-KAYED

....

ACKNOWLEDGMENT

I would like to thank Dr. Omar M. Eidous for his invaluable assistance, patience, and encouragement throughout my graduate work, including this thesis. Also, I would like to thank all others who loved me and I loved so much.

TABLE OF CONTENTS

LIST OF TABLES

Table No. **Page**

ABSTRACT

Al-Kayead, Sami Faleh. Additive Kernel Estimator for the Probability Density Function. Master of Science Thesis, Department of Statistics, Yarmouk University, 2011 (Supervisor: Dr. Omar Eidous).

One way of improving the rate of convergent of classical kernel estimator is to use higher –order kernel function. In this thesis, we suggest a new method called additive kernel to estimate a smooth *pdf* *f(x)*.The proposed method produces many estimators for *f(x),* which are simple and interpretable as the higher –order estimator. The asymptotic properties of the proposed estimators are derived and formula for the smoothing parameter is also given based on minimizing the asymptotic mean square error (AMSE). Theoretical asymptotic results show the good potential properties of the proposed estimators over the higher –order kernel estimator.

Keyword:

Kernel method, higher –order kernel estimator, smoothing parameter, bias rate of convergence.

ملخص

الكايد ، سامي فالح "مقدر النواة الجمعي لاقتران الكثافة الاحتمالية" رسالة ماجستير بجامعة اليرموك، 2011 (المشرف: الدكتور عمر محمد اعدوس).

احدى طرق تحسين نسبة تقارب تقدير النواة التقليدي هي استخدام تقدير النواة ذي الدرجه الاعلى. وفي هذه الاطروحة اقترحنا طريقه نواة تجميعيه جديدة لتقدير اقتران الكثافة الاحتمالي التقليدي حيث تعطي الطريقة الجديدة عدة مقدرات تمتاز ببساطتها وسهولة تفسيرها كمقدر النواة ذي الدرجه الاعلى.

قمنا كذلك باشتقاق خصائص التقارب للمقدرات الجديدة ومقدر النواة الاعلى متضمناً ايضاً اشتقاقاً لمعلمة التنعيم وذلك بالاعتماد على تقليل القيمة التقريبية لمعدل مربع الخطأ.

لقد اظهرت نتائج خصائص تقارب المقدر المقترح افضليه على خصائص تقارب مقدر النواة ذي الدرجه الاعلى.

الكلمات المفتاحية:

طريقة النواة ، اقتران النواة ذي الدرجه الاعلى، معلمة التنعيم، معدل التحيز للتقارب.

CHAPTER ONE

INTRODUCTION

1.1 Introduction

There are two approaches to estimate an unknown probability density function (*pdf*) $f(x)$ for a random sample $X_1, \dots, X_n$. One can either choose the parametric approach, which requires the assumption that the random sample belongs to a known parametric family of distributions with unknown parameter(s). Then the unknown parameter(s) are estimated based on a random sample $X_1, \dots, X_n$. An obvious pitfall of the parametric approach is that important data structure can be masked when there is no previous knowledge of the sample to assist in the choice of the parametric family of distributions (Jeter 2005). Alternatively, we can determine the *pdf* using the nonparametric approach, which requires no assumptions about the functional form of density. Thus, a nonparametric approach is a good choice to estimate the unknown *pdf* $f(x)$ if the functional form of the *pdf* is unknoun.

The most popular nonparametric method in density estimation is called "kernel density estimation". The estimator obtained by using this method will be called "classical kernel density estimator",

1.2 Classical Kernel Method

Let $X_1, \ldots, X_n$ be a random sample from a continuous univariate *pdf* $f(x)$. The nonparametric classical kernel density estimator of $f(x)$ is (Silverman, 1986):

$$\hat{f}(x) = (nh)^{-1} \sum_{i=1}^{n} K_{(2)}\{(x - X_i)/h\} \tag{1.1}$$

where $K_{(2)}(.)$ is kernel function (second-order kernel function) and h is a positive real number, called the bandwidth or the smoothing parameter. The kernel function $K_{(2)}$ is assumed to be symmetric function satisfying the following conditions:

$$\int_{-\infty}^{\infty} K_{(2)}(u)du = 1, \int_{-\infty}^{\infty} uK_{(2)}(u)du = 0, \int_{-\infty}^{\infty} u^2 K_{(2)}(u)du = k_2 \neq 0 < \infty \tag{1.2}$$

If we assume that h_n is chosen for that $nh_n \to 0$ as $n \to \infty$, then, under the conditions(1.2), the mean of $\hat{f}(x)$ can be expressed as follows: (Silverman,1986)

$$E\hat{f}(x) = f(x) + \frac{1}{2}h^2\mu_2(K_{(2)})\, f''(x) + o(h^2). \tag{1.3}$$

Thus, the bias of $\hat{f}(x)$ is given by

$$bias\left(\hat{f}(x)\right) = \frac{1}{2}h^2\mu_2\left(K_{(2)}\right)f''(x) + o(h^2) \;, \tag{1.4}$$

where $\mu_2\left(K_{(2)}\right) = \int z^2 K_{(2)}(z)dz$. According to Silverman (1986), the variance of $\hat{f}(x)$ is

$$var\{\hat{f}(x)\} = (nh)^{-1} \int K_{(2)}{}^2(z) f(x - hz)dz,$$

$$=(nh)^{-1}R(K_{(2)})f(x)+o((nh)^{-1}), \qquad (1.5)$$

where $R(g)=\int_{-\infty}^{\infty} g^2(z)dz$ Therefore, the mean square error (MSE) of $\hat{f}(x)$ is

$$MSE(\hat{f}(x)) = E(\hat{f}(x)-f(x))^2,$$

$$=Var\left(\hat{f}(x)\right)+(E\hat{f}(x)-f(x))^2,$$

$$=(nh)^{-1}R(K_{(2)})f(x)+\frac{1}{4}h^4\mu_2(K_{(2)})^2 f''(x)^2+o((nh)^{-1}$$

$$+h^4). \qquad (1.6)$$

The MSE combines the variance and the squared bias of $\hat{f}(x)$ at a fixed point x to measure the quality of $\hat{f}(x)$. The mean integrated squared error (MISE) is another global measure which is used to measure the quality of $\hat{f}(x)$ over the entire real line. It is simply defined by the equation (Silverman, 1986):

$$MISE=\int MSE\left(\hat{f}(x)\right)dx.$$

Therefore, the asymptotic mean integrated square error (AMISE) of $\hat{f}(x)$ is (Wand and Jones, 1995)

$$AMISE\{\hat{f}(x)\}=(nh)^{-1}R(K_{(2)})+\frac{1}{4}h^4\mu_2(K_{(2)})^2 R(f''). \qquad (1.7)$$

1.3 Smoothing Parameter Selection

The choice of a bandwidth h in the kernel density estimator is very important because the value of the h plays a vital role in the accuracy of the kernel density estimator (Wand and Jones 1995). As we can see from (1.4) and (1.5), if the bandwidth h is chosen too large, then the bias of $\hat{f}(x)$ increases and the variance decreases. On the other hand, if the bandwidth is too small, then the bias of $\hat{f}(x)$ decreases and the variance increases. Usually, the bandwidth is chosen by finding the value of h that minimizes the AMISE. This value of h is known as the optimal bandwidth with respect to AMISE, which is given by Jeter (2005),

$$h = \left[\frac{R(K_{(2)})}{\mu^2{}_2(K_{(2)})R(f'')n}\right]^{1/5}. \tag{1.8}$$

Formula (1.8) demonstrates that h approaches to zero (as $n \to \infty$) but at a rate slower than n^{-1}.

1.4 Kernel Function

The selection of kernel function K is discussed in Silverman (1986). He reported that all functions $K(\mathrm{z})$ that satisfy $\int_{-\infty}^{\infty} K(\mathrm{z})d\mathrm{z} = 1$, $\int_{-\infty}^{\infty} \mathrm{z}\, K(\mathrm{z})d\mathrm{z} = 0$ & $\int_{-\infty}^{\infty} \mathrm{z}^2\, K(\mathrm{z})d\mathrm{z} \neq 0 < \infty$ perform about the same as each other in estimating $f(x)$. For example, the following functions are candidate as kernel function:

(a) $K(\mathrm{z}) = \frac{1}{2}$, $|\mathrm{z}| < 1$, (Rectangular kernel)

(b) $K(\mathrm{z}) = \frac{3}{4}\left(1 - \frac{1}{5}\mathrm{z}^2\right)/\sqrt{5}$, $|\mathrm{z}| < \sqrt{5}$, (Epanechnikov kernel)

(c) $K(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$, $-\infty < z < \infty$, (Gaussian kernel).

Usually, the kernel function K is chosen to be a unimodal and symmetric density to yield unbiased estimates using a symmetric distribution of the weights on both sides of the point of estimate. However, the selection of h is more efficient on the performance of $\hat{f}(x)$ rather than K (Ali, 1998).

1.5 Thesis Objectives

The main objective of this Thesis is to propose a new kernel density estimator for $f(x)$ and then to investigate its asymptotic properties. The proposed estimator is denoted by $\tilde{f}(x)$ to this end, we will follow the steps below

(1) Derive the asymptotic mean and the asymptotic variance of proposed estimator

(2) Derive a rule for the optimal value of the smoothing parameter b of proposed estimator

(3) Explore the relationship between the asymptotic properties of proposed estimator and that of the higher-order kernel estimator.

(4) Make a mathematical comparison between the properties of proposed estimator and that of the higher-order kernel estimator.

1.6 The Rest of this Thesis is Organized as Follows

According to our objectives, this thesis is organized as follows:

Chapter Two describes the higher-order kernel estimator and the proposed estimator of $f(x)$. The asymptotic statistical properties (the bias, the variance and the mean square error, MSE) of both estimators are also presented in this chapter .

In Chapter Three a mathematical comparison study between the asymptotic properties of the higher order kernel density estimator and the proposed estimator is presented. The conclusions of the study and the suggestions for further research are given in Chapter Four.

CHAPTER TWO

HIGHER-ORDER KERNEL AND PROPOSED ESTIMATOR

2.1Introduction

In this chapter, we present the higher-order kernel estimator of $f(x)$ and its asymptotic properties. also, Its asymptotic properties are derived and a formula for the optimal smoothing parameter of the proposed estimator is given. The relationship between the properties of two estimators are also derived in the final section of this chapter.

2.2 Higher–Order Kernel Density Estimator

One way of improving the accuracy of the ordinary kernel estimator (1.1) is to use higher-order kernel function. Let $\mu_j\left(K_{(r)}\right) = \int z^j \, K_{(r)}(z)dz$, then the second-order kernel, $K_{(2)}$ requires $\mu_0\left(K_{(2)}\right) = 1,\ \mu_1\left(K_{(2)}\right) = 0$, and $\mu_2\left(K_{(2)}\right) = k_2 \neq 0 < \infty$, to ensure that the bias of estimator (1.1) has a rate of convergence $O(h^2)$, while the convergence rate of the variance is $O(nh)^{-1}$.

To improve the bias convergence rate of estimator (1.1), one can construct higher-order kernels (Wand and Jones, 1995). Let $K_{(r)}$ be the r^{th}-order kernel assumed to be symmetric about zero (even) function, then $K_{(r)}$ is defined to have

$$\begin{cases} \mu_0(K_{(r)}) = 1 \\ \mu_j(K_{(r)}) = 0, \quad j = 1,2,3,\ldots,r-1 \\ \mu_r(K_{(r)}) = k_r \neq 0 < \infty \end{cases} . \tag{2.1}$$

For example, the fourth-order kernel ($r = 4$) is the kernel function $K_{(4)}$ that should satisfy the conditions: $\mu_0(K_{(4)}) = 1$, $\mu_1(K_{(4)}) = \mu_2(K_{(4)}) = \mu_3(K_{(4)}) = 0$ and $\mu_4(K_{(4)}) = k_4 \neq 0 < \infty$.

Wand and Jones (1995) gave a rule that can be adopted to construct higher-order kernels as the following:

Let $\emptyset(x)$ be the standard normal pdf and take $K_{(2)}(x) = \emptyset(x)$, then the r^{th}-order kernel, $K_{(r)}$ is obtained by using:

$$K_{(r)}(x) = \frac{3}{2}K_{(r-2)}(x) + \frac{1}{2}xK'_{(r-2)}(x) \quad , \quad r = 4, 6, 8, \ldots$$

For example, the fourth-order kernel, $K_{(4)}(x)$ is

$$K_{(4)}(x) = \frac{3}{2}K_{(2)}(x) + \frac{1}{2}xK'_{(2)}(x)$$

$$= \frac{3}{2}\ \emptyset(x) + \frac{1}{2}x\ \emptyset'(x)$$

$$= \frac{3}{2}\ \emptyset(x) - \frac{1}{2}x^2\ \emptyset(x)$$

$$= \frac{1}{2}\ \emptyset(x)(3 - x^2).$$

Wand and Schucany (1990) gave a class of higher-order kernels that satisfies condition (2.1). They derived the following rule to construct $K_{(r)}(x)$ based on the Gaussian kernel function

$$K_{(r)}(x) = \frac{(-1)^{r/2}\ \emptyset^{(r-1)}(x)}{2^{r/2-1}(r/2-1)!\,x} \quad r = 2, 4, 6, \ldots$$

where $\emptyset(x)$ is the standard normal distribution and $\emptyset^{(m)}(x)$ is the m^{th} derivative of $\emptyset(x)$. The first five even-order kernels are given in Table (2.1), which are taken from Wand and Schucany (1990). In addition, we added the value of $R(K_{(r)})$ of the corresponding even-order kernel and the rate of convergence for the corresponding bias.

Jones and Foster (1993) demonstrated that a higher-order kernel (r = 4,6,...) may take some negative terms and the density estimator based on them may be negative, especially in areas where the data are sparse, but this might be tolerated if improved bias ensues.

© Arabic Digital Library - Yarmouk University

The question arises here is; why r takes only even values? The answer of this question can be illustrated as follows:

Since $K_{(r)}$ is assumed to be symmetric function, then $\mu_j(K_{(r)}) = 0, \forall\, j = 1,3,5,...$ Therefore, if one try to obtain $O(h^{*r})$ bias for $= 1,3,5,...$, then he automatically obtains $O(h^{*r+1})$ bias. More clarification can be detected by inspecting the bias of $\hat{f}^*(x)$ in the next section.

Table (2.1): Gaussian-kernel of order 2,4,6,8 and 10 (Wand and Schucany, 1990)

r	$K_{(r)}$	$R[K_{(r)}]$	bias rate of convergence
2	$\emptyset(x)$	0.2821	$O(h^{*2})$
4	$\frac{1}{2}(3-x^2)\emptyset(x)$	0.4760	$O(h^{*4})$
6	$\frac{1}{8}(15-10x^2+x^4)\emptyset(x)$	0.6240	$O(h^{*6})$
8	$\frac{1}{48}(105-105x^2+21x^4-x^6)\emptyset(x)$	0.7479	$O(h^{*8})$
10	$\frac{1}{384}(945-1260x^2+378x^4-36x^6+x^8)\emptyset(x)$	0.8565	$O(h^{*10})$

2.2.1 The Expected and Bias of Higher-Order Kernel

Let $X_1, \dots, X_n$ be a random sample from a continuous *pdf* $f(x)$ and assume that an r^{th}-order kernel that satisfies the conditions (2.1) is obtained. Then the r^{th}-order kernel estimator of $f(x)$ is

$$\hat{f}^*(x) = (nh^*)^{-1} \sum_{i=1}^{n} K_{(r)}\{(x - X_i)/h^*\}, -\infty < x < \infty. \tag{2.2}$$

If the r^{th}-order derivatives of $f(x)$ are continuous, then the expected value of $\hat{f}^*(x)$, is Wand and Jones (1995)

$$E\hat{f}^*(x) = \int K_{(r)}(z) \sum_{\iota=0}^{r} (-h^* z)^{\iota} (\iota!)^{-1} f^{(\iota)}(x) dz + o(h^{*r}),$$

$$= f(x) - h^* f^{(1)}(x) \mu_1(K_{(r)}) + \frac{h^{*2}}{2} f^{(2)}(x) \mu_2(K_{(r)}) + \cdots +$$

$$(-1)^r \left\{\frac{\mu_r(K_{(r)})}{r!}\right\} h^{*r} f^{(r)}(x) + o(h^{*r}).$$

$$= f(x) + (-1)^r \left\{\frac{\mu_r(K_{(r)})}{r!}\right\} h^{*r} f^{(r)}(x) + o(h^{*r}). \tag{2.3}$$

Therefore, the bias of $\hat{f}^*(x)$ is

$$B_x(\hat{f}^*(x)) = E\hat{f}^*(x) - f(x).$$

$$= (-1)^r \left\{\frac{\mu_r(K_{(r)})}{r!}\right\} h^{*r} f^{(r)}(x) + o(h^{*r})$$

$$= Q_r h^{*r} f^{(r)}(x) + o(h^{*r}), \tag{2.4}$$

which indicates that the convergence rate of the bias of $\hat{f}^*(x)$ is $O(h^{*r})$. The values of $Q_r = (-1)^r \left\{\frac{\mu_r(K_{(r)})}{r!}\right\}$ for $r = 2, 4, 6, 8, 10$ are given in Table (2.2) where $K_{(2)}$is standard normal.

2.2.2 The Variance of Higher-Order

The variance of $\hat{f}^*(x)$ is

$$Var(\hat{f}^*(x)) = Var\{(nh)^{-1} \sum_{i=1}^{n} K\{(x - X_i)/h^*\}$$

$$= \frac{1}{nh^{*2}} E[K\{(x - X)/h^*\}]^2 - \frac{1}{nh^2} [EK\{(x - X)/h^*\}]^2$$

$$= \frac{1}{nh^*} f(x) R[K_{(r)}] + o(n^{-1}), \tag{2.5}$$

where $R(g) = \int_{-\infty}^{\infty} g^2(x)dx$. The values of $R[K_{(r)}]$ for $r = 2, 4, 6, 8, 10$ are given in Table (2.2). Note that, the rate of convergence for the variance of $\hat{f}^*(x)$ remains $O(nh^*)^{-1}$, the same as that of Estimator (1.1)

2.2.3 The Mean Square Error and the Smoothing Parameter of $\hat{f}^*(x)$

The MSE combines the variance and the squared bias of $\hat{f}^*(x)$ at a fixed point x to measure the quality of $\hat{f}^*(x)$, which is given by

$$\text{MSE}(\hat{f}^*(x)) = E(\hat{f}^*(x) - f(x))^2$$

$$= Var(\hat{f}^*(x)) + (E\hat{f}^*(x) - f(x))^2$$

$$\cong \frac{1}{nh^*} f(x) R[K_{(r)}] + [(-1)^r \left\{\frac{\mu_r(K_{(r)})}{r!}\right\} h^{*r} f^{(r)}(x)]^2$$

$$\cong \frac{1}{nh^*} f(x) R[K_{(r)}] + \left\{\frac{\mu_r(K_{(r)})}{r!}\right\}^2 h^{*2r} [f^{(r)}(x)]^2. \qquad (2.6)$$

The value of h^* is obtained by considering the MSE as a function of h^* (say D(h^*)), differentiate $D(h^*)$ with respect to h^* and equating to zero, then solve the resulting equation with respect to h^* as the following:

$$\frac{-1}{nh^{*2}} f(x) R[K_{(r)}] + 2rh^{*2r-1} \left\{\frac{\mu_r(K_{(r)})}{r!}\right\}^2 [f^{(r)}(x)]^2 = 0.$$

The solution of the above equation gives,

$$\frac{1}{nh^{*2}} f(x) R[K_{(r)}] = 2rh^{*2r-1} \left\{\frac{\mu_r(K_{(r)})}{r!}\right\}^2 [f^{(r)}(x)]^2.$$

If we multiply both sides of the last equation by h^{*2} then we obtain

$$h^{*2r+1} = \frac{(r!)^2 f(x) R(K_{(r)})}{n(2r)[f^{(r)}(x)]^2 \mu^2{}_r(K_{(r)})}.$$

Therefore,

$$h^* = \left(\frac{(r!)^2 f(x) R(K_{(r)})}{n(2r)[f^{(r)}(x)]^2 \mu^2{}_r(K_{(r)})}\right)^{1/(2r+1)}$$

$$= V_r \left(\frac{f(x)}{n[f^{(r)}(x)]^2}\right)^{1/(2r+1)}, \qquad (2.7)$$

where $V_r = [\frac{(r!)^2 R(K_{(r)})}{(2r)\mu^2{}_r(K_{(r)})}]^{1/(2r+1)}$. The values of V_r for $r = 2, 4, 6, 8, 10$ are given in Table (2.2). Substituting the expression of h^* back into Formula (2.6) leads to the smallest possible value of MSE for $\hat{f}^*(x)$, which is given by,

$$MSE\hat{f}^*(x) = \frac{f(x)R[K_{(r)}]}{n\left[\frac{(r!)^2 f(x)R(K_{(r)})}{n(2r)[f^{(r)}(x)]^2 \mu^2{}_r(K_{(r)})}\right]^{\frac{1}{2r+1}}} + \left\{\frac{\mu_r(K_{(r)})}{r!}\right\}^2 \left\{\left[\frac{(r!)^2 f(x)R(K_{(r)})}{n(2r)[f^{(r)}(x)]^2 \mu^2{}_r(K_{(r)})}\right]^{\frac{1}{2r+1}}\right\}^{2r} [f^{(r)}(x)]^2$$

$$= \left[\frac{2r+1}{\{2r\}^{2r/(2r+1)}\{r!\}^{2/(2r+1)}}\right][f(x)R(K_{(r)})]^{\frac{2r}{2r+1}} [f^{(r)}(x))^2 \mu^2{}_r(K_{(r)})]^{\frac{1}{2r+1}} [n]^{\frac{-2r}{2r+1}}$$

$$= T_r[f(x)]^{\frac{2r}{2r+1}} [f^{(r)2}(x)]^{\frac{1}{2r+1}} [n]^{\frac{-2r}{2r+1}}, \tag{2.8}$$

where $T_r = \frac{2r+1}{\{2r\}^{2r/(2r+1)}\{r!\}^{2/(2r+1)}} [R(K_{(r)})]^{\frac{2r}{2r+1}} [\mu^2{}_r(K_{(r)})]^{\frac{1}{2r+1}}$. The values of T_r for $r = 2, 4, 6, 8, 10$ are given in Table (2.2). Note that the functions $K_{(r)}$, $r = 2, 4, 6, 8, 10$ are given in Table (2.1).

We note here that, if the asymptotic mean integrated squared error ($MISE$) is to be used instead of the asymptotic MSE then the right hand side of equations (2.4), (2.5), (2.7) and (2.8) are slightly changed and they become

$Q_r h^{*r} \int_{-\infty}^{\infty} f^{(r)}(x) + o(h^{*r})$, $\frac{1}{nh^*} R[K_{(r)}] + o(n^{-1})$, $V_r [\frac{1}{n \int_{-\infty}^{\infty} [f^{(r)}(x)]^2 dx}]^{1/(2r+1)}$ and $T_r \{ \int_{-\infty}^{\infty} [f^{(r)}(x)]^2 dx \}^{\frac{1}{2r+1}} [n]^{\frac{-2r}{2r+1}}$ respectively. For purpose of comparison, the using of $MISE$ or MSE are equivalent.

Table (2.2): The values of $\mu_r(K_{(r)})$, Q_r , $R[K_{(r)}]$, V_r and T_r for $r = 2, 4, 6, 8, 10$.

r	$K_{(r)}$	$\mu_r(K_{(r)})$	Q_r	$R[K_{(r)}]$	V_r	T_r
2	$K_{(2)}$	1	0.5	0.2821	0.7764	0.4542
4	$K_{(4)}$	-3	-0.125	0.4760	1.1602	0.4616
6	$K_{(6)}$	15	0.020833	0.6240	1.4451	0.4678
8	$K_{(8)}$	-105	-0.00260417	0.7479	1.6818	0.4725
10	$K_{(10)}$	945	0.000260417	0.8565	1.8889	0.4761

Note that

$$\mu_r(K_{(r)}) = \int_{-\infty}^{\infty} z^r K_{(r)}(z)dz,\ \ Q_r = (-1)^r \left\{\frac{\mu_r(K_{(r)})}{r!}\right\},\ \ R[K_{(r)}] = \int_{-\infty}^{\infty} K_{(r)}{}^2(z)dz\ ,$$

$$V_r = \left[\frac{(r!)^2 R(K_{(r)})}{(2r)\mu^2{}_r(K_{(r)})}\right]^{1/(2r+1)},\ T_r = \frac{2r+1}{\{2r\}^{2r/(2r+1)}\{r!\}^{2/(2r+1)}}[R(K_{(r)})]^{\frac{2r}{2r+1}}\ [\mu^2{}_r(K_{(r)})]^{\frac{1}{2r+1}}$$

and $K_{(r)}, r = 2, 4, 6, 8, 10$ are given in Table (2.1).

2.3 The Proposed Estimator

Let $X_1, \ldots, X_n$ be a random sample of size n from a probability density function $f(x)$ that assumes to have r^{th} continuous derivatives at x. We propose the following kernel estimator for $f(x)$:

$$\tilde{f}(x) = \frac{1}{nb}\sum_{j=1}^{d}\sum_{i=1}^{n} a_j K\left(\frac{x - X_i}{jb}\right), \quad -\infty < x < \infty. \tag{2.9}$$

which we shall call it "additive kernel estimator ", where $\sum_{j=1}^{d} j a_j = 1$, K is the second-order kernel function satisfies condition (1.2) , b is the smoothing parameter, d and a_j , $j = 1, 2, 3, \ldots, d$ are constants introduced to control the magnitude of bias and variance of $\tilde{f}(x)$.

We can show that the integration of the proposed estimator over the range $(-\infty, \infty)$ equals one as follows:

$$\int_{-\infty}^{\infty} \tilde{f}(x)dx = \int_{-\infty}^{\infty} (nb)^{-1} \sum_{j=1}^{d} \sum_{i=1}^{n} a_j K\{(x - X_i)/jb\}\, dx$$

$$= (nb)^{-1} \sum_{j=1}^{d} \sum_{i=1}^{n} a_j \int_{-\infty}^{\infty} K\left\{\frac{x - X_i}{jb}\right\} dx$$

Now, let $u = \left\{\frac{x - X_i}{jb}\right\}$ then

$\int_{-\infty}^{\infty} K\left\{\frac{x - X_i}{jb}\right\} dx = jb \int_{-\infty}^{\infty} K(u)du = jb$. Thus

$\int_{-\infty}^{\infty} \tilde{f}(x)dx = (nb)^{-1} \sum_{j=1}^{d} \sum_{i=1}^{n} a_j (jb) = (nb)^{-1}(nb) \sum_{j=1}^{d} j a_j = 1$ because

$\sum_{j=1}^{d} j a_j = 1$

A special case of (2.9) was studied by Eidous (2011). He derived the asymptotic bias and variance of $\tilde{f}(x)$ when $x=0$ and for the non-negative values of the random sample applying for line transect data. The asymptotic properties of $\tilde{f}(x)$ are derived in the following subsections.

2.3.1 The Expected Value and Bias of Proposed Estimator

The expected value of $\tilde{f}(x)$ is:

$$E(\tilde{f}(x)) = E\left[(nb)^{-1} \sum_{j=1}^{d} \sum_{i=1}^{n} a_j K\{(x - X_i)/jb\}\right]$$

$$= (nb)^{-1} \sum_{j=1}^{d} a_j n E K\{(x - X_i)/jb\}.$$

Now, the expected value of $K\left\{\frac{x-X_i}{jb}\right\}$ is

$$E\left(K\left\{\frac{x-X_i}{jb}\right\}\right) = \int_{-\infty}^{\infty} K\left\{\frac{x-X_i}{jb}\right\} f(x)dx .$$

Let $u = \left\{\frac{x-X_i}{jb}\right\}$ and use the Taylor's series expansion of $f(x - jbu)$ around x, then,

$$E\left(K\left\{\frac{x-X_i}{jb}\right\}\right) = jb\int_{-\infty}^{\infty} K(u)f(x-jbu)du,$$

$$= jb\int_{-\infty}^{\infty} K(u)[f(x) - jbuf'(x) + \frac{j^2b^2u^2f''(x)}{2!} - \frac{j^3b^3u^3f'''(x)}{3!} + \cdots]du,$$

$$= jbf(x) - j^2b^2f'(x)\int_{-\infty}^{\infty} uK(u)du + \frac{j^3b^3f''(x)}{2}\int_{-\infty}^{\infty} u^2\, K(u)du - \cdots,$$

$$= jbf(x) - j^2b^2f'(x)\mu_1(K) + \frac{j^3b^3f''(x)}{2}\mu_2(K) + \cdots,$$

where $\mu_i(K) = \int_{-\infty}^{\infty} u^i\, K(u)du$. Now, suppose that the smoothing parameter and the sample size are related such that $b \to 0$ when $n \to \infty$, then the expected value of $\tilde{f}(x)$ is:

$$E(\tilde{f}(x)) = E((nb)^{-1}\sum_{j=1}^{d}\sum_{i=1}^{n} a_j K\{(x-X_i)/jb\})$$

$$= (b)^{-1}\sum_{j=1}^{d} a_j E(K\{(x-X_i)/jb\}$$

$$= \sum_{j=1}^{d} a_j\, (jf(x) - j^2bf'(x)\mu_1(K) + \frac{j^3b^2f''(x)\mu_2(K)}{2} + \cdots)$$

$$= f(x)\sum_{j=1}^{d} ja_j - bf'(x)\mu_1(K)\,\sum_{j=1}^{d} j^2a_j + \frac{b^2f''(x)}{2}\mu_2(K)\sum_{j=1}^{d} j^3a_j$$

$$+ \ldots + \frac{b^r f^{(r)}(x)}{r!}\mu_r(K)\sum_{j=1}^{d} j^{r+1}a_j + o(b^r).$$

$$= f(x)A_1 - bf^{(1)}(x)\mu_1(K)\,A_2 + \frac{b^2}{2}f^{(2)}(x)\mu_2(K)A_3$$

$$+ \cdots + \frac{b^r f^{(r)}(x)}{r!}\mu_r(K)A_{r+1} + o(b^r)\,, \qquad (2.10)$$

where $A_l = \sum_{j=1}^{d} j^l a_j$ and $f^{(r)}(x)$ is the r^{th} derivative of $f(x)$. Therefore, under the constraints $A_1 = 1$ and $A_3 = A_5 = A_7 = \cdots = A_{r-1} = 0$, the bias of $\tilde{f}(x)$ is

$$B_x(\tilde{f}(x)) = E\tilde{f}(x) - f(x).$$

$$= \frac{b^r f^{(r)}(x)\mu_r(K)A_{r+1}}{r!} + o(b^r)$$

$$= QQ_r b^r f^{(r)}(x) + o(b^r), \quad (2.11)$$

where $QQ_r = \frac{\mu_r(K)A_{r+1}}{r!}$. Because the second order kernel K is assumed to be even function, then $\mu_r(K) = \int_{-\infty}^{\infty} u^r K(u)du = 0$, for odd values of r. That is $\mu_1(K) = \mu_3(K) = \mu_5(K) = \cdots = 0$. Therefore, the terms $bf^{(1)}(x)\mu_1(K)A_2$, $\frac{b^3 f^{(3)}(x)}{3!}\mu_3(K)A_4$, $\frac{b^5 f^{(5)}(x)}{5!}\mu_5(K)A_6, \ldots$ are vanishing without assuming $A_2 = A_4 = A_6 = \cdots = 0$. In other words and since r is even value, the constraints $A_1 = 1$ and $A_3 = A_5 = A_7 = \cdots = A_{r-1} = 0$ should be valid to obtain $O(b^r)$ bias for $\tilde{f}(x)$.

2.3.2 The Variance of Proposed Estimator

In this subsection, we derive the asymptotic variance of $\tilde{f}(x)$ under the assumption $nb \to \infty$ when $n \to \infty$. The variance of $\tilde{f}(x)$ is

$$Var(\tilde{f}(x)) = Var((nb)^{-1} \sum_{j=1}^{d} \sum_{i=1}^{n} a_j K\{(x - X_i)/jb\})$$

$$= \frac{1}{nb^2} Var[\sum_{j=1}^{d} a_j (K\{(x - X_i)/jb\})]$$

$$= \frac{1}{nb^2} \left[E(\sum_{j=1}^{d} a_j K\{(x - X_i)/jb\})^2 - \left(\sum_{j=1}^{d} a_j EK\{(x - X_i)/jb\}\right)^2 \right].$$

By substituting the expression of $EK[(x - X_i)/jb]$ in the second term of the last expression then we obtain

$$Var(\tilde{f}(x)) = \frac{1}{nb^2} E(\sum_{j=1}^{d} a_j (K[\frac{x-X_i}{jb}]))^2 + o(n^{-1})$$

$$= \frac{1}{nb^2} E(\sum_{j=1}^{d} \sum_{l=1}^{d} a_j a_l K\left[\frac{x-X_i}{jb}\right] K\left[\frac{x-X_i}{lb}\right]) + o(n^{-1})$$

$$= \frac{1}{nb^2} E(\sum_{j=1}^{d} a_j^2 K^2 \left\{\frac{x-X_i}{jb}\right\} + 2\sum_{j<l}\sum a_j a_l K\left[\frac{x-X_i}{jb}\right] K\left[\frac{x-X_i}{lb}\right]) + o(n^{-1})$$

$$= \frac{1}{nb^2} (\sum_{j=1}^{d} a_j^2 EK^2 \left\{\frac{x-X_i}{jb}\right\} + 2\sum_{j<l}\sum a_j a_l E(K\left[\frac{x-X_i}{jb}\right] K\left[\frac{x-X_i}{lb}\right])) + o(n^{-1}).$$

Now,

$$EK^2\left\{\frac{x-X_i}{jb}\right\} = \int_{-\infty}^{\infty} K^2\left\{\frac{x-X_i}{jb}\right\} f(x)\, dx$$

$$= jb \int_{-\infty}^{\infty} K^2(u) f(x - jbu)\, du$$

$$= jb \int_{-\infty}^{\infty} K^2(u)[f(x) - jbuf'(x) + \frac{j^2b^2u^2f''(x)}{2} + \cdots]du$$

$$= jbf(x)\int_{-\infty}^{\infty} K^2(u)du - j^2b^2f'(x)\int_{-\infty}^{\infty} uK^2(u) + \cdots,$$

and

$$\frac{1}{nb^2}\left(\sum_{j=1}^{d} a_j{}^2 EK^2\left\{\frac{x-X_i}{jb}\right\}\right) =$$

$$\frac{1}{nb^2}\sum_{j=1}^{d} a_j{}^2\,[jbf(x)\int_{-\infty}^{\infty} K^2(u)du - j^2b^2 f'(x)\int_{-\infty}^{\infty} uK^2(u) + \cdots]$$

$$= \frac{1}{nb} f(x)\int_{-\infty}^{\infty} K^2(u)du \sum_{j=1}^{d} ja_j{}^2 - \frac{1}{n} f'(x)\int_{-\infty}^{\infty} uK^2(u)\sum_{j=1}^{d} j^2 a_j{}^2 + o(\frac{h}{n})$$

$$= \frac{1}{nb} f(x)\int_{-\infty}^{\infty} K^2(u)\,du \sum_{j=1}^{d} ja_j{}^2 + o(n^{-1}) .$$

Now,

$$EK\left[\frac{x-X_i}{jb}\right] K\left[\frac{x-X_i}{lb}\right] = \int_{-\infty}^{\infty} K\left[\frac{x-X_i}{jb}\right] K\left[\frac{x-X_i}{lb}\right] f(x)dx$$

$$= b\int_{-\infty}^{\infty} K\left[\frac{u}{j}\right] K\left[\frac{u}{l}\right] f(x-bu)du$$

$$= b\int_{-\infty}^{\infty} K\left[\frac{u}{j}\right] K\left[\frac{u}{l}\right] [f(x) - buf'(x) + \frac{b^2u^2f''(x)}{2} + \cdots]du$$

$$= bf(x)\int_{-\infty}^{\infty} K\left[\frac{u}{j}\right] K\left[\frac{u}{l}\right] du - b^2 f'(x)\int_{-\infty}^{\infty} uK\left[\frac{u}{j}\right] K\left[\frac{u}{l}\right] du + \cdots$$

$$= bf(x)D_1 - b^2 f'(x)D_2 + \cdots ,$$

where $D_1 = \int_{-\infty}^{\infty} K\left[\frac{u}{j}\right] K\left[\frac{u}{l}\right] du$ and $D_2 = \int_{-\infty}^{\infty} uK\left[\frac{u}{j}\right] K\left[\frac{u}{l}\right] du$. Now,

$$\frac{2}{nb^2}\sum_{j<l}\sum a_j a_l EK\left[\frac{x-X_i}{jb}\right] EK\left[\frac{x-X_i}{lb}\right]$$

$$= \frac{2}{nb^2} \sum_{j<l} \sum a_j a_l \left[b f(x) D_1 - b^2 f'(x) D_2 + \cdots \right]$$

$$= \frac{2f(x)}{nb} \sum_{j<l} \sum a_j a_l D_1 - \frac{2}{n} f'(x) \sum_{j<l} \sum a_j a_l D_2 + o(\frac{b}{n})$$

$$= \frac{2f(x)}{nb} \sum_{j<l} \sum a_j a_l \int_{-\infty}^{\infty} K\left[\frac{u}{j}\right] K\left[\frac{u}{l}\right] du + o(n^{-1}).$$

Therefore, the asymptotic variance of $\tilde{f}(x)$ is

$Var(\tilde{f}(x)) =$

$$\frac{f(x)}{nb} \int_{-\infty}^{\infty} K^2(u)\, du \sum_{j=1}^{d} j a_j{}^2 + \frac{2f(x)}{nb} \sum_{j<l} \sum a_j a_l \int_{-\infty}^{\infty} K\left[\frac{u}{j}\right] K\left[\frac{u}{l}\right] du + o(n^{-1})$$

$$= \frac{f(x)}{nb} \{ \int_{-\infty}^{\infty} K^2(u)\, du \sum_{j=1}^{d} j a_j{}^2 + 2 \sum_{j<l} \sum a_j a_l \int_{-\infty}^{\infty} K\left[\frac{u}{j}\right] K\left[\frac{u}{l}\right] du \} + o(n^{-1})$$

$$= \frac{f(x)}{nb} S(a_1, a_2, \ldots, a_d) + o(n^{-1}), \qquad (2.12)$$

where $S(a_1, a_2, \ldots, a_d) = \int_{-\infty}^{\infty} K^2(u)\, du \sum_{j=1}^{d} j a_j{}^2 + 2 \sum_{j<l} \sum a_j a_l \int_{-\infty}^{\infty} K\left[\frac{u}{j}\right] K\left[\frac{u}{l}\right] du \}$. Note that the convergence rate for variance of $\tilde{f}(x)$ is $O(nb)^{-1}$, the same as that of the classical kernel estimator and that of higher-order kernel estimator.

2.3.3 The Mean Square Error and the Smoothing Parameter of $\tilde{f}(x)$

The asymptotic mean square error (MSE) of $\tilde{f}(x)$ is

$$\text{MSE}\,(\tilde{f}(x)) = E(\tilde{f}(x) - f(x))^2$$

$$= Var(\tilde{f}(x)) + (Bias\ \tilde{f}(x))^2$$

$$= \frac{f(x)\, S(a_1,a_2,\dots,a_d)}{nb} + \left(\frac{b^r f^{(r)}(x)\mu_r(K)A_{r+1}}{r!}\right)^2$$

$$= \frac{f(x)\, S(a_1,a_2,\dots,a_d)}{nb} + \frac{b^{2r} {f^{(r)}}^2(x){\mu_r}^2(K)A_{r+1}^2}{(r!)^2}. \qquad (2.13)$$

The value of b can be obtained by considering the MSE of $\tilde{f}(x)$ as a function of b (say $T(b)$), differentiate $T(b)$ with respect to b and equating to zero, then solve the resulting equation with respect to b to obtain,

$$\frac{-f(x)\, S(a_1,\ a_2,\ \dots,\ a_d)}{nb^2} + \frac{2rb^{2r-1} {f^{(r)}}^2(x){\mu_r}^2(K)A_{r+1}^2}{(r!)^2} = 0$$

which implies

$$\frac{f(x)\, S(a_1,\ a_2,\ \dots,\ a_d)}{nb^2} = \frac{2rb^{2r-1} {f^{(r)}}^2(x){\mu_r}^2(K)A_{r+1}^2}{(r!)^2}.$$

Multiple both sides by b^2, then we obtain,

$$\frac{f(x)\, S(a_1,a_2,\dots,a_d)}{n} = \frac{2rb^{2r+1} {f^{(r)}}^2(x){\mu_r}^2(K)A_{r+1}^2}{(r!)^2},$$

and then,

$$b = \left[\frac{f(x)\, S(a_1,a_2,\dots,a_d)(r!)^2}{n(2r){f^{(r)}}^2(x){\mu_r}^2(K)A_{r+1}^2}\right]^{\frac{1}{2r+1}}$$

$$= VV_r \left[\frac{f(x)}{{f^{(r)}}^2(x)}\right]^{\frac{1}{2r+1}} [n]^{\frac{-1}{2r+1}}, \qquad (2.14)$$

where $VV_r = [\frac{S(a_1,a_2,\dots,a_d)(r!)^2}{(2r)(x)\mu_r{}^2(K)A^2_{r+1}}]^{\frac{1}{2r+1}}$. Note that, the smoothing parameter $b \longrightarrow 0$ when $n \longrightarrow \infty$. Also, the term $VV_r\, [\frac{f(x)}{f^{(r)^2}(x)}]^{\frac{1}{2r+1}}\ [n]^{\frac{-1}{2r+1}} \longrightarrow 0$ as $n \longrightarrow \infty$ but at a rate slower than n^{-1} because $\frac{1}{2r+1} < 1$. Now, by substituting the above expression of b back into Formula (2.13) leads to the smallest possible MSE of $\tilde{f}(x)$.

$$MSE(\tilde{f}(x)) =$$

$$\frac{f(x)\, S(a_1,a_2,\dots,a_d)}{\left[\frac{f(x)\, S(a_1,a_2,\dots,a_d)(r!)^2}{n(2r)f^{(r)^2}(x)\mu_r{}^2(K)A^2_{r+1}}\right]^{\frac{1}{2r+1}}} + \frac{\left(\left[\frac{f(x)\, S(a_1,a_2,\dots,a_d)(r!)^2}{n(2r)f^{(r)^2}(x)\mu_r{}^2(K)A^2_{r+1}}\right]^{\frac{1}{2r+1}}\right)^{2r} [f^{(r)^2}(x)][\mu_r{}^2(K)][A^2_{r+1}]}{(r!)^2}.$$

$$= [\frac{2r+1}{\{2r\}^{\frac{2r}{2r+1}}\{r!\}^{\frac{2}{2r+1}}}][f(x)S(a_1,a_2,\dots,a_d)]^{\frac{2r}{2r+1}}[f^{(r)^2}(x)\mu_r{}^2(k)A^2{}_{r+1}]^{\frac{1}{2r+1}}[n]^{\frac{-2r}{2r+1}}$$

$$= TT_r\ [f(x)]^{\frac{2r}{2r+1}}\ [f^{(r)2}(x)]^{\frac{1}{2r+1}}\ [n]^{\frac{-2r}{2r+1}} \qquad (2.15)$$

where $TT_r = \frac{2r+1}{\{2r\}^{2r/(2r+1)}\{r!\}^{2/(2r+1)}}[S(a_1,a_2,\dots,a_d)]^{\frac{2r}{2r+1}}\ [\mu_r{}^2(k)A^2{}_{r+1}]^{\frac{1}{2r+1}}$. Equation (2.15) implies that, if the *pdf* $f(x)$ has r continuous derivatives and by picking suitable values of $a_1, a_2, \dots, a_d$ such that $A_1 = 1$ and $A_3 = A_5 = A_7 = \dots = A_{r-1} = 0$, the speed of convergence can be improved over that $O(n)^{-4/5}$ found for $r = 2$ (the convergence rate of MSE for the classical kernel estimator). For example, when $r = 4$, the MSE converges at rate $O(n)^{-8/9}$, which is better than that $O(n)^{-4/5}$. An inspection of the asymptotic MSE of the higher-order kernel estimator (Equation 2.8) leads us to the same conclusion provided that the higher-order kernel function $K_{(r)}$ that satisfies (2.1) is obtained. The relationship between the smoothing parameters, asymptotic biases, asymptotic variances and asymptotic mean square errors of the higher-order kernel estimator, $\hat{f}^*(x)$ and the proposed estimator, $\tilde{f}(x)$ are presented in Section (2.4). Despite that the two estimators $\hat{f}^*(x)$ and $\tilde{f}(x)$ can achieve $O(n)^{\frac{-2r}{2r+1}}$, $r = 2, 4, 6, \ldots$, a preference of $\tilde{f}(x)$ over $\hat{f}^*(x)$ can be obtained as a comparison study in Chapter (3) demonstrated.

2.4. Relationship Between the Asymptotic Properties of $\tilde{f}(x)$ and $\hat{f}^*(x)$

The optimal formula of the smoothing parameter b for the proposed estimator is

$$b = VV_r \left[\frac{f(x)}{{f^{(r)}}^2(x)}\right]^{\frac{1}{2r+1}} [n]^{\frac{-1}{2r+1}}.$$

Where $VV_r = \left(\frac{(r!)^2 S(a_1,a_2,\ldots,a_d)}{2r\,\mu^2{}_r(K) A^2_{r+1}}\right)^{\frac{1}{2r+1}}$. Therefore b is related to $h^*$where h^* is the smoothing parameter of the higher-order kernel estimator by

$$b \cong \frac{VV_r}{V_r} h^*.$$

Substituting formula VV_r in the last equation yields

$$b \cong \left(\frac{S(a_1,a_2,\ldots,a_d)}{\mu^2{}_r(K) A^2_{r+1}} \frac{\mu^2{}_r(K_r)}{R(K_r)}\right)^{\frac{1}{2r+1}} h^*. \qquad (2.16)$$

Note that ,if $r = 2$ then $\mu^2{}_r(K) = \mu^2{}_r(K_r)$.

Suppose that b and h^* are related according to Formula (2.16), then the asymptotic bias ,variance and MSE of $\tilde{f}(x)$ and $\hat{f}^*(x)$ are related as follows:

I. Let $QQ_r = \frac{\mu_r(K)A_{r+1}}{r!}$. Then

$$B_x\left(\tilde{f}(x)\right) \cong QQ_r b^r f^{(r)}(x),$$

$$= QQ_r \left(\frac{S(a_1,a_2,\dots,a_d)}{\mu^2{}_r(K)A^2_{r+1}} \frac{\mu^2{}_r(K_r)}{R(K_r)}\right)^{\frac{r}{2r+1}} h^{*r} f^{(r)}(x),$$

$$= \frac{QQ_r}{Q_r} \left(\frac{S(a_1,a_2,\dots,a_d)}{\mu^2{}_r(K)A^2_{r+1}} \frac{\mu^2{}_r(K_r)}{R(K_r)}\right)^{\frac{r}{2r+1}} Bias(\hat{f}^*(x)),$$

$$= \frac{\mu_r(K)A_{r+1}}{(-1)^r \mu_r(K_r)} \left(\frac{S(a_1,a_2,\dots,a_d)}{\mu^2{}_r(K)A^2_{r+1}} \frac{\mu^2{}_r(K_r)}{R(K_r)}\right)^{\frac{r}{2r+1}} B_x(\hat{f}^*(x)). \qquad (2.17)$$

II. $Var(\tilde{f}(x)) \cong \frac{f(x)S(a_1,a_2,\dots,a_d)}{nb}$

$$= \frac{f(x)S(a_1,a_2,\dots,a_d)}{nh^*} \left(\frac{S(a_1,a_2,\dots,a_d)}{\mu^2{}_r(K)A^2_{r+1}} \frac{\mu^2{}_r(K_r)}{R(K_r)}\right)^{\frac{-1}{2r+1}}$$

$$= [S^{2r}(a_1,a_2,\dots,a_d)\, A^2_{r+1}]^{\frac{1}{2r+1}} \left[\frac{\mu^2{}_r(K)}{\mu^2{}_r(K_r)R^{2r}(K_r)}\right]^{\frac{1}{2r+1}} Var(\hat{f}^*(x)) \qquad (2.18)$$

III. Let $TT_r = (2r+1)[(2r)^{2r}(r!)^2]^{\frac{-1}{2r+1}} \left[S^{2r}(a_1,a_2,\dots,a_d)A^2_{r+1}\mu^2{}_r(K)\right]^{\frac{1}{2r+1}}$

Then the asymptotic mean square error of $\tilde{f}(x)$ is

$$MSE(\tilde{f}(x)) \cong TT_r \left(f^{(r)^2}(x) f^{2r}(x) n^{-2r}\right)^{\frac{1}{2r+1}}$$

$$\cong \frac{TT_r}{T_r} MSE(\hat{f}^*(x))$$

$$= \left(\frac{S^{2r}(a_1,a_2,\dots,a_d)A_{r+1}^2\mu^2{}_r(K)}{\mu^2{}_r(K_r)R^{2r}(K_r)}\right)^{\frac{1}{2r+1}} MSE(\hat{f}^*(x)) \qquad (2.19)$$

To simplify formulas (2.16),(2.17),(2.18) and (2.19),let us consider the five values of $r = 2,4,6,8$ and 10 and take the second order kernel to be N(0,1) (i.e. $K = K_2 = N(0,1)$).Then $\mu_2(K) = 1$, $\mu_4(K) = 3$, $\mu_6(K) = 15$, $\mu_8(K) = 105$ and $\mu_{10}(K) = 945$. Also $\mu_r(K_r)$ and $R(K_r)$ for $r = 2,4,6,8$ and 10 are given in Table (2.2).

1- If $r = 2$ then

$$b \cong 1.2880\left(\frac{S(a_1,a_2,\dots,a_d)}{A_3^2}\right)^{\frac{1}{5}} h^*,$$

$$B_x\left(\tilde{f}(x)\right) \cong 1.6590A_3\left(\frac{S(a_1,a_2,\dots,a_d)}{A_3^2}\right)^{\frac{2}{5}} B_x(\hat{f}^*(x))$$

$$\cong 1.6590A_3{}^{\frac{1}{5}}(S(a_1,a_2,\dots,a_d))^{\frac{2}{5}} B_x(\hat{f}^*(x))$$

$$Var(\tilde{f}(x)) \cong 2.7522A_3{}^{\frac{2}{5}}(S(a_1,a_2,\dots,a_d))^{\frac{4}{5}}Var(\hat{f}^*(x))$$

and

$$MSE\left(\tilde{f}(x)\right) \cong 2.7522(S^4(a_1,a_2,\dots,a_d)A_3^2)^{\frac{1}{5}}MSE(\hat{f}^*(x))$$

2- If $r = 4$ then

$$b \cong 1.0860\left(\frac{S(a_1,a_2,\dots,a_d)}{A_5^2}\right)^{\frac{1}{9}} h^*.$$

$$B_x\left(\tilde{f}(x)\right) \cong -1.3909A_5\left(\frac{S(a_1,a_2,\dots,a_d)}{A_5^2}\right)^{\frac{4}{9}} B_x(\hat{f}^*(x))$$

$$= -1.3909(A_5S^4(a_1,a_2,\dots,a_d))^{\frac{1}{9}} B_x(\hat{f}^*(x))$$

$$Var(\tilde{f}(x)) \cong 1.9345(A_5^2S^8(a_1,a_2,\dots,a_d))^{\frac{1}{9}} Var(\hat{f}^*(x))$$

and

$$MSE\left(\tilde{f}(x)\right) \cong 1.9345(A_5^2S^8(a_1,a_2,\dots,a_d))^{\frac{1}{9}} MSE(\hat{f}^*(x))$$

3- If $r = 6$ then

$$b \cong 1.03694\left(\frac{S(a_1,a_2,\dots,a_d)}{A_7^2}\right)^{\frac{1}{13}} h^*.$$

$$B_x\left(\tilde{f}(x)\right) \cong 1.2432A_7\left(\frac{S(a_1,a_2,\dots,a_d)}{A_7^2}\right)^{\frac{6}{13}} B_x(\hat{f}^*(x))$$

$$= 1.2432(A_7S^6(a_1,a_2,\dots,a_d))^{\frac{1}{13}} B_x(\hat{f}^*(x))$$

$$Var(\tilde{f}(x)) \cong 1.5455(A_7^2S^{12}(a_1,a_2,\dots,a_d))^{\frac{1}{13}} Var(\hat{f}^*(x))$$

and

$$MSE\left(\tilde{f}(x)\right) \cong 1.5455(A_7^2S^{12}(a_1,a_2,\dots,a_d))^{\frac{1}{13}} MSE(\hat{f}^*(x))$$

4- If $r = 8$ then

$$b \cong 1.0172\left(\frac{S(a_1,a_2,\dots,a_d)}{A_9^2}\right)^{\frac{1}{17}} h^*.$$

$$B_x\left(\tilde{f}(x)\right) \cong -1.1465A_9\left(\frac{S(a_1,a_2,\dots,a_d)}{A_9^2}\right)^{\frac{8}{17}} B_x(\hat{f}^*(x))$$

$$= -1.1465(A_9S^8(a_1,a_2,\dots,a_d))^{\frac{1}{17}} B_x(\hat{f}^*(x))$$

$$Var(\tilde{f}(x)) \cong 1.3144(A_9^2S^{16}(a_1,a_2,\dots,a_d))^{\frac{1}{17}}Var(\hat{f}^*(x))$$

and

$$MSE\left(\tilde{f}(x)\right) \cong 1.3144(A_9^2S^{16}(a_1,a_2,\dots,a_d))^{\frac{1}{17}}MSE(\hat{f}^*(x))$$

5- If $r = 10$ then

$$b \cong 1.0074\left(\frac{S(a_1,a_2,\dots,a_d)}{A_{11}^2}\right)^{\frac{1}{21}} h^*.$$

$$B_x\left(\tilde{f}(x)\right) \cong 1.0766A_{11}\left(\frac{S(a_1,a_2,\dots,a_d)}{A_{11}^2}\right)^{\frac{10}{21}} B_x(\hat{f}^*(x))$$

$$= 1.0766(A_{11}S^{10}(a_1,a_2,\dots,a_d))^{\frac{1}{21}} B_x(\hat{f}^*(x))$$

$$Var(\tilde{f}(x)) \cong 1.1590(A_{11}^2S^{20}(a_1,a_2,\dots,a_d))^{\frac{1}{21}}Var(\hat{f}^*(x))$$

and

$$MSE\left(\tilde{f}(x)\right) \cong 1.590(A_{11}^2S^{20}(a_1,a_2,\dots,a_d))^{\frac{1}{21}}MSE(\hat{f}^*(x))$$

CHAPTER THREE

CONSTRUCTION OF THE PROPOSED ESTIMATOR AND ASYMPTOTIC COMPARISON

3.1 Introduction

In this chapter, we compare the estimator $\tilde{f}(x)$ and $\hat{f}^{*}(x)$ in terms of bias and asymptotic MSE. This required to determine the proposed estimator weights $a_1, a_2, \ldots, a_d$ and the value of the constant d. The global measure; asymptotic mean square error (MSE) is considered for purpose of comparison, which is equivalent to use the asymptotic mean integrated squares error $(MISE)$.

3.2 Construction of the Proposed Estimator

Recall that the expected value of the proposed estimator is (Formula 2.10)

$$E\left(\tilde{f}(x)\right) = f(x)A_1 - bf^{(1)}(x)\mu_1(K)\,A_2 + \frac{b^2}{2}f^{(2)}(x)\mu_2(K)A_3$$

$$+\cdots+\frac{b^r f^{(r)}(x)}{r!}\mu_r(K)A_{r+1} + o(b^r),$$

where $A_l = \sum_{j=1}^{d} j^l a_j$, $r = 2, 4, 6, \ldots$ and

$$\mu_r(K) = \begin{cases} \int_{-\infty}^{\infty} z^r K(z) dz, & r = 2, 4, 6, \dots, \\ 0 & , \quad r = 1, 3, 5, \dots. \end{cases}$$

We obtain the $O(b^r)$ bias for $\tilde{f}(x)$, under the constraints $A_1 = 1$ and $A_3 = A_5 = A_7 = \dots = A_{r-1} = 0$.For example, to obtain $O(b^2)$, only $A_1 = 1$ is required, while to obtain $O(b^4)$ the constraints $A_1 = 1$ and $A_3 = 0$ are required and so on. In addition to the above constraints and for each even value of r, the term A_{r+1} should not equal zero (i.e. $A_{r+1} \neq 0$). If $A_{r+1} = 0$ then we obtain $O(b^{r+2})$ bias not $O(b^r)$. This also can be concluded by inspection the formula of the smoothing parameter b (Formula 2.14), which is no longer applicable if $A_{r+1} = 0$. Therefore and in addition to the above constraints we need to introduce another constraint about the value of A_{r+1}. In this study, we chose to fix the value of A_{r+1} at a specific value λ chosen to be 0.5 for all different values of r.

To construct the proposed estimator, the weights $a_1, a_2, \dots, a_d$ and the constant d should be determined. If an $O(b^r)$ bias is required, then we need to find $a_1, a_2, \dots, a_d$ under the constraints $A_1 = 1$, $A_3 = A_5 = A_7 = \dots = A_{r-1} = 0$ and $A_{r+1} = \lambda$. Because r is even value then the number of constraints is $\frac{r}{2} + 1$. This requires the integer value of d to be $\frac{r}{2} + 1$ (i.e. $d = \frac{r}{2} + 1$). For example, if $r = 2$ then the required constraints on $a_1, a_2, \dots, a_d$ are $A_1 = a_1 + 2a_2 + \dots + da_d = 1$ and $A_3 = a_1 + 8a_2 + \dots + d^3 a_d = \lambda$. The unique solution for the two equations can be obtained if $d = \frac{r}{2} + 1$ (i.e. $d = 2$). The convergence rate for a bias of $\tilde{f}(x)$ and the corresponding required constraints are given in Table (3.1).

Now, the other more logical method to determine $a_1, a_2, \dots, a_d$ is to find their values that minimize the asymptotic MSE of $\tilde{f}(x)$ under the above mentioned constraints. The word minimization technique requires d to be greater than $\frac{r}{2} + 1$ (i.e. $d > \frac{r}{2} + 1$). We note here that, when d equals the number of constraints then the values of $a_1, a_2, \dots, a_d$ are determined without minimizing the MSE, while the increasing in the value of d over the number of constraints gives many sets of candidate values for the weights and we need to determine the set of values that minimize MSE.

In the following sections, we will consider a $O(b^r)$ bias for both estimators; $\hat{f}^*(x)$ and $\tilde{f}(x)$, where the five even values $r = 2, 4, 6, 8, 10$ are studied separately. For each value of r, the asymptotic properties of the proposed estimator $\tilde{f}(x)$ are investigated by choosing the three values of d; $d = \frac{r}{2} + 1$ (no minimization for MSE), $d = \frac{r}{2} + 2$ and $d = \frac{r}{2} + 3$. The value of λ is taken to be 0.5 for all different cases and the second order kernel function $K_{(2)} = K$ is $N(0, 1)$. The minimization problems in this thesis are solved by using mathematica (6) .

3.3 The Case $O(b^2)$ Bias

As we mentioned above, given that $O(b^r)$ bias is desirable, there are several ways to construct the proposed estimator, $\tilde{f}(x)$, i.e. to determine d and $a_1, a_2, \ldots, a_d$. If the order of bias is $O(b^2)$ (i.e. $r = 2$) then we will consider three values of d; $d = 2, 3, 4$ according to the above rules. Therefore, we can obtain three proposed estimators, which are

$$\tilde{f}(x) = \frac{1}{nb}\sum_{j=1}^{d}\sum_{i=1}^{n} a_j K\left(\frac{x - X_i}{jb}\right), \quad d = 2, 3, 4.$$

The values of $a_1, a_2, \ldots, a_d$ are determined under the two constraints:

$$A_1 = \sum_{j=1}^{d} j a_j = 1 \text{ and } A_3 = \sum_{j=1}^{d} j^3 a_j = (\lambda = 0.5). \quad (3.1)$$

The three cases of d are treated as the following:

(1) If $d = 2$, then the solution of the two equations in (3.1) gives

$$a_1 = 1.1667 \text{ and } a_2 = -0.0833.$$

In addition,

$QQ_2 = 0.25, S(a_1, a_2) = 0.3185, VV_2 = 1.0496$ and $TT_2 = 0.3793$.

(2) If $d = 3$, then the minimization of $S(a_1, a_2, a_3)$ subject to the constraints of (3.1) gives

$$a_1 = -0.6355, \quad a_2 = 1.358 \text{ and } a_3 = -0.3604,$$

which yields,

$QQ_2 = 0.25, S(a_1, a_2, a_3) = 0.1722\,, VV_2 = 0.9281$ and $TT_2 = 0.2319.$

(3) If $d = 4$, then the minimization of $S(a_1, a_2, a_{3,} a_4)$ subject to the constraints of (3.1) gives

$$a_1 = 0.3898\ ,\ a_2 = -1.5648\ ,\ a_3 = 2.2480 \text{ and } a_4 = -0.7510,$$

and then,

$QQ_2 = 0.25, S(a_1, a_2, a_3, a_4) = 0.1126\,, VV_2 = 0.8526$ and $TT_2 = 0.1651$

The above three values of d corresponds three different estimators based on $\tilde{f}(x)$. Now, let $b_r^{(d)}$ and $MSE_r^{(d)}$ be the smoothing parameter and the asymptotic mean square error of $\tilde{f}(x)$, then based on the above values of r, d and $a_1, a_2, \dots, a_d$, we obtain,

$$b_2^{(2)} = 1.0496\,[\frac{f(x)}{f^{(2)^2}(x)}]^{1/5}\ [n]^{-1/5} \quad , \quad MSE_2^{(2)} = 0.3793[f(x)]^{4/5}[f^{(2)^2}(x)]^{1/5}[n]^{-4/5}$$

$$b_2^{(3)} = 0.9281\,[\frac{f(x)}{f^{(2)^2}(x)}]^{1/5}\ [n]^{-1/5} \quad , \quad MSE_2^{(3)} = 0.2319[f(x)]^{4/5}[f^{(2)^2}(x)]^{1/5}[n]^{-4/5}$$

and

$$b_2^{(4)} = 0.8526\,[\frac{f(x)}{f^{(2)^2}(x)}]^{1/5}\ [n]^{-1/5} \quad , \quad MSE_2^{(4)} = 0.1651[f(x)]^{4/5}[f^{(2)^2}(x)]^{1/5}[n]^{-4/5}.$$

For $r = 2$ and based on Table (2.2), the smoothing parameter h^* and the MSE of the higher-order kernel estimator $\hat{f}^*(x)$ are respectively

$$h^* = 0.7764\,[\frac{f(x)}{f^{(2)^2}(x)}]^{1/5}\ [n]^{-1/5} \quad , \quad MSE = 0.4542[f(x)]^{4/5}[f^{(2)^2}(x)]^{1/5}[n]^{-4/5}.$$

By comparing MSE($\hat{f}^*(x)$),$\text{MSE}_2^{(2)}$, $MSE_2^{(3)}$ and $MSE_2^{(4)}$ we see that

$$\text{MSE}(\hat{f}^*(x))=1.1974 MSE_2^{(2)}=1.9586 MSE_2^{(3)}= 2.751 MSE_2^{(4)} \quad (3.2)$$

Also, it follows from the formulas of h^*, $b_2^{(2)}$, $b_2^{(3)}$ and $b_2^{(4)}$ that the bandwidth of the different estimators are related such that

$$h^*=0.7397 b_2^{(2)}=0.8365 b_2^{(3)}= 0.9106 b_2^{(4)} \quad (3.3)$$

Formulas (3.2)and (3.3) state that, if the smoothing parameter of the different estimators are computed according to Formula(3.3), then the asymptotic MSE of the proposed estimator with d=2,3 and 4 are smaller than that of the higher-order kernel estimator.

3.4 The Case $O(b^4)$ Bias

The case $O(b^4)$ bias for $\tilde{f}(x)$ indicates $r = 4$ and therefore the suggested three values of d are $d = 3, 4, 5$. The corresponding proposed estimator is

$$\tilde{f}(x)=\frac{1}{nb}\sum_{j=1}^{d}\sum_{i=1}^{n}a_j K\left(\frac{x-X_i}{jb}\right), \quad d = 3, 4, 5.$$

The values of $a_1, a_2, \ldots, a_d$ are determined under the three constraints:

$$A_1 = \sum_{j=1}^{d} j a_j = 1, A_3 = \sum_{j=1}^{d} j^3 a_j = 0 \text{ and } A_5 = \sum_{j=1}^{d} j^5 a_j = (\lambda = 0.5) \quad (3.4)$$

(1) If $d = 3$, we need to solve the three equations of (3.2) simultaneously. This gives

$$a_1 = 1.5208 \;, \quad a_2 = -0.3167 \quad \text{and} \quad a_3 = 0.0375\,,$$

Based on these values of a_1, a_2, a_3, we obtain,

$QQ_4 = 0.0625, S(a_1, a_2, a_3) = 0.3939\,, VV_4 = 1.3252$ and $TT_4 = 0.3344$.

Also,

$b_4^{(3)} = 1.3252\,[\frac{f(x)}{f^{(4)^2}(x)}]^{1/9}\;[n]^{-1/9}$ and $MSE_4^{(3)} = 0.3344[f(x)]^{8/9}[f^{(4)^2}(x)]^{1/9}[n]^{-8/9}$

(2) If $d = 4$, then we want to minimize $S(a_1, a_2, a_3, a_4)$ subject to the three constraints of (3.2), which gives

$$a_1 = -0.7091 \;, \quad a_2 = 1.9132 \quad , \quad a_3 = -0.9182 \quad \text{and} \quad a_4 = 0.1593.$$

These values and the values $r = 4$ and $d = 4$ give,

$QQ_4 = 0.0625, \quad S(a_1, a_2, a_3, a_4) = 0.2088\,, \quad VV_4 = 1.2350$ and $TT_4 = 0.1902$.

Also,

$b_4^{(4)} = 1.235\,[\frac{f(x)}{f^{(4)^2}(x)}]^{1/9}\;[n]^{-1/9}$ and $MSE_4^{(4)} = 0.1902[f(x)]^{8/9}[f^{(4)^2}(x)]^{1/9}[n]^{-8/9}$

(3) If $d = 5$, then

$a_1 = 0.4176\,, a_2 = -1.9654\,,\; a_3 = 3.6927, a_4 = -2.2146$ and $a_4 = -2.2146$,

which are obtained by minimizing $S(a_1, a_2, a_3, a_4, a_5)$ subject to the constraints of (3.2). Accordingly, we obtain,

$QQ_4 = 0.0625,\ S(a_1, a_2, a_3, a_4,, a_5) = 0.1403$, $VV_4 = 1.1816$ and $TT_4 = 0.1336$.

Also,

$$b_4^{(5)} = 1.1816\, [\frac{f(x)}{f^{(4)^2}(x)}]^{1/9}\ [n]^{-1/9} \text{and } MSE_4^{(5)} = 0.1336[f(x)]^{8/9}[f^{(4)^2}(x)]^{1/9}[n]^{-8/9}.$$

If $r = 4$, we can use Table (2.2) to produce the smoothing parameter and the asymptotic (MSE) of the higher-order kernel which are respectively given by

$$h^* = 1.1602\, [\frac{f(x)}{f^{(4)^2}(x)}]^{1/9}\ [n]^{-1/9} \quad \text{and} \quad MSE = 0.4616[f(x)]^{8/9}[f^{(4)^2}(x)]^{1/9}[n]^{-8/9}$$

By comparing MSE($\hat{f}^*(x)$),$MSE_4^{(3)}$, $MSE_4^{(4)}$ and $MSE_4^{(5)}$ we see that

$$\text{MSE}(\hat{f}^*(x))\text{=}1.1974MSE_4^{(3)}\text{=}1.9586MSE_4^{(4)}\text{=}\ 2.751MSE_4^{(5)} \qquad (3.5)$$

Also, it follows from h^*,$b_4^{(3)}$, $b_4^{(4)}$ and $b_4^{(5)}$ that the bandwidth of the different estimator are related such that

$$h^*\text{=}0.7397b_4^{(3)}\text{=}0.8365b_4^{(4)}\text{=}\ 0.9106b_4^{(5)} \qquad (3.6)$$

For any value of h^* taking $b_4^{(3)} = 1.3519h^*$ makes $MSE_4^{(3)} = 0.835\text{MSE}\ (\hat{f}^*(x))$, taking $b_4^{(4)} = 1.195h^*$ makes $MSE_4^{(4)} = 0.51\text{MSE}(\hat{f}^*(x))$ and taking $b_4^{(5)} = 1.098h^*$ makes $MSE_4^{(5)} = 0.363\text{MSE}(\ \hat{f}^*(x))$.

3.5 The Case $O(h^6)$ Bias

This case indicates that $r = 6$ and therefore the three values of d are $d = 4, 5,6$.

The proposed estimator is represented as

$$\tilde{f}(x)=\frac{1}{nb}\sum_{j=1}^{d}\sum_{i=1}^{n}a_jK\left(\frac{x-X_i}{jb}\right),\quad d = 4,5,6.$$

The values of $a_1, a_2, \ldots, a_d$ are determined under the four constraints:

$A_1 = \sum_{j=1}^{d} ja_j = 1, A_3 = \sum_{j=1}^{d} j^3 a_j = 0 , A_5 = \sum_{j=1}^{d} j^5 a_j = 0$ and

$$A_7 = \sum_{j=1}^{d} j^7 a_j = (\lambda = 0.5) \tag{3.7}$$

(1) If $d = 4$ then

$$a_1 = 1.5986\,,\quad a_2 = -1.9654\,,\quad a_3 = 0.0756 \text{ and } a_4 = -0.0070$$

This value are obtained by solving equation in (3.3) . Based on these values, we obtain

$$QQ_6 = 0.0104, S(a_1, a_2, a_3, a_4) = 0.4064\,, VV_6 = 1.5556 \text{ and } TT_6 = 0.2831.$$

Also,

$$b_6^{(4)} = 1.5556\,[\frac{f(x)}{f^{(6)^2}(x)}]^{1/13}[n]^{-1/13} \text{ and } MSE_6^{(4)} = 0.2831[f(x)]^{12/13}[f^{(6)^2}(x)]^{1/13}[n]^{-12/13}$$

(2) If $d = 5$, then we want to minimize $S(a_1, a_2, \ldots, a_5)$ subject to the four constraints of (3.3), which gives

$$a_1 = -0.7298\,,\quad a_2 = 2.2624\,,\quad a_3 = -1.4212\,,\quad a_4 = 0.4365 \text{ and}$$

$$a_5 = -0.0554$$

These values and the values $r = 6$ and $d = 5$ give,

$$QQ_6 = 0.0104,\quad S(a_1, a_2, \ldots, a_5) = 0.2265\,,\quad VV_6 = 1.4871 \quad\text{and}\quad TT_6 = 0.1650$$

Also,

$b_6^{(5)} = 1.4871\,[\frac{f(x)}{f^{(4)^2}(x)}]^{1/13}\ [n]^{-1/13}$ and $MSE_6^{(5)} = 0.1650[f(x)]^{12/13}[f^{(6)^2}(x)]^{1/13}[n]^{-12/13}$

(3) If $d = 6$, then

$a_1 = 0.4258$, $a_2 = -2.2227$, $a_3 = 4.9473$, $a_4 = -4.0029$ and $a_5 = 1.5064$,

$$a_6 = -0.2238$$

which are obtained by minimizing $S(a_1, a_2, \dots, a_6)$ subject to the constraints of (3.3). Accordingly, we obtain,

$QQ_6 = 0.0104$, $S(a_1, a_2, \dots, a_6) = 0.1570$, $VV_6 = 1.4458$ and $TT_6 = 0.1177$.

Also,

$b_6^{(6)} = 1.4458\,[\frac{f(x)}{f^{(6)^2}(x)}]^{1/13}\ [n]^{-1/13}$ and $MSE_6^{(6)} = 0.1177[f(x)]^{12/13}[f^{(6)^2}(x)]^{1/13}[n]^{-12/13}$.

For $r = 6$ and based on Table (2.2), the smoothing parameter h^* and the MSE of the higher-order kernel estimator $\hat{f}^*(x)$ are respectively

$h^* = 1.4451\,[\frac{f(x)}{f^{(6)^2}(x)}]^{1/13}\ [n]^{-1/13}$ and $MSE = 0.4678[f(x)]^{12/13}[f^{(6)^2}(x)]^{1/13}[n]^{-12/13}$

by comparing MSE($\hat{f}^*(x)$),$MSE_6^{(4)}$, $MSE_6^{(5)}$ and $MSE_6^{(6)}$ we see that

$$\text{MSE}(\hat{f}^*(x)) = 1.6524 MSE_6^{(4)} = 2.8351 MSE_6^{(5)} = 3.974 MSE_6^{(6)} \qquad (3.8)$$

Also, it follows from h^*,$b_6^{(4)}$, $b_6^{(5)}$ and $b_6^{(6)}$ that the bandwidth of the different estimator are related such that

© Arabic Digital Library - Yarmouk University

$$h^{*}=0.9282b_6^{(4)}=0.7917b_6^{(5)}=0.9995b_6^{(6)} \qquad (3.9)$$

Equation (3.8) indicates that the proposed estimators $\tilde{f}_6^4(x)$, $\tilde{f}_6^5(x)$ and $\tilde{f}_6^6(x)$ are more than 39% , 64% and 74% better than estimator $\hat{f}^*(x)$ respectively in MSE. The other cases for any convergent rate of bias can be treated in the same manner. The results of the tow cases $O(b^8)$ and $O(b^{10})$ are given in Tables (3.1) and (3.2) together with the results of the previous section.

Table (3.1): Convergence rate of $O(b^r)$ bias and constraints for the proposed estimator and the values of weights for different values of r and d.

r	d	$a_1, a_2, \dots, a_d$	Constraints	$O(b^r)$
2	2	$a_1 = 1.1667$, a_2=-0.0833	$A_1 = 1,\ A_3 = 0.5$	$O(b^2)$
	3	$a_1 = -0.6355$, a_2=1.3584 a_3=-0.3604		
	4	$a_1 = 0.3898$, a_2=-1.5648 a_3=2.2480, a_4=-0.7510		
4	3	$a_1 = 1.5208$, a_2=-0.3167 a_3= 0.0375	$A_1 = 1,\ A_3 = 0$ $A_5 = 0.5$	$O(b^4)$
	4	$a_1 = -0.7091$, a_2=1.9132 a_3=-0.9182, a_4=0.1593		
	5	$a_1 = 0.4176$, a_2=-1.9654 a_3=3.6927 a_4=-2.2146, a_5= 0.4587		
6	4	$a_1 = 1.5986$, a_2=-0.3986 a_3=0.0756, a_4=-0.0070	$A_1 = 1,\ A_3 = 0$ $A_5 = 0,\ A_7 = 0.5$	$O(b^6)$
	5	$a_1 = -0.7298$ a_2=2.2624 a_3=-1.4212 a_4=0.4365, a_5=-0.0554		
	6	$a_1 = 0.4258$, a_2=-2.2227 a_3=4.9473 a_4=-4.0029, a_5=1.5064, a_6=-0.2238		
8	5	$a_1 = 1.6667$ a_2=-0.4763 a_3=0.1191 a_4=-0.0199, a_5=0.0016	$A_1 = 1,\ A_3 = 0$ $A_5 = 0,\ A_7 = 0$ $A_9 = 0.5$	$O(b^8)$
	6	$a_1 = -0.7445$, a_2=2.5377 a_3=-1.8902 a_4=0.7839, a_5=-0.1811, a_6=-0.0183		
	7	$a_1 = 0.4290$ a_2=-2.4132 a_3=6.0559 a_4=-5.9421, a_5=3.0752, a_6=-0.8434 a_7=0.0975		
10	6	$a_1 = 1.7143$ a_2=-0.5357 a_3= 0.1587 a_4=-0.0357 a_5=0.0052 a_6=-0.0004	$A_1 = 1,\ A_3 = 0$ $A_5 = 0,\ A_7 = 0$ $A_9 = 0,\ A_{11} = 0.5$	$O(b^{10})$
	7	$a_1 = -0.7546$ a_2=2.7562 a_3=-2.3102 a_4=1.1613, a_5=-0.3689 a_6=0.0687 a_7=-0.0058		
	8	$a_1 = 0.4149$ a_2=-2.5013 a_3=6.9279 a_4=-7.8105 a_5=4.9805, a_6=-1.9036 a_7=0.4107 a_8=-0.0388		

Note that: $A_l = \sum_{j=1}^{d} j^l a_j$.

Table (3.2): The values of r, d and the corresponding values of $QQ_r, S(a_1, a_2, \dots, a_d)$ TT_r and VV_r

r	d	QQ_r	$S(a_1, a_2, \dots, a_d)$	VV_r	TT_r
	2		0.3185	1.0496	0.3793
2	3	0.2500	0.1722	0.9281	0.2319
	4		0.1126	0.8526	0.1651
	3		0.3939	1.3252	0.3344
4	4	0.0625	0.2088	1.2350	0.1902
	5		0.1403	1.1816	0.1336
	4		0.4064	1.5556	0.2831
6	5	0.0104	0.2265	1.4871	0.1650
	6		0.1570	1.4458	0.1177
	5		0.4171	1.7631	0.2514
8	6	0.0013	0.2385	1.7061	0.1485
	7		0.1680	1.6776	0.1073
	6		0.4243	1.9514	0.2283
10	7	0.00013	0.2471	1.9018	0.1364
	8		0.1775	1.8721	0.0996

Note that:

$$QQ_r = \frac{\mu_r(K)A_{r+1}}{r!} \quad , \quad S(a_1,\ a_2,\ \dots,\ a_d) = \int_{-\infty}^{\infty} K^2(u)\, du \sum_{j=1}^{d} j a_j{}^2 + 2\sum_{j<l}\sum a_j a_l \int_{-\infty}^{\infty} K\left[\frac{u}{j}\right] K\left[\frac{u}{l}\right] du$$

$$VV_r = \left[\frac{(r!)^2 S(a_1, a_2, \dots, a_d)}{(2r)\mu^2{}_r(K)A_{r+1}^2}\right]^{\frac{1}{2r+1}} , \qquad TT_r = \frac{2r+1}{\{2r\}^{2r/(2r+1)}\{r!\}^{2/(2r+1)}} [S(a_1, a_2, \dots, a_d)]^{\frac{2r}{2r+1}} \left[\mu_r{}^2(k)A^2{}_{r+1}\right]^{\frac{1}{2r+1}}$$

and $A_{r+1} = 0.5$ for $r = 2, 4, 6, 8, 10$.

CHAPTER FOUR

CONCLUSIONS AND SUGGESTION FOR FUTURE WORK

4.1 Conclusions

On the basis of the previous two chapter, we conclude the following:

A general new estimator for the continuous *pdf* *f(x)* is proposed. The asymptotic properties of the proposed estimator are derived. The mathematical comparisons between the proposed estimator and the higher order kernel estimator shows the superiority of the proposed estimator over the higher order kernel estimator. The proposed estimator depends on same quantities; $d, a_1, a_2, \ldots, a_d$ which were under the user control. The choice of these quantities plays a vital role in the bias convergence rate of the proposed estimator and in its variance quantity. Despite that the asymptotic properties of the proposed estimator were derived for $x \in (a, \infty)$, where $a = -\infty$, it is straightforward to derive its asymptotic properties for any finite value of a.

It is worthwhile to mention here that we performed a limited simulation study for finite sample size. The simulation study results give good performances for the proposed estimator over that of the higher-order kernel estimator.

4.2 suggestions for further research

How is the performance of the proposed estimator $\tilde{f}(x)$ compared to the classical kernel estimator $\hat{f}(x)$ and the higher order kernel estimator $\hat{f}^{*}(x)$ for finite samples? To answer this question a simulation study needs to be addressed. Marron and Wand (1992) gave different normal mixture densities ,which commonly used to study the nonparametric estimator for the probability density function *f(x)*. the Marron and Wand's densities represent symmetric, kurtosis, unimodalies, bimodal, trimodal, skewed, and strongly skewed distribution. These densities can be used to study and to compare the finite sample properties of the estimator $\tilde{f}(x)$, $\hat{f}(x)$ and $\hat{f}^{*}(x)$ in the future research . In addition, the value of $A_{r+1} = \lambda$ is fixed at 0.5 in this thesis. A possible future work can be adopted to study the influence of choosing λ on the performance of $\tilde{f}(x)$.

REFERENCES

Ali, A. (1998). Nonparametric Spatial Rainfall Characterization Using Adaptive Kernel Estimator. Journal of Geographic Information and Decision Analysis, vol. 2, no. 2, pp. 34-43.

Eidous.O.(2011). New Reflection kernel model using line transect sampling. Submitted to Computational Statistics and Data Analysis .

Jeter, J. (2005). Comparison Between the Exact and the approximate Bandwidths in the Kernel Density Estimation. Master's thesis, Southern Illinois University .

Jones, M.C. and Foster, P. J. (1993). Generalized Jackknifing and higher order. *Nonparametric Statistics*, 3, 81-94.

Marron,J.S. Wand,M.P.(1992).Exact mean integrated squared error. Annals of statistics ,20,712-736.

Silverman, B. W. (1986). Density Estimation for Statistics and Data Analysis. Chapman &hall. London.

Wand M. P. and Jones. M. C. (1995). Kernel smoothing. Chapman & hall.

Wand M.P. and Schucany W.R. (1990). Gaussian-based kernel. The Canadian Journal of Statistics, 18(3), 197-204.