8-9-2019

On gamma kernel function in recursive density estimation

Xiaoxiao Ma

On gamma kernel function in recursive density estimation

By

Xiaoxiao Ma

A Thesis
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Master of Science
in Statistics
in the Department of Mathematics and Statistics

Mississippi State, Mississippi

August 2019

On gamma kernel function in recursive density estimation

By

Xiaoxiao Ma

Approved:

Prakash N. Patil
(Major Professor)

Jonathan R. Woody
(Committee Member)

Tung-Lung Wu
(Committee Member)

Mohammad Sepehrifar
(Graduate Coordinator)

Rick Travis
Dean
College of Arts & Sciences

Name: Xiaoxiao Ma

Date of Degree: August 9, 2019

Institution: Mississippi State University

Major Field: Statistics

Major Professor: Prakash N. Patil

Title of Study: On gamma kernel function in recursive density estimation

Pages of Study: 58

Candidate for Degree of Master of Science

In this thesis we investigate the convergence rate of gamma kernel estimators in recursive density estimation. Unlike the traditional symmetric and fixed function, the gamma kernel is a kernel function with bounded support and varying shapes. Gamma kernels have been used to address the boundary bias problem which occurs when a symmetric kernel is used to estimate a density which has support on $[0, \infty)$.

The recursive density estimation is useful when an 'additional data' (on-line) comes from the population density which we want to estimate. We utilize the ideas and results from the adaptive kernel estimation to show that the L_2 convergence rate of the recursive kernel density estimators which use gamma kernels is $n^{-4/5}$.

Key words: kernel estimation, bandwidth, convergence rate, standard kernel, boundary bias, gamma kernel, recursive estimation

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Patil. This is my first academic research pertaining to Statistics. Because I did not have much experience writing papers or doing research in statistics before writing this thesis, I encountered many difficulties when I began this research. Dr. Patil has guided my learning and has inspired my thinking not only of concepts related to this thesis but also of the broad range of concepts within the field of statistics at large. I am grateful for his help, which will greatly impact my future study. Words cannot express my appreciation for Dr. Patil. I would also like to thank the commitee members, Dr. Woody and Dr. Wu for their suggestions and guidance.

I would like to thank all the professors who taught me. Whether it was for four classes or just for one day, they all helped me grow from a student who did not know anything about statistics to a student who has the ability to do research. I would also like to thank the Department of Mathematics and Statistics and Mississippi State University. What I have achieved so far has been possible because of my acceptance as a graduate student two years ago.

Finally, I would like to thank my family. There were several times when I did not think that I could complete the thesis. They unconditionally encouraged me and believed that I could do it. Without their support, I could never have made it.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION

A histogram is a classical density estimator, see for example, Scott [10] for the properties of histogram estimation. It is known that the convergence rate of $MISE$ of a histogram estimator is $n^{-2/3}$, which is too slow compared to n^{-1}, the parametric convergence rate. Aside from this issue, one encounters a couple of other problems. First, bias is larger near the edge of the bins as opposed to at the center of the bin. Second, for different starting points for bins of the same bin width, one receives different histograms. This leads to another method, referred to as kernel method, which improves $MISE$ convergence rate, and to a certain extent also addresses the problems metioned.

The kernel method is one of the most commonly used density estimators. There is a vast amount of literature on kernel estimation, such as Silverman [11], Wand and Jones [12], Rao [9], et cetera. For a kernel density estimate one usually chooses symmetric, non-negative density function as a kernel function. Even though symmetric kernel functions perform well in many situations, when the density has a bounded support, the problem of boundary bias appears. In order to resolve this, the bounded kernel function with the support matching the density and varying shape could be used. For this purpose Chen [5] introduces the gamma kernel estimator and explores its properties.

It should be noted that the kernel estimator mentioned in the last paragraph assumes that all data observations are available before we compute the estimator. If extra data, for example, on-line data, is to be used to construct an estimator of the density, we may need to consider the recursive density estimation and not non-recursive estimation. For basic concepts and notations for resursive estimation, see Rao [9]. Hall and Patil [7] study the performance of on-line kernel density estimators of a general form. From discussion there, it is clear that the recursive estimators are of the special form of the on-line density estimators. The standard kernel function when used in recursive estimation also causes the boundary bias for the density with the support on [0, ∞). Since in non-recursive estimation, gamma kernels address the boundary bias problem reasonably well, e.g., the resulting estimators have L_2 convergence rate of $n^{-4/5}$, here, interest is to explore the use of gamma kernels in recursive density estimation and the effect on the boundary bias. That is, interest is to compare the L_2 convergence rates of estimators which use the standard kernel and the gamma kernels.

This thesis is organized as follows: In chapter 2, we introduce basic ideas of the kernel estimation and its properties. For example, results providing asymptotical MSE and $MISE$, optimal bandwidth, convergence rate and choice of kernel functions for a classical kernel density estimator are stated. Chapter 3 gives the performance of the representative symmetric and non-negative kernel function, referred to as a standard kernel. In chapter 4, we introduce the boundary bias caused by a symmetric and fixed kernel in the bounded density with the support on $[0, \infty)$. We discuss the use of the gamma kernel in density estimation and in particular provide discussion on how it addresses the problem of boundary bias in chapter 5. In chapter 6, we consider the recursive density estimators and study the convergence rate of the standard kernel recursive estimator. In the last chapter, we replace the standard kernel with the gamma kernel in recursive density estimator and study the convergence rate of the resulting estimators.

CHAPTER II

KERNEL ESTIMATION

In order to address the two problems mentioned in the histogram estimation and improve the convergence rate, the first idea is to consider every X_i as the center of the bin. This way, no data observation is closed to the edge of the bin and one gets a histogram in each X_i. This means that the trouble caused by different starting points of the bin is addressed. Therefore, this estimator does not suffer from the couple of problems mentioned in Chapter 1 as much as the histogram estimator and it is $\hat{f}(x) = n^{-1}\sum_{i=1}^{n} h^{-1} I_{[-\frac{1}{2},\frac{1}{2}]}(\frac{x-X_i}{h})$. Nevertheless, one creates another problem that the estimator is too rough. Hence a smooth kernel function $K(x)$ is used in place of the function $I_{[-\frac{1}{2},\frac{1}{2}]}$.

2.1 General Idea

Silverman [11] and Wand and Jones [12] all illustrate the kernel method in density estimation. Let $X_1, X_2, ..., X_n$ be random variables from a probability density function $f(x)$ which we want to estimate. We define the kernel estimator as follows:

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K\left(\frac{x - X_i}{h}\right) \tag{2.1}$$

where K is the kernel function, which is a probability density function. So the kernel function satisfies $\int_{-\infty}^{+\infty} K(x)dx = 1$. h is called bandwidth or smoothing parameter. As h

tends to zero, the estimator shows more details and becomes rougher. When h is larger and larger, the curve which we estimate is smoother. Usually, the kernel function is symmetric and non-negative. A standard kernel function satisfies the properties

$$\int t^k K(t)dt \begin{cases} = 1, & \text{if } k = 0 \\ = 0, & \text{if } k = 1 \\ > 0, & \text{if } k = 2. \end{cases} \tag{2.2}$$

2.2 The Expectation and Bias of the Kernel Estimator

Mean square error (MSE) is used to assess the accuracy of the estimator. In density estimation, one could also use MSE to study the performance of a kernel estimator. It is known that $MSE = Var + Bias^2$. So in this section and 2.3, we derive the $E[\hat{f}(x)]$, $Bias(\hat{f}(x))$ and $Var(\hat{f}(x))$. For the kernel estimator in (2.1),

$$\begin{aligned} E[\hat{f}(x)] &= E\left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K\left(\frac{x-X_i}{h}\right)\right] \\ &= E\left[\frac{1}{h}K\left(\frac{x-X_i}{h}\right)\right] \\ &= \int \frac{1}{h}K\left(\frac{x-y}{h}\right)f(y)dy. \end{aligned} \tag{2.3}$$

By letting $y = x - ht$ and Taylor expansion, we obtain

$$\begin{aligned} E[\hat{f}(x)] &= \int K(t)f(x-ht)dt \\ &= \int K(t)\left[f(x) - htf'(x) + \frac{1}{2}f''(x)h^2t^2 + O(h^3)\right]dt \end{aligned}$$

where $f(x-ht) = f(x) + \frac{f'(x)}{1!}(x-ht-x) + \frac{f''(x)}{2!}(x-ht-x)^2 + O(h^3)$. Then,

$$
\begin{aligned}
Bias(\hat{f}(x)) &= E[\hat{f}(x)] - f(x) \\
&= \int K(t)\left[f(x) - htf'(x) + \frac{1}{2}f''(x)h^2t^2 + O(h^3)\right]dt - f(x) \\
&= \int K(t)\left[f(x) - htf'(x) + \frac{1}{2}f''(x)h^2t^2 + O(h^3)\right]dt - \int K(t)dt \cdot f(x) \\
&= \int K(t)\left[-htf'(x) + \frac{1}{2}f''(x)h^2t^2 + O(h^3)\right]dt
\end{aligned}
$$

where $\int K(t)dt = 1$ determined by the prerequisite that $K(x)$ is a density function. Also, the kernel function $K(x)$ is defined in (2.2), in which $\int tK(t)dt = 0$. So the bias is

$$
\begin{aligned}
Bias(\hat{f}(x)) &= -hf'(x)\int tK(t)dt + \frac{1}{2}f''(x)h^2\int t^2K(t)dt + O(h^3) \\
&= \frac{1}{2}f''(x)h^2\int t^2K(t)dt + O(h^3).
\end{aligned}
$$

According to the result, $E[\hat{f}(x)]$ is not equal to $f(x)$, which indicates that the kernel estimator is biased. On the other hand, let $h \to 0$, $Bias(\hat{f}(x))$ tends to zero. Therefore, the kernel estimator $\hat{f}(x)$ is asymptotically unbiased. Furthermore, it should be noted that the first term includes $f''(x)$ which measures the roughness of $f(x)$. Accordingly, the rougher the density, the larger the bias.

2.3 The Variance of the Kernel Estimator

$$\begin{aligned}
Var(\hat{f}(x)) &= Var\left(n^{-1}\sum_{i=1}^{n} h^{-1}K\left(\frac{x-X_i}{h}\right)\right)\\
&= (nh^2)^{-1}\int K^2\left(\frac{x-y}{h}\right)f(y)dy - n^{-1}\left\{h^{-1}\int K\left(\frac{x-y}{h}\right)f(y)dy\right\}^2\\
&= (nh^2)^{-1}\int K^2\left(\frac{x-y}{h}\right)f(y)dy - n^{-1}\left\{E[\hat{f}(x)]\right\}^2\\
&= (nh^2)^{-1}\int K^2\left(\frac{x-y}{h}\right)f(y)dy - n^{-1}\left\{f(x)+Bias(\hat{f}(x))\right\}^2\\
&= (nh)^{-1}\int K^2(t)f(x-ht)dt - n^{-1}\left\{f(x)+O(h^2)\right\}^2\\
&= (nh)^{-1}\int K^2(t)\left\{f(x)-htf'(x)+......\right\}dt + O(n^{-1})
\end{aligned}$$

where $f(x-ht) = f(x) - htf'(x) + ...$ by Taylor expansion. In addition, we have obtained $Bias(\hat{f}(x)) = h^2\left\{\frac{1}{2}f''(x)\int t^2K(t)dt\right\} + O(h^3)$ in Section 2.2. One notices that h is included from the second term in the $R.H.S.$ of the Taylor expansion. Therefore, when $h \to 0$,

$$Var(\hat{f}(x)) = (nh)^{-1}f(x)\int K^2(t)dt + O(n^{-1}).$$

2.4 MSE and MISE of the Kernel Estimator

In this section, we use the bias and the variance obtained in previous sections to derive MSE, through which one could investigate the performance of the kernel estimators. However, MSE is dependent on the point x. It is obvious that we could learn about the performance of the kernel estimator by studying the performance of $\hat{f}(x)$ over the whole curve. Silverman [11], as well as many textbooks and academic literatures, has used different metrics to study the performance of $\hat{f}(x)$ over the whole curve. The most widely used criteria is mean integrated square error,

$$\begin{aligned} MISE(\hat{f}(x)) &= \int E\left[\hat{f}(x) - f(x)\right]^2 dx \\ &= \int \left\{ var(\hat{f}(x)) + \left[bias(\hat{f}(x))\right]^2 \right\} dx \\ &= \int MSE(\hat{f}(x)) dx. \end{aligned}$$

The other possible ways to investigate the performance of $\hat{f}(x)$ over the whole curve are ISE (Integrated Square Error) and IAE (Integrated Absolute Error), which are defined respectively as

$$\begin{aligned} ISE(\hat{f}(x)) &= \int \left[\hat{f}(x) - f(x)\right]^2 dx \\ IAE(\hat{f}(x)) &= \int \left|\hat{f}(x) - f(x)\right| dx. \end{aligned}$$

Here, we derive the MSE and $MISE$ to investigate the performance of $\hat{f}(x)$.

2.4.1 MSE (Mean Square Error)

In section 2.2 and 2.3, we have obtained the Bias and variance of $\hat{f}(x)$. Now

$$\begin{aligned} MSE(\hat{f}(x)) &= (nh)^{-1} f(x) \int K^2(t)dt + O(n^{-1}) \\ &\qquad + \left\{ \frac{1}{2} f''(x) h^2 \int t^2 K(t)dt + O(h^3) \right\}^2 \\ &= (nh)^{-1} f(x) \int K^2(t)dt + O(n^{-1}) \\ &\qquad + \frac{1}{4} \{f''(x)\}^2 h^4 \left\{ \int t^2 K(t)dt \right\}^2 + O(h^5). \end{aligned}$$

Let $C_1 = f(x) \int K^2(t)dt$ and $C_2 = \frac{1}{4} \{f''(x)\}^2 \left\{\int t^2 K(t)dt\right\}^2$. Because C_1 and C_2 are fixed values, then $MSE(\hat{f}(x))$ depends on n and h. Therefore, we have

$$MSE(\hat{f}(x)) \sim (nh)^{-1} C_1 + C_2 h^4. \tag{2.4}$$

When $n \to \infty$, $h \to 0$ such that $(nh)^{-1} \to 0$, one can see that $MSE(\hat{f}(x)) \to 0$. This means that $\hat{f}(x)$ is consistent of $f(x)$. In addition, bandwidth h influences both the Var and $Bias$. The power of h is negative in variance term and is positive in bias term in (2.4). So if one tries to decrease either of variance and bias, the other one will be increased. At the same time, bandwidth h is smooth parameter. This means that as h tends to larger the estimate curve becomes smoother and one gets the result of larger bias and smaller variance, and vice versa.

One wants MSE as small as possible. Thus minimizing the $MSE(\hat{f}(x))$ respect to h, the optimal bandwidth h is,

$$\begin{aligned} \frac{dMSE(\hat{f}(x))}{dh} &= 0 \\ n^{-1}C_1(-h^{-2}) + C_2 4h^3 &= 0 \\ h_{opt} &= \left(\frac{C_1}{4nC_2}\right)^{1/5} \\ h_{opt} &= \left\{\frac{f(x)\int K^2(t)dt}{[f''(x)]^2[\int t^2K(t)dt]^2}\right\}^{1/5} \cdot n^{-1/5}. \end{aligned} \tag{2.5}$$

The expression of the optimal bandwidth in (2.5) includes $f(x)$ and the term $f''(x)$, which are from the probability density function $f(x)$ that we want to estimate. Therefore, we cannot know the optimal bandwidth h in practice. Obviously, if we knew the probability density function, we do not need to estimate it. However, we still can investigate the properties of the kernel estimators by (2.5). It could be noted from (2.5) that h_{opt} depends on $n^{-1/5}$ or

$$h_{opt} \sim n^{-1/5}. \tag{2.6}$$

Formula (2.6) indicates that as the number of data observations increases, the bandwidth h tends to zero. And the rate of $h \to 0$ is slower because of the power of n is $-1/5$.

After obtaining the optimal bandwidth h, we can derive the optimal mean square error by putting the optimal bandwidth h into (2.4).

$$\begin{aligned} MSE(\hat{f}(x)) &\sim n^{-1}C_1\left(\frac{C_1}{4nC_2}\right)^{-1/5} + C_2\left(\frac{C_1}{4nC_2}\right)^{4/5} \\ &\sim n^{-4/5}C_1^{4/5}4^{1/5}C_2^{1/5} + C_2^{1/5}C_1^{4/5}n^{-4/5}4^{-4/5} \\ &\sim n^{-4/5}C_1^{4/5}C_2^{1/5}(4^{1/5}+4^{-4/5}) \\ &\sim \frac{5}{4}\{f(x)\}^{4/5}\left\{\int K^2(t)dt\right\}^{4/5} \\ &\qquad \cdot \{f''(x)\}^{2/5}\left\{\int t^2K(t)dt\right\}^{2/5} n^{-4/5}. \end{aligned} \tag{2.7}$$

In terms of the result of (2.7), given the same samples size n and the optimal bandwidth h, the MSE's depend on the kernel function. One could let MSE be small by choosing the appropriate kernel function.

Although the above convergence rate is not as good as n^{-1}, the parameter convergence rate, it is better than $n^{-2/3}$, the convergence rate of MSE of histogram estimator.

Note also that compared to the histogram method, the convergence rate of optimal bandwidth h of the kernel estimator, $n^{-1/5}$ which is slower than $n^{-1/3}$, the convergence rate of optimal bandwidth h of histogram estimator.

2.4.2 MISE (Mean Integrated Square Error)

We have known that $\hat{f}(x)$ is consistent of $f(x)$ asymptotically. And in the beginning of this section, we mentioned that $MISE$ is used to study the performance of $\hat{f}(x)$ over the whole curve and $MISE = \int MSE(\hat{f}(x))dx$. Therefore,

$$\begin{aligned}
MISE(\hat{f}(x)) &= \int MSE(\hat{f}(x))dx \\
&= \int \left\{ Var(\hat{f}(x)) + \left[Bias(\hat{f}(x)) \right]^2 \right\} dx \\
&= \int \left\{ (nh)^{-1} f(x) \int K^2(t)dt + O(n^{-1}) \right\} dx \\
&\qquad + \int \left\{ \left[\frac{1}{2} f''(x) h^2 \int t^2 K(t)dt + O(h^3) \right]^2 \right\} dx \\
&= \int \left\{ (nh)^{-1} f(x) \int K^2(t)dt + O(n^{-1}) \right\} dx \\
&\qquad + \int \left\{ \frac{1}{4} [f''(x)]^2 h^4 \left[\int t^2 K(t)dt \right]^2 + O(h^5) \right\} dx \\
&= (nh)^{-1} \left[\int K^2(t)dt \right] + O(n^{-1}) \\
&\qquad + \frac{1}{4} h^4 \int [f''(x)]^2 dx \left[\int t^2 K(t)dt \right]^2 + O(h^5).
\end{aligned} \tag{2.8}$$

Let $C_3 = \int K^2(t)dt$ and $C_4 = \frac{1}{4} \int [f''(x)]^2 dx \cdot \left[\int t^2 K(t)dt \right]^2$. Because C_3 and C_4 are fixed values, then $MISE(\hat{f}(x))$ depends on n and h. That is

$$MISE \sim (nh)^{-1} C_3 + h^4 C_4. \tag{2.9}$$

This form is the same as the one of MSE in (2.4).

Let $n \to \infty$, $h \to 0$ such that $(nh)^{-1} \to 0$, then $MISE(\hat{f}(x)) \to 0$. This means that $\hat{f}(x)$ is consistent of $f(x)$ asymptotically. By differentiating $MISE(\hat{f}(x))$ and solving the equation,

$$\begin{aligned} \frac{dMISE(\hat{f}(x))}{dh} &= 0 \\ n^{-1}C_3(-h^{-2}) + 4h^3C_4 &= 0 \\ h^*_{opt} &= \left(\frac{C_3}{4nC_4}\right)^{1/5} \\ h^*_{opt} &= \left\{\frac{\int K^2(t)dt}{\int [f''(x)]^2 dx [\int t^2 K(t)dt]^2}\right\}^{1/5} \cdot n^{-1/5}. \qquad (2.10) \end{aligned}$$

All terms on the right hand side of (2.10) are fixed values, then, the optimal bandwidth h depends on $n^{-1/5}$. The optimal bandwidth h of $MISE$ thus decreases to zero at the rate of $n^{-1/5}$.

Comparing this with (2.6), it is clear that the optimal bandwidth h's for $MISE$ and MSE have the same convergence rate, $n^{-1/5}$. It is still slower than the one of histogram estimator. Observe that $\int [f''(x)]^2 dx$ which, to a certain extent, measures the intensity of roughness of $f(x)$, is included in (2.10). If the density is rougher, which means that the absolute value of $f''(x)$ is larger, then, the denominator of optimal bandwidth h is larger and the h_{opt} is smaller.

Putting the optimal bandwidth h given in (2.10) into (2.9), the optimal mean integrated square error is

$$
\begin{aligned}
MISE(\hat{f}(x)) &\sim n^{-1}C_3\left(\frac{C_3}{4nC_4}\right)^{-1/5} + C_4\left(\frac{C_3}{4nC_4}\right)^{4/5} \\
&\sim n^{-4/5}C_3^{4/5}4^{1/5}C_4^{1/5} + C_4^{1/5}C_3^{4/5}n^{-4/5}4^{-4/5} \\
&\sim n^{-4/5}C_3^{4/5}C_4^{1/5}(4^{1/5}+4^{-4/5}) \qquad (2.11) \\
&\sim \frac{5}{4}\left\{\int K^2(t)dt\right\}^{4/5} \\
&\qquad \cdot\left\{\int [f''(x)]^2dx\right\}^{1/5}\left\{\int t^2K(t)dt\right\}^{2/5}n^{-4/5},
\end{aligned}
$$

which shows the convergence rate is $n^{-4/5}$. The whole curve has the optimal convergence rate which is the same as the convergence rate for a fixed point. It is also faster than the convergence rate, $n^{-2/3}$ of $MISE$ for histogram estimator and slower than the parameter convergence rate, n^{-1}.

In addition, for the same samples of size n and the optimal bandwidth h, a kernel function determines the value of $MISE$. Thus one could obtain a small value of MSE by choosing a kernel.

The greatly similar form of MSE and $MISE$ in (2.4) and (2.9) indicates that they have identical properties proved in sections 2.4.1 and 2.4.2. One could represent the MSE and $MISE$ in one form:

$$\tau(n) \sim (nh)^{-1}C^* + h^4C^{**}. \qquad (2.12)$$

For the formula in (2.12), when $(C^*, C^{**}) = (C_1, C_2)$, $\tau(n)$ represents the mean square error, and when $(C^*, C^{**}) = (C_3, C_4)$, $\tau(n)$ represents the mean integrated square error.

2.5 Choice of Kernel Functions

In section 2.4, we have investigated the MSE and $MISE$. Formulas (2.7) and (2.11) showed the optimal convergence rate for the MSE and $MISE$, $n^{-4/5}$. There we mentioned that given the same samples of size n and the optimal bandwidth h, the values of MSE and $MISE$ depend on the kernel function $K(x)$. Therefore, choice of an appropriate kernel function $K(x)$ will minimize the MSE and $MISE$. Now, the interest is how to choose the kernel function $K(x)$. The MSE and $MISE$ are as follows,

$$MSE(\hat{f}(x)) \sim (nh)^{-1} f(x) \int K^2(t)dt + \frac{1}{4}h^4 \{f''(x)\}^2 \left\{ \int t^2 K(t)dt \right\}^2 \tag{2.13}$$

and

$$MISE(\hat{f}(x)) \sim (nh)^{-1} \left\{ \int K^2(t)dt \right\} + \frac{1}{4}h^4 \int [f''(x)]^2 dx \left\{ \int t^2 K(t)dt \right\}^2. \tag{2.14}$$

Based on the structures of the terms of the MSE and $MISE$, one notes that the problem for choosing an appropriate kernel function $K(x)$ is not a simple problem, because choices of h and the kernel function $K(x)$ are coupled with each other. We could see Wand and Jones [12] about this problem. It also should be noted that one cannot define a kernel function well. In order to address these problems, Wand and Jones [12] propose the following idea. Firstly, rescale the kernel function $K(x)$ in another form, that is,

$$\begin{aligned} K(x) &= \frac{1}{h} K\left(\frac{x}{h}\right) \\ &= \left(\frac{1}{(h/\delta)\cdot\delta}\right) K\left(\frac{x}{(h/\delta)\cdot\delta}\right) \\ &= \frac{1}{h_\delta} K\left(\frac{x}{h_\delta}\right) \end{aligned} \tag{2.15}$$

where $h_\delta = h/\delta$, $K_\delta(x) = \frac{1}{\delta} K\left(\frac{x}{\delta}\right)$. Through choice of δ, we can define the kernel function $K(x)$ well, so one of the two problems is addressed. Now we can choose class $\{K_\delta; \delta > 0\}$

to rescale the kernel function instead of choosing $K(x)$. Also, we can address the other problem of coupled h and the kernel function $K(x)$ by choosing δ. Let $t = x/\delta$, then $\int \{K(x)\}^2 dx$ and $\int x^2 K(x)dx$ can be derived to the new forms,

$$\begin{aligned} \int \{K(x)\}^2 dx &= \int \{K_\delta(x)\}^2 dx \\ &= \int \frac{1}{\delta^2} \left\{K\left(\frac{x}{\delta}\right)\right\}^2 dx \\ &= \frac{1}{\delta} \int \{K(t)\}^2 dt \end{aligned} \tag{2.16}$$

and

$$\begin{aligned} \int x^2 K(x)dx &= \int x^2 K_\delta(x)dx \\ &= \int \frac{x^2}{\delta} K\left(\frac{x}{\delta}\right) dx \\ &= \delta^2 \int t^2 K(t)dt. \end{aligned} \tag{2.17}$$

In (2.16) and (2.17), the new forms derived from $\int \{K(x)\}^2 dx$ and $\int x^2 K(x)dx$ are based on the rescaling method in (2.15). Next, we firstly put the two new forms into the equation of MSE in (2.13). Then, the expression is:

$$\begin{aligned} MSE(\hat{f}(x)) \sim (nh)^{-1} f(x) &\left\{\frac{1}{\delta} \int \{K(t)\}^2 dt\right\} \\ &+ \frac{1}{4} h^4 \{f''(x)\}^2 \left\{\delta^2 \int t^2 K(t)dt\right\}^2. \end{aligned} \tag{2.18}$$

In order to uncouple h and the kernel function $K(x)$, we need to derive (2.18) to the form constituted as a product of two terms including only h and $K(x)$ respectively. Therefore, we let (2.16) be equal to (2.17) and to derive this equation.

$$\begin{aligned} \frac{1}{\delta} \int \{K(t)\}^2 dt &= \left\{\delta^2 \int t^2 K(t)dt\right\}^2 \\ \delta^* &= \left\{\frac{\int [K(t)]^2 dt}{[\int t^2 K(t)dt]^2}\right\}^{1/5}. \end{aligned} \tag{2.19}$$

Wand and Jones [12] call $K_{\delta^*}(x)$, which is $K_\delta(x)$ when $\delta = \delta^*$, as canonical kernel. Because it is known in (2.19) that only if $\delta = \delta^*$, h and $K(x)$ can be uncoupled, we put δ^* into (2.18) to obtain the MSE for $\delta = \delta^*$,

$$\begin{aligned} MSE_{\delta^*}(\hat{f}(x)) &\sim (nh)^{-1} f(x) \left\{ \left[\frac{\int [K(t)]^2 dt}{[\int t^2 K(t) dt]^2} \right]^{-1/5} \int \{K(t)\}^2 dt \right\} \\ &\quad + \frac{1}{4} h^4 \{f''(x)\}^2 \left\{ \left[\frac{\int [K(t)]^2 dt}{[\int t^2 K(t) dt]^2} \right]^{2/5} \int t^2 K(t) dt \right\}^2 \\ &\sim \left\{ (nh)^{-1} f(x) + \frac{1}{4} h^4 [f''(x)]^2 \right\} \\ &\quad \cdot \left\{ \left(\int [K(t)]^2 dt \right)^{4/5} \left(\int t^2 K(t) dt \right)^{2/5} \right\}. \end{aligned}$$

The same steps could be taken in $MISE$ to obtain $MISE$ for $\delta = \delta^*$. Put (2.16) and (2.17) into $MISE$ in (2.14), then

$$\begin{aligned} MISE(\hat{f}(x)) &\sim (nh)^{-1} \left\{ \frac{1}{\delta} \int \{K(t)\}^2 dt \right\} \\ &\quad + \frac{1}{4} h^4 \int [f''(x)]^2 dx \left\{ \delta^2 \int t^2 K(t) dt \right\}^2. \end{aligned} \tag{2.20}$$

(2.20) is as similar as (2.18). So one still lets (2.16) be equal to (2.17) to uncouple h and $K(x)$, and puts δ^* into (2.20) to derive the $MISE$ for $\delta = \delta^*$:

$$\begin{aligned} MISE_{\delta^*}(\hat{f}(x)) &\sim (nh)^{-1} \left\{ \left[\frac{\int [K(t)]^2 dt}{[\int t^2 K(t) dt]^2} \right]^{-1/5} \int \{K(t)\}^2 dt \right\} \\ &\quad + \frac{1}{4} h^4 \int [f''(x)]^2 dx \left\{ \left[\frac{\int [K(t)]^2 dt}{[\int t^2 K(t) dt]^2} \right]^{2/5} \int t^2 K(t) dt \right\}^2 \\ &\sim \left\{ (nh)^{-1} + \frac{1}{4} h^4 \int [f''(x)]^2 dx \right\} \\ &\quad \cdot \left\{ \left(\int [K(t)]^2 dt \right)^{4/5} \left(\int t^2 K(t) dt \right)^{2/5} \right\}. \end{aligned}$$

According to the forms of MSE_{δ^*} and $MISE_{\delta^*}$, it is clear that we uncouple h and the kernel function $K(x)$. Let $C(K) = \left\{ \left(\int [K(t)]^2 dt \right)^{4/5} \left(\int t^2 K(t) dt \right)^{2/5} \right\}$, $C(K)$ is invariant under rescaling of $K(x)$ and minimizing $C(K)$ can give the smallest MSE or $MISE$, see Wand and Jones [12] for this. Given that $K(x)$ is subject to

$$\int t^k K(t)dt \begin{cases} = 1, & \text{if } k = 0 \\ = 0, & \text{if } k = 1 \\ > 0, & \text{if } k = 2. \end{cases}$$

Hodges and Lehmann [8] show that the kernel function that minimizes $C(K)$ is

$$K^a(x) = \frac{3}{4}\left(1 - x^2/5a^2\right)/(5^{1/2}a)I_{[-5^{1/2}a,\, 5^{1/2}a]}.$$

Let $a^2 = 1/5$, we have

$$K^*(x) = \frac{3}{4}(1 - x^2)I_{[-1,\, 1]}.$$

This function is an optimal kernel first suggested by Epanechnikov [6]. Then, it is usually called $Epanechnikov\ kernel$. Furthermore, Wand and Jones [12] and Silverman [11] present the table of the efficiencies of some other kernel functions, where eff$(K) = \{C(K^*)/C(K)\}^{4/5}$.

Table 2.1: Efficiencies for kernels

Kernel	$K(x)$	Efficiency
Epanechnikov	$\frac{3}{4}(1-x^2/5)/b^{1/2}I_{[-1,1]}$	1.000
Biweight	$\frac{15}{16}(1-x^2)^2 I_{[-1,1]}$	0.994
Triangular	$1-\|x\|I_{[-1,1]}$	0.987
Normal	$\frac{1}{\sqrt{2\pi}}e^{(1/2)x^2}$	0.951
Uniform	$\frac{1}{2}I_{[-1,1]}$	0.930

The efficiency is comparing a kernel function with the Epanechnikov kernel. In the table one can find that the efficiencies for these kernels are very closed to 1. There is almost no difference between them. This indicates that choice of different kernels affects MSE and $MISE$ very little.

CHAPTER III

STANDARD KERNEL FUNCTION IN DENSITY ESTIMATION

In chapter 2, we investigated the properties of the kernel estimation. In terms of the kernel estimator $\hat{f}(x)$ in (2.1), we derived the expectation and bias of $\hat{f}(x)$ as well as the variance of $\hat{f}(x)$. With the variance and bias, we obtained the $MSE(\hat{f}(x))$ and $MISE(\hat{f}(x))$. Even though it is difficult to understand them directly, by letting $n \to \infty$, $h \to 0$ such that $(nh)^{-1} \to 0$, one checked the asymptotical properties of the $MSE(\hat{f}(x))$ and $MISE(\hat{f}(x))$. We obtained the results that $MSE(\hat{f}(x)) \to 0$ and $MISE(\hat{f}(x)) \to 0$ asymptotically, which meant $\hat{f}(x)$ tends to $f(x)$ asymptotically. Also, by minimizing $MSE(\hat{f}(x))$ and $MISE(\hat{f}(x))$, we obtained the optimal bandwidth h. After putting this optimal h into $MSE(\hat{f}(x))$ and $MISE(\hat{f}(x))$, we derived the optimal L_2 convergence rate, $n^{-4/5}$. By investigating the choice of an optimal kernel we showed that choice of a kernel has a very little or no effect on MSE or $MISE$.

In the kernel estimation, the symmetric kernel function is usually chosen, and one of the classical kernels is the standard kernel function. Even though its efficiency is very closed to 1 see Table 2.1, it is still worthy to investigate its performance in this chapter.

3.1 General Idea

In Chapter 2, we investigated the performance of the kernel estimator in (2.1) and defined that the kernel function $K(x)$ is subject to

$$\int t^k K(t)dt \begin{cases} = 1, & \text{if } k = 0 \\ = 0, & \text{if } k = 1 \\ > 0, & \text{if } k = 2. \end{cases}$$

This shows that the kernel function is a symmetric density function with $mean = 0$ and $Var \neq 0$, and covers the situation of the standard kernel. Therefore, the properties that we have studied in Chapter 2 can be applied to all symmetric kernel functions with $mean = 0$ and $Var \neq 0$, including the most representative kernel, normal or standard normal kernel function. On the other hand, the standard kernel is a special form of symmetric kernel functions and one realizes that it is a little different from the general form. Then, when we investigate the performance of the standard kernel, we also should discover the details not the same as the kernel estimator. It is known that the standard normal distribution is $X_i \sim N(0, 1)$ and $Var(x) = E[x^2] - \{E[x]\}^2$. So through simple derivation, one gets $E[x^2] = 1$. Then, we define the kernel function $K(x)$ is subject to

$$\int t^k K(t)dt = \begin{cases} 1, & \text{if } k = 0 \\ 0, & \text{if } k = 1 \\ 1, & \text{if } k = 2. \end{cases} \tag{3.1}$$

According to the kernel estimator in (2.1), we define the standard kernel estimator is

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K_{z(0,1)}\left(\frac{x-X_i}{h}\right)$$

where $K_{z(0,1)}(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$ and the form of the standard kernel estimator is equal to (2.1). Then, we derive the MSE and $MISE$.

3.2 The Expectation and Bias of the Standard Kernel estimator

In Section 2.2, by using Taylor expansion and letting $y = x - ht$, we derived the expectation of the kernel estimator. Here by the same method,

$$\begin{aligned}
E[\hat{f}(x)] &= E\left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K_{z(0,1)}\left(\frac{x-X_i}{h}\right)\right] \\
&= \int \frac{1}{h}K_{z(0,1)}\left(\frac{x-y}{h}\right)f(y)dy \\
&= \int K_{z(0,1)}(t)f(x-ht)dt \\
&= \int K_{z(0,1)}(t)\left[f(x) - htf'(x) + \frac{1}{2}f''(x)h^2t^2 + \frac{1}{6}f'''(x)h^3t^3 + O(h^4)\right]dt.
\end{aligned}$$

Then,

$$\begin{aligned}
Bias(\hat{f}(x)) =& E[\hat{f}(x)] - f(x) \\
=& \int K(t)\left[f(x) - htf'(x) + \frac{1}{2}f''(x)h^2t^2 - \frac{1}{6}f'''(x)h^3t^3\right]dt \\
& + O(h^4) - f(x) \\
=& -hf'(x)\int tK(t)dt + \frac{1}{2}f''(x)h^2\int t^2K(t)dt \\
& - \frac{1}{6}f'''(x)h^3\int t^3K(t)dt + O(h^4) \\
=& -hf'(x)\cdot 0 + \frac{1}{2}f''(x)h^2\cdot 1 - \frac{1}{6}f'''(x)h^3\cdot 0 + O(h^4) \\
=& \frac{1}{2}f''(x)h^2 + O(h^4).
\end{aligned} \tag{3.2}$$

Unlike the kernel estimator which uses the kernel in (2.2), the form of the bias of the kernel estimator which uses the kernel in (3.1) is different. By letting $h \to 0$, $Bias(\hat{f}(x)) \to 0$. Therefore, the standard kernel estimator $\hat{f}(x)$ is also asymptotically unbiased.

3.3 The Variance of the Standard Kernel estimator

Because the derivation of the variance does not involve the term $\int t^2 k(t)dt$, the variance of the estimator which uses the kernel in (3.1) is completely the same as the one of the general form. The variance is,

$$\begin{aligned}
Var(\hat{f}(x)) &= Var\left(n^{-1}\sum_{i=1}^{n} h^{-1} K_{z(0,1)}\left(\frac{x-X_i}{h}\right)\right) \\
&= (nh^2)^{-1}\int\left\{K_{z(0,1)}\left(\frac{x-y}{h}\right)\right\}^2 f(y)dy \\
&\qquad - n^{-1}\left\{h^{-1}\int K_{z(0,1)}\left(\frac{x-y}{h}\right) f(y)dy\right\}^2 \\
&= (nh)^{-1}\int\left\{K_{z(0,1)}(t)\right\}^2\left\{f(x) - htf'(x) + ...\right\}dt + O(n^{-1}).
\end{aligned}$$

By Taylor Expansion and letting $y = x - ht$, when $h \to 0$, the variance is

$$Var(\hat{f}(x)) = (nh)^{-1} f(x)\int\left\{K_{z(0,1)}(t)\right\}^2 dt + O(n^{-1}). \tag{3.3}$$

3.4 MSE and MISE of the Standard Kernel estimator

In terms of (3.2) and (3.3), we can derive $MSE(\hat{f}(x))$ and $MISE(\hat{f}(x))$,

$$
\begin{aligned}
MSE(\hat{f}(x)) &= (nh)^{-1} f(x) \int \left\{K_{z(0,1)}(t)\right\}^2 dt + O(n^{-1}) + \left\{\frac{1}{2} f''(x) h^2 + O(h^4)\right\}^2 \\
&= (nh)^{-1} f(x) \int \left\{K_{z(0,1)}(t)\right\}^2 dt + O(n^{-1}) + \frac{1}{4}\left\{f''(x)\right\}^2 h^4 + O(h^6), \\
MISE(\hat{f}(x)) &= \int \left\{(nh)^{-1} f(x) \int \left\{K_{z(0,1)}(t)\right\}^2 dt + O(n^{-1})\right\} dx \\
&\qquad + \int \left\{\frac{1}{4}[f''(x)]^2 h^4 + O(h^6)\right\} dx \\
&= (nh)^{-1} \left\{\int \left[K_{z(0,1)}(t)\right]^2 dt\right\} + O(n^{-1}) + \frac{h^4}{4} \int [f''(x)]^2 dx + O(h^6).
\end{aligned}
$$

Letting $C_5 = f(x) \int \left\{K_{z(0,1)}(t)\right\}^2 dt$, $C_6 = \frac{1}{4}\left\{f''(x)\right\}^2$, $C_7 = \int \left\{K_{z(0,1)}(t)\right\}^2 dt$ and $C_8 = \frac{1}{4}\int [f''(x)]^2 dx$, and because these four terms are all fixed values, then

$$MSE(\hat{f}(x)) \sim (nh)^{-1} C_5 + C_6 h^4,$$

$$MISE(\hat{f}(x)) \sim (nh)^{-1} C_7 + C_8 h^4$$

which shows that MSE and $MISE$ of the standard kernel estimator can be represented as the same form as (2.13), that is

$$\tau_{z(0,1)}(n) \sim (nh)^{-1} C^* + h^4 C^{**}$$

where, $(C^*, C^{**}) = (C_5, C_6)$ indicates that $\tau_{z(0,1)}(n)$ is MSE, and $(C^*, C^{**}) = (C_7, C_8)$ indicates that $\tau_{z(0,1)}(n)$ is $MISE$. When $n \to \infty$, $h \to 0$ such that $(nh)^{-1} \to 0$, $\tau_{z(0,1)}(n)$ representing MSE or $MISE$ tends to zero. Therefore, $\hat{f}(x)$ is consistent for $f(x)$.

The optimal bandwidth h for MSE and $MISE$ is derived by minimizing $\tau_{z(0,1)}(n)$ with respect to h.

$$\frac{d\tau_{z(0,1)}(n)}{dh} = 0$$
$$n^{-1}C^*(-h^{-2}) + C^{**}4h^3 = 0 \tag{3.4}$$
$$h_{opt} = \left(\frac{C^*}{4C^{**}}\right)^{1/5} n^{-1/5}.$$

Put the optimal bandwidth h into $\tau_{z(0,1)}(n)$.

$$\tau_{z(0,1)}(n) = C^*\left(\frac{C^*}{4C^{**}}\right)^{-1/5} n^{-4/5} + C^{**}\left(\frac{C^*}{4C^{**}}\right)^{4/5} n^{-4/5}$$
$$= n^{-4/5}C^{*4/5}C^{**1/5}(4^{1/5} + 4^{-4/5}).$$

Based on the expressions of C^* and C^{**}, we derive the optimal MSE and $MISE$.

$$\begin{aligned} MSE(\hat{f}(x)) &= \frac{5}{4}\{f(x)\}^{4/5}\left\{\int K^2(t)dt\right\}^{4/5}\{f''(x)\}^{2/5} n^{-4/5}, \\ MISE(\hat{f}(x)) &= \frac{5}{4}\left\{\int K^2(t)dt\right\}^{4/5}\left\{\int [f''(x)]^2 dx\right\}^{1/5} n^{-4/5}. \end{aligned} \tag{3.5}$$

It is known that C_5, C_6, C_7 and C_8 are fixed values. So from (3.4) and (3.5), the optimal bandwidth h depends on $n^{-1/5}$ and the optimal MSE and $MISE$ depend on $n^{4/5}$.

$$h_{opt} \sim n^{-1/5},$$
$$MSE \sim n^{-4/5},$$
$$MISE \sim n^{-4/5}.$$

These results also mean that the convergence rate for bandwidth h_{opt} decreases to zero at the rate of $n^{-1/5}$, and the standard kernel estimator has the optimal convergence rate, $n^{-4/5}$, which are as same as that of the kernel estimator in Chapter 2. This was expected, because the standard kernel is one of the symmetrical kernel functions with $mean = 0$ and

$Var \neq 0$. Here the only difference between the general form of kernel estimators and the standard kernel is $\int t^2K(t)dt = 1$. It is should be noted that $\int t^2K(t)dt$ is the fixed value, then this does not affect the convergence rate.

On the other hand, the $bias$, MSE, $MISE$ and optimal bandwidth h_{opt} of the standard kernel estimator does not include the term, $\int t^2K(t)dt$, or in another word, $\int t^2K(t)dt$ appearing in the standard kernel estimation equals to 1. Meanwhile, $\int t^2K(t)dt$ is a value greater than 0 and less than infinity in the kernel estimator in Chapter 2. Therefore, the optimal bandwidth h_{opt}, the optimal MSE and the optimal $MISE$ are different between the general form of kernel estimators and the standard kernel estimator.

CHAPTER IV

BOUNDARY BIAS

In Section 2.5, we showed the table of efficiencies for symmetrical kernel functions provided by Wand and Jones [12] and Silverman [11]. The efficiencies in Table 2.1 are all closed to 1, which indicates that the choice of symmetrical kernels does not affect MSE_δ and $MISE_\delta$ much.

However, in some situations symmetrical kernels cannot estimate a density function well. If the density has the bounded support, which is that data observations from the population density have bounded range, symmetrical kernel functions are not the suitable kernels for density estimation. Because, for the data closed to the boundary or the edge, the symmetric kernel will allocate the weight on the area in which there is no density. For example, there are the data of time and we want to estimate the density. It is the common sense that time is non-negative. So the density with data of time has the support from 0 to ∞. When the symmetrical kernel, such as the standard kernel mentioned in Chapter 3, is used to estimate the density, it puts the weight outside the support on $[0, \infty)$ when the data appears in the region closed to zero. It is obvious that there exists more severe bias when the data is more closed to zero. This problem is called boundary bias. In this chapter, we illustrate this problem.

Wand and Jones [12] introduce the ideas about boundary bias. We define that the kernel function $K(x)$ is a symmetric density function and has the support on [-1, 1]. Also, we assume that the density which we want to estimate is of the form

$$f(x) \begin{cases} = 0, & \text{if } x < 0 \\ > 0, & \text{if } x \geq 0. \end{cases}$$

For the kernel estimator, $\hat{f}(x) = n^{-1} \sum_{i=1}^{n} h^{-1} K\left((x - X_i)h^{-1}\right)$, in (2.1), by letting $y = x - ht$ as we did in previous chapters, its expectation is

$$\begin{aligned} E[\hat{f}(x)] &= E\left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K\left(\frac{x - X_i}{h}\right)\right] \\ &= \int \frac{1}{h}K\left(\frac{x-y}{h}\right) f(y)dy \\ &= \int_{-1}^{1} K(t)f(x - ht)dt. \end{aligned} \tag{4.1}$$

The density has the support from 0 to ∞. For the data observations far away from the boundary, there is no exception and one can derive the expectation as we did in Chapter 2.

However, for the data observations closed to the boundary, the situation is different. Because the support of $K(x)$ is from -1 to 1, when X_i is too closed to the boundary, zero, for example, the distance from X_i to 0 is less than h, the kernel function puts some weight outside the boundary, zero. Since the density is supported on $[0, \infty)$, it is expected that the kernel estimator should be zero in the region $x < 0$. Therefore, in (4.1) it seems that, by adjusting the upper and lower bounds of the integration, one may be able to resolve this. Wand and Jones [12] consider this approach. After letting $y = x - ht$, the upper and lower bounds exchange their positions. Therefore, x/h replaces the upper bound 1, and the lower bound keeps -1.

$$E[\hat{f}(x)] = \int_{-1}^{x/h} K(t)f(x-ht)dt. \tag{4.2}$$

In (4.2) for the data far away from 0, which is $x \geq h$, it is obvious that x/h equals to 1 and we can normally derive the expectation and obtain the result of being asymptotically unbiased, which conforms to the situation that we considered as above. When $x < h$, let $\phi = x/h$ for $\phi \in [0, 1)$. We put ϕ into (4.2) to derive the expectation as follows by Taylor expansion,

$$\begin{aligned}
E[\hat{f}(x)] &= \int_{-1}^{\phi} K(t)f(x-ht)dt \\
&= \int_{-1}^{\phi} K(t)\left\{f(x) - htf'(x) + \frac{1}{2}f''(x)h^2t^2 + O(h^3)\right\}dt \\
&= f(x)\int_{-1}^{\phi} K(t)dt - hf'(x)\int_{-1}^{\phi} tK(t)dt + \frac{1}{2}h^2f''(x)\int_{-1}^{\phi} t^2K(t)dt + O(h^3).
\end{aligned}$$

When $x \geq h$, there is $\int_{-1}^{x/h} K(t)dt = \int_{-1}^{1} K(t)dt = 1$. So the first term, $f(x)\int_{-1}^{\phi} K(t)dt$, equals to $f(x)$. Let $h \to 0$, $E[\hat{f}(x)]$ tends to $f(x)$. This means that $\hat{f}(x)$ is unbiased asymptotically for the data far away from the boundary. For $x < h$, because $[-1, \phi]$ is just the proportion of the support on $[-1, 1]$ of the kernel $K(x)$, $\int_{-1}^{x/h} K(t)dt = \int_{-1}^{\phi} K(t)dt \neq 1$. Then, the the first term, $f(x)\int_{-1}^{\phi} K(t)dt$, does not equal to $f(x)$. Therefore, even though letting $h \to 0$, the kernel estimator is still asymptotically biased.

Also, the data at the boundary means $x = 0$ and $\phi = x/h = 0$. Considering that the kernel function is symmetric which indicates $\int_{-1}^{0} tK(t)dt = \frac{1}{2}$, we obtain the expectation when $x = 0$,

$$\begin{aligned} E[\hat{f}(0)] &= f(0)\int_{-1}^{0} K(t)dt - hf'(0)\int_{-1}^{0} tK(t)dt \\ &\qquad\qquad + \frac{1}{2}h^2 f''(0)\int_{-1}^{0} t^2 K(t)dt + O(h^3) \\ &= \frac{1}{2}f(0) - hf'(0)\int_{-1}^{0} tK(t)dt \\ &\qquad\qquad + \frac{1}{2}h^2 f''(0)\int_{-1}^{0} t^2 K(t)dt + O(h^3). \end{aligned} \tag{4.3}$$

When $h \to 0$, $E[\hat{f}(0)] \to \dfrac{1}{2}f(0)$. That is, the expectation of the kernel estimator at the boundary is the half of the true value. This is because the symmetric kernel function puts weight on the both sides at the boundary.

In order to resolve the problem of boundary bias, Wand and Jones [12] mention the method, referred to as boundary kernel. For the density function in this section, one will try to find a kernel function such that when data is closed to boundary 0, the kernel does not put the weight to the area where there is no probability mass. Because a symmetric kernel indicates $\int_{-1}^{0} tK(t)dt = \frac{1}{2}$, intuitively, in (4.3) if one could find some asymmetric kernel leading to $\int_{-1}^{0} tK(t)dt = 1$, one can address the boundary bias problem. Considering the situation that the density has the support on $[0, \infty)$, one will try the asymmetric kernel function with the same support on $[0, \infty)$. And the most popular kernel with such support is the gamma kernel. Therefore, in next chapter, we will investigate the properties of gamma kernels.

CHAPTER V

GAMMA KERNELS IN DENSITY ESTIMATION

The problem of boundary bias occurs when the symmetric kernel is used to estimate density function with a bounded support. In Chapter 4, we considered the density functions which have support on $[0, \infty)$.The consequence of that was the asymptotically biased kernel density estimator for the data closed to the boundary, where $x < h$. Also, we showed that the bias of the kernel estimator at boundary 0 is half of the true value. In order to resolve the problem of boundary bias discussed in the last chapter, the interest here is to use gamma kernels. Chen [5] introduces gamma kernels for the density function with the bounded support. Now we study the performance of a kernel estimator which uses gamma kernels.

5.1 General Idea

In previous chapters the kernel estimator is defined as follows,

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K\left(\frac{x - X_i}{h}\right).$$

The kernel used in the $\hat{f}(x)$ is symmetric and has a fixed shape, which causes the problem of boundary bias. Unlike this kernel, Chen [5] introduces estimators which use gamma kernel functions. The gamma kernels have two characters which are same as the ones described by Chen [4] about beta kernel, another boundary kernel used to address boundary bias. The first one is that gamma kernel estimators have varying shapes according to different data observations, X_i, which could adjust the amount of smoothing. The second one is that the supports of the kernel function and the density function which we want to estimate are same. This character also reduces the variance of the estimator. We define the density is $f(x)$ which has the support on $[0, \infty)$.

5.2 The Expectation and Bias of Gamma Kernel estimators

Referring to the two characters proposed in Chen [4], Chen [5] introduces $Gamma(\frac{x}{h}+1, h)$, i.e. $f(u) = \frac{1}{\Gamma(\frac{x}{h}+1)h^{(\frac{x}{h}+1)}} u^{x/h} e^{-u/h}$, $0 \le u < \infty$, $(\frac{x}{h}+1), h > 0$, as the kernel function, where h is bandwidth, and defines the first gamma kernel estimator as

$$\hat{f}_1(x) = \frac{1}{n} \sum_{i=1}^{n} K_{\frac{x}{h}+1,h}(X_i)$$

where $K_{\frac{x}{h}+1,h}(u) = \frac{1}{\Gamma(\frac{x}{h}+1)h^{(\frac{x}{h}+1)}} u^{x/h} e^{-u/h}$, $0 \le u < \infty$, $(\frac{x}{h}+1), h > 0$. For this kernel estimator,

$$E[\hat{f}_1(x)] = \int_0^\infty K_{\frac{x}{h}+1,h}(y) f(y) dy = E[f(\psi_x)]. \tag{5.1}$$

Note that the ψ_x is the random variable from $Gamma(\frac{x}{h}+1, h)$ distribution in (5.1). Then, Chen [5] derives the expectation by Taylor expansion,

$$\begin{aligned} E[\hat{f}_1(x)] = E[f(\psi_x)] &= f(E[\psi_x]) + \frac{1}{2}f''(x)Var(\psi_x) + o(h) \\ &= f(x+h) + \frac{1}{2}f''(x)\left(xh + h^2\right) + o(h) \\ &= f(x) + f'(x)h + o(h) + \frac{1}{2}f''(x)xh + \frac{1}{2}f''(x)h^2 + o(h) \\ &= f(x) + f'(x)h + o(h) + \frac{1}{2}f''(x)xh + O(h^2) + o(h) \\ &= f(x) + h\left\{f'(x) + \frac{1}{2}xf''(x)\right\} + o(h) \end{aligned} \quad (5.2)$$

where $E[\psi_x] = x + h$ and $Var(\psi_x) = xh + h^2$ because of $f(\psi_x) \sim Gamma\left(\frac{x}{h} + 1, h\right)$. When $h \to 0$, $E[\hat{f}_1(x)]$ tends to $f(x)$. That is, the gamma kernel estimator is asymptotically unbiased. Meanwhile, note that there is no $\int K(t)dt$ in the first term of (5.2). Therefore, for the data observations either closed to 0 or far away from 0, the estimator does not suffer from the problem of boundary bias.

However, $E[\psi_x] = x + h$ leads to $f(x + h)$ appearing in first term of $E[\hat{f}_1(x)]$ so that the term including $f'(x)$ is involved in (5.2). One usually wants to avoid this situation. So one should use a gamma kernel with $E[\psi_x] = x$. Then we consider $Gamma\left(\frac{x}{h}, h\right)$ as the kernel function. But we create another problem. When $x \to 0$, the kernel function is

$$\begin{aligned} K_{\frac{x}{h},h}(t) &= \frac{1}{\Gamma(x/h)h^{x/h}}t^{\frac{x}{h}-1}e^{-t/h} \\ &\approx \frac{1}{\Gamma(0)h^0}t^{-1}e^{-t/h}. \end{aligned}$$

When t is smaller and smaller, $K_{\frac{x}{h},h}(t)$ tends to infinity. To overcome this, Chen [5] proposes to use different parameters for the gamma kernel when X_i is closed to 0 and far away from 0. Because a gamma kernel is not a symmetrical kernel function, we considers $x < 2h$ as being closed to 0. Chen [5] introduces the second gamma kernel estimator,

$$\hat{f}_2(x) = \frac{1}{n}\sum_{i=1}^{n} K_{\theta_h(x),h}(X_i)$$

where $\theta_h(x)$ represents the different parameters in terms of different X_i, that is

$$\theta_h(x) = \begin{cases} \frac{x}{h}, & \text{if } x \geq 2h \\ \frac{1}{4}\left(\frac{x}{h}\right)^2 + 1, & \text{if } x \in [0, 2h). \end{cases}$$

By Taylor expansion, Chen [5] derives the expectation of the second gamma kernel estimator,

$$E[\hat{f}_2(x)] = \begin{cases} f(x) + \frac{1}{2}xf''(x)h + o(h), & \text{if } x \geq 2h \\ f(x) + \psi_h(x)hf'(x) + o(h), & \text{if } x \in [0, 2h) \end{cases} \quad (5.3)$$

where $\psi_h(x) = \frac{(1-x)[\theta_h(x) - \frac{x}{h}]}{[1 + h\theta_h(x) - x]}$. In (5.3), for $x \geq 2h$ there is no $f'(x)$ appearing in any term. Even though $f'(x)$ appears when $0 \leq x < 2h$, no $f''(x)$ appears when $x \in [0, 2h)$.

In addition, also note that except the first term, $f(x)$, h is included in all terms in (5.3). Then, letting $h \to 0$, $E[\hat{f}_2(x)]$ also tends to $f(x)$ just like the situation of $E[\hat{f}_1(x)]$. Therefore, the second gamma kernel estimator is also asymptotically unbiased. Also, it is apparent that this estimator does not suffer from boundary bias problem.

According to (5.2) and (5.3), we obtain the biases of the two gamma kernel estimators. Chen [5] introduces $\int_0^\infty [xf''(x)]^2 dx < \infty$. This means that $xf''(x)$ tends to zero as x tends to ∞. Considering that the biases of the two gamma estimators include $xf''(x)$, they decrease as x is farther and farther from 0. The biases of the two gamma kernel estimators are

$$Bias(\hat{f}_1(x)) = h\left\{f'(x) + \frac{1}{2}xf''(x)\right\} + o(h),$$

$$Bias(\hat{f}_2(x)) = \begin{cases} \frac{1}{2}xf''(x)h + o(h), & \text{if } x \geq 2h \\ \\ \psi_h(x)hf'(x) + o(h), & \text{if } x \in [0, 2h). \end{cases}$$

5.3 The Variance of Gamma Kernel estimators

Because the variances of $\hat{f}_1(x)$ and $\hat{f}_2(x)$ have very similar forms, here we consider the variance of $\hat{f}_1(x)$ only, which is introduced by Chen [5].

$$\begin{aligned} Var(\hat{f}_1(x)) &= \frac{1}{n}Var\left(K_{\frac{x}{h}+1,h}(X_i)\right) \\ &= \frac{1}{n}E\left[K^2_{\frac{x}{h}+1,h}(X_i)\right] - \frac{1}{n}\left\{E[K_{\frac{x}{h}+1,h}(X_i)]\right\}^2 \\ &= \frac{1}{n}E\left[K^2_{\frac{x}{h}+1,h}(X_i)\right] + O(n^{-1}) \\ &= \frac{1}{n}H_h(x)E[f(\zeta_x)] + O(n^{-1}) \\ &= \frac{1}{n}\left[\frac{h^{-1}\Gamma(\frac{2x}{h}+1)}{2^{\frac{2x}{h}+1}\Gamma^2(\frac{x}{h}+1)}\right]E[f(\zeta_x)] + O(n^{-1}) \\ &= \frac{1}{n}\left[\frac{h^{-\frac{1}{2}}x^{-\frac{1}{2}}R^2\left(\frac{x}{h}\right)}{2\sqrt{\pi}R(\frac{2x}{h})}\right]E[f(\zeta_x)] + O(n^{-1}) \end{aligned} \tag{5.4}$$

where ζ_x follows $Gamma\left(\frac{2x}{h}, h\right)$ distribution and $R(y) = \frac{\sqrt{2\pi}e^{-y}\cdot y^{y+\frac{1}{2}}}{\Gamma(y+1)}$ for $y \geq 0$. As shown in Brown and Chen [3], as y tends to ∞, $R(y)$ tends to 1, $R(y) < 1$ for $y > 0$. Accordingly, $H_h(x)$ in (5.4) is,

$$H_h(x) \sim \begin{cases} \frac{1}{2\sqrt{\pi}}h^{-\frac{1}{2}}x^{-\frac{1}{2}}, & \text{if } \frac{x}{h} \to \infty \\ \\ \frac{h^{-1}\Gamma(2m+1)}{2^{1+2m}\Gamma^2(m+1)}, & \text{if } \frac{x}{h} \to m \end{cases} \tag{5.5}$$

where m is a non-negative constant. Putting (5.5) into (5.4) the variance is,

$$Var(\hat{f}_1(x)) \sim \begin{cases} \frac{1}{2\sqrt{\pi}} n^{-1} h^{-\frac{1}{2}} x^{-\frac{1}{2}} f(x), & \text{if } \frac{x}{h} \to \infty \\ \frac{n^{-1}h^{-1}\Gamma(2m+1)}{2^{1+2m}\Gamma^2(m+1)} f(x), & \text{if } \frac{x}{h} \to m. \end{cases}$$

The variance of $\hat{f}_2(x)$ has the similar form. The only difference is about the term $\Gamma(2m+1)/[2^{1+2m}\Gamma^2(m+1)]$.

5.4 MSE and MISE of Gamma Kernel estimators

By using the bias and the variance of the second gamma kernel and letting $x/h \to \infty$, we derive the MSE of $\hat{f}_2(x)$,

$$\begin{aligned} MSE(\hat{f}_2(x)) &= \frac{1}{2\sqrt{\pi}} n^{-1} h^{-\frac{1}{2}} x^{-\frac{1}{2}} f(x) + O(n^{-1}) + \frac{1}{4}\left[x f''(x)\right]^2 h^2 + o(h^2) \\ &\sim C_9 n^{-1} h^{-\frac{1}{2}} + C_{10} h^2 \end{aligned} \tag{5.6}$$

where $C_9 = \frac{1}{2\sqrt{\pi}} x^{-\frac{1}{2}} f(x)$ and $C_{10} = \frac{1}{4}\left[x f''(x)\right]^2$. Letting $h \to 0$, $n \to \infty$ such that $n^{-1}h^{-\frac{1}{2}} \to 0$, the MSE tends to zero. So $\hat{f}_2(x)$ is consistent of $f(x)$ asymptotically. And by taking derivative, the optimal bandwidth h is

$$\begin{aligned} \frac{dMSE(\hat{f}_2(x))}{dh} &= 0 \\ C_9 \cdot n^{-1}(-1/2)(h^{-3/2}) + C_{10} \cdot 2h &= 0 \\ h_{opt} &= \left(\frac{C_9}{4C_{10}}\right)^{2/5} \cdot n^{-2/5}. \end{aligned}$$

Note that the $MISE$ optimal bandwidth decreases to zero at the rate of $n^{-2/5}$. Thus, the optimal MSE of the second gamma kernel estimator is,

$$MSE(\hat{f}_2(x)) = \frac{5}{4}\left\{\frac{1}{2\sqrt{\pi}} f(x)\right\}^{4/5} \{f''(x)\}^{2/5} n^{-4/5}.$$

Clearly, the MSE of $\hat{f}_2(x)$ converges to zero at the rate, $n^{-4/5}$ which is same as that of the standard kernel estimator. Therefore, the performance of the estimator which uses the gamma kernel is same as that of the standard kernel estimator for the corresponding support. To evaluate the performance of the gamma kernel estimator for the whole $f(x)$, the $MISE$'s now can be obtained. For example, Chen [5] shows that,

$$\begin{aligned} MISE(\hat{f}_1(x)) &= h^2 \int_0^\infty \left\{ f'(x) + \frac{1}{2} x f''(x) \right\}^2 dx \\ &\quad + \frac{1}{2\sqrt{\pi}} n^{-1} h^{-\frac{1}{2}} \int_0^\infty x^{-\frac{1}{2}} f(x) dx + o(n^{-1}h^{-1/2} + h^2) \\ &\sim h^2 C_{11} + n^{-1} h^{-1/2} C_{12} \end{aligned} \tag{5.7}$$

where $C_{11} = \int_0^\infty \left\{ f'(x) + \frac{1}{2} x f''(x) \right\}^2 dx$ and $C_{12} = \frac{1}{2\sqrt{\pi}} \int_0^\infty x^{-\frac{1}{2}} f(x) dx$, and

$$\begin{aligned} MISE(\hat{f}_2(x)) &= \frac{1}{4} h^2 \int_0^\infty \left\{ x f''(x) \right\}^2 dx \\ &\quad + \frac{1}{2\sqrt{\pi}} n^{-1} h^{-\frac{1}{2}} \int_0^\infty x^{-\frac{1}{2}} f(x) dx + o(n^{-1}h^{-1/2} + h^2) \\ &\sim h^2 C_{13} + n^{-1} h^{-1/2} C_{12} \end{aligned} \tag{5.8}$$

where $C_{12} = \frac{1}{2\sqrt{\pi}} \int_0^\infty x^{-\frac{1}{2}} f(x) dx$ and $C_{13} = \frac{1}{4} \int_0^\infty \left\{ x f''(x) \right\}^2 dx$.

Letting $h \to 0$, $n \to \infty$ such that $n^{-1} h^{-\frac{1}{2}} \to 0$, the $MISE$'s tend to zero. So $\hat{f}_1(x)$ and $\hat{f}_2(x)$ are consistent of $f(x)$ asymptotically over the whole curve. By differentiating the $MISE$'s, the optimal bandwidths are

$$h_1^* = \frac{\left[\frac{1}{2\sqrt{\pi}} \int_0^\infty x^{-\frac{1}{2}} f(x) dx \right]^{2/5}}{4^{2/5} \left[\int_0^\infty \left\{ x f'(x) + \frac{1}{2} x f''(x) \right\}^2 dx \right]^{2/5}} \cdot n^{-2/5},$$

$$h_2^* = \frac{\left[\frac{1}{2\sqrt{\pi}} \int_0^\infty x^{-\frac{1}{2}} f(x) dx \right]^{2/5}}{4^{2/5} \left[\int_0^\infty \left\{ \frac{1}{2} x f''(x) \right\}^2 dx \right]^{2/5}} \cdot n^{-2/5}.$$

Thus, putting the optimal bandwidths into (5.7) and (5.8) the optimal $MISE$'s are,

$$\begin{aligned} MISE^*(\hat{f}_1(x)) = & \frac{5}{4^{4/5}} \left[\frac{1}{2\sqrt{\pi}} \int_0^\infty x^{-1/2} f(x) dx \right]^{4/5} \\ & \cdot \left[\int_0^\infty \left\{ x f'(x) + \frac{1}{2} x f''(x) \right\}^2 dx \right]^{1/5} n^{-4/5}, \\ MISE^*(\hat{f}_2(x)) = & \frac{5}{4^{4/5}} \left[\frac{1}{2\sqrt{\pi}} \int_0^\infty x^{-1/2} f(x) dx \right]^{4/5} \\ & \cdot \left[\int_0^\infty \left\{ \frac{1}{2} x f''(x) \right\}^2 dx \right]^{1/5} n^{-4/5}. \end{aligned}$$

Note that the $MISE$ optimal bandwidths decrease to zero at the rate of $n^{-2/5}$ same as that of the MSE optimal bandwidth, which is different from the standard kernel estimator. However, the L_2 convergence rate of the gamma kernel estimators is identical with that of the standard kernel estimator, that is, both of them converge at the rate of $n^{-4/5}$. Also, one can find that the only difference between h_1^* and h_2^*, and between $MISE$'s is $\int_0^\infty \left\{ x f'(x) + \frac{1}{2} x f''(x) \right\}^2 dx$ and $\int_0^\infty \left\{ \frac{1}{2} x f''(x) \right\}^2 dx$. Accordingly, it is not difficult to discover that the former is larger than the latter. Then, $MISE^*(\hat{f}_1(x))$ is larger than $MISE^*(\hat{f}_2(x))$. h_1^* is smaller than h_2^*. Therefore, the performance of the second gamma kernel estimator with the larger optimal bandwidth is better than the first one.

CHAPTER VI

STANDARD KERNEL IN RECURSIVE DENSITY ESTIMATION

In Chapter 5, we studied the performance of gamma kernels in density estimation. Because the standard kernel is fixed and symmetric, this causes the problem of boundary bias for the density function with bounded support. Meanwhile, the gamma kernel with varying shapes and the support on $[0, \infty)$ can address the boundary bias problem for the density function also on $[0, \infty)$. Chen [5] introduces the two gamma kernel estimators, in which the second one avoids $f'(x)$ appearing in the expectation and bias for $x \geq 2h$. The gamma kernel estimators have the optimal L_2 convergence rate, $n^{-4/5}$ which is same as that of the standard kernel estimator. This indicates that they have the equivalent performance in density estimations in $[0, +\infty)$ and $(-\infty, +\infty)$ respectively. On the other hand, the optimal bandwidth h of the gamma kernel estimators decrease to zero at the rate of $n^{-2/5}$, which is faster than $n^{-1/5}$ of the standard kernel estimator.

Now consider the situation where $n - 1$ data observations from the population density function that we want to estimate are available, and $\hat{f}_{n-1}(x)$ is computed. Our interest now is to update $\hat{f}_{n-1}(x)$ when nth data observation becomes available. In order to address this, in this chapter, we investigate recursive kernel density estimation.

6.1 General Idea

As introduced by Rao [9] in recursive density estimation, one derives the kernel estimator based on n data observations from the kernel estimator based on $n-1$ data points. That is,

$$\begin{aligned}\hat{f}_n(x) &= \frac{1}{n}\sum_{j=1}^{n}\frac{1}{h_j}K\left(\frac{x-X_j}{h_j}\right)\\ &= \frac{n-1}{n}\cdot\frac{1}{n-1}\sum_{j=1}^{n-1}\frac{1}{h_j}K\left(\frac{x-X_j}{h_j}\right)+\frac{1}{nh_n}K\left(\frac{x-X_n}{h_n}\right) \qquad (6.1)\\ &= \frac{n-1}{n}\hat{f}_{n-1}(x)+\frac{1}{nh_n}K\left(\frac{x-X_n}{h_n}\right).\end{aligned}$$

This shows that one obtains $\hat{f}_n(x)$ from $\hat{f}_{n-1}(x)$ by the simple calculations. This is the idea of obtaining the recursive density estimators after only $O(1)$ calculations mentioned in Hall and Patil [7] where on-line kernel density estimators are studied. The recursive density estimators are the special forms of on-line kernel density estimators. Define $K(x)$ satisfying

$$\int t^k K(t)dt = \begin{cases} 1, & \text{if } k=0\\ 0, & \text{if } 1\le k\le r-1\\ (-1)^r r!\kappa, & \text{if } k=r\end{cases}$$

where $K(x)$ is a bounded kernel function and κ is a constant. A standard kernel means $r=2$ and $\kappa=1/2$. Then $K(x)$ is subject to

$$\int t^k K(t)dt = \begin{cases} 1, & \text{if } k=0\\ 0, & \text{if } k=1\\ 1, & \text{if } k=2.\end{cases}$$

Hall and Patil [7] introduce the definition of on-line kernel estimators,

$$\hat{f}_n(x) = \sum_{i=1}^{k} p_i \hat{f}_{ni}(x) \tag{6.2}$$

where

$$\hat{f}_{ni}(x) = b_i(n) \sum_{j=N_{i-1}}^{N_i} a_j h_j^{-1} K\left(\frac{x - X_j}{h_j}\right). \tag{6.3}$$

In (6.2) and (6.3), let a_j for $j \geq 1$ be the positive constants, $k \geq 1$ be a fixed integer, $1 = N_0 < N_1 < ... < N_k = n$ which are also integers, $\sum p_i = 1$ with $p_1, p_2, ..., p_k > 0$ and $b_i(n) = (\sum_{N_{i-1} \leq j \leq N_i} a_j)^{-1}$. Hall and Patil [7] introduce that when $k = 1$, (6.2) and (6.3) give us the recursive estimators and when $k \geq 2$, we obtain the on-line kernel estimators.

For recursive estimators, we let a_j for $j \geq 1$ be still the positive constants, $1 = N_0 < N_1 = n$ which are also integers, $p_1 = 1$ and $b_1(n) = (\sum_{j=1}^{n} a_j)^{-1} = \frac{1}{a_1 + ... + a_n}$. Therefore, the recursive density estimators are,

$$\begin{aligned} \hat{f}_n(x) &= \sum_{i=1}^{k} p_i \hat{f}_{ni}(x) \\ &= \sum_{i=1}^{1} p_1 \hat{f}_{n1}(x) \\ &= b_1(n) \sum_{j=1}^{n} a_j h_j^{-1} K\left(\frac{x - X_j}{h_j}\right). \end{aligned} \tag{6.4}$$

6.2 The Expectation and Bias of the Standard Kernel Recursive Estimators

With the recursive estimators in (6.4), we derive the expectation,

$$\begin{aligned} E[\hat{f}_n(x)] &= E\left[b_1(n) \sum_{j=1}^{n} a_j h_j^{-1} K\left(\frac{x - X_j}{h_j}\right)\right] \\ &= b_1(n) \sum_{j=1}^{n} a_j h_j^{-1} E\left[K\left(\frac{x - X_j}{h_j}\right)\right]. \end{aligned} \tag{6.5}$$

By Taylor expansion and letting $y = x - h_j t$ for $j \geq 1$, we have

$$\begin{aligned} E\left[K\left(\frac{x - X_j}{h_j}\right)\right] &= \int K\left(\frac{x-y}{h_j}\right) f(y)dy \\ &= \int K(t) f(x - h_j t) h_j dt \\ &= h_j \left[f(x) + \frac{1}{2} h_j^2 f''(x) + O(h_j^4)\right]. \end{aligned}$$

Put this result into (6.5), with $b_1(n) \sum_{j=1}^{n} a_j = 1$, then the expectation and the bias of the recursive estimators are

$$\begin{aligned} E[\hat{f}_n(x)] &= b_1(n) \sum_{j=1}^{n} a_j h_j^{-1} \left\{ h_j \left[f(x) + \frac{1}{2} h_j^2 f''(x) + O(h_j^4) \right] \right\} \\ &= f(x) + b_1(n) \sum_{j=1}^{n} a_j \left[\frac{1}{2} h_j^2 f''(x)\right] + O(h_j^4), \\ Bias(\hat{f}_n(x)) &= b_1(n) \sum_{j=1}^{n} a_j \left[\frac{1}{2} h_j^2 f''(x)\right] + O(h_j^4). \end{aligned} \tag{6.6}$$

Letting $n \to \infty$ and $h_j \to 0$ such that $b_1(n) \sum_{j=1}^{n} a_j h_j^2 \to 0$, $E[\hat{f}_n(x)]$ tends to $f(x)$ and $Bias(\hat{f}_n(x))$ tends to zero. Therefore, the recursive estimators are asymptotically unbiased.

6.3 The Variance of the Standard Kernel Recursive Estimators

In terms of (6.4), the variance of $\hat{f}_n(x)$ is,

$$\begin{aligned} Var(\hat{f}_n(x)) &= Var\left(b_1(n) \sum_{j=1}^{n} a_j h_j^{-1} K\left(\frac{x - X_j}{h_j}\right)\right) \\ &= b_1^2(n) \sum_{j=1}^{n} a_j^2 h_j^{-2} Var\left(K\left(\frac{x - X_j}{h_j}\right)\right). \end{aligned} \tag{6.7}$$

By letting $y = x - h_j t$ for $j \geq 1$ and $h_j \to 0$, and Taylor expansion. We derive

$$
\begin{aligned}
Var\left(K\left(\frac{x-X_j}{h_j}\right)\right) &= \int \left\{K\left(\frac{x-y}{h_j}\right)\right\}^2 f(y)dy - \left\{\int K\left(\frac{x-X_j}{h_j}\right) f(y)dy\right\}^2 \\
&= \int [K(t)]^2 \left\{f(x) - h_j t f'(x) + ...\right\} h_j dt \\
&\qquad - \left\{\int K\left(\frac{x-X_j}{h_j}\right) f(y)dy\right\}^2 \\
&= h_j f(x) \int [K(t)]^2 dt - \{E[\hat{f}_n(x)]\}^2.
\end{aligned}
$$

The variance is

$$
\begin{aligned}
Var(\hat{f}_n(x)) = b_1^2(n) \sum_{j=1}^{n} a_j^2 h_j^{-2} &\left\{h_j f(x) \int [K(t)]^2 dt\right\} \\
&- b_1^2(n) \sum_{j=1}^{n} a_j^2 h_j^{-2} \{E[\hat{f}_n(x)]\}^2 \qquad (6.8) \\
= b_1^2(n) \sum_{j=1}^{n} a_j^2 h_j^{-1} f(x) &\int [K(t)]^2 dt + o\left(b_1^2(n) \sum_{j=1}^{n} a_j^2 h_j^{-1}\right).
\end{aligned}
$$

6.4 The Convergence Rate of MSE and MISE

Based on the bias in (6.6) and the variance in (6.8), we derive the MSE and $MISE$ of the recursive estimators,

$$
\begin{aligned}
MSE(\hat{f}_n(x)) &= b_1^2(n) \sum_{j=1}^{n} a_j^2 h_j^{-1} f(x) \int [K(t)]^2 dt + o\left(b_1^2(n) \sum_{j=1}^{n} a_j^2 h_j^{-1}\right) \\
&\qquad + \left\{b_1(n) \sum_{j=1}^{n} a_j \left[\frac{1}{2} h_j^2 f''(x)\right]\right\}^2 + o(h_j^6) \\
&= b_1^2(n) \sum_{j=1}^{n} a_j^2 h_j^{-1} \cdot f(x) \int [K(t)]^2 dt + \left\{b_1(n) \sum_{j=1}^{n} a_j h_j^2\right\}^2 \cdot \left[\frac{1}{2} f''(x)\right]^2 \\
&\qquad + o\left(b_1^2(n) \sum_{j=1}^{n} a_j^2 h_j^{-1} + h_j^6\right)
\end{aligned}
$$

and

$$
\begin{aligned}
MISE(\hat{f}_n(x)) &= \int MSE dx \\
&= \int \left\{ b_1^2(n) \sum_{j=1}^{n} a_j^2 h_j^{-1} f(x) \int [K(t)]^2 dt \right\} dx + o\left(b_1^2(n) \sum_{j=1}^{n} a_j^2 h_j^{-1} \right) \\
&\qquad + \int \left\{ b_1(n) \sum_{j=1}^{n} a_j \left[\frac{1}{2} h_j^2 f''(x) \right] \right\}^2 dx + o(h_j^6) \\
&= b_1^2(n) \sum_{j=1}^{n} a_j^2 h_j^{-1} \int [K(t)]^2 dt + \left\{ b_1(n) \sum_{j=1}^{n} a_j h_j^2 \right\}^2 \int \left[\frac{1}{2} f''(x) \right]^2 dx \\
&\qquad + o\left(b_1^2(n) \sum_{j=1}^{n} a_j^2 h_j^{-1} + h_j^6 \right).
\end{aligned}
$$

By letting $n \to \infty$ and $h_j \to 0$ such that $b_1^2(n) \sum_{j=1}^n a_j^2 h_j^{-1} \to 0$ and $b_1(n) \sum_{j=1}^n a_j h_j^2 \to 0$, $MSE(\hat{f}_n(x))$ and $MISE(\hat{f}_n(x))$ both tend to zero. The recursive estimators are consistent of $f(x)$ asymptotically. Because the MSE and $MISE$ have the same form, let $\delta_1 = b_1^2(n) \sum_{j=1}^n a_j^2 h_j^{-1}$, $\delta_2 = b_1(n) \sum_{j=1}^n a_j h_j^2$, $d_1 = f(x) \int [K(t)]^2 dt$, $d_2 = \frac{1}{2} f''(x)$, $D_1 = \int [K(t)]^2 dt$ and $D_2^2 = \int \left[\frac{1}{2} f''(x)\right]^2 dx$. Then, $MSE = d_1\delta_1 + d_2^2\delta_2^2 + o(\delta_1 + h_j^6)$ and $MISE = D_1\delta_1 + D_2^2\delta_2^2 + o(\delta_1 + h_j^6)$. Therefore, set $\delta(n) = c_1\delta_1 + c_2\delta_2^2$ and note that it is the asymptotic MSE when $c_1 = d_1$ and $c_2 = d_2^2$. And it is asymptotic $MISE$ when $c_1 = D_1$ and $c_2 = D_2^2$.

Because of this, there are many h_j, one cannot obtain asymptotically optimal MSE h_j or $MISE$ h_j by differentiating $\delta(n)$ and solving multiple equations. Therefore, Hall and Patil [7] first assume

$$
\delta(n) \sim (\gamma_1 c^{-1} + \gamma_2^2 c^4) n^{-4/5} \tag{6.9}
$$

where $\gamma_1 = 30c_1$ and $\gamma_2 = \frac{5}{3}c_2^{1/2}$. Then, minimizing $\delta(n)$ means minimizing $(\gamma_1 c^{-1}+\gamma_2^2 c^4)$. By differentiating $(\gamma_1 c^{-1}+\gamma_2^2 c^4)$, we obtain the $c = \left(\frac{\gamma_1}{4\gamma_2}\right)^{1/5}$. Then, we derive the optimal $\delta(n)$ by putting c into (6.9).

$$\begin{aligned}
\delta(n) &= \left\{\gamma_1 \left(\frac{\gamma_1}{4\gamma_2}\right)^{-1/5} + \gamma_2^2 \left(\frac{\gamma_1}{4\gamma_2}\right)^{4/5}\right\} n^{-4/5} \\
&= \left\{\gamma_1^{4/5}\gamma_2^{2/5}4^{1/5} + \gamma_1^{4/5}\gamma_2^{2/5}4^{-4/5}\right\} n^{-4/5} \\
&= \frac{5}{4^{4/5}}\gamma_1^{4/5}\gamma_2^{2/5} \cdot n^{-4/5}.
\end{aligned}$$

This indicates that the L_2 convergence rate of MSE and $MISE$ of the standard kernel recursive estimators is $n^{-4/5}$. This is the optimal convergence rate and is same as that of the standard kernel estimator for non-recursive estimation.

Like the standard kernel in non-recursive density estimation, standard kernel in recursive density estimation also suffers from the problem of boundary bias because of being symmetric and fixed. We need a kernel function with appropriate support to address the problem just like the gamma kernels in chapter 5. In the next chapter, we will explore gamma kernels in recursive density estimation.

CHAPTER VII

GAMMA KERNELS IN RECURSIVE DENSITY ESTIMATION

In Chapter 6, we investigated the standard kernel recursive estimators. Even though non-recursive density estimators have good properties, the disadvantage is that all data observations have to be available before one estimates the density function. For the situation that there is another data observation coming from the population density function, recursive estimators should be considered. Rao [9] shows the ideas of recursive estimators. Also, Hall and Patil [7] introduce the properties of on-line kernel estimators. Based on these results, we studied the performance of the standard kernel recursive estimators and obtained the optimal L_2 convergence rate, $n^{-4/5}$. It is equal to the standard kernel in non-recursive estimation.

On the other hand, as a symmetric and fixed kernel function, the standard kernel brings the problem of boundary bias that we have to face for a bounded density. In the expectation in (6.6), the first term is $f(x)$ derived from $f(x)\int K(t)dt$ with $\int K(t)dt = 1$. For the data closed to boundary, $\int K(t)dt$ is not equal to 1 and $f(x)\int K(t)dt \neq f(x)$ again. The boundary bias problem occurs. In Chapter 5, we used gamma kernels to address the problem of boundary bias in non-recursive density estimation. Therefore, the interest is whether or not gamma kernels address the same problem in recursive estimation. In this chapter, we investigate the performance of the gamma kernel recursive estimators.

7.1 General Idea

We first define that the density function which we want to estimate has the support on $[0, +\infty)$. Then, Chen [5] introduces that the two features of the boundary kernel help to resolve the boundary bias problem. They are varying shapes for different data observations and the support matching with the density function. Also, Chen [5] introduces the two gamma kernels for non-recursive density estimation. And note that in recursive density estimators there is h_j, not h. So we change h to h_j in the two gamma kernels introduced by Chen [5] when the gamma kernels are used in recursive estimation. Furthermore, we have shown the recursive estimators in (6.4) based on the notations in Hall and Patil [7]. Then, combining all the notations together, the two gamma kernel recursive estimators are

$$\hat{f}_{n_1}(x) = b_1(n) \sum_{j=1}^{n} a_j K_{\frac{x}{h_j}+1, h_j}(X_i),$$

$$\hat{f}_{n_2}(x) = b_1(n) \sum_{j=1}^{n} a_j K_{\theta_{h_j}(x), h_j}(X_i)$$

where let $B_1(n) = (\sum_{j=1}^{n} a_j)^{-1} = \frac{1}{a_1 + \ldots + a_n}$ and a_j for $j \geq 1$ be the positive constants. Also, $\theta_{h_j}(x)$ is defined as

$$\theta_{h_j}(x) = \begin{cases} \frac{x}{h_j}, & \text{if } x \geq 2h_j \\ \frac{1}{4}\left(\frac{x}{h_j}\right)^2 + 1, & \text{if } x \in [0, 2h_j). \end{cases}$$

7.2 The Expectation and Bias of the Gamma Kernel Recursive Estimators

Chen [5] introduces the expectation of the first gamma kernel estimator that we mentioned in Chapter 5. According to the definition of $b_1(n)$ and a_j, it is known that $b_1(n)\sum_{j=1}^{n} a_j = 1$. Therefore, the expectation of the first gamma kernel recursive estimators are

$$\begin{aligned}
E[\hat{f}_{n_1}(x)] &= E\left[b_1(n)\sum_{j=1}^{n} a_j K_{\frac{x}{h_j}+1,h_j}(X_i)\right] \\
&= b_1(n)\sum_{j=1}^{n} a_j E\left[K_{\frac{x}{h_j}+1,h_j}(X_i)\right] \\
&= b_1(n)\sum_{j=1}^{n} a_j \left\{f(x) + h_j\left[f'(x) + \frac{1}{2}xf''(x)\right] + o(h_j)\right\} \\
&= f(x) + b_1(n)\sum_{j=1}^{n} a_j h_j \left[f'(x) + \frac{1}{2}xf''(x)\right] + o(h_j).
\end{aligned}$$

The first term is $f(x)$ when either the data observation is closed to 0 or moves away from 0. So the first gamma kernel recursive estimators are free from boundary bias problem. When $h_j \to 0$, $E[\hat{f}_{n_1}(x)]$ tends to $f(x)$. $\hat{f}_{n_1}(x)$ is asymptotically unbiased. Then, we derive the bias of the first gamma kernel recursive estimators,

$$Bias[\hat{f}_{n_1}(x)] = b_1(n)\sum_{j=1}^{n} a_j h_j \left[f'(x) + \frac{1}{2}xf''(x)\right] + o(h_j).$$

Like the situation in Chapter 5, $f'(x)$ should be avoided to appear in the bias. And Chen [5] shows us the second gamma kernel estimators to avoid this situation. Referring to Chen

[5] and the expectation of the second gamma kernel non-recursive estimator in Chapter 5, the expectation of the second gamma kernel recursive estimators is

$$
\begin{aligned}
& E[\hat{f}_{n_2}(x)] \\
&= E\left[b_1(n)\sum_{j=1}^{n} a_j K_{\theta_{h_j}(x),h_j}(X_i)\right] \\
&= b_1(n)\sum_{j=1}^{n} a_j E\left[K_{\theta_{h_j}(x),h_j}(X_i)\right] \\
&= \begin{cases}
f(x) + b_1(n)\sum_{j=1}^{n} a_j h_j\left[\frac{1}{2}xf''(x)\right] + o(h_j), & \text{if } x \geq 2h_j \\
f(x) + b_1(n)\sum_{j=1}^{s} a_j h_j\left[\frac{1}{2}xf''(x)\right] & \\
\quad +b_1(n)\sum_{j=s+1}^{n} a_j h_j \psi_{h_j}(x) f'(x) + o(h_j), & \text{if } x \geq 2h_{1,\ldots,s}, x \in [0, 2h_{s+1,\ldots,n}) \\
f(x) + b_1(n)\sum_{j=1}^{n} a_j h_j \psi_{h_j}(x) f'(x) + o(h_j), & \text{if } x \in [0, 2h_j)
\end{cases}
\end{aligned}
$$

where $\psi_{h_j}(x) = \frac{(1-x)[\theta_{h_j}(x) - \frac{x}{h_j}]}{[1+h_j\theta_{h_j}(x) - x]}$. Even though $f'(x)$ is involved in the region where data observations are near zero, no $f''(x)$ appears there. For $x \geq 2h_j$, $f'(x)$ is not involved. Also, let $h_j \to 0$, $E[\hat{f}_{n_2}(x)]$ tends to $f(x)$ and $\hat{f}_{n_2}(x)$ is asymptotically unbiased. The bias of the second gamma recursive estimators is

$$
\begin{aligned}
& Bias[\hat{f}_{n_2}(x)] \\
&= E[\hat{f}_{n_2}(x)] - f(x) \\
&= \begin{cases}
b_1(n)\sum_{j=1}^{n} a_j h_j\left[\frac{1}{2}xf''(x)\right] + o(h_j), & \text{if } x \geq 2h_j \\
b_1(n)\sum_{j=1}^{s} a_j h_j\left[\frac{1}{2}xf''(x)\right] + o(h_j) & \\
\quad +b_1(n)\sum_{j=s+1}^{n} a_j h_j \psi_{h_j}(x) f'(x), & \text{if } x \geq 2h_{1,\ldots,s}, x \in [0, 2h_{s+1,\ldots,n}) \\
b_1(n)\sum_{j=1}^{n} a_j h_j \psi_{h_j}(x) f'(x) + o(h_j), & \text{if } x \in [0, 2h_j).
\end{cases}
\end{aligned}
$$

7.3 The Variance of the Gamma Kernel Recursive Estimators

In Chapter 5, we introduced the variance of the two gamma kernel estimators from Chen [5]. In terms of this and the gamma kernel recursive estimators, one derives the variance. As mentioned in Chapter 5, the forms of the variances of the two gamma kernel estimators are almost same, therefore, the variance of the first ones can be applied to the second ones here. The variance of the first gamma kernel resursive estimators is

$$\begin{aligned}
Var\left(\hat{f}_{n_1}(x)\right) &= Var\left(b_1(n)\sum_{j=1}^{n} a_j K_{\frac{x}{h_j}+1,h_j}(X_i)\right) \\
&= b_1^2(n)\sum_{j=1}^{n} a_j^2 Var\left(K_{\frac{x}{h_j}+1,h_j}(X_i)\right) \\
&\sim \begin{cases} b_1^2(n)\sum_{j=1}^{n} a_j^2 h_j^{-\frac{1}{2}} \cdot \frac{1}{2\sqrt{\pi}} x^{-\frac{1}{2}} f(x), & \text{if } \frac{x}{h_j} \to \infty \\ b_1^2(n)\sum_{j=1}^{n} a_j^2 h_j^{-1} \cdot \frac{\Gamma(2m_j+1)}{2^{1+2m_j}\Gamma^2(m_j+1)} f(x), & \text{if } \frac{x}{h_j} \to m_j. \end{cases}
\end{aligned}$$

7.4 The Convergence Rate of MSE and MISE

In Section 7.2 and 7.3, we have derived the bias and the variance. By letting $\frac{x}{2h_j} \to \infty$, we obtain the MSE of the second gamma recursive estimators,

$$\begin{aligned}
MSE(\hat{f}_{n_2}(x)) = b_1^2(n)\sum_{j=1}^{n} a_j^2 h_j^{-\frac{1}{2}} \cdot \frac{1}{2\sqrt{\pi}} x^{-\frac{1}{2}} f(x) \\
+ b_1^2(n)\left[\sum_{j=1}^{n} a_j h_j\right]^2 \left[\frac{1}{2} x f''(x)\right]^2 + o(h_j).
\end{aligned}$$

Here we encounter the same problem that we have faced in Chapter 6 when we want to minimize the MSE. There are many bandwidth h_j so that we cannot derive the effective result by taking derivative. For the standard kernel recursive estimators, Hall and Patil [7] first assume the notation in (6.9) to compute the convergence rate of MSE and $MISE$.

However, this notation cannot be applied to this circumstance, because the standard kernel in Chapter 6 conforms to the kernel function defined by Hall and Patil [7]. And that definition of $K(x)$ is different from the gamma kernels. Therefore, we need another method to derive the convergence rate.

It is known that the recursive estimators include different bandwidth h_j. This situation also occurs in adaptive kernel estimators. Silverman [11] introduces this important method. Based on various data observations, different bandwidths are used in adaptive kernel estimation. Meanwhile, in recursive estimators, we also have different h_j in terms of data observations. So the ideas and results of adaptive kernel estimators are used to derive MSE and $MISE$ of the gamma kernel recursive estimators.

Silverman [11] considers the adaptive kernel estimators and the concepts. We define $h_j = \lambda_j h$, where h is bandwidth and $\lambda_j = \left\{\frac{\tilde{f}(X_j)}{g}\right\}^{-\alpha}$. Also, $\tilde{f}(X_j) > 0$ is the pilot estimate. g is the geometric mean of $\tilde{f}(X_j)$. α is the sensitivity parameter satisfying $0 \leq \alpha \leq 1$. Abramson [1] sets $\alpha = 1/2$ and Silverman [11] uses λ_j equal to $\frac{1}{\sqrt{f(X_j)}}$. Furthermore, Hall and Patil [7] mention $a_j = 1$ used in Wolverton and Wagner [13]. With all of them, we derive the expectation and bias. For $x/2h_j \to \infty$,

$$\begin{aligned} E[\hat{f}_{n_2}(x)] &= f(x) + b_1(n)\sum_{j=1}^{n} a_j h_j \left[\frac{1}{2}xf''(x)\right] + o(h_j) \\ &= f(x) + b_1(n)\sum_{j=1}^{n} a_j \lambda_j h \left[\frac{1}{2}xf''(x)\right] + o(h) \end{aligned} \tag{7.1}$$

where

$$b_1(n)\sum_{j=1}^{n} a_j\lambda_j = \frac{1}{n}\sum_{j=1}^{n}\frac{1}{\sqrt{f(X_j)}}.$$

Let $\{X_j \in [s_1, s_2], s_1 > 0$ and $s_2 < \infty\}$. Because $f(X_j)$ is non-negative and continuous in $[s_1, s_2]$, then there must exist a value m such that $f_{min}(X_j) = f(m)$. And $\sum_{j=1}^{n} \frac{1}{\sqrt{f(X_i)}} \leq \sum_{j=1}^{n} \frac{1}{\sqrt{f(m)}} = \frac{n}{\sqrt{f(m)}}$. Also, because $f(x)$ is a density function in $[s_1, s_2]$ and $\int_{s_1}^{s_2} f(x)dx = 1$, $\sum_{j=1}^{n} \frac{1}{\sqrt{f(X_i)}} = nA_2$, where $A_2 \leq \frac{1}{\sqrt{f(m)}}$ is a fixed value. Therefore, $b_1(n)\sum_{j=1}^{n} a_j\lambda_j = \frac{1}{n}\cdot nA_2 = A_2$. Put this into (7.1), the expectation and bias of the second gamma recursive estimators for $x/(2\lambda_j h) \to \infty$ are

$$E[\hat{f}_{n_2}(x)] = f(x) + A_2 h\left[\frac{1}{2}xf''(x)\right] + o(h),$$

$$Bias(\hat{f}_{n_2}(x)) = A_2 h\left[\frac{1}{2}xf''(x)\right] + o(h).$$

With $b_1(n)\sum_{j=1}^{n} a_j\lambda_j = A_2$, we also derive the expectation and bias of the first gamma kernel recursive estimators,

$$\begin{aligned}
E[\hat{f}_{n_1}(x)] &= f(x) + b_1(n)\sum_{j=1}^{n} a_j h_j\left[f'(x) + \frac{1}{2}xf''(x)\right] + o(h_j) \\
&= f(x) + b_1(n)\sum_{j=1}^{n} a_j\lambda_j h\left[f'(x) + \frac{1}{2}xf''(x)\right] + o(h) \\
&= f(x) + A_2 h\left[f'(x) + \frac{1}{2}xf''(x)\right] + o(h),
\end{aligned}$$

$$Bias(\hat{f}_{n_1}(x)) = A_2 h\left[f'(x) + \frac{1}{2}xf''(x)\right] + o(h).$$

Except the first term, $f(x)$, other terms of $E[\hat{f}_{n_1}(x)]$ and $E[\hat{f}_{n_2}(x)]$ all include h. So when $h \to 0$, the expectations tend to $f(x)$. So the two gamma recursive estimators are asymptotcally unbiased.

The variance for $x/2h_j \to \infty$ is

$$\begin{aligned}
Var(\hat{f}_{n_1}(x)) &\sim b_1^2(n)\sum_{j=1}^{n} a_j^2 h_j^{-\frac{1}{2}} \cdot \frac{1}{2\sqrt{\pi}} x^{-\frac{1}{2}} f(x) \\
&\sim b_1^2(n)\sum_{j=1}^{n} a_j^2 \lambda_j^{-\frac{1}{2}} h^{-\frac{1}{2}} \cdot \frac{1}{2\sqrt{\pi}} x^{-\frac{1}{2}} f(x)
\end{aligned} \tag{7.2}$$

where

$$b_1^2(n)\sum_{j=1}^{n} a_j^2\lambda_j^{-\frac{1}{2}} = \frac{1}{n^2}\sum_{j=1}^{n}\{f(X_i)\}^{1/4}.$$

Because of $\{X_j \in [s_1, s_2], s_1 > 0$ and $s_2 < \infty\}$ assumed in the derivation of expectation and bias, and the situation that $f(X_j)$ is a density function in $[s_1, s_2]$, there also must exist a value M such that $f_{max}(X_j) = f(M)$. Accordingly, $\sum_{j=1}^{n}\{f(X_i)\}^{\frac{1}{4}} \leq \sum_{j=1}^{n}\{f(M)\}^{\frac{1}{4}} = n\{f(M)\}^{\frac{1}{4}}$. Also, $\sum_{j=1}^{n}\{f(X_i)\}^{1/4} = nA_3$, where $A_3 \leq \{f(X_i)\}^{1/4}$ is a fixed value. Therefore, $b_1^2(n)\sum_{j=1}^{n} a_j^2\lambda_j^{-\frac{1}{2}} = \frac{1}{n^2}\cdot nA_3 = \frac{1}{n}A_3$. Putting this into (7.2), the variance is

$$Var(\hat{f}_{n_1}(x)) \sim \frac{A_3}{n}\cdot h^{-\frac{1}{2}}\cdot\frac{1}{2\sqrt{\pi}}x^{-\frac{1}{2}}f(x).$$

Because the variance of the first gamma kernel recursive estimators can be applied to the second ones, we derive MSE of $\hat{f}_{n_2}(x)$,

$$\begin{aligned} MSE(\hat{f}_{n_2}(x)) &\sim A_3n^{-1}\cdot h^{-\frac{1}{2}}\cdot\frac{1}{2\sqrt{\pi}}x^{-\frac{1}{2}}f(x) + A_2^2h^2\left[\frac{1}{2}xf''(x)\right]^2 \\ &\sim A_3n^{-1}\cdot h^{-\frac{1}{2}}\cdot C_9 + A_2^2h^2C_{10}. \end{aligned} \tag{7.3}$$

In (5.6), we have let $C_9 = \frac{1}{2\sqrt{\pi}}x^{-\frac{1}{2}}f(x)$ and $C_{10} = \frac{1}{4}\left[xf''(x)\right]^2$. By taking derivative and solving the equation, the optimal bandwidth h_{opt} is $\left(\frac{A_3C_9}{4A_2^2C_{10}}\right)^{2/5}\cdot n^{-2/5}$. After we put the optimal bandwidth into (7.3), we obtain the optimal MSE, which has the optimal convergence rate, $n^{-4/5}$,

$$MSE(\hat{f}_2(x)) = \frac{5}{4}\left\{A_3\frac{1}{2\sqrt{\pi}}f(x)\right\}^{4/5}\{A_2f''(x)\}^{2/5}\cdot n^{-4/5}.$$

Then, we investigate the performance of the two gamma kernel recursive estimators over the whole curve based on the bias and variance which we have derived. In (5.7) and

(5.8) we let $C_{11} = \int_0^\infty \left\{f'(x) + \frac{1}{2}xf''(x)\right\}^2 dx$, $C_{12} = \frac{1}{2\sqrt{\pi}}\int_0^\infty x^{-\frac{1}{2}}f(x)dx$ and $C_{13} = \frac{1}{4}\int_0^\infty \{xf''(x)\}^2 dx$. Therefore, the $MISE$'s are

$$\begin{aligned}
MISE(\hat{f}_1(x)) &\sim h^2 A_2^2 \int_0^\infty \left\{f'(x) + \frac{1}{2}xf''(x)\right\}^2 dx \\
&\qquad\qquad + n^{-1}h^{-\frac{1}{2}}A_3 \frac{1}{2\sqrt{\pi}}\int_0^\infty x^{-\frac{1}{2}}f(x)dx \\
&\sim h^2 A_2^2 C_{11} + n^{-1}h^{-1/2}A_3 C_{12}, \\
MISE(\hat{f}_2(x)) &\sim h^2 A_2^2 \int_0^\infty \left\{\frac{1}{2}xf''(x)\right\}^2 dx \\
&\qquad\qquad + n^{-1}h^{-\frac{1}{2}}A_3 \frac{1}{2\sqrt{\pi}}\int_0^\infty x^{-\frac{1}{2}}f(x)dx \\
&\sim h^2 A_2^2 C_{13} + n^{-1}h^{-1/2}A_3 C_{12}.
\end{aligned} \tag{7.4}$$

Letting $h \to 0$ and $n \to \infty$ such that $n^{-1}h^{-1/2} \to 0$, the MSE and $MISE$ tend to zero. So the gamma kernel recursive estimators are consistent of $f(x)$ asymptotically. Then, by minimizing the $MISE$'s with respect to h, we obtain the optimal bandwidths $h_1^* = (\frac{A_3 C_{12}}{4A_2^2 C_{11}})^{2/5} \cdot n^{-2/5}$ and $h_2^* = (\frac{A_3 C_{12}}{4A_2^2 C_{13}})^{2/5} \cdot n^{-2/5}$. With the optimal bandwidths we derive the optimal $MISE$'s through putting them into (7.4),

$$\begin{aligned}
MISE^*(\hat{f}_1(x)) &= \frac{5}{4^{4/5}}\left[A_3 \frac{1}{2\sqrt{\pi}}\int_0^\infty x^{-1/2}f(x)dx\right]^{4/5} \\
&\qquad \cdot \left[\int_0^\infty A_2^2 \left\{xf'(x) + \frac{1}{2}xf''(x)\right\}^2 dx\right]^{1/5} \cdot n^{-4/5}, \\
MISE^*(\hat{f}_2(x)) &= \frac{5}{4^{4/5}}\left[A_3 \frac{1}{2\sqrt{\pi}}\int_0^\infty x^{-1/2}f(x)dx\right]^{4/5} \\
&\qquad \cdot \left[\int_0^\infty A_2^2 \left\{\frac{1}{2}xf''(x)\right\}^2 dx\right]^{1/5} \cdot n^{-4/5}.
\end{aligned}$$

The relationship between the two $MISE$'s of the gamma kernel recursive estimators are similar with the one of gamma kernel non-recursive estimators. The $MISE(\hat{f}_1(x))$ is larger than the $MISE(\hat{f}_2(x))$. This indicates that the performance of the second ones is better than the first ones. And the L_2 convergence rates of $MISE$ and MSE are both the optimal convergence rate, $n^{-4/5}$. This is also the same as that of non-recursive estimators. Also, compared to the L_2 convergence rate of the standard kernel recursive estimators, they are both $n^{-4/5}$, which means that the gamma kernel recursive estimators and the standard kernel recursive estimators both have optimal convergence rate. These two recursive estimators perform equally well corresponding to the density functions which they are used to estimate.

CHAPTER VIII

CONCLUSIONS

8.1 Final Comments

We have investigated the performance of kernel estimators, for either non-recursive or recursive density estimators. We also discussed the specific commonly used kernel functions, the standard kernel function and the gamma kernel functions. The result shows that the L_2 convergence rate of the gamma kernel recursive estimators is $n^{-4/5}$. It is as same as the L_2 convergence rate of the standard kernel recursive estimators. This means that the gamma kernels in $[0, \infty)$ is equivalent with the standard kernel in $(-\infty, +\infty)$ for recursive density estimation. Furthermore, the L_2 convergence rates of the standard kernel non-recursive estimator and the gamma kernel non-recursive estimators are also $n^{-4/5}$. All of them have the same L_2 convergence rate. At last, gamma kernels in recursive estimation can address the problem of boundary bias well for the density with the support on $[0, \infty)$ as it did in non-recursive estimation.

8.2 Future Work

Gamma kernels with the support on $[0, \infty)$ and varying shapes can address the boundary bias in the density function with the same support. In Chapter 4, we took an example that time is a kind of non-negative data observations. It is obvious that gamma kernels could be applied to the density with the data of time to avoid boundary bias. In practice we have this density in survival analysis. On the other hand, we are more interested in the hazard rate function, $h(x) = \frac{f(x)}{1-F(x)}$, than the density function because the hazard rate function shows more information in survival analysis. It is known that the hazard rate function is constituted by the density function and the cumulative distribution function, referred to as $F(x) = \int_{-\infty}^{x} f(t)dt$. This indicates that the hazard rate function depends on the density function. Therefore, the density estimation is useful in the hazard rate estimation. Bouezmarni, El Ghouch and Mesfioui [2] introduce the performance of the gamma kernel estimators in hazard rate estimation. Also, we have investigated the performance of the gamma kernel recursive estimators. Therefore, in the next step the performance of gamma kernel recursive hazard rate estimators will be explored.

REFERENCES

[1] I. S. Abramson, "On Bandwidth Variation in Kernel Estimates-A Square Root Law," *The Annals of Statistics*, vol. 10, no. 4, 1982, pp. 1217–1223.

[2] T. Bouezmarni, A. El Ghouch, and M. Mesfioui, "Gamma Kernel Estimators for Density and Hazard Rate of Right Censored Data," *Journal of Probability and Statistics*, vol. 2011, 2011, pp. 1–16.

[3] B. M. Brown and S. X. Chen, "Beta-Bernstein Smoothing for Regression Curves with Compact Support," *Scandinavian Journal of Statistics*, vol. 26, no. 1, 1999, pp. 47–59.

[4] S. X. Chen, "Beta Kernel Estimators for Density Functions," *Comput. Statist. Data Anal.*, vol. 31, no. 2, 1999, pp. 131–145.

[5] S. X. Chen, "Probability Density Function Estimation Using Gamma Kernels," *Annals of the Institute of Statistical Mathematics*, vol. 52, no. 3, Sep 2000, pp. 471–480.

[6] V. K. Epanechnikov, "Non-Parametric Estimation of A Multivariate Probability Density," *Theory Prob. Appl.*, vol. 14, 1969, pp. 153–158.

[7] P. Hall and P. Patil, "On the Efficiency of On-Line Density Estimators," *IEEE Trans. Inf. Theor.*, vol. 40, no. 5, Sept. 1994, pp. 1504–1512.

[8] J. L. Hodges and E. L. Lehmann, "The Efficiency of Some Nonparametric Competitors of the t-Test," *The Annals of Mathematical Statistics*, vol. 27, no. 2, 1956, pp. 324–335.

[9] B. Rao, *Nonparametric Functional Estimation*, NATO Conference Series. Academic Press, 1983.

[10] D. W. Scott, "On Optimal and Data-Based Histograms," *Biometrika*, vol. 66, no. 3, 12 1979, pp. 605–610.

[11] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*, Chapman & Hall, London, 1986.

[12] M. Wand and M. Jones, *Kernel Smoothing*, Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis, 1995.

[13] C. T. Wolverton and T. J. Wagner, "Asymptotically Optimal Discriminant Functions for Pattern Classification," *Trans. Inform. Theory*, vol. 15, 1969, pp. 258–265.