ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE ENGINEERING AND TECHNOLOGY

ON ESTIMATION OF PROBABILITY DENSITY FUNCTION

Ph.D. THESIS

Elif ERÇELİK

Department of Mathematical Engineering

Mathematical Engineering Programme

JUNE 2019

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE
ENGINEERING AND TECHNOLOGY

ON ESTIMATION OF PROBABILITY DENSITY FUNCTION

Ph.D. THESIS

Elif ERÇELİK
(509142203)

Department of Mathematical Engineering

Mathematical Engineering Programme

Thesis Advisor: Prof. Dr. Mustafa NADAR

JUNE 2019

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

OLASILIK YOĞUNLUK FONKSİYONU TAHMİNİ ÜZERİNE

DOKTORA TEZİ

Elif ERÇELİK
(509142203)

Matematik Mühendisliği Anabilim Dalı

Matematik Mühendisliği Programı

Tez Danışmanı: Prof. Dr. Mustafa NADAR

HAZİRAN 2019

Elif ERÇELİK, a Ph.D. student of ITU Graduate School of Science Engineering and Technology 509142203 successfully defended the thesis entitled "ON ESTIMATION OF PROBABILITY DENSITY FUNCTION", which he/she prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor :	**Prof. Dr. Mustafa NADAR** Istanbul Technical University	....
Jury Members :	**Prof. Dr. Mansur İSGENDEROĞLU** Gebze Technical University	....
	Prof. Dr. Müjgan TEZ Marmara University	....
	Assoc. Prof. Dr. Ergün YARANERİ Istanbul Technical University	....
	Assist. Prof. Dr. Gül İNAN Istanbul Technical University	....

Date of Submission : **22 May 2019**
Date of Defense : **26 June 2019**

To my beloved parents,

FOREWORD

I would like to express my sincere thanks to my supervisor Prof.Dr. Mustafa Nadar for his patience, insightful comments and continuous guidance during my whole university life and during the development of this thesis.

I owe great deal of thanks to Prof.Dr. Mansur İsgenderoğlu for sharing his wealth of knowledge during my university life. I also owe a thank to Assoc. Prof.Dr. Ergün Yaraneri for his guidance and valuable time.

I would like to thank Prof.Dr. Müjgan Tez and Assist. Prof.Dr. Gül İnan for being the examiners on my PhD viva.

I want to express my sincere thanks to Prof.Dr. Fatma Özdemir for her guidance and encouragement. I would like to thank Prof.Dr. Neşe Özdemir and Prof.Dr. Kamil Oruçoğlu for their support. Thanks also go to the staff of the mathematical engineering department. I am very grateful to have had the opportunity to work among such supportive scientist. I would also like to thank Prof.Dr. Tahir Azeroğlu for his valuable suggestions.

I gratefully acknowledge the financial support provided by TUBİTAK 2211/A Graduate Scholarship Programme.

I also thank to The Scientific Research Projects Unit (PAB) at Istanbul Technical University (ITU) for the support under grant (Project No: 40198).

And, most importantly, I thank to my mother Nurhan, and my father, Sait, for always believing in me and for all the love, encouragement and support throughout my life. I will be forever grateful and thankful.

June 2019 Elif ERÇELİK

TABLE OF CONTENTS

Page

ABBREVIATIONS

r.v.	: Random Variable
pdf	: Probability Density Function
i.i.d.	:Independent and Identically Distributed
MSE	: Mean Squared Error
MISE	: Mean Integrated Squared Error
ISE	: Integrated Square Error
SE	: Squared Error
KDE	: Kernel Density Estimation
LSCV	: Least Squares Cross-Validation
LCV	: Likelihood Cross-Validation
Gam1	: Gamma1
Gam2	: Gamma2
IG	: Inverse Gaussian
RIG	: Reciprocal Inverse Gaussian
BS-PE	: Birnbaum Saunders Power-Exponential

LIST OF TABLES

Page

LIST OF FIGURES

ON ESTIMATION OF PROBABILITY DENSITY FUNCTION

SUMMARY

Density estimation is one of the most fundamental problem in statistics. It can be simply determined as the construction of an estimate of the density function from the observed data when these observed data assumed to be a sample from an unknown probability density function (pdf).

There are two approaches to density estimation problem: parametric and nonparametric. Under parametric approach, shape of the density is assumed to be unknown. Nonparametric approach relaxes this assumption since it relies solely on the data and allows the "data speaks for itself". Nonparametric density estimation problem arises in many fields, including economics, banking, genetics, climatology, hydrology, etc. That is why, the literature about density estimation methods are vast. The kernel method, orthogonal series method and delta sequence method have the major interest among many other density estimation methods.

In the first chapter, a brief introduction about density estimation problem is given. The purpose and scope of this dissertation are introduced. Some of the most used methods are introduced and studies about these methods are mentioned as a literature summary.

In the second chapter, background and some basic definitions used in this thesis are given.

In the third chapter, delta sequence method is studied. Many work related to density estimation impose smoothness conditions on the density function f and its derivatives although there are applications in which discontinuities in f are natural. However, the assumptions of smoothness condition restricts the class of densities, so the weakening of any conditions on the density is of considerable interest in application. For this purpose, the conditions on the density functions are written by using the second order modulus of continuity type majorants. Stronger local convergency rate of the mean squared error (MSE) corresponding to d-variate delta sequence based density estimator is obtained for both univariate and multivariate cases when compared with the convergency rate of the MSE of the density estimators defined by the first order finite differences.

In the fourth chapter, orthogonal series method is considered. Density function is studied by means of Hermite functions and convergency rate of the mean integrated square error (MISE) of density estimators by using delta sequences is obtained when the support of the density function is infinite. Then, convergency rate of the MSE and MISE of estimator are obtained for the densities having compact support. The results of former publications about rate of convergence of estimators based on Hermite series are improved.

In the fifth chapter, the kernel method is examined. In this method, a kernel is usually considered as symmetric and it is widely believed that kernel is of minor importance

than the smoothing bandwidth. But, when the estimated density has compact or semi-infinite support, classical kernel estimators give rise to boundary bias problem. To avoid boundary bias problem, a new asymmetric kernel estimator is proposed by using beta prime distribution as kernel. Finite sample properties investigated and comparisons are made with other asymmetric kernel estimators in terms of average ISE via Monte Carlo simulations. In addition, adaptive Bayesian bandwidth selection with Lindley approximation method proposed which is new for the asymmetric kernel estimators. Then, it was shown that, the average ISE of the new estimator with this new approach has better performance in comparison to the classical least squared cross-validation method. Also, real data applications are performed to illustrate the potential usefulness of the proposed estimator.

In the sixth chapter, asymmetric kernel density estimation method is studied for the densities defined on the positive real line. Scaled inverse chi-squared density function is used to construct a new kernel estimator. The adaptive Bayesian bandwidth selection with Lindley approximation which is proposed in the previous chapter is used for the numerical studies. Then, the average ISE comparisons are made using different methods for the kernel estimators under consideration. Real data applications are made to illustrate potential usefulness of the scaled inverse chi-squared estimator. Those applications demonstrated that the proposed estimator is capable to reproduce the shoulder near zero, unlike the beta prime estimator. Therefore, it can be used as an alternative to beta prime kernel estimator for this kind of data sets.

Finally, last chapter devoted to the conclusions.

OLASILIK YOĞUNLUK FONKSİYONU TAHMİNİ ÜZERİNE

ÖZET

Yoğunluk fonksiyonu kestirimi istatistiğin en temel problemlerinden biridir. Yoğunluk fonksiyonu kestirimi basitçe dağılımı bilinmeyen bir veri seti için yoğunluk fonksiyonu oluşturulması problemi olarak tanımlanabilir.

Yoğunluk fonksiyonu kestirimi için parametrik ve parametrik olmayan yaklaşımlar mevcuttur. Parametrik yaklaşımda, yoğunluk fonksiyonunun birkaç parametreye kadar bilindiği varsayılmaktadır. Böylece, bilinmeyen parametreler için kestiriciler kurmak, parametrik kestirim yaklaşımı için yeterlidir. Parametrik olmayan yoğunluk fonksiyonu kestirimi yaklaşımında bu varsayım hafifletilmiştir. Parametrik olmayan yaklaşım sadece verilere dayanır ve "verinin kendi adına konuşmasına izin verir". Parametrik olmayan yoğunluk fonksiyonu kestirimi ekonomi, bankacılık, genetik, klimatoloji, hidroloji gibi çok çeşitli alanlarda karşımıza çıkar. Bu nedenle, parametrik olmayan yaklaşım ile ilgili literatürde birçok çalışma mevcuttur ve yoğunluk fonksiyonu kestirimi için çeşitli metodlar önerilmiştir. Bu metodlardan çekirdek kestirimi, ortogonal kestirim metodu ve delta dizileri metodu en çok kullanılan metodlardır.

Tezin ilk bölümünde, yoğunluk fonksiyonu kestirim problemi hakkında kısa bir giriş yapılarak, kullanım alanları açıklanmıştır. Daha sonra, çekirdek kestirimi, ortogonal kestirim metodu ve delta dizileri metodu ile ilgili literatür özeti verilmiştir. Son olarak, tezin amacı açıklanarak ilk bölüm tamamlanmıştır.

İkinci bölümde, bu tezde kullanılan temel tanımlar ve metodlar verilmiştir. Yoğunluk fonksiyonu kestirimi için en çok kullanılan metodlar açıklanmıştır. Ayrıca kestiricinin performansını ölçmek için gerekli ve kullanışlı metodlar tanıtılmıştır. Daha sonra, yoğunluk fonksiyonu kestirimi için çok önemli olan bant genişliği seçimi metodlarından bahsedilmiştir.

Üçüncü bölümde, delta dizileri metodu çalışılmıştır. Yoğunluk fonksiyonu kestirimi ile ilgili literatürdeki çalışmalarda, yoğunluk fonksiyonu ve türevleri üzerine düzgünlük koşulu yazılmaktadır. Ancak, yoğunluk fonksiyonunun süreksiz olduğu noktaların var olduğu birçok uygulama mevcuttur. Dolayısıyla, düzgünlük koşulu yoğunluk fonksiyonu sınıflarını kısıtlar ve bu kısıtın kaldırılması ya da hafifletilmesi uygulamada oldukça önemlidir. Bu amaçla, yoğunluk fonksiyonu üzerindeki koşullar ikinci dereceden süreklilik modülü majorantları cinsinden yazılarak literatürdeki çalışmalarda çoğunlukla kullanılan ikinci dereceden diferansiyellenebilme koşulu hafifletilmiştir. Ayrıca, tek değişkenli ve çok değişkenli durumlar için d-değişkenli delta dizileri yardımıyla yazılmış kestiricilerin bir noktada ortalama karesel hata yakınsaklık hızı incelenmiş ve birinci dereceden sonlu farklar yardımıyla yazılmış yoğunluk fonksiyonu kestiricileri için daha önce elde edilen sonuçlar iyileştirilmiştir.

Dördüncü bölümde, ortogonal seri kestirim metodu çalışılmıştır. Yoğunluk fonksiyonu kestiricileri Hermite serileri kullanılarak delta dizileri yardımıyla yazılmıştır. Böylece sonsuz destekli yoğunluk fonksiyonları için kestiricilerin bütünleşik hata kareleri ortalaması (MISE) yakınsama hızı elde edilmiştir. Daha sonra, kompakt destekli yoğunluk fonksiyonları için yazılmış kestiricilerin bütünleşik hata kareleri ortalaması ve ortalama karesel hata (MSE) yakınsaklık hızları incelenmiştir. Delta dizilerinin kullanılması, literatürde daha önce Hermite serileri yardımıyla yazılmış kestiriciler için elde edilen sonuçların iyileştirilmesini sağlamıştır.

Beşinci bölümde ise çekirdek metodu çalışılmıştır. Çekirdek metodu parametrik olmayan kestirim metodlarından en yaygın olanıdır. Bu metodda, çekirdek genellikle simetrik olup, çekirdek seçiminin band genişliği seçiminden daha az önemli olduğu düşünülmektedir. Ancak, yoğunluk fonksiyonu kompakt ya da yarı sonlu desteğe sahip olduğunda klasik simetrik çekirdekler ile yazılmış kestiriciler sınır yanlılığı sorununa neden olmaktadır. Bu problemin çözümü için literatürde birçok yöntem mevcuttur. Son zamanlarda önerilen bir yöntem ise klasik simetrik çekirdek ile yazılmış kestirici yerine asimetrik çekirdek ile yazılmış kestirici kullanmaktır. Bu tezde klasik kestirici yerine, beta prime yoğunluk fonksiyonu uygun parametrelerle çekirdek yerine kullanılarak, yeni bir asimetrik çekirdek kestiricisi önerilmiştir. Önerilen yeni kestiricinin, sınır yanlılığı problemini çözdüğü ve optimal ortalama karesel hata ve bütünleşik hata kareleri ortalaması yakınsama hızına sahip olduğu gösterilmiştir. Diğer asimetrik kestiricilerde olduğu gibi, düzlemenin yapıldığı noktadan uzaklaştıkça varyansın azaldığı gözlenmiştir. Bu da gözlemlerin seyrek olduğu yoğunluk fonksiyonu kestiriminde avantaj sağlayan bir özelliktir. Ayrıca, simülasyon çalışmaları yardımıyla, bu kestiricinin sonlu örnek özellikleri incelenmiş ve bu kestirici ile literatürde var olan asimetrik kestiricilerin ortalama bütünleşik karesel hataları (ISE) karşılaştırılmıştır. Kalın kuyruklu yoğunluk fonksiyonları için klasik bant genişliği seçim metodlarının yetersiz kaldığı bilinmektedir. Bu nedenle, kalın kuyruklu yoğunluk fonksiyonları kestirimi için klasik bant genişliği seçim metodları yerine uyarlamalı Bayesian bant genişliği seçim metodu, asimetrik kestiricilerin bant genişliği için daha önce kullanılmamış bir metot olan Lindley yaklaşımı yardımıyla kullanılmıştır. Bu yaklaşımdan elde edilen bant genişlikleri ile klasik en küçük kareler çapraz geçerleme (LSCV) metodundan elde edilen bant genişliklerinden alınan ortalama bütünleşik karesel hataları karşılaştırılarak önerilen metodun kullanışlılığı gösterilmiştir. Daha sonra, elde edilen sonuçlar gerçek veriler kullanılarak örneklenmiştir.

Altıncı bölümde ise ölçeklendirilmiş ters ki kare yoğunluk fonksiyonu kullanılarak yeni bir asimetrik çekirdek kestirici önerilmiştir. Bu kestiricinin asimptotik özellikleri incelenerek ortalama karesel hata ve bütünleşik hata kareleri ortalaması optimal yakınsama hızına sahip olduğu gösterilmiştir. Ölçeklendirilmiş ters ki kare kestiricisi için, bir önceki bölümde önerilen Lindley yaklaşımı yardımıyla uyarlamalı Bayesian bant genişliği seçim metodu ile elde edilen bant genişliklerinden alınan ortalama bütünleşik karesel hataları değerinin en küçük kareler çapraz geçerleme ile elde edilen bant genişliklerinden elde edilenden çok daha küçük olduğu gözlenmiştir. Simulasyon çalışmalarında ayrıca, yeni önerilen kestirici ile beta prime kestiricileri ve Birnbaum Saunders power-exponential çekirdek kestiricileri ortalama bütünleşik karesel hataları karşılaştırmaları yapılmıştır. Ayrıca, gerçek veri uygulamalarıyla yeni kestiricinin performansı incelenmiştir. Beta prime çekirdek kestiriciler için yapılan çalışmalarda sınırda omuz şekline sahip verilerin ("shoulder data") uygun olmadığı buna karşı yeni önerilen kestiricinin uygun olduğu gösterilmiştir. Böylece yeni kestiricinin beta prime kestiricisine alternatif olarak kullanılabileceği düşüncesi ortaya çıkmıştır.

Son bölümde ise bu tezde elde edilen sonuçlar açıklanarak, gelecekte yapılabilecek çalışmalardan bahsedilmiştir.

1. INTRODUCTION

Density estimation is one of the fundamental research topic of statistics since the late 1950's. First, density estimation was considered from theoretical point of view. However, the development in technology of computing created extensive interest not only in theoretical but also in practical aspect. Thereafter, nonparametric density estimation is used in different fields such as economics, banking, genetics, hydrology, climatology as well as many branches of statistics.

Density estimation is the construction of an accurate and a robust estimator of an unknown density function from the observed data. There are two approaches of density estimation: parametric and nonparametric. In parametric approach, the density of the underlying data is drawn from one of the known parametric family of distributions, namely the shape of the density assumed to be known. Then, one can create a parametric density estimate of a density f underlying the data by plugging in estimators for those parameters. On the other hand, nonparametric approach relies solely on the data and allows the "data speaks for itself" without any assumption about the shape of the underlying density.

The studies on nonparametric density estimation problem was introduced by pioneering work of [1]. After that, a lot of studies are performed. In those studies, observations were generally assumed to be independent and identically distributed (i.i.d.) random variables. Depending on the assumptions of the density function to be estimated, different results for the local and global convergency rate of the mean integrated square error (MISE) and mean square error (MSE) are obtained.

The kernel method for the estimation of the i.i.d. observations with the continuous and symmetric density functions is one of the main interest in the statistics and it first appeared in paper [1]. Then, the kernel method for the univariate case was studied and it was shown that the convergency rate of the MSE is optimal when the density function has two continuous derivatives as one can see in paper [2]. The assumptions on the density functions were written in terms of the derivatives and Lipschitz condition in paper [3]. In that paper, more general density estimation problem which is called delta sequence method is studied and different delta sequence estimators depending on the local and global properties of the density functions are considered. For the results obtained in [3], the dominant terms in expansions of the MSE and MISE for the trigonometric series density estimators is obtained in [4]. After that, some asymptotic properties of the delta sequence based density estimators for the multivariate case were studied in [5].

There is also a vast literature about an orthogonal series density estimation and in those studies, estimators are written in terms of the classical orthogonal polynomials including Legendre, Jacobi and Hermite. The choice of orthogonal polynomial depends on the support of the density function. If the support is real line or the half line then the Hermite or Laguerre series are useful. On the other hand, if the support is compact, then Jacobi series or trigonometric series are recommended to use. Orthogonal series estimation with the estimators based on the Hermite functions is studied in [6]. They obtained the consistency and the rate of the convergence for the MSE of the univariate and multivariate densities by writing the additional conditions on the density functions. Also, Hermite series estimators for the estimation of the density function, its derivatives and characteristic function were studied in [7]. After that, the orthogonal series estimation was studied in [8]. They proposed an estimator for the derivatives of the density function by using Hermite series and obtained better convergency rate of the MISE and the convergency of the MSE than the former studies.

In paper [9], the asymptotic properties of the estimators based on the Jacobi and Legendre polynomials by using delta sequence method was studied. In that paper, the certain summability methods were used to avoid the negative values of orthogonal series estimators and also, the MSE and MISE convergency rate were obtained for the densities having compact support. Recently, the rate of convergence of the MSE of the estimators for the multivariate case based on delta sequence method is investigated in [10]. Unlike the former studies, the assumptions on the density function were written in terms of the first order modulus of continuity type majorants.

In the literature, it is widely believed that the choice of smoothing bandwidth is of crucial importance than the choice of kernel functions. Therefore, most of the studies about kernel density estimation problem considered symmetric kernels. However, when the support of the density function is half line or compact then classical symmetric kernel estimators yield the boundary bias problem. Boundary bias problem occurs when kernels with infinite support are used for data with semi infinite or compact support, since this would lead to a leakage of probability mass. Many methods are proposed to overcome the boundary bias problem including data reflection method discussed in [11], boundary kernel method studied in [12], [13], hybrid method suggested by [14], the local linear estimator given in [15], empirical transformation method proposed in [16] and generating pseudodata developed in [17].

As an alternative method to remove the boundary bias problem, the use of asymmetric beta distribution as kernel when estimating densities with compact support was proposed by [18]. In this method, support of the asymmetric kernel matches the support of the density to be estimated and the amount of smoothing are controlled by the suitable parametrization chosen for the kernel functions. Then, for the densities with semi infinite support, two new asymmetric kernel estimator was developed by using gamma distribution as kernel in paper [19]. After that, this method was used to develop the lognormal and Birnbaum-Saunders kernel estimators given in [20]. In that paper, it was shown that these two kernel estimators are suitable for the high frequency or ultra frequency data via simulation studies and real data application of the high frequency financial intraday time duration data. Then, inverse gaussian (IG) and reciprocal inverse gaussian (RIG) kernel estimators are proposed in paper [21]. As more recent studies generalized Birnbaum-Saunders, skewed generalized Birnbaum-Saunders kernel estimators are proposed by [22], [23]. In paper [24], the estimators studied in papers [20] and [21] was reformulated. Also, inverse gamma kernel estimator was developed by [25] and then, this new estimator was reformulated and its asymptotic properties were studied in [26]. Moreover, in paper [27], weighted distributions are used to propose a new class of lognormal kernel estimators which is first studied in paper [20].

All of the papers discussed above used classical global bandwidth selection methods. However, for some distributions, global bandwidth selection methods result in unsatisfactory outcomes. Estimators obtained with global bandwidth selectors tend to under or over smooth density functions. Because of this reason, the adaptive Bayesian bandwidth selection method for the univariate symmetric kernel estimators was proposed in paper [28]. Unlike the classical methods, the bandwidth is considered as a parameter of the model. In study [29], the adaptive Bayesian method was used for the asymmetric Birnbaum-Saunders power exponential (BS-PE) kernel estimators to estimate the heavy tailed densities. Also, the adaptive Bayesian bandwidth selection for multivariate discrete associate kernel estimator based on finite differences are studied in paper [30].

Studies on nonparametric density estimation with censored data was initiated by the work of [31]. In paper [32], comprehensive review about the earlier density estimation methods for the censored data was given. The density estimation using asymmetric kernels was studied in [33] for the right censored case. In that paper, a data driven Bayesian local bandwidth selection method was used. Then, Gamma kernel estimator discussed in [19] is adapted to the right censored case for the density and hazard rate functions by [34].

In this thesis, the density estimation based on the delta sequence method, the orthogonal series method and the asymmetric kernel method are of major interest. The basic methodological approach of the theory is to obtain closeness of the estimator to the true density in various ways. The most studied measures of discrepancy is the MSE and MISE. So, in this dissertation, the rate of convergency of MSE of an estimator is derived when estimating densities at a single point. Moreover, for the global accuracy of an estimator the rate of convergency of MISE is investigated. In Chapter 3, the delta sequence method is considered for both univariate and multivariate cases. The motivation was to find an answer whether the MSE rate of convergence of an estimator is improved when using densities belonging to the class of functions defined by second order finite differences over the class of functions defined by first order finite differences. For this purpose, the conditions on density function are written in terms of the second order modulus of continuity type majorants. Moreover, the second order differentiability assumption is weakened by utilizing second order modulus of continuity type majorants. It is an advantage since there is applications in which the discontinuity of the density function is natural. As a result, the MSE rate of convergency is obtained better than the one obtained by using first order finite differences. In Chapter 4, the orthogonal series method is considered and the delta sequence estimators based on Hermite polynomials are studied instead of classical

Hermite series estimators. Then, the MSE and MISE rate of convergence show that, the delta sequence estimators based on Hermite polynomials gives better estimates than the former studies. In Chapters 5 and 6, the asymmetric kernel method is studied to fix the boundary bias problem. A new estimator is proposed based on beta prime distribution function in Chapter 5. Then, for the theoretical treatment the MSE and MISE rate of convergence of beta prime estimator are discussed. After that, the finite sample properties of the beta prime estimator are investigated via Monte Carlo simulation studies. Furthermore, adaptive Bayesian bandwidth selection method is used with Lindley's approximation for the heavy tailed density functions. This method is new for the asymmetric kernel estimators. It is shown that the bandwidths obtained from adaptive Bayesian bandwidth selection method yields better estimates than the one obtained from the classical least squares cross validation method. Then, real data examples are given to illustrate the findings. In Chapter 6, a new asymmetric kernel estimator is proposed by using scaled inverse chi-squared density estimator. Similar to the existing kernel estimators, it is shown that, the proposed estimator is free of boundary bias problem and achieves the optimal rate of convergence of MSE and MISE. Numerical studies are conducted to compare the average ISE performance of bandwidths obtained from LSCV method with the bandwidths obtained from adaptive Bayesian method. Moreover, real data applications demonstrated that the scaled inverse chi-squared estimator is suitable to capture the bumps of the models. Also, it is suitable to use this proposed estimator when the estimated density has a shoulder near zero unlike the beta prime estimator.

2. BACKGROUND AND BASIC DEFINITIONS

In this chapter, the preliminary definitions and background of this dissertation are given.

2.1 Notation

It is assumed that the observations $X_1, X_2, ..., X_n$ are independent and identically distributed (i.i.d.) random variables (r.v.) with probability density function (pdf) f. The symbol $\widehat{f}$ will be used to denote the estimator of density functions which are under consideration.

2.2 Definitions and Background

Definition 1. *Let F be a collection of subsets of a nonempty set Ω. Then F is called a $\sigma-algebra$ if it satisfies the following properties:*

i. *The empty set $\emptyset \in F$,*

ii. *If $A \in F$, then the complement $A^c \in F$,*

iii. *If $A_1, A_2, ...$ is a sequence of elements of F, then their union $\bigcup\limits_{i=1}^{\infty} A_i \in F$.*

A pair (Ω, F) consisting of a set Ω and a $\sigma-algebra$ F is called a measurable space. The elements of F are called measurable sets or events.

Definition 2. *Let (Ω, F, P) be arbitrary probability space, and let X be a real valued function on Ω; X is a random variable (r.v.) if $X = X(w)$ is a F-measurable function, or equivalently $X^{-1}(U) = \{w \in \Omega : X(w) \in U\} \in F$.*

There are two types of random variables, discrete and continuous. A discrete r.v. takes on only countable number of distinct values. A continuous r.v takes an infinite number of possible values (i.e. its range is closed or open interval).

2.3 Density Estimation

The problem of constructing an estimator for a set of observed data points based on unknown pdf is called density estimation. There are a lot of approaches to density estimation problem. Among many available approaches, the most used are histogram, kernel density estimator, orthogonal series estimator and delta sequence estimator, see [35].

2.3.1 Histogram

The oldest and simplest approach in density estimation is the histogram. If we have an origin x_0 and a bin width h, then the bins of the histogram yields to be the intervals $[x_0 + mh, x_0 + (m+1)h]$ for integers m. In such case, the density estimator will be,

$$\widehat{f}(x) = \frac{1}{n} \frac{\text{number of observations } X_i \text{ in the same bin}}{\text{length of the bin}} \tag{2.1}$$

Although the histogram is the simplest and useful way to estimate a density function, it is often necessary to use more sophisticated method. Because, histogram method can have drawback when derivatives of the density estimates are required. Also, the histogram method substantially depends on the choice of origin. So, alternative methods to histogram are proposed.

2.3.2 Kernel Density Estimator

The kernel density estimator (KDE) is one of the most used method for density estimation. The kernel estimator is defined as

$$\widehat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) \tag{2.2}$$

where, $h > 0$ is the smoothing bandwidth, $K(x)$ is kernel function that is always nonnegative, generally assumed to be symmetric and smooth function, see Figure 2.1

The kernel K satisfies the following conditions:

$$\int K(u)du = 1, \quad \int uK(u)du = 0, \quad \int u^2 K(u)du = k_2 < \infty \tag{2.3}$$

Since kernel K usually considered as symmetric pdf, the constant k_2 will then be variance of the distribution with this density. Also, the kernel K is everywhere nonnegative and integrates to unity, i.e. probability density function, then from the

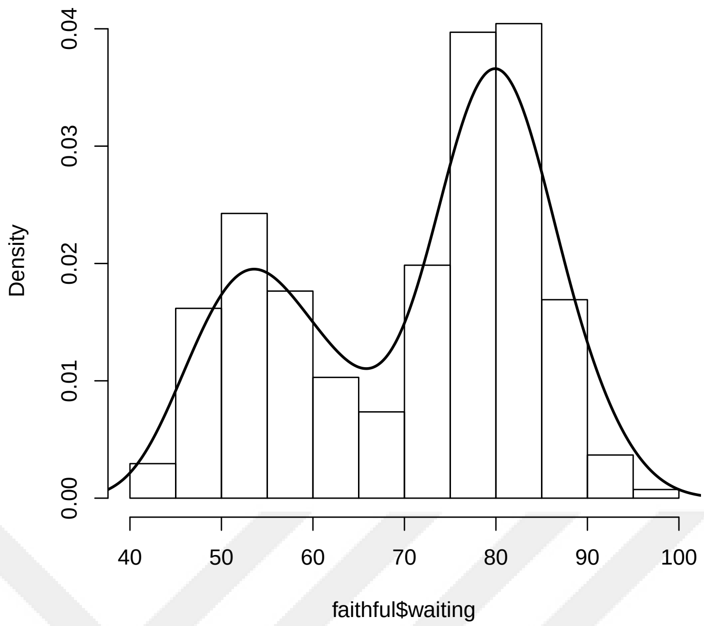

Figure 2.1 : KDE for faithful data.

definition, $\widehat{f}$ itself be a pdf. Moreover, $\widehat{f}$ inherits all the continuity and differentiability properties of the kernel K.

Basically, the KDE smoothes each data point X_i into a small density bumps and then by adding all these small bumps together it constructs the final density estimate, see Figure 2.2. The kernel function K determines the shape of the bumps while the bandwidth h determines their width. If the bandwidth h is chosen too large then all the detail of the distribution obscured. Otherwise, if the bandwidth h is chosen too small the structure of the distribution spurious, see Figure 2.3. That's why a lot of methods proposed to choose the bandwidth in the literature.

2.3.3 Orthogonal Series Estimators

Suppose that a density function $f(x)$ is given by

$$f(x) = \sum_{i=0}^{\infty} a_i \phi_i(x), x \in I \tag{2.4}$$

where $a_i = \int_I f(x) \phi_i(x) dx$ and $\{\phi_i\}_{i \geq 0}$ be a complete set of orthonormal basis functions on an interval I. Then, the orthogonal series estimator is

$$\widehat{f}(x) = \sum_{i=0}^{m} \widehat{a}_i \phi_i(x) \tag{2.5}$$

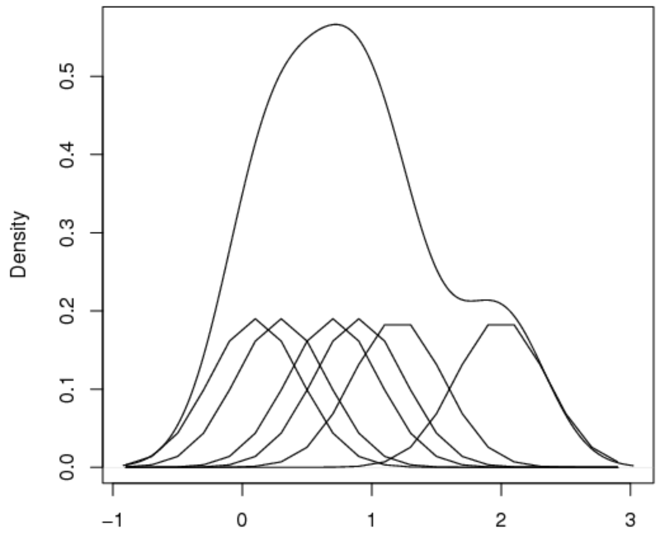

Figure 2.2 : Kernel estimation with Gaussian kernel.

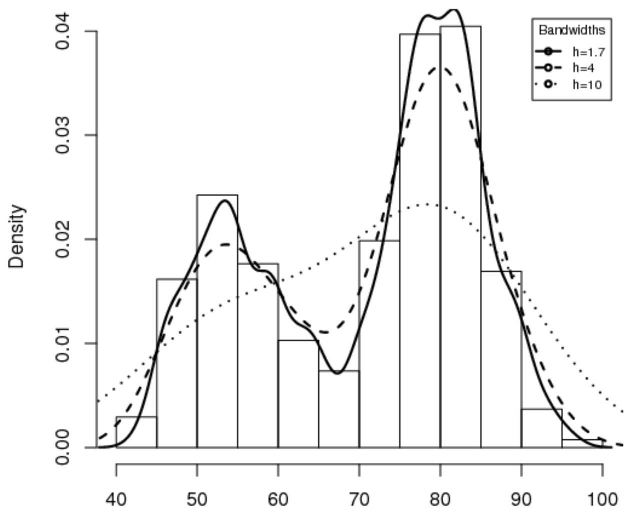

Figure 2.3 : Effect of bandwidths values to KDE.

for some integer $m \geq 0$, and $\widehat{a_i}$ is an unbiased estimator of a_i, since

$$\begin{aligned} a_i &= \int_I f(x)\,\phi_i(x)\,dx = E\left[\phi_i(x)\right] \\ &\approx \frac{1}{n}\sum_{j=1}^{n}\phi_i\left(X_j\right) \equiv \widehat{a_i} \end{aligned} \tag{2.6}$$

where $X_1, ..., X_n$ is a sequence of i.i.d. random variables.

2.3.4 Delta Sequence density estimation method

Let J be an open interval of the real line R. A sequence $\{\delta_m(x,t)\}$ of bounded, measurable function on $J \times J$ is a delta sequence on J, if for each $x \in J$ and each

$C^\infty-$function ϕ with support in J,

$$\lim_{m\to\infty}\int_J \delta_m(x,t)\phi(t)dt = \phi(x). \tag{2.7}$$

Then, delta sequence density estimator can be defined as

$$\widehat{f}(x) = \frac{1}{n}\sum_{i=1}^{n}\delta_m(x,X_i) \tag{2.8}$$

Definition 3. *A delta sequence $\{\delta_m\}$ is said to be a positive type delta sequence if for* ***x**, **t*** $\in R^d$

(a) $\delta_m(\mathbf{x},\mathbf{t}) \geq 0$*;*

(b) $\int\limits_{R^d} \delta_m(\mathbf{x},\mathbf{t})d\mathbf{t} = 1$*;*

(c) $\delta_m(\mathbf{x},\mathbf{t}) \leq cm^d \frac{1}{1+(m|\mathbf{t}-\mathbf{x}|)^{d+2}}$*, where c is a positive constant.*

2.4 Mean Square Error and Mean Integrated Square Error

Various measures have been studied to discuss the closeness of the estimator $\widehat{f}$ to the true density f. A natural measures of discrepancy is mean square error (MSE) when estimating densities at a single point. MSE is defined by

$$MSE\left(\widehat{f}(x)\right) = E\left(\widehat{f}(x) - f(x)\right)^2 = \left(E\left(\widehat{f}(x)\right) - f(x)\right)^2 + var\left(\widehat{f}(x)\right) \tag{2.9}$$

the sum of squared bias and variance at x. There is a relation between bias and variance terms in equation (2.9). The bias can be reduced at the expense of increasing variance, and vice versa, by adjusting the amount of smoothing. For the global accuracy of the density estimator, the most widely used measure of discrepancy is the mean integrated square error (MISE), and it is defined by

$$MISE\left(\widehat{f}(x)\right) = E\int\left\{\widehat{f}(x) - f(x)\right\}^2 dx \tag{2.10}$$

There are other global measures of performance of estimators such as mean integrated absolute error, but due to its mathematical tractability the MISE criterion is widely used in the literature. MISE can be expressed in an another way due to negative integrand in (2.10). When order of integration and expectation reversed by Fubini theorem, the MISE can be expressed in terms of its bias and variance, such as;

$$\begin{aligned} MISE\left(\widehat{f}(x)\right) &= \int E\left\{\widehat{f}(x) - f(x)\right\}^2 dx = \int MSE_x\left(\widehat{f}\right)dx \\ &= \int\left\{E\widehat{f}(x) - f(x)\right\}^2 dx + \int var\left(\widehat{f}(x)\right)dx \end{aligned} \tag{2.11}$$

Some approximate properties are the followings;

$$bias_h(x) = E\widehat{f}(x) - f(x) = \int \frac{1}{h} K\left(\frac{x-y}{h}\right) f(y)dy - f(x) \tag{2.12}$$

This equation can be used to obtain an approximate expression for the bias. Change of variable $(x-y)/h = t$ gives

$$bias_h(x) = \int K(t) f(x-ht)dt - f(x) = \int K(t)\left[f(x-ht) - f(x)\right]dt \tag{2.13}$$

Then, Taylor series expansion yields

$$f(x-ht) = f(x) - htf'(x) + \frac{1}{2}h^2t^2 f''(x) + .. \tag{2.14}$$

and equation gives

$$\begin{aligned} bias_h(x) &= -hf'(x)\int tK(t)dt + \frac{1}{2}h^2 f''(x)\int t^2 K(t)dt + ... \\ &= \frac{1}{2}h^2 f''(x)k_2 + \text{higher order terms in h} \end{aligned} \tag{2.15}$$

$$\int bias_h(x)^2 dx \approx \frac{1}{4}h^4 k_2^2 \int f''(x)^2 dx \tag{2.16}$$

Now, variance term can be written as

$$\begin{aligned} var\widehat{f}(x) &= \frac{1}{n}\int \frac{1}{h^2} K\left(\frac{x-y}{h}\right)^2 f(y)dy - \frac{1}{n}\left\{f(x) + bias_h(x)\right\}^2 \\ &= \frac{1}{nh}\int f(x-ht)K(t)^2 dt - \frac{1}{n}\left\{f(x) + O(h^2)\right\}^2 \end{aligned} \tag{2.17}$$

Then, using Taylor series expansion variance term can be obtained as

$$\begin{aligned} var\widehat{f}(x) &\approx \frac{1}{nh}\int \left\{f(x) - htf'(x) + ...\right\} K(t)^2 dt + O\left(\frac{1}{n}\right) \\ &= \frac{1}{nh} f(x)\int K(t)^2 dt + O\left(\frac{1}{n}\right) \approx \frac{1}{nh} f(x)\int K(t)^2 dt \end{aligned} \tag{2.18}$$

Since f is a pdf, then integrated variance term is

$$\int var\widehat{f}(x)dx = \frac{1}{nh}\int K(t)^2 dt \tag{2.19}$$

Therefore, combining integrated bias and variance, MISE can be written as

$$MISE\left(\widehat{f}(x)\right) = \frac{1}{4}h^4 k_2^2 \int f''(x)^2 dx + \frac{1}{nh}\int K(t)^2 dt. \tag{2.20}$$

The optimal bandwidth minimizing MISE is,

$$h_{opt} = k_2^{-2/5}\left\{\int K(t)^2 dt\right\}^{1/5}\left\{\int f''(x)^2 dx\right\}^{-1/5} n^{-1/5}. \tag{2.21}$$

For detailed information see [35].

2.5 Bandwidth Selection

The problem of choosing the smoothing bandwidth is of crucial importance in density estimation. The purpose of the estimation can be influential in the appropriate choice of bandwidth. If the purpose is just to explore data in order to suggest possible models and hypothesis, then it is sufficient to subjectively choose the bandwidth by looking at the density estimates produced by a range of bandwidths. However, for many applications this approach is impractical, so the requirement for automatic choice of smoothing bandwidths arises. Different methods are proposed to choose the optimal bandwidth such as plug-in, least squares cross-validation (LSCV), likelihood cross-validation (LCV), etc.

2.5.1 Plug-in method

The intuitive and simple way to obtain bandwidths is plug-in method. Since the $\|f''\|_2^2$ is unknown in (2.21), it was proposed to assign a value to the $\|f''\|_2^2$ for the ideal bandwidth in [35]. For example, if a Gaussian kernel is being used, then

$$\begin{aligned}\int f''(x)^2 dx &= \sigma^{-5} \int \phi''(x)^2 dx \\ &= \frac{3}{8}\pi^{-1/2}\sigma^{-5} \approx 0.212\sigma^{-5}\end{aligned} \tag{2.22}$$

So, substituting this into the equation (2.21), optimal bandwidth can be obtained as

$$h_{MISE} = (4\pi)^{-1/10}\left(\frac{3}{8}\pi^{-1/2}\right)^{-1/5}\sigma n^{-1/5} = \left(\frac{4}{3}\right)^{1/5}\sigma n^{-1/5} = 1.06\sigma n^{-1/5} \tag{2.23}$$

For obtaining the better results, interquartile range can be used as a robust alternative for a standard deviation. This modified version (see [35])

$$h_{robust} = 1.06 \min\left(\text{standard deviation}, \frac{\text{interquartile range}}{1.34}\right) n^{-1/5} \tag{2.24}$$

2.5.2 Least Squares Cross-Validation method

Given any estimator $\widehat{f_h}$ of a density f, we know ISE can be written as

$$ISE(\widehat{f_h}(x)) = \int \left\{\widehat{f_h}(x) - f(x)\right\}^2 dx = \int \widehat{f_h}(x)^2 dx - 2\int \widehat{f_h}(x) f(x)\, dx + \int f(x)^2 dx. \tag{2.25}$$

Since the last term of ISE does not depend on $\widehat{f_h}$, so the ideal choice of bandwidth corresponds to the choice which minimizes the quantity defined by

$$R(\widehat{f_h}) = \int \widehat{f_h}(x)^2 dx - 2\int \widehat{f_h}(x) f(x)\, dx. \tag{2.26}$$

The idea of LSCV is to construct an estimate of $R(\widehat{f}_h)$ from the data themselves and then minimize this estimate over h to give the choice of window width. Define $\widehat{f}_{h,-i}$ to be the density estimate constructed from all the data points except X_i, that is

$$\widehat{f}_{h,-i}(x) = \frac{1}{(n-1)h} \sum_{j \neq i} K\left(\frac{x - X_j}{h}\right), \tag{2.27}$$

Now, define

$$M(h) = \int \widehat{f}_h(x)^2 dx - \frac{2}{n} \sum_i \widehat{f}_{h,-i}(X_i). \tag{2.28}$$

The score M depends only on the data. So, the basic principle of LSCV is to minimize M over h.

2.5.3 Likelihood Cross-Validation

The likelihood cross-validation choice of h is the value of h which maximizes the function $CV(h)$

$$CV(h) = \frac{1}{n} \sum_i \log \widehat{f}_{h,-i}(X_i) \tag{2.29}$$

for the given data. This method, does not present severe computational difficulties. Maximizing CV(h) should yield a density estimate which is close to the true density in terms of Kullback-Leibler information distance, defined by

$$I(f, \widehat{f}_h) = \int f(x) \log \left\{ f(x) / \widehat{f}_h(x) \right\} dx. \tag{2.30}$$

3. DELTA SEQUENCE DENSITY ESTIMATION USING SECOND ORDER MODULUS OF CONTINUITY TYPE MAJORANTS

3.1 Purpose

In this chapter, the local convergence rate of MSE corresponding to d-variate delta sequence density estimator are obtained for both univariate and multivariate cases. To weaken the differentiability conditions used in the former studies, the assumptions on the density function are written using the second order modulus of continuity type majorants.

Throughout this chapter, $\mathbf{x} = (x_1, x_2, ..., x_d)$ denotes a point in the d-dimensional Euclidean space R^d and $d\mathbf{x}$ denotes $\prod_{i=1}^{d} dx_i$.

3.2 Second Order Modulus of Continuity Type Majorants

This section mainly stems from the use of second order modulus of continuity type majorants. So, the definition of higher order modulus of continuity and its some of the useful properties are given below.

Definition 4. *The modulus of continuity of order $k \geq 1$ of a function $f \in C([a,b])$ is defined as follows:*

$$w_k(t) = w_k(f;a,b;t) = \sup_{x\in[a,b],\ x+kh\in[a,b],\ |h|\leq t} \left|\Delta_h^k f(x)\right| \tag{3.1}$$

where

$$\Delta_h^k f(x) = \sum_{\nu=0}^{k} (-1)^{k-\nu} \binom{k}{\nu} f(x+\nu h) \tag{3.2}$$

defined for non-negative values of $t \leq \frac{b-a}{k}$.

In this dissertation second order modulus of continuity is used, so $k = 2$. For detailed discussion of higher order modulus of continuity, one can refer to [36].

Definition 5. *A function $w_k(t) : [0,1] \to [0,\infty)$ which satisfies the conditions:*

(a) $w_k(0) = 0$,

(b) w_k *is nondecreasing,*

(c) w_k *is continuous,*

(d) $t_2^{-k} w_k(t_2) \leq 2^k t_1^{-k} w_k(t_1), \quad 0 \leq t_1 \leq t_2$

is called kth order modulus of continuity type majorant.

Modulus of smoothness generally used in approximation theory, Fourier analysis and their applications. It describes the structural properties of functions; in particular, they describe the measure of smoothness of the function via the k-th difference $\Delta_h^k f(x)$. In fact, for functions belonging to the Lebesgue space L^p, $1 \leq p < \infty$ or the space of continuous functions C, the classical k-th modulus of continuity has turned out to be a rather good measure for determining the rate of convergence of best approximation.

Following inequalities are necessary for the derivation of the proof.

$$w_2(nt) \leq n^2 w_2(t), \quad n \in N \tag{3.3}$$

$\forall t \geq 1, \forall \delta > 0$

$$w_2(t\delta) \leq (2t)^2 w_2(\delta). \tag{3.4}$$

In particular when $t = \frac{1}{\delta}$ then

$$w_2(1) \leq \frac{2^2}{\delta^2} w_2(\delta). \tag{3.5}$$

Then, from the property (d) of the Definition 5

$$w_2(\delta) = w_2(\delta)\frac{\delta^2}{\delta^2} = \frac{w_2(\delta)}{\delta^2} \int_0^\delta 2t dt \leq c \int_0^\delta \frac{w_2(t)}{t} dt. \tag{3.6}$$

The following operator is also useful for the derivation of the proof.

Definition 6. *(see [37]). For the second order modulus of continuity type majorant* w_2,

$$Z_2(w_2, \delta) = \int_0^\delta \frac{w_2(t)}{t} dt + \delta^2 \int_\delta^1 \frac{w_2(t)}{t^3} dt \tag{3.7}$$

is called Zygmund operator.

Now, let $\{\mathbf{X}_n\}$ be a sequence of i.i.d. random vectors in R^d with density function $f(\mathbf{x})$. The delta sequence density estimator of $f(\mathbf{x})$ defined by

$$\widehat{f}_{n,m}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n}\delta_m(\mathbf{x},\mathbf{X}_i) \tag{3.8}$$

where $\{\delta_m\}$ with $m = m(n)$ is a positive type delta sequence.

3.3 MSE of Delta Sequence Based Density Estimators Using Second Order Modulus of Continuity Type Majorants

In this subsection, the bias and variance of the estimator are investigated separately, since MSE of an estimator is defined as the sum of squared bias and variance of estimator at a point.

Theorem 1. *Let $f \in L^p(R^d), 1 \le p < \infty$ and $\{\delta_m\}$ be a delta sequence of positive type.*

(i) $\int \delta_m(\mathbf{x},\mathbf{t})f(\mathbf{t})d\mathbf{t} \to f(\mathbf{x})$ a.e. Lebesgue$(\mathbf{x})$;

(ii) If there exist $\eta > 0$ such that $|f(\mathbf{x}+\mathbf{t})+f(\mathbf{x}-\mathbf{t})-2f(\mathbf{x})| \le cw_2(|\mathbf{t}|)$, $|\mathbf{t}| \le \eta < 1$, then the order of bias term

$$\left|\int \delta_m(\mathbf{x},\mathbf{t})f(\mathbf{t})d\mathbf{t} - f(\mathbf{x})\right| = O\left(Z_2\left(w_2,\frac{1}{m}\right)\right). \tag{3.9}$$

Proof. Part (i) is similar to Theorem 1.25 of [38], so its proof is not given. For (ii), by using change of variable and adding and subtracting some terms, it is obtained as

$$\begin{aligned}\left|\int_{-\infty}^{+\infty} \delta_m(\mathbf{x},\mathbf{t})f(\mathbf{t})d\mathbf{t} - f(\mathbf{x})\right| &= \left|\frac{1}{2}\int_{-\infty}^{+\infty} \delta_m(\mathbf{x},\mathbf{x}+\mathbf{t})\left[f(\mathbf{x}+\mathbf{t})+f(\mathbf{x}-\mathbf{t})-2f(\mathbf{x})\right]d\mathbf{t}\right| \\ &\le \left|\int_{|\mathbf{t}|\le\eta}\right| + \left|\int_{|\mathbf{t}|\ge\eta}\right| = I_1 + I_2\end{aligned} \tag{3.10}$$

By a change of variable to polar coordinates and using properties of positive type delta sequence gives for $m \ge \eta^{-1}$:

$$\begin{aligned}I_1 &\le c_1\int_0^{\eta}\frac{m^d}{1+(mr)^{d+2}}r^{d-1}w_2(r)dr \\ &= c_1\left\{\int_0^{\frac{1}{m}}\frac{m^d}{1+(mr)^{d+2}}r^{d-1}w_2(r)dr + \int_{\frac{1}{m}}^{\eta}\frac{m^d}{1+(mr)^{d+2}}r^{d-1}w_2(r)dr\right\} = c_1\left(I_1' + I_1''\right).\end{aligned} \tag{3.11}$$

Let's investigate $I_1^{'}$ and $I_1^{''}$ separately. Since $0 \leq r \leq \frac{1}{m}$, a bound for $I_1^{'}$ can be obtained as

$$\begin{aligned} I_1^{'} &= \int_0^{\frac{1}{m}} \frac{m^d}{1+(mr)^{d+2}} r^{d-1} w_2(r) dr \\ &\leq c_2 \int_0^{\frac{1}{m}} \frac{w_2(r)}{r} dr. \end{aligned} \tag{3.12}$$

For $I_1^{''}$

$$I_1^{''} = \int_{\frac{1}{m}}^{\eta} \frac{m^d}{1+(mr)^{d+2}} r^{d-1} w_2(r) dr \leq \frac{c_3}{m^2} \int_{\frac{1}{m}}^{\eta} \frac{w_2(r)}{r^3} dr. \tag{3.13}$$

Hence by combining $I_1^{'}$ and $I_1^{''}$, a bound for I_1 can be obtained as

$$I_1 \leq c_4 \left(\int_0^{\frac{1}{m}} \frac{w_2(r)}{r} dr + \frac{1}{m^2} \int_{\frac{1}{m}}^{\eta} \frac{w_2(r)}{r^3} dr \right) = O\left(Z_2 \left(w_2, \frac{1}{m} \right) \right). \tag{3.14}$$

For I_2 the cases $d > 1$ and $d = 1$ should be investigated separately. When $d > 1$ using Hölder inequalities,

$$\begin{aligned} I_2 &= \left| \int_{|\mathbf{t}| \geq \eta} \delta_m(\mathbf{x}, \mathbf{x}+\mathbf{t}) \left[f(\mathbf{x}+\mathbf{t}) + f(\mathbf{x}-\mathbf{t}) - 2f(\mathbf{x}) \right] d\mathbf{t} \right| \\ &\leq \left| \int_{|\mathbf{t}| \geq \eta} \delta_m(\mathbf{x}, \mathbf{x}+\mathbf{t}) f(\mathbf{x}+\mathbf{t}) d\mathbf{t} \right| + \left| \int_{|\mathbf{t}| \geq \eta} \delta_m(\mathbf{x}, \mathbf{x}+\mathbf{t}) f(\mathbf{x}-\mathbf{t}) d\mathbf{t} \right| \\ &+ 2 \left| \int_{|\mathbf{t}| \geq \eta} \delta_m(\mathbf{x}, \mathbf{x}+\mathbf{t}) f(\mathbf{x}) d\mathbf{t} \right| \\ &\leq 2 \|f\|_p \left\| \psi_\eta \delta_m \right\|_q + 2 |f(\mathbf{x})| \left\| \psi_\eta \delta_m \right\|_1 \end{aligned} \tag{3.15}$$

where $\psi_\eta = \chi_{\{\mathbf{t} \in R^d : |\mathbf{t}| > \eta\}}$.

Now, to obtain an order for the $\left\|\psi_\eta \delta_m\right\|_q$ term, it can be written as follows

$$
\begin{aligned}
\left\|\psi_\eta \delta_m\right\|_q &= \left[\int |\psi_\eta \delta_m|^q \, d\mathbf{t}\right]^{\frac{1}{q}} = \left[\int\limits_{|\mathbf{t}|\geq\eta} (\delta_m(\mathbf{x},\mathbf{x}+\mathbf{t}))^q d\mathbf{t}\right]^{\frac{1}{q}} \\
&\leq c_5 \left[\int\limits_{\eta}^{\infty} \left(\frac{m^d}{(mr)^{d+2}} r^{d-1}\right)^q dr\right]^{\frac{1}{q}} \\
&= c_5 \left[\left(\frac{1}{m^2}\right)^q \int\limits_{\eta}^{\infty} \left(r^{-3q}\right) dr\right]^{\frac{1}{q}} = O\left(\frac{1}{m^2}\right).
\end{aligned}
\tag{3.16}
$$

For $\left\|\psi_\eta \delta_m\right\|_1$

$$
\begin{aligned}
\left\|\psi_\eta \delta_m\right\|_1 &= \left[\int |\psi_\eta \delta_m| \, d\mathbf{t}\right] \leq c_6 \int\limits_{r\geq\eta} \frac{m^d}{(mr)^{d+2}} r^{d-1} dr \\
&= \frac{c_6}{m^2} \int\limits_{r\geq\eta} \frac{dr}{r^3} = O\left(\frac{1}{m^2}\right).
\end{aligned}
\tag{3.17}
$$

Therefore, (3.16) and (3.17) imply

$$
I_2 = O\left(\frac{1}{m^2}\right). \tag{3.18}
$$

When $d = 1$:

$$
\begin{aligned}
I_2 &= \left| \int\limits_{|t|\geq\eta} \delta_m(x,x+t)\left[f(x+t)+f(x-t)-2f(x)\right] dt \right| \\
&\leq \int\limits_{|t|\geq\eta} |\delta_m(x,x+t)|\,|f(x+t)|\, dt + \int\limits_{|t|\geq\eta} |\delta_m(x,x+t)|\,|f(x-t)|\, dt \\
&\quad + 2|f(x)| \int\limits_{|t|\geq\eta} |\delta_m(x,x+t)|\, dt
\end{aligned}
\tag{3.19}
$$

For the first term of (3.19),

$$
\begin{aligned}
\int\limits_{|t|\geq\eta} |\delta_m(x,x+t)|\,|f(x+t)|\, dt &\leq \max_{t\geq\eta} \delta_m(x,x+t) \int\limits_{|t|\geq\eta} |f(x+t)|\, dt \\
&\leq \frac{1}{m^2\eta^3} c_7 = O\left(\frac{1}{m^2}\right).
\end{aligned}
\tag{3.20}
$$

For the other terms, the same order can be obtained similarly. Hence, for $d > 1$ and $d = 1$

$$
I_2 = O\left(\frac{1}{m^2}\right) \tag{3.21}
$$

Therefore, by combining (3.14) and (3.21)

$$\left|\int \delta_m(\mathbf{x},\mathbf{t})f(\mathbf{t})d\mathbf{t} - f(\mathbf{x})\right| \leq c_8 \left\{ \frac{1}{m^2} + \int\limits_0^{1/m} \frac{w_2(r)}{r}dr + \frac{1}{m^2}\int\limits_{1/m}^{\eta} \frac{w_2(r)}{r^3}dr \right\} \tag{3.22}$$

Then, using the properties of second order modulus of continuity type majorants a bound for the $\frac{1}{m^2}$ term can be obtained as

$$\frac{1}{m^2} \leq \frac{4}{w_2(1)} w_2(\frac{1}{m}) \leq c_9 \int\limits_0^{\frac{1}{m}} \frac{w_2(r)}{r}dr. \tag{3.23}$$

Finally,

$$\left|\int \delta_m(\mathbf{x},\mathbf{t})f(\mathbf{t})d\mathbf{t} - f(\mathbf{x})\right| = O\left(Z_2\left(w_2,\frac{1}{m}\right)\right) \tag{3.24}$$

□

To obtain MSE of the estimator, first let us investigate the variance of the estimator.

$$nVar\left\{\widehat{f}(\mathbf{x})\right\} = Var\left\{\delta_m(\mathbf{x},\mathbf{t})\right\} = \int \delta_m^2(\mathbf{x},\mathbf{t})f(\mathbf{t})d\mathbf{t} - \left(\int \delta_m(\mathbf{x},\mathbf{t})f(\mathbf{t})d\mathbf{t}\right)^2 \tag{3.25}$$

For the variance term, it is sufficient to obtain the rate for $\int \delta_m^2(\mathbf{x},\mathbf{t})f(\mathbf{t})d\mathbf{t}$. Since $\int \delta_m(\mathbf{x},\mathbf{t})f(\mathbf{t})d\mathbf{t} \leq \|f\|_\infty$ and $\|\delta_m(\mathbf{x},\mathbf{t})\|_\infty = O(m^d)$ then

$$\int \delta_m^2(\mathbf{x},\mathbf{t})f(\mathbf{t})d\mathbf{t} \leq \sup_{\mathbf{x}} \delta_m(\mathbf{x},\mathbf{t}) \int \delta_m(\mathbf{x},\mathbf{t})f(\mathbf{t})d\mathbf{t} = O\left(m^d\right) \tag{3.26}$$

Finally,

$$Var\left\{\widehat{f}(\mathbf{x})\right\} = O\left(\frac{m^d}{n}\right). \tag{3.27}$$

Hence using the orders for the bias (3.24) and variance (3.27), the local convergency rate for the MSE of the estimator is

$$\begin{aligned} MSE\left\{\widehat{f}(\mathbf{x})\right\} &= Var\left\{\widehat{f}(\mathbf{x})\right\} + Bias^2\left\{\widehat{f}(\mathbf{x})\right\} \\ &= O\left(\frac{m^d}{n} + Z_2^2\left(w_2,\frac{1}{m}\right)\right). \end{aligned} \tag{3.28}$$

Corollary 1. *Let* $w_2(\mathbf{t}) = |\mathbf{t}|^\alpha$, $1 < \alpha < 2$, $m = n^{\frac{1}{2\alpha+d}}$, *then*

$$MSE\left\{\widehat{f}(\mathbf{x})\right\} = O\left(n^{-\frac{2\alpha}{2\alpha+d}}\right). \tag{3.29}$$

Corollary 2. *Let* $w_2(\mathbf{t}) = |\mathbf{t}|^2$, $m = n^{\frac{1}{5}}$ *then*

$$MSE\left\{\widehat{f}(\mathbf{x})\right\} = O\left(n^{-\frac{4}{d+4}}(\ln n)^2\right) \tag{3.30}$$

In the following example, it is shown that the rate of the MSE of $\widehat{f}(0)$ is greater than or equal to $c_2 n^{-\frac{4}{5}} (\ln n)^2$ and the rate of the MSE of $\widehat{f}(0)$ is also less than or equal to $c_2 n^{-\frac{4}{5}} (\ln n)^2$ by the Corollary 2 when $d = 1$, which proves the MSE of $\widehat{f}(x)$ at $x = 0$ has the best possible convergency rate.

Example 1. *Let*

$$f(x) = \begin{cases} x(x+|x|), & -1 \leq x \leq 1 \\ 0, & elsewhere \end{cases}$$

and $w_2(t) = t^2$, *and*

$$\delta_m(x,0) = \frac{cm}{1+(m|x|)^3}. \tag{3.31}$$

It can be seen that, f belongs to the function classes which is defined by second order finite differences. Let us look at the rate of convergence at $x = 0$. For the bias term we have

$$\begin{aligned} \int_{-\infty}^{+\infty} \delta_m(x,0) f(x) dx - f(0) &= \int_{-1}^{1} \frac{cm}{1+(m|x|)^3} x^2 dx + \int_{-1}^{1} \frac{cm}{1+(m|x|)^3} x|x| dx \\ &= \frac{c}{m^2} \ln\left|1+m^3\right| \geq \frac{c_1}{m^2} \ln m \end{aligned} \tag{3.32}$$

where $m > 1$. Hence, when $m = n^{\frac{1}{5}}$ we have

$$MSE\left\{\widehat{f}(0)\right\} = Var\left\{\widehat{f}(0)\right\} + Bias^2\left\{\widehat{f}(0)\right\} \geq Bias^2\left\{\widehat{f}(0)\right\} = c_2 n^{-\frac{4}{5}} (\ln n)^2. \tag{3.33}$$

By using second order modulus of continuity type majorants the faster convergency rate of the MSE of the densities that belong to the class which is defined by second order finite differences is obtained when compared to the one obtained by using the first order finite differences.

4. DENSITY ESTIMATION BASED ON HERMITE POLYNOMIALS

In this chapter, delta sequence density estimation method based on Hermite polynomials is proposed. The convergency rate of MSE and MISE of the density estimators are obtained using Hermite polynomials for the densities having compact or infinite support. By using delta sequence density estimation method, it is shown that the convergency rate of the MISE is better than the rates obtained in papers [6] and [7]. Moreover, for the density functions which have at least third order derivative, the convergency rate of the MISE for the proposed estimator is faster than the rate obtained in paper [8]. On the order hand, orthogonal series estimators based on Hermite functions are useful in applied work. It requires considerably less computational time than kernel estimators for large N since the computations are only based on the recurrence relations for the Hermite functions. However, orthogonal series estimators have a drawback since those estimators can take negative values as oppose to modified Jakobi polynomials (see [9]). Hence, the problem of obtaining a nonnegative orthogonal series estimator based on Hermite functions may be a challenging work for future studies.

4.1 Delta Sequence Density Estimators Based on Hermite Polynomials

The Hermite orthonormal system over the real line R given by

$$h_k(x) = \left(2^k k! \pi^{\frac{1}{2}}\right)^{-\frac{1}{2}} H_k(x) e^{-\frac{x^2}{2}}, \quad k = 0, 1, 2, \ldots \tag{4.1}$$

where

$$H_k(x) = (-1)^k e^{x^2} \left(\frac{d^k}{dx^k}\right) e^{-x^2} \tag{4.2}$$

is the kth Hermite polynomial.

It is known that the normalized Hermite functions $\{h_k\}$ are the complete orthonormal system in $L_2(-\infty, \infty)$ and they satisfy the recurrence formulas

$$x h_k = \left(\frac{k}{2}\right)^{\frac{1}{2}} h_{k-1} + \left(\frac{k+1}{2}\right)^{\frac{1}{2}} h_{k+1}, \quad k = 1, 2, \ldots \tag{4.3}$$

and

$$\frac{d}{dx}h_k = \left(\frac{k}{2}\right)^{\frac{1}{2}} h_{k-1} - \left(\frac{k+1}{2}\right)^{\frac{1}{2}} h_{k+1}, \quad k = 1,2,... \tag{4.4}$$

and satisfy the following inequalities (see [39]);

$$|h_k(x)| \leq \frac{c_1}{(k+1)^{1/12}}, \ x \in (-\infty,\infty), \quad k = 0,1,2... \tag{4.5}$$

and

$$|h_k(x)| \leq \frac{c_2}{(k+1)^{1/4}}, \ x \in (-a,a), \quad k = 0,1,2... \tag{4.6}$$

where a is any nonnegative integer and the constants c_1 and c_2 are independent of x and k.

Let $X_1, X_2, ..., X_N$ be a sequence of i.i.d. random variables with unknown density function $f(x)$. Then, an unknown density function f can be written by means of the Hermite series

$$f(x) = \sum_{k=0}^{\infty} a_k h_k(x) \tag{4.7}$$

with the Hermite coefficients defined by

$$a_k = \int f(x) h_k(x) dx. \tag{4.8}$$

Throughout this chapter, we shall assume that $f(x)$ is square integrable and we use c or c_i, $i = 1,2,...,m$ for any positive constant, independent of f. Now, by using Hermite polynomials delta sequence density estimator can be defined as

$$\widehat{f}_{N,n}(x) = \frac{1}{N}\sum_{i=1}^{N} \delta_n(x, X_i) \tag{4.9}$$

where

$$\delta_n(x,X_i) = \sum_{k=0}^{n} \frac{H_k(x)e^{-\frac{x^2}{2}}}{\left(2^k k! \pi^{\frac{1}{2}}\right)^{\frac{1}{2}}} \frac{H_k(X_i)e^{-\frac{x^2}{2}}}{\left(2^k k! \pi^{\frac{1}{2}}\right)^{\frac{1}{2}}} = \sum_{k=0}^{n} h_k(x) h_k(X_i). \tag{4.10}$$

Since orthogonal series density estimate can take negative values, then it is proposed that density estimate at x is the max $\left[0, \widehat{f}_{N,n}(x)\right]$. For the discussion of the MISE and MSE, following two lemmas, proved by [6], are necessary.

Lemma 1. *Assume that the function* $\left(x - \frac{d}{dx}\right)^r f \in L_2(-\infty,\infty)$ *for some integer* $r > 0$. *Then the coefficients* a_k, $k = 1,2,...$ *satisfy the bound*

$$|a_k| \leq \frac{c_3}{(2k)^{\frac{r}{2}}} \tag{4.11}$$

where c_3 *is the* L_2 *norm of* $\left(x - \frac{d}{dx}\right)^r f$.

Lemma 2. *Let f(x) be continuous, of bounded variation, L_1 and L_2 in $(-\infty,\infty)$. Then, the series in*

$$f(x)=\sum_{k=0}^{\infty}a_k h_k(x) \tag{4.12}$$

converges uniformly in any interval interior to $(-\infty,\infty)$.

4.1.1 Convergency rate of MISE of estimators for densities having infinite support

Theorem 2. *Let $\left(x-\frac{d}{dx}\right)^j f \in L_2(-\infty,\infty)$ for $j=1,2,...,r$; then the MISE rate of (4.9) satisfies*

$$MISE(\widehat{f}_{N,n}(x)) = O\left(N^{-\frac{12r}{6r+23}}\right). \tag{4.13}$$

Proof. First, lets investigate the integrated variance term

$$N\int_{-\infty}^{\infty} var(\widehat{f}_{N,n}(x))dx = \int_{-\infty}^{\infty}\left\{\int_{-\infty}^{\infty}\delta_n^2(x,t)f(t)dt-\left(\int_{-\infty}^{\infty}\delta_n(x,t)f(t)dt\right)^2\right\}dx \tag{4.14}$$

it is sufficient to investigate the rate of grows of $\int\int\delta_n^2(x,t)f(t)dtdx$.

$$\begin{aligned}
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\delta_n^2(x,t)f(t)dtdx &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\sum_{k=0}^{n}h_k(x)h_k(t)\right)^2 f(t)dtdx\\
&\le \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\sum_{k=0}^{n}h_k^2(x)\sum_{k=0}^{n}h_k^2(t)\right)f(t)dtdx\\
&= \int_{-\infty}^{\infty}\sum_{k=0}^{n}h_k^2(x)dx\int_{-\infty}^{\infty}\sum_{k=0}^{n}h_k^2(t)f(t)dt\\
&\le \int_{-\infty}^{\infty}\sum_{k=0}^{n}h_k^2(x)dx\sum_{k=0}^{n}\frac{1}{(k+1)^{\frac{1}{12}}}\int_{-\infty}^{\infty}h_k(t)f(t)dt
\end{aligned} \tag{4.15}$$

From Lemma 1, it can be deduced that

$$|a_k|\le\frac{c}{(2k+2)^{r/2}},\quad k=0,1,2,... \tag{4.16}$$

Since

$$\int_{-\infty}^{\infty}e^{-x^2}H_k^2(x)dx=\sqrt{\pi}2^n n! \tag{4.17}$$

and using above bound, an order for the integrated variance term can be obtained as

$$\int_{-\infty}^{\infty}var\left(\widehat{f}_{N,n}(x)\right)dx = O\left(\frac{n^{\frac{23-6r}{12}}}{N}\right) \tag{4.18}$$

Now, the squared bias term can be written as

$$
\begin{aligned}
Bias^2\left(\widehat{f}_{N,n}(x)\right) &= \left\{\int\limits_{-\infty}^{\infty} \delta_n(x,t)f(t)dt - f(x)\right\}^2 \\
&= \left\{\int\limits_{-\infty}^{\infty} \sum_{k=0}^{n} h_k(x)h_k(t)f(t)dt - \sum_{k=0}^{\infty} a_k h_k(x)\right\}^2 \\
&= \left\{\sum_{k=n+1}^{\infty} a_k h_k(x)\right\}^2
\end{aligned} \tag{4.19}
$$

since Hermite functions are orthonormal in L_2, then the integrated squared bias is written as follows

$$
\int\limits_{-\infty}^{\infty} Bias^2\left(\widehat{f}_{N,n}(x)\right) = \sum_{k=n+1}^{\infty} a_k^2 \tag{4.20}
$$

In order to find a bound for integrated squared bias, the inequality $a_k^2 \leq k^{-r} b_{k+r}^2$ is used which is derived in paper [7] under the condition $(x - d/dx)^r f(x)$. Where b_k is the kth coefficients of the expansion of $\left(x - \frac{d}{dx}\right)^r f$ in the Hermite series and the series $\sum b_k^2$ converges. So, by virtue of the result of the paper [7],

$$
\int\limits_{-\infty}^{\infty} Bias^2\left(\widehat{f}_{N,n}(x)\right) = \sum_{k=n+1}^{\infty} a_k^2 = O(n^{-r}). \tag{4.21}
$$

Since MISE can be expressed as the sum of integrated squared bias and integrated variance, by combining (4.18) and (4.21)

$$
MISE(\widehat{f}_{N,n}(x)) = O\left(N^{\frac{-12r}{6r+23}}\right). \tag{4.22}
$$

□

Remark 1. *Hermite series method was used in [6] and [7] and the convergency rate of MISE of the density estimator was obtained as* $O\left(N^{-\frac{(r-1)}{r}}\right)$ *in paper [6] and for the estimate of the pth derivative and assuming that the density has r derivatives where* $0 \leq p < r$, *the MISE rate obtained as* $O\left(N^{-\frac{6(r-p)-5}{6r}}\right)$ *in paper [7]. Note that, for comparison reasons take* $p = 0$ *for the proposed estimator and also for the estimator used in paper [6]. The hypothesis are the same but faster result are obtained using delta sequence method.*

Remark 2. *For* $r > 2$, *the rate of convergency of MISE is better than* $O\left(N^{-\frac{2(r-p)}{2r+1}}\right)$ *which was obtained in paper [8].*

4.2 Convergency Rate of MSE and MISE of Estimators for Densities Having Compact Support

In this section, the rate of convergence of MSE and MISE of an estimator of densities having compact support and based on Hermite functions are investigated. Since the density function has compact support then the hypothesis weaken a little.

Theorem 3. *Let f have compact support and suppose $\left(\frac{d}{dx}\right)^j f \in L_2$ for $j = 1,2,...,r$; then the MSE of the estimate (4.9) satisfies*

$$MSE(\widehat{f}_{N,n}(x)) = O(N^{-\frac{2r+1}{2r+2}}) \tag{4.23}$$

Proof. First, lets obtain bound for delta sequence by using (4.6)

$$|\delta_n(x,t)| = \left|\sum_{k=0}^{n} h_k(x)h_k(t)\right| \leq \sum_{k=0}^{n} |h_k(x)|\,|h_k(t)| \leq \sum_{k=0}^{n} \frac{1}{(1+k)^{1/2}} \tag{4.24}$$

If a convenient integral is used as the upper bound for the (4.24), then

$$|\delta_n(x,t)| = O\left(n^{1/2}\right) \tag{4.25}$$

For the variance term, since $\int \delta_m(x,t)f(t)dt \leq \|f\|_\infty$

$$Var(\widehat{f}_{N,n}(x)) \leq \frac{1}{N}\int \delta_n^2(x,t)f(t)dt \leq \frac{c_4}{N}\sup|\delta_n(x,t)| \leq c_5\frac{n^{1/2}}{N}. \tag{4.26}$$

The bias term can be written as

$$\begin{aligned}
bias^2(\widehat{f}_{N,n}(x)) &= \left(\int \delta_n(x,t)f(t)dt - f(x)\right)^2 \\
&= \left(\int \sum_{k=0}^{n} h_k(x)h_k(t)f(t)dt - \sum_{k=0}^{\infty} a_k h_k(x)\right)^2 = \left(\sum_{k=n+1}^{\infty} a_k h_k(x)\right)^2 \\
&\leq \left(\sum_{k=n+1}^{\infty} b_{k+r}\frac{1}{(2k)^{\frac{r}{2}}}\frac{1}{(k+1)^{\frac{1}{4}}}\right)^2 \leq \left(\sum_{k=n+1}^{\infty} b_{k+r}k^{\frac{-2r-1}{4}}\right)^2 \\
&\leq \left((n+1)^{-\frac{2r+1}{4}}\sum_{k=n+1+r}^{\infty} b_k\right)^2
\end{aligned} \tag{4.27}$$

Note that, since f has compact support and $D^r f \in L^2$, $x^p D^s f \in L^2$ for all integers $p \geq 0$ and $0 \leq s \leq r$. So, it follows that $(x-D)^r f \in L^2$. Then, the bounds for $\left|a_k^2\right|$ obtained by Walter (1977) can be used to obtain an order for the squared bias term

$$bias^2(\widehat{f}_{N,n}(x)) = O(n^{-\frac{2r+1}{2}}). \tag{4.28}$$

So, MSE is

$$MSE(\widehat{f}_{N,n}(x)) = O\left(\frac{n^{\frac{1}{2}}}{N} + \frac{1}{n^{\frac{2r+1}{2}}}\right). \tag{4.29}$$

If $n = N^{\frac{1}{r+1}}$ is chosen, then the MSE rate of estimator based on Hermite functions is obtained as below

$$MSE(\widehat{f}_{N,n}(x)) = O(N^{-\frac{2r+1}{2r+2}}) \tag{4.30}$$

$\square$

Theorem 4. *Let f have compact support and suppose $\left(\frac{d}{dx}\right)^j f \in L_2$ for $j = 1,2,...,r$; then the MISE of the estimate (4.9) satisfies*

$$MISE(\widehat{f}_{N,n}(x)) = O(N^{-\frac{2r+1}{2r+2}}). \tag{4.31}$$

Proof. The proof is similar to earlier one. Notice that the bounds of delta sequence in (4.25) is used for the integrated variance term. $\square$

Remark 3. *The rates obtained in this study are better than those reported by [7] for the densities having compact support. Moreover, the rates of convergence obtained in this work is also better than those suggested by [9] who used delta sequence method to obtain rate of convergence of estimator based on Jakobi polynomials. In paper [9], it is reported that, the convergency rate of MISE and MSE as $O\left(N^{-\frac{1}{3}}\right)$. They are considerably slower than the rates obtained in this study. However, in paper [9], the negativity problem of orthogonal series estimators based on Jakobi polynomials was solved by using certain summability methods. In this thesis, the negativity problem of Hermite series estimators could not be solved, so it is assumed that density estimate at x is the* max $\left[0, \widehat{f}_{N,n}(x)\right]$ *to avoid the negative values of the estimator. So, the problem of obtaining a nonnegative orthogonal series estimator based on Hermite functions may be a future study.*

5. BETA PRIME DENSITY ESTIMATOR

It is generally thought that, the bandwidth choice is more crucial than the kernel function in density estimation problem. However, when the density function has closed or semi infinite support, then classical symmetric kernel estimator has edge effect (or boundary bias problem) since it causes the leakage of probability mass. In this chapter, to avoid the edge effect problem, a new asymmetric kernel estimator is established to estimate the densities having support on $[0,\infty)$. Beta prime distribution function is used as a kernel instead of classical symmetric kernels. It is shown that, beta prime kernel estimator is free of boundary bias problem. Also, similar to existing asymmetric kernel estimators, the variance of the beta prime kernel estimator reduces as the position, where the smoothing is made, moves away from the boundary. Then, the expressions for the bandwidths that minimize the asymptotic approximation for the MSE and MISE are obtained. Furthermore, simulation studies are conducted to show the superior performance of the beta prime estimator over some existing asymmetric kernel estimators in terms of average ISE. For bandwidth selection problem, adaptive Bayesian with Lindley approximation method is proposed. Lindley approximation method have not been used before for the asymmetric kernel estimators. To show the efficiency of this method, a comparison is made between bandwidths obtained from adaptive Bayesian with Lindley approximation method and bandwidths obtained from global LSCV method by using simulation studies. Moreover, real data applications are made to demonstrate the usefulness of the beta prime estimator and new bandwidth selection method.

5.1 Beta Prime Kernel Estimator

Let X_i, $i = 1,2,...,n$ be a random sample from a distribution with an unknown pdf f having support on the positive real line. Lets assume

i. f is twice continuously differentiable,

ii. $\int_0^\infty (1+x)^{-\frac{1}{2}} f(x)dx < \infty$

iii. $\int_0^\infty \left\{f^{'}(x)\right\}^2 dx < \infty, \quad \int_0^\infty \left\{(1+x)f^{''}(x)\right\}^2 dx < \infty$.

These assumptions are necessary for the Taylor expansion and finiteness of the integrated bias and integrated variance.

Let $K_{B(\lambda,\mu)}$ be the density of a $Betaprime(\lambda,\mu)$ distributed random variable Y defined as:

$$K_{B(\lambda,\mu)}(y) = \frac{y^{\lambda-1}(1+y)^{-\lambda-\mu}}{\beta(\lambda,\mu)}, \quad y > 0,\ \lambda > 0,\ \mu > 0. \tag{5.1}$$

The mean and variance of Y are equal to

$$E(Y) = \frac{\lambda}{\mu-1},\ \mu > 1, \quad Var(Y) = \frac{\lambda(\lambda+\mu-1)}{(\mu-1)^2(\mu-2)}, \quad \mu > 2. \tag{5.2}$$

The beta prime kernel can be written as

$$K_{B(\frac{x^2}{b}+x+1,\frac{x}{b}+\frac{1}{x+b}+1)}(y) = \frac{y^{\frac{x^2}{b}+x}(1+y)^{-(\frac{x^2}{b}+x+\frac{x}{b}+\frac{1}{x+b}+2)}}{\beta(\frac{x^2}{b}+x+1,\frac{x}{b}+\frac{1}{x+b}+1)}, \quad y > 0 \tag{5.3}$$

where β is the beta function, b is a smoothing parameter satisfying the condition that $b \to 0$ and $nb \to \infty$ as $n \to \infty$, and x is the point where the density is estimated. So, the beta prime kernel estimator defined as, for $x \in [0,\infty)$,

$$\widehat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} K_{B(\frac{x^2}{b}+x+1,\frac{x}{b}+\frac{1}{x+b}+1)}(X_i). \tag{5.4}$$

It is similar to standard kernel estimator, only replaces fixed kernel with beta prime kernel. Figure 5.1 demonstrates the kernel shapes of IG, RIG, Gam2 and beta prime depend on the value of x and smoothing parameter b. The amount of smoothing applied by the illustrated kernel estimators are controlled by the chosen parameters. Note that, the choice of parametrization is not unique.

The two Gamma kernel estimators denoted by Gam1 and Gam2 of [19],the IG and RIG kernel estimators of [21] and Birnbaum Saunders-power-exponential (BS-PE) kernel estimator of [29] are listed below for the comparison purpose:

$$\widehat{f}_{Gam1}(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{(X_i)^{x/b} e^{-\frac{X_i}{b}}}{b^{x/b+1}\Gamma(x/b+1)} \tag{5.5}$$

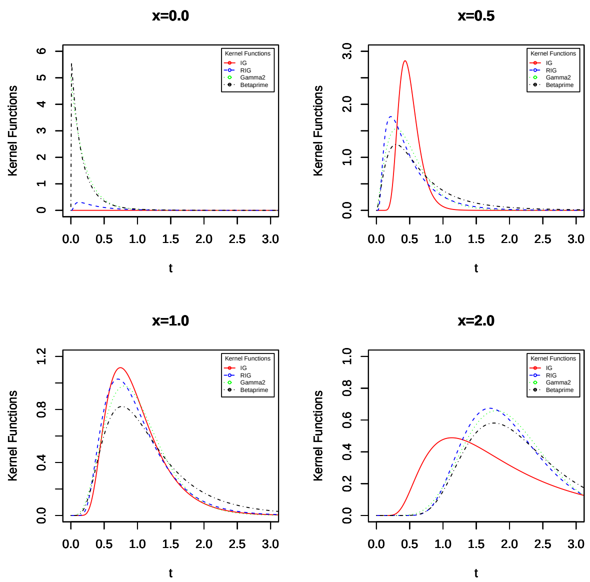

Figure 5.1 : Shapes of the kernel functions for different x values and b=0.2.

$$\widehat{f}_{Gam2}(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{(X_i)^{\rho_b(x)-1} e^{-\frac{X_i}{b}}}{b^{\rho_b(x)}\Gamma(\rho_b(x))}, \tag{5.6}$$

where

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

and

$$\widehat{f}_{IG}(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{b^{-\frac{1}{2}}}{\left(2\pi X_i^3\right)^{1/2}} \exp\left\{-\frac{1}{2bx}\left(\frac{X_i}{x} - 2 + \frac{x}{X_i}\right)\right\} \tag{5.7}$$

$$\widehat{f}_{RIG}(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{b^{-\frac{1}{2}}}{(2\pi X_i)^{1/2}} \exp\left\{-\frac{x-b}{2b}\left(\frac{X_i}{x-b} - 2 + \frac{x-b}{X_i}\right)\right\}. \tag{5.8}$$

$$\widehat{f}_{BS-PE}(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{v}{2^{1/2v}\Gamma\left(\frac{1}{2v}\right)\sqrt{4h_i}} \left(\frac{1}{\sqrt{xX_i}} + \sqrt{\frac{x}{X_i^3}}\right) \exp\left(-\frac{1}{2h_i^v}\left(\frac{X_i}{x} + \frac{x}{X_i} - 2\right)^v\right), \quad x > 0. \tag{5.9}$$

Now, lets investigate the integrated squared bias and integrated variance terms to obtain the MISE of the beta prime estimator.

Proposition 1. *The bias of the estimator is*

$$Bias\left(\widehat{f}(x)\right) = b\left[f'(x) + \frac{(1+x)}{2} f''(x)\right] + o(b), \tag{5.10}$$

where b satisfies the condition that $b \to 0$ and $nb \to \infty$ as $n \to \infty$.

Proof. To prove the bias of the estimator, note that

$$E\left\{\widehat{f}(x)\right\} = \int_0^\infty K_{B(\frac{x^2}{b}+x+1, \frac{x}{b}+\frac{1}{x+b}+1)}(y) f(y) dy = E\left\{f(\xi_x)\right\} \tag{5.11}$$

here ξ_x is the $Betaprime(\frac{x^2}{b}+x+1, \frac{x}{b}+\frac{1}{x+b}+1)$ random variable.

By standard moments properties of beta prime distribution,

$$E\left(\xi_x\right) = x + xb + O\left(b^2\right), \quad Var\left(\xi_x\right) = b(x^2+x) + O\left(b^2\right). \tag{5.12}$$

From Taylor expansion and equation (5.12), it can be obtained as

$$\begin{aligned} E\left(\widehat{f}(\xi_x)\right) &= f(x) + f^{'}(x)E\left[f\left(\xi_x\right) - x\right] + \frac{f^{''}(x)}{2}E\left[f\left(\xi_x\right) - x\right]^2 + ... \\ &= f(x) + bf^{'}(x) + b(1+x)\frac{f^{''}(x)}{2} + o(b). \end{aligned} \tag{5.13}$$

Then, the bias of the beta prime estimator is

$$Bias\left\{\widehat{f}(x)\right\} = b\left(f^{'}(x) + (1+x)\frac{f^{''}(x)}{2}\right) + o(b). \tag{5.14}$$

Since the bias is $O(b)$ near the origin as well as in the interior, the beta prime estimator is free of boundary bias. □

Proposition 2. *(Variance) The variance of beta prime estimator is*

$$Var\left\{\widehat{f}(x)\right\} = \frac{1}{2\sqrt{\pi}}b^{-\frac{1}{2}}n^{-1}(1+x)^{-\frac{1}{2}}f(x) + o(b^{-\frac{1}{2}}n^{-1}). \tag{5.15}$$

Proof.

$$\begin{aligned} Var\left\{\widehat{f}(x)\right\} &= \frac{1}{n}Var\left\{K_{B(\frac{x^2}{b}+x+1,\frac{x}{b}+\frac{1}{x+b}+1)}(X_i)\right\} \\ &= \frac{1}{n}\left[E\left(K_{B(\frac{x^2}{b}+x+1,\frac{x}{b}+\frac{1}{x+b}+1)}(X_i)\right)^2\right] + O(\frac{1}{n}) \end{aligned} \tag{5.16}$$

and by multiplying both numerator and denominator by $\beta(\frac{2x^2}{b}+2x+1, \frac{2x}{b}+\frac{2}{x+b}+2)$, it is obtained that

$$\begin{aligned} E\left(K_{B(\frac{x^2}{b}+x+1,\frac{x}{b}+\frac{1}{x+b}+1)}(X_i)\right)^2 &= \int_0^\infty \left(\frac{y^{\frac{x^2}{b}+x}(1+y)^{-(\frac{x^2}{b}+x+\frac{x}{b}+\frac{1}{x+b}+2)}}{\beta(\frac{x^2}{b}+x+1,\frac{x}{b}+\frac{1}{x+b}+1)}\right)^2 f(y)dy \\ &= A_b\int_0^\infty \frac{y^{\frac{2x^2}{b}+2x}(1+y)^{-(\frac{2x^2}{b}+2x+\frac{2x}{b}+\frac{2}{x+b}+3)}}{\beta(\frac{2x^2}{b}+2x+1,\frac{2x}{b}+\frac{2}{x+b}+2)}\frac{f(y)}{(y+1)}dy \\ &= A_bE\left[(1+\xi_x)^{-1}f(\xi_x)\right] \end{aligned} \tag{5.17}$$

where

$$A_b = \frac{\beta(\frac{2x^2}{b}+2x+1,\frac{2x}{b}+\frac{2}{x+b}+2)}{\beta^2(\frac{x^2}{b}+x+1,\frac{x}{b}+\frac{1}{x+b}+1)}, \tag{5.18}$$

β is the beta function and ξ_x is a $Betaprime\left(\frac{2x^2}{b}+2x+1, \frac{2x}{b}+\frac{2}{x+b}+2\right)$ distributed random variable.

By using Stirling formula $\Gamma(z+1)=\sqrt{2\pi}\exp(-z)z^{z+\frac{1}{2}}/R(z)$, where $R(z)$ converges to 1 as $z\to\infty$ and $R(z)<1$ for any $z>0$, an order for the equation (5.18) can be

$$A_b \approx \frac{1}{2\sqrt{\pi}} b^{-1/2}(x+1)^{\frac{1}{2}}. \tag{5.19}$$

Then, applying Taylor expansion again,

$$E\left[(1+\xi_x)^{-1} f(\xi_x)\right] = (1+x)^{-1} f(x) + o(b). \tag{5.20}$$

Therefore, variance term is

$$Var\left\{\widehat{f}(x)\right\} = \frac{1}{2\sqrt{\pi}} b^{-\frac{1}{2}} n^{-1}(1+x)^{-\frac{1}{2}} f(x) + o(b^{-\frac{1}{2}} n^{-1}). \tag{5.21}$$

□

Since MSE is the sum of squared bias and variance of the estimator, by combining equations (5.14) and (5.21), MSE is obtained as

$$MSE\left(\widehat{f}(x)\right) = b^2\left(f^{'}(x)+(1+x)\frac{f^{''}(x)}{2}\right)^2 + \frac{1}{2\sqrt{\pi}} b^{-\frac{1}{2}} n^{-1}(1+x)^{-\frac{1}{2}} f(x) + o(b^2+b^{-\frac{1}{2}}n^{-1}) \tag{5.22}$$

Then, the optimal bandwidth that minimizes the MSE is

$$b^*_{MSE} = \left(\frac{1}{8\sqrt{\pi}} \frac{(1+x)^{-\frac{1}{2}} f(x)}{\left(f'(x)+\frac{1}{2}(1+x)f''(x)\right)^2}\right)^{\frac{2}{5}} n^{-\frac{2}{5}} \tag{5.23}$$

So, the corresponding optimal MSE is

$$MSE(\widehat{f}(x))^* = \frac{5}{(8\sqrt{\pi})^{\frac{4}{5}}} \left\{(1+x)^{-\frac{1}{2}} f(x)\right\}^{\frac{4}{5}} \left\{f'(x)+\frac{1}{2}(1+x)f''(x)\right\}^{\frac{2}{5}} n^{-\frac{4}{5}} \tag{5.24}$$

Similarly, the optimal MISE based on

$$b^*_{MISE} = \left(\frac{\frac{1}{4\sqrt{\pi}}\int_0^\infty (1+x)^{-\frac{1}{2}} f(x)dx}{2\int_0^\infty \left[f^{'}(x)+\frac{(1+x)}{2}f^{''}(x)\right]^2 dx}\right)^{\frac{2}{5}} n^{-\frac{2}{5}}. \tag{5.25}$$

is

$$\begin{aligned} MISE^*\left\{\widehat{f}(x)\right\} = \frac{5}{2^{8/5}} &\left[\frac{1}{2\sqrt{\pi}}\int_0^\infty (1+x)^{-\frac{1}{2}} f(x)dx\right]^{4/5} \\ &\left[\int_0^\infty \left[(f^{'}(x)+\frac{(1+x)}{2}f^{''}(x)\right]^2 dx\right]^{1/5} n^{-4/5}. \end{aligned} \tag{5.26}$$

The optimal bandwidths (5.23) and (5.25) obtained for MSE and MISE depend on the unknown density function f. For this reason, global automatic bandwidth selection methods like LSCV, LCV are available in the literature. However, for some distribution functions this global methods give rise to unsatisfactory results. Therefore, adaptive Bayesian bandwidth selection method is proposed as an alternative way to this methods. For small sample sizes, adaptive Bayesian approach has good smoothing quality. That's why, there has been a considerable interest in this approach, recently (see [28]; [29]; [40]).

5.2 Adaptive Bayesian Bandwidth Selection Method Using Lindley's Approximation

Following the papers [28] and [40], the adaptive asymmetric kernel estimator of f is given by

$$\widehat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} K_{x,h_i}(X_i) \tag{5.27}$$

where K_{x,h_i} is the adaptive asymmetric kernel and h_i is the variable bandwidth associated with each observation x_i. Therefore, $f(x_i)$ can be estimated by the adaptive asymmetric associated kernel estimator based on all points except x_i. So, the leave one out estimator is given by the formula

$$\widehat{f}_{-i}(x_i) = \widehat{f}(x_i|\{x_{-i}\},h_i) = \frac{1}{n-1}\sum_{j=1,i\neq j}^{n} K_{x_i,h_i}(X_j) \tag{5.28}$$

where $\{x_{-i}\}$ is the set of observations excluding x_i. Let $\pi(h_i)$ be the prior distribution of h_i, then the posterior of each variable bandwidth h_i takes the form

$$\pi(h_i|x_i) = \frac{\widehat{f}(x_i|\{x_{-i}\},h_i)\pi(h_i)}{\int \widehat{f}(x_i|\{x_{-i}\},h_i)\pi(h_i)dh_i}. \tag{5.29}$$

Under squared error (SE) loss functions, the Bayes estimates of h_i is of the form

$$\widehat{h}_i = \int h_i \pi(h_i|x_i)dh_i. \tag{5.30}$$

Usually, these expression cannot be obtained in a simple closed form. Alternatively, Lindley's approximation (see [41]) method for the computation of these equations can be used. Lindley's approximation is a method to obtain Taylor series expansion of the function involved in posterior moment,

$$E\{u(h)|\underline{x}\} = \frac{\int u(h)v(h)\exp(L(h))dh}{\int v(h)\exp(L(h))dh} \tag{5.31}$$

where $u(h)$ and $v(h)$ are arbitrary functions of h and $L(h)$ is the log-likelihood function. For a single parameter case,

$$E\{h|\underline{x}\} = h+\rho\left(-\frac{1}{L_2}\right)+\frac{1}{2}\left(-\frac{1}{L_2}\right)^2 L_3+O\left(\frac{1}{n^2}\right). \tag{5.32}$$

where ρ is derivative of the logarithm of prior function with respect to h, L_2 and L_3 are second and third derivative of log-likelihood function of h, respectively.

Now, lets define the beta prime kernel estimator with variable bandwidth h_i and the approximate Bayes estimates of h_i under the SE loss function. The conditional distribution of x_i excluding the observation x_i and the beta prime kernel with variable bandwidth h_i are given respectively by

$$\widehat{f}(x_i|\{x_{-i}\},h_i) = \frac{1}{n-1}\sum_{j=1,i\neq j}^{n} K_{x_i,h_i}(X_j) \tag{5.33}$$

and

$$K_{x_i,h_i}(X_j) = \frac{(X_j)^{\frac{x_i^2}{h_i}+x_i}(1+X_j)^{-(\frac{x_i^2}{h_i}+x_i+\frac{x_i}{h_i}+\frac{1}{x_i+h_i}+2)}}{B(\frac{x_i^2}{h_i}+x_i+1,\frac{x_i}{h_i}+\frac{1}{x_i+h_i}+1)} \tag{5.34}$$

where β is Beta function. Therefore, the approximate Bayes estimates of h_i under the SE loss functions is obtained as

$$\widehat{h}_i^L = h_i+\rho\left(-\frac{1}{L_2^*}\right)+\frac{1}{2}\left(-\frac{1}{L_2^*}\right)^2 L_3^* \Big|_{h_i=\widetilde{h}_i} \tag{5.35}$$

where $\widetilde{h}_i$ is the posterior mode obtained from the equation$Q = \log(\widehat{f}_{-i}(x_i)) + \log(\pi(h_i))$ by equating $\frac{\partial Q}{\partial h_i}$ to zero. L_2^* and L_3^* are the second and third derivative of $\log(\widehat{f}_{-i}(x_i))$ with respect to h_i, respectively. Note that, the closed form of the L_2^*, L_3^* cannot be obtained in a simple form so it is not listed here.

5.3 Simulation Results

In this section, by using different underlying distribution functions, the finite sample performance of the two gamma kernels, IG and RIG kernel estimators are compared with the beta prime kernel estimator in terms of average ISE's of the estimators. Then, to show the effectiveness of the adaptive Bayesian bandwidth selection method with Lindley's approximation, the performance of the bandwidths obtained from this method and the bandwidths obtained from global LSCV method are compared through Monte Carlo simulation studies. All simulation studies are performed using R statistical software.

5.3.1 Simulation studies to compare the average ISE performance of the estimators

The two gamma kernel estimators, IG and RIG kernel estimators and beta prime kernel estimator are considered to investigate their finite sample properties in terms of the average ISE. Data sets are generated from various distributions (see Figure 5.2) described in A-E given below. Some of these distributions have been also studied in [21] and [27].

A. Beta prime density: $f(x) = \frac{x^{(\alpha-1)}}{(1+x)^{(\alpha+\lambda)}\beta(\alpha,\lambda)}$, with parameters $(\alpha,\lambda) = (2,1)$.

B. Weibull density: $f(x) = \frac{\alpha x^{(\alpha-1)}e^{-(x/\lambda)^{\alpha}}}{\lambda^{\alpha}}$, with parameters $(\alpha,\lambda) = (3,1)$.

C. Gamma density: $f(x) = \frac{x^{(\mu-1)}e^{-(x/\lambda)}}{\lambda^{\mu}\Gamma(\mu)}$, with parameters $(\mu,\lambda) = (1,3)$.

D. Mixture Gamma density: $f(x) = 0.5\frac{x^{(\mu_1-1)}e^{-\frac{x}{\lambda_1}}}{\lambda_1{}^{\mu_1}\Gamma(\mu_1)} + 0.5\frac{x^{(\mu_2-1)}e^{-\frac{x}{\lambda_2}}}{\lambda_2^{\mu_2}\Gamma(\mu_2)}$, with parameters $(\mu_1,\lambda_1) = (1,3), (\mu_2,\lambda_2) = (2,3)$.

E. Mixture Weibull density: $f(x) = 0.5\frac{\mu_1 x^{(\mu_1-1)}e^{-\left(\frac{x}{\lambda_1}\right)^{\mu_1}}}{\lambda_1^{\mu_1}} + 0.5\frac{\mu_2 x^{(\mu_2-1)}e^{-\left(\frac{x}{\lambda_2}\right)^{\mu_2}}}{\lambda_2^{\mu_2}}$, with parameters $(\mu_1,\lambda_1) = (3,1), (\mu_2,\lambda_2) = (5,3)$.

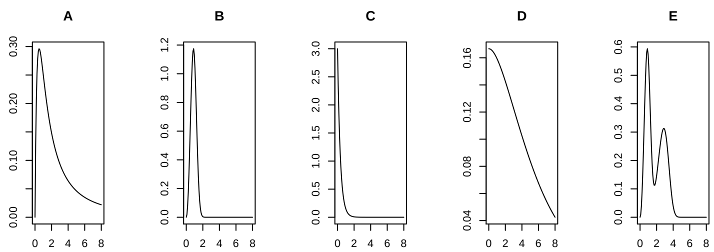

Figure 5.2 : Density plots of the cases A-E.

1500 replications of sample size n=50, 100, 200 and 500 are generated for each density defined in cases A-E. Numerical integrations are done by using Gauss Legendre quadrature with 96 knots. The performance of the estimators are compared in terms of the ISE criterion which is defined by

$$ISE = \int_0^a \left\{\widehat{f}(x) - f(x)\right\}^2 dx. \tag{5.36}$$

where a is chosen for each underlying distribution in such a way that the densities having virtually zero values outside of $[0,a]$. For each replication, the ISE of each competitive kernel estimators are calculated from a grid of bandwidth values such that the end points of the sequence of b are chosen in such a way that no minimum ISE is achieved at the end point. For example, the sequence for b starts from 0.001 with increment 0.001 up to 2 for IG, Gam1 and Gam2 kernels and it starts from 2 with increment 0.001 up to 4 for RIG and beta prime kernels for the case A in Table 5.1. Then, the minimum average ISE and corresponding smoothing parameter b for each kernel estimator are reported in Table 5.1. For the other cases similar idea is followed. The sequence for b created with this procedure covers the sequence used in [21]. From Table 5.1, one can see that, the average ISEs of all estimators decrease as the sample size n increases. Beta prime kernel estimator outperforms others except for the case D. In this case, the differences are almost negligible and it may be due to the small sample size. The IG kernel estimator is dominated by others in all cases. As explained in [21], when the shape parameter of Gamma distribution is less than 1.5, then $MISE_{RIG}$ and $MISE_{IG}$ are not well defined. That is why, in case C, RIG and IG kernel estimators do not perform well. However, for this case beta prime kernel estimator yields the smallest average ISE. Figure (5.3) illustrates the pointwise bias, variance and MSE of

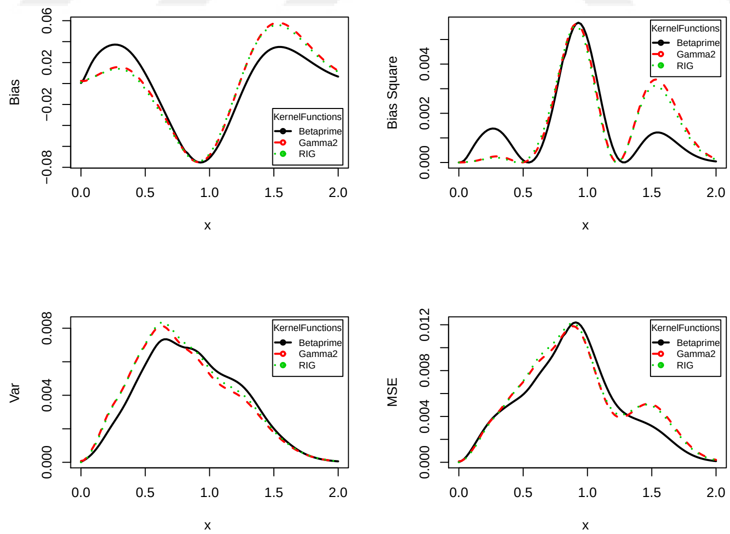

Figure 5.3 : Bias, Variance and MSE comparison of kernel estimators.

RIG, Gam2 and beta prime kernel for $x \in [0,2]$, for the Weibull(3,1) density when the sample size is 200. As stated in [19], Gam2 kernel has better global performance due to the smaller MISE so only Gam2 kernel is used in the comparison. When $x > 1.5$, the beta prime estimator has smaller MSE compared with other estimators. On the other hand, there are no clear comparison among estimates for $x < 1.5$. Also, from Figure

Table 5.1 : Average ISE of different kernel estimators.

	n	IG	RIG	Gam1	Gam2	Beta Prime
A	50	0.0098	0.0017	0.0015	0.0043	0.0009
		(0.006)	(3.6)	(1.05)	(0.32)	(3.58)
	100	0.0058	0.0011	0.0009	0.0030	0.0005
		(0.006)	(3.59)	(0.87)	(0.24)	(3.1)
	200	0.0031	0.0008	0.0005	0.0016	0.0003
		(0.005)	(3.58)	(0.69)	(0.18)	(2.67)
	500	0.0014	0.0006	0.0003	0.0008	0.0001
		(0.004)	(3.58)	(0.51)	(0.13)	(2.29)
B	50	0.0575	0.0282	0.0301	0.0278	0.0261
		(0.034)	(0.036)	(0.033)	(0.037)	(0.018)
	100	0.0352	0.0172	0.0181	0.0170	0.0159
		(0.023)	(0.026)	(0.024)	(0.027)	(0.013)
	200	0.0218	0.0107	0.0110	0.0106	0.0098
		(0.016)	(0.019)	(0.018)	(0.020)	(0.009)
	500	0.0108	0.0054	0.0057	0.0054	0.0050
		(0.011)	(0.013)	(0.012)	(0.013)	(0.006)
C	50	0.3073	0.0177	0.0057	0.0060	0.0056
		(0.167)	(0.185)	(0.481)	(0.747)	(0.594)
	100	0.2163	0.0114	0.0035	0.0037	0.0033
		(0.132)	(0.116)	(0.360)	(0.582)	(0.422)
	200	0.1257	0.0074	0.0022	0.0023	0.0020
		(0.089)	(0.064)	(0.271)	(0.439)	(0.298)
	500	0.0543	0.0040	0.0011	0.0012	0.0010
		(0.055)	(0.028)	(0.179)	(0.289)	(0.189)
D	50	0.0877	0.0087	0.0031	0.0031	0.0035
		(0.07)	(0.284)	(0.796)	(1.211)	(0.888)
	100	0.1092	0.0057	0.0018	0.0018	0.0019
		(0.094)	(0.198)	(0.642)	(1.089)	(0.705)
	200	0.0834	0.0037	0.0011	0.0012	0.0011
		(0.064)	(0.128)	(0.507)	(0.944)	(0.543)
	500	0.0343	0.0021	0.0006	0.0008	0.0005
		(0.036)	(0.069)	(0.365)	(0.648)	(0.368)
E	50	0.0319	0.0160	0.0169	0.0159	0.0158
		(0.039)	(0.048)	(0.046)	(0.050)	(0.027)
	100	0.0198	0.0095	0.0097	0.0094	0.0092
		(0.021)	(0.035)	(0.034)	(0.036)	(0.019)
	200	0.0142	0.0059	0.0062	0.0059	0.0054
		(0.014)	(0.029)	(0.027)	(0.029)	(0.016)
	500	0.0066	0.0030	0.0031	0.0030	0.0029
		(0.009)	(0.018)	(0.017)	(0.018)	(0.009)

(5.3), one can see that the variance of the estimators reduces as one moves away from the boundary as oppose to bias terms, this is considered as an advantage if estimated density has sparse areas.

5.3.2 Comparison of adaptive Bayesian analysis using Lindley's approximation with LSCV method

In this subsection, for the purpose of showing the usefulness of the Lindley's approximation for the heavy tailed distribution, a comparison is made between average ISE's of beta prime estimator obtained from adaptive Bayesian method using Lindley's approximation and average ISE's of beta prime estimator obtained from global LSCV method. In the paper [29], the Bayesian adaptive approach under quadratic loss function based on BS-PE kernel estimator is studied, then exact expression for the variable bandwidths h_i is obtained. Therefore, the average ISE's of beta prime estimator obtained from adaptive Bayesian method using Lindley's approximation and LSCV method is also compared with the average ISE's of the BS-PE kernel estimator obtained from the Bayesian adaptive approach under quadratic loss function.

As a prior density function of bandwidth h_i, beta prime distribution is chosen with parameters λ and μ:

$$\pi(h_i) = \frac{(h_i)^{\lambda-1}}{(1+h_i)^{\lambda+\mu}} \frac{1}{\beta(\lambda,\mu)}, \quad \lambda > 0, \quad \mu > 2, \tag{5.37}$$

where mean and variance of the prior are

$$E(h_i) = \frac{\lambda}{\mu-1}, \quad \mu > 1; \quad Var(h_i) = \frac{\lambda(\lambda+\mu-1)}{(\mu-1)^2(\mu-2)}, \quad \mu > 2 \tag{5.38}$$

Note that, the prior selection and its parametrization are not unique for the Lindley's approximation method. By following the idea of [42] and [29], the prior parameters are chosen as $\lambda = 1$ and $\mu = n^{4/5}$, since $E(h_i) > 0$ for $\lambda > 0$ and $\mu > 1$, $Var(h_i) > 0$ for $\lambda > 0$ and $\mu > 2$, also for large values of μ, prior of h_i is concentrated at zero. Note that in practice, $\mu = n^{4/5}$ may not be satisfactory for the smoothing quality.

For comparison purpose of findings with those of [29], the data is simulated from heavy tailed distributions which are Burr, lognormal, mixture of gamma and Levy distributions. For each density, 1500 replications of sample size 25, 50, 100 and 200 are generated and the results are given in Table 5.2.

Table 5.2 : Comparison of average ISE's of adaptive Bayesian and LSCV methods.

	n	LSCV_BPR	Lindley_BPR	Quadratic_BS-PE
A	25	0.0128	0.0106	0.0161
	50	0.0072	0.0061	0.0096
	100	0.0042	0.0036	0.0058
	200	0.0027	0.0021	0.0034
B	25	0.0598	0.0277	0.0329
	50	0.0339	0.0183	0.0214
	100	0.0140	0.0111	0.0127
	200	0.0085	0.0071	0.0079
C	25	0.0111	0.0089	0.0113
	50	0.0091	0.0054	0.0068
	100	0.0047	0.0034	0.0041
	200	0.0018	0.0015	0.0025
D	25	0.0771	0.0272	0.0328
	50	0.0594	0.0237	0.0272
	100	0.0577	0.0215	0.0231
	200	0.0517	0.0197	0.0206

A. Lognormal density:$f(x) = \frac{1}{x\lambda\sqrt{2\pi}} \exp\left(-\frac{1}{2\lambda^2}(\ln x - \mu)^2\right)$with parameters$(\mu, \lambda) = (1, 1)$.

B. Burr density:$f(x) = \frac{\mu x^{\mu-1}}{(1+\lambda x^{\mu})^{\lambda+1}}$ with parameters $(\mu, \lambda) = (3, 1)$.

C. Mixture of Gamma density:$f(x) = 0.5\frac{x^{\mu_1-1}\exp(-x)}{\Gamma(\mu_1)} + 0.5\frac{x^{\mu_2-1}\exp(-x)}{\Gamma(\mu_2)}$ with parameters $(\mu_1, \mu_2) = (2.5, 10)$.

D. Levy density:$f(x) = \sqrt{\frac{\lambda}{2\pi}} \frac{1}{(x-\mu)^{\frac{3}{2}}} \exp\left(-\frac{\lambda}{2(x-\mu)}\right)$, $x > \mu$, with parameters $(\mu, \lambda) = (0, \frac{1}{2})$.

Table 5.2 illustrates that, adaptive Bayesian method with Lindley approximation dominates BS-PE adaptive kernel estimator used in [29], despite the fact that they use exact values of bandwidth h_i for heavy tailed distribution function. The bandwidths obtained from adaptive Bayesian with Lindley approximation method yields smaller average ISE than the bandwidths obtained from global LSCV method. Moreover, Lindley approximation method is suitable for selection of different priors. On the other hand, for the light tailed distributions, Lindley approximation with beta prime prior do not perform well, but choosing different prior one can obtain good results. For example, inverse gamma prior leads a better average ISE for Gam(1,3) distribution.

5.4 Real Data Applications

Two data sets studied to illustrate the performance of the beta prime, Gam2 and RIG kernel estimators. Since IG kernel estimator is not suitable for the data sets, it is not displayed in the figures. The first data set is the daily ozone level measurements in New York, May to September, 1973 and it consists of 116 observations This data set have been studied earlier in the paper [43]. The second data set is about snowfall collected for Grand Rapids, MI, going back to 1893. In this thesis, only the data collected in December is used. This data set consists of 119 observations of the inches of snow. It is available at http://www.crh.noaa.gov/grr/climate/data/grr/snowfall/. For the second graphics, in Figure (5.4) and Figure (5.5), optimal global bandwidths are obtained by minimizing the LSCV criterion for the kernel estimators. Beta prime kernel estimator captures modes and bumps of the models and it can be considered satisfactory for this kind of data sets. Beta prime, Gam2 and RIG estimators shows similar performance for the data sets. For the last graphics in Figure (5.4) and Figure (5.5), bandwidths obtained by using adaptive Bayesian with Lindley approximation method. Prior distribution with parameters $\lambda = 1$ and $\mu = n^{1/5}$ is used for ozone data and snowfall data. In those graphics, the solid line represents the beta prime estimator with bandwidth selected by the adaptive Bayesian method using Lindley approximation and the dashed line represents the beta prime estimator with bandwidth selected by the LSCV method. It can be seen that both adaptive Bayesian and LSCV method successfully captures the bumps and unimodality of the models.

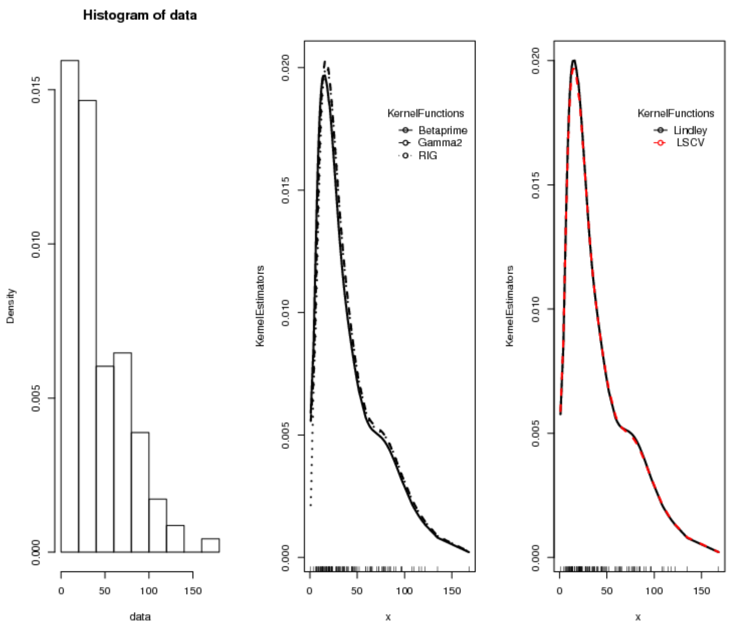

Figure 5.4 : Density estimates for ozone data.

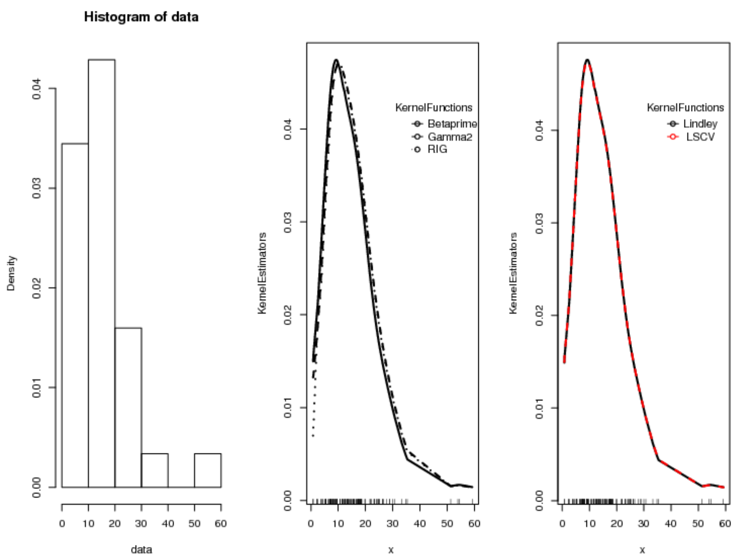

Figure 5.5 : Density estimates for snow data.

6. SCALED INVERSE CHI-SQUARED KERNEL ESTIMATOR

In this chapter, the scaled inverse chi-squared kernel estimator is proposed as an new kernel estimator for a density with support $[0,\infty)$ and therefore, the class of asymmetric kernel estimators is extended. Scaled inverse chi-squared distributions are closely related to the inverse chi-squared distribution and the inverse gamma distribution. Also, it can be used as a conjugate prior for the variance parameter of a normal distribution in Bayesian statistics. It is showed that, scaled inverse chi-squared kernel estimator is free of boundary bias, has flexible shape, always nonnegative and achieve the optimal rate of convergence for the MSE and MISE similar to the other asymmetric kernel density estimators. For the selection of bandwidths, the adaptive Bayesian bandwidth selection method with Lindley approximation is used for the proposed estimator. Then, numerical studies are conducted to compare the performance of bandwidths obtained from global LSCV method with the bandwidths obtained from adaptive Bayesian bandwidth selection method with Lindley approximation. Furthermore, real data applications illustrate that it is suitable to use this proposed estimator when the estimated density has a shoulder near zero and it captures the bumps and unimodality of the models. Note that, neither beta prime kernel estimator nor BS-PE kernel estimator is appropriate for shoulder data. Therefore, scaled inverse chi-squared kernel estimator can be used as an alternative to beta prime kernel estimator for this kind of data.

6.1 Scaled Inverse Chi-Squared Estimator

Let $X_1, X_2, ..., X_n$ be an i.i.d. random sample from a distribution with an unknown probability density function f defined on the positive real line.

The following assumptions are made for the Taylor expansion and the finiteness of integrated squared bias and integrated variance terms.

i. f is twice differentiable

ii. $\int_0^\infty x^{-\frac{1}{2}} f(x)dx < \infty$

iii. $\int_0^\infty f'(x)^2 dx < \infty$ and $\int_0^\infty (xf''(x))^2 dx < \infty$.

Scaled inverse chi- squared distribution is denoted by $SI\_\chi^{-2}(\nu,\tau^2)$ with parameters ν, τ^2. Let $K_{SI\_\chi^{-2}(\nu,\tau^2)}$ be the density of a $SI\_\chi^{-2}(\nu,\tau^2)$ distributed random variable Y, it is defined as:

$$K_{SI\_\chi^{-2}(\nu,\tau^2)}(y) = \frac{\left(\tau^2\nu/2\right)^{\nu/2}}{\Gamma(\nu/2)} \frac{\exp\left(-\frac{\nu\tau^2}{2y}\right)}{y^{1+\frac{\nu}{2}}}, \quad y > 0. \tag{6.1}$$

Then, the mean and variance of Y are

$$E(Y) = \frac{\nu\tau^2}{(\nu-2)}, \quad \nu > 2, \quad Var(Y) = \frac{2\nu^2\tau^4}{(\nu-2)^2(\nu-4)}, \quad \nu > 4. \tag{6.2}$$

So, the scaled inverse chi- squared kernel estimator considered in this thesis can be defined as

$$\widehat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} K_{SI\_\chi^{-2}\left(\frac{x}{b}+5,\frac{(x+b)(x+3b)}{(x+5b)}\right)}(X_i), \tag{6.3}$$

where b is smoothing bandwidth satisfying the condition $b \to 0$ and $nb \to \infty$ as $n \to \infty$, and $x \in [0,\infty)$ is the point where the density is estimated.

Proposition 3. *(Bias) The bias of the proposed kernel estimate is equal to*

$$Bias\left(\widehat{f}(x)\right) = b\left[f'(x) + xf''(x)\right] + o(b). \tag{6.4}$$

where b satisfies the condition that $b \to 0$ and $nb \to \infty$ as $n \to \infty$.

Proof. As in previous section $E\left(\widehat{f}(x)\right)$ can be written as

$$E\left(\widehat{f}(x)\right) = \int_0^\infty K_{SI\_\chi^{-2}\left(\frac{x}{b}+5,\frac{(x+b)(x+3b)}{(x+5b)}\right)}(y) f(y)dy = E\left(f\left(\xi_x\right)\right) \tag{6.5}$$

where ξ_x is the $SI\_\chi^{-2}(\frac{x}{b}+5, \frac{(x+b)(x+3b)}{(x+5b)})$ random variable. By Taylor expansion and standard properties of scaled inverse chi-squared distribution

$$E\left(f\left(\xi_x\right)\right) = f(x) + bf'(x) + 2bx\frac{f''(x)}{2} + o(b). \tag{6.6}$$

Then, the bias term is

$$Bias\left(\widehat{f}(x)\right) = b\left[f'(x) + xf''(x)\right] + o(b). \tag{6.7}$$

□

Proposition 4. *(Variance) The variance of the proposed kernel estimate is equal to*

$$Var\left(\widehat{f}(x)\right) = \frac{1}{2^{\frac{3}{2}}\sqrt{\pi}} b^{-\frac{1}{2}} x^{-\frac{1}{2}} n^{-1} f(x) + o(b^{-\frac{1}{2}} n^{-1}). \tag{6.8}$$

Proof. The variance term for the scaled inverse chi squared estimator can be written as

$$\begin{aligned} Var\left(\widehat{f}(x)\right) &= \frac{1}{n} Var\left(K_{SI\_\chi^{-2}(\frac{x}{b}+5, \frac{(x+b)(x+3b)}{(x+5b)})}(X_i)\right) \\ &= \frac{1}{n} E\left(K_{SI\_\chi^{-2}(\frac{x}{b}+5, \frac{(x+b)(x+3b)}{(x+5b)})}(X_i)\right)^2 + O\left(\frac{1}{n}\right) \end{aligned} \tag{6.9}$$

and

$$\begin{aligned} E\left(K_{SI\_\chi^{-2}(\frac{x}{b}+5, \frac{(x+b)(x+3b)}{(x+5b)})}(X_i)\right)^2 &= \int_0^\infty \frac{\left(\frac{(x+b)(x+3b)}{(x+5b)})\frac{\left(\frac{x}{b}+5\right)}{2}\right)^{(\frac{x}{b}+5)} \exp(-\frac{v\frac{(x+b)(x+3b)}{(x+5b)})}{y})}{\Gamma^2\left(\frac{\left(\frac{x}{b}+5\right)}{2}\right) y^{2(1+\frac{\left(\frac{x}{b}+5\right)}{2})}} f(y)dy \\ &= \frac{\Gamma\left(\frac{x+5b}{2b}\right)}{\Gamma\left(\frac{x+5b}{b}\right)^2}\left(\frac{1}{2}\right)^{\frac{x+5b}{2b}} \int_0^\infty \frac{1}{y} K_{SI\_\chi^{-2}(2(\frac{x}{b}+5), \frac{(x+b)(x+3b)}{(x+5b)})}(y) f(y)dy \\ &= B_b E(\xi_x^{-1} f(\xi_x)). \end{aligned} \tag{6.10}$$

where ξ_x is the $SI\_\chi^{-2}\left(2\left(\frac{x}{b}+5\right), \frac{(x+b)(x+3b)}{(x+5b)}\right)$ distributed random variable and

$$B_b = \frac{\Gamma\left(\frac{x+5b}{2b}\right)}{\Gamma\left(\frac{x+5b}{b}\right)^2}\left(\frac{1}{2}\right)^{\frac{x+5b}{2b}}. \tag{6.11}$$

By using Stirling formula $\Gamma(z) = \sqrt{2\pi}\exp(-z)z^{z-\frac{1}{2}}/R(z)$, where $R(z)$ converges to 1 as $z \to \infty$ and $R(z) < 1$ for any $z > 0$, it is obtained as

$$B_b \approx \frac{1}{2^{\frac{3}{2}}\sqrt{\pi}} \frac{x^{1/2}}{b^{1/2}}. \tag{2.15}$$

Then, using Taylor expansion, it gives

$$E(\xi_x^{-1} f(\xi_x)) = x^{-1} f(x) + o(b). \tag{6.12}$$

So,

$$Var\left(\widehat{f}(x)\right) = \frac{1}{2^{\frac{3}{2}}\sqrt{\pi}} b^{-\frac{1}{2}} x^{-\frac{1}{2}} n^{-1} f(x) + o(b^{-\frac{1}{2}} n^{-1}). \tag{6.13}$$

□

Then, using bias and variance terms MISE is obtained as

$$MISE\left(\widehat{f}(x)\right)=b^2\int_0^\infty\left[f'(x)+xf''(x)\right]^2dx+\frac{1}{2^{\frac{3}{2}}\sqrt{\pi}}b^{-\frac{1}{2}}n^{-1}\int_0^\infty x^{-\frac{1}{2}}f(x)dx+o\left(b^2+b^{-\frac{1}{2}}n^{-1}\right). \tag{6.14}$$

The optimal bandwidth minimizing the MISE obtained above is

$$b^*_{MISE}=\left(\frac{1}{2^{\frac{7}{2}}\sqrt{\pi}}\frac{n^{-1}\int_0^\infty x^{-\frac{1}{2}}f(x)dx}{\int_0^\infty\left[f'(x)+xf''(x)\right]^2dx}\right)^{\frac{2}{5}}. \tag{6.15}$$

Then, the corresponding optimal MISE can be can be obtained as

$$MISE\left(\widehat{f}(x)\right)=\left(\frac{5}{\pi^{\frac{2}{5}}2^{\frac{14}{5}}}\right)\left(\int_0^\infty x^{-\frac{1}{2}}f(x)dx\right)^{\frac{4}{5}}\left(\int_0^\infty\left[f'(x)+xf''(x)\right]^2dx\right)^{\frac{1}{5}}n^{-\frac{4}{5}}. \tag{6.16}$$

Also, the optimal bandwidth minimizing the MSE is

$$b^*_{MSE}=\left(\frac{1}{2^{\frac{7}{2}}\sqrt{\pi}}\frac{x^{-\frac{1}{2}}f(x)}{\left[f'(x)+xf''(x)\right]^2}\right)^{\frac{2}{5}}n^{-\frac{2}{5}}. \tag{6.17}$$

Then, by using the optimal bandwidth, one can obtain the optimal MSE as

$$MSE\left(\widehat{f}(x)\right)=\left(\frac{5}{\pi^{\frac{2}{5}}2^{\frac{14}{5}}}\right)\left(x^{-\frac{1}{2}}f(x)\right)^{\frac{4}{5}}\left[f'(x)+xf''(x)\right]^{\frac{2}{5}}n^{-\frac{4}{5}}. \tag{6.18}$$

6.1.1 Simulation study for comparison of MSE's

1500 replications of sample size $n=500$ are generated from Gamma(3,1) distribution function. The pointwise bias, variance and MSE of Gam2 kernel, BS-PE kernel and scaled inverse chi-squared kernel estimators are illustrated in Figure (6.1). From Figure (6.1), it can be seen that the variance of the kernel estimators reduces when one moves away from zero. When, $x>4$ the MSE of scaled inverse chi squared distribution is better than others. When $x<4$, there are no clear comparison among estimators.

6.2 Numerical Studies for Bayesian analysis with Lindley Approximation

In this section, the average ISE's of the adaptive Bayesian approach using Lindley approximation method with the classical LSCV bandwidth selection method are compared. Moreover, we compare the average ISE's with those of [29] Bayesian adaptive approach under quadratic loss function based on BS-PE kernel estimator and beta prime kernel estimator proposed in previous chapter with bandwidths obtained

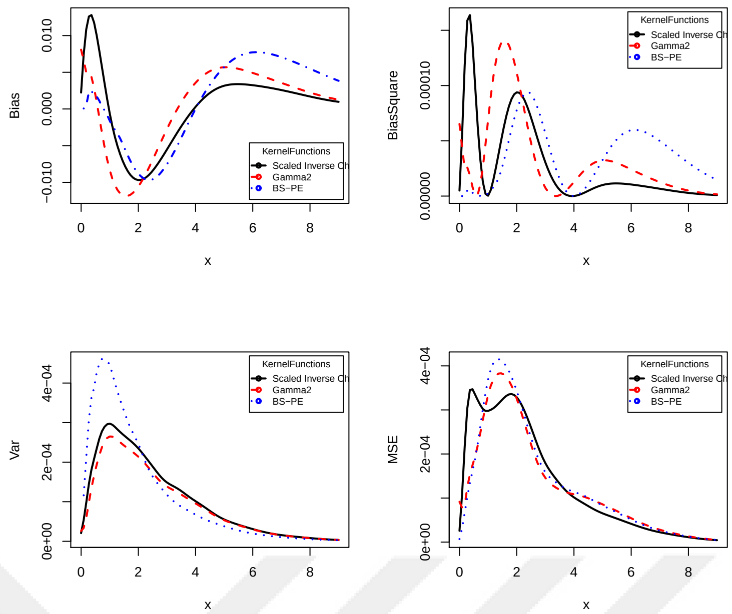

Figure 6.1 : MSE comparison of some kernels.

from Bayesian approach using Lindley approximation. The inverse gamma distribution is chosen as a prior for the variable bandwidth h_i,

$$\pi(h_i) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} (h_i)^{-\alpha-1} \exp\left(-\frac{\beta}{h_i}\right). \tag{6.19}$$

where mean and variance of the prior are

$$E(h_i) = \frac{\beta}{\alpha-1}, \quad \alpha > 1, \quad Var(h_i) = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)}, \quad \alpha > 2. \tag{6.20}$$

The prior parameters are chosen as $\alpha = 2.5$ and $\beta = 0.1$, since $E(h_i) > 0$ for $\beta > 0$ and $\alpha > 1$, $Var(h_i) > 0$ for $\beta > 0$ and $\alpha > 2$. Note that, in practice, this parameter values are not necessarily the best choice for obtaining the best smoothing quality.

As in previous chapter, for the densities described below 1000 replications of sample size 25, 50, 100 are generated and the results are given in Table 6.1.

A. Lognormal density: $f(x) = \frac{1}{x\lambda\sqrt{2\pi}} \exp\left(-\frac{1}{2\lambda^2}(\ln x-\mu)^2\right)$ with parameters $(\mu,\lambda) = (1,1)$.

B. Burr density: $f(x) = \frac{\mu x^{\mu-1}}{(1+\lambda x^{\mu})^{\lambda+1}}$ with parameters $(\mu,\lambda) = (3,1)$.

Table 6.1 : Average ISE comparisons of some kernel estimators.

	n	LSCV_SI_CS	Lindley_SI_CS	BS-PE	Beta Prime
A	25	0.0186	0.0141	0.0163	0.0106
	50	0.0093	0.0073	0.0097	0.0061
	100	0.0046	0.0040	0.0058	0.0036
B	25	0.0398	0.0326	0.0357	0.0277
	50	0.0207	0.0187	0.0214	0.0183
	100	0.0149	0.0123	0.0134	0.0111
C	25	0.0135	0.0092	0.0094	0.0089
	50	0.0108	0.0057	0.0058	0.0054
	100	0.0029	0.0027	0.0036	0.0034
D	25	0.0703	0.0648	0.0675	0.0272
	50	0.0674	0.0599	0.0600	0.0237
	100	0.0589	0.0554	0.0561	0.0215

C. Mixture of Gamma density:$f(x) = 0.5\frac{x^{\mu_1-1}\exp(-x)}{\Gamma(\mu_1)} + 0.5\frac{x^{\mu_2-1}\exp(-x)}{\Gamma(\mu_2)}$ with parameters $(\mu_1, \mu_2) = (2.5, 10)$.

D. Levy density:$f(x) = \sqrt{\frac{\lambda}{2\pi}}\frac{1}{(x-\mu)^{\frac{3}{2}}}\exp\left(-\frac{\lambda}{2(x-\mu)}\right)$, $x > \mu$, with parameters $(\mu, \lambda) = \left(0, \frac{1}{2}\right)$.

Note that, in Table 6.1 LSCV_SI_CS represents the average ISE of scaled inverse chi-squared kernel estimators with bandwidth obtained from LSCV method, Lindley_SICS represents the average ISE of scaled inverse chi- squared kernel estimators with bandwidth obtained from Lindley approximation method, BS-PE represents the average ISE of BS-PE kernel estimator with bandwidths obtained from Bayesian adaptive approach under quadratic loss function and Beta Prime represents the average ISE of beta prime kernel estimators with bandwidth obtained from Lindley approximation method.

It can be seen from Table 6.1 that the beta prime kernel estimator gives better estimates in terms of average ISE for heavy tailed distribution functions. As the sample size n increases, the average ISEs of the estimators decrease as expected. The average ISE of scaled inverse chi-squared kernel estimator by using adaptive Bayesian method with Lindley approximation is smaller than BS-PE adaptive kernel estimator used in [29]. The bandwidths obtained by using adaptive Bayesian with Lindley approximation method gives smaller average ISE than the bandwidths obtained from global LSCV method.

6.3 Real Data Applications

In this section, two data sets are studied. The first data set is daily ozone level measurements studied in the previous chapter. The second data set was collected to estimate the abundance of Southern Bluefin Tuna in Great Australian Bight (see [44]). It consists of 64 observations on the perpendicular distance (in miles) of tuna schools to transect line.

In Figure (6.2), the solid line represents the scaled inverse chi-squared estimator with bandwidth selected by the adaptive Bayesian method using Lindley approximation and the dashed line represents the scaled inverse chi-squared estimator with bandwidth selected by the LSCV method for ozone data. From Figure (6.2) and Figure (6.3), it can be said that, the scaled inverse chi-squared kernel with the bandwidths obtained from the two different method successfully captures modes and bumps of the models, so it can be considered satisfactory for those data sets. In Figure (6.3), panel (a) and (b) represent the scaled inverse chi-squared kernel estimator and beta prime kernel estimator, respectively. In those panels, bandwidths obtained from adaptive Bayesian with Lindley approximation and LSCV methods. For the selection of bandwidths by using adaptive Bayesian with Lindley approximation method, it is chosen that the prior for bandwidths follows the inverse gamma distribution with parameter values $\alpha = n^{4/5}$ and $\beta = 0.1$ for the scaled inverse chi-squared kernel estimator and beta prime prior with parameters $\alpha = 1$ and $\beta = n^{1/5}$ are employed for the beta prime kernel estimator. Panel (c) represents the BS-PE estimator proposed in paper [29] with bandwidths obtained from Bayesian adaptive approach under quadratic loss function and LSCV method. In (c), inverse gamma distribution with parameter values $\alpha = 2.5$ and $\beta = 1$ are used as a prior for the bandwidths obtained from Bayesian adaptive approach. According to practitioners the tuna data has a shoulder near the $x = 0$ (see [44]). Panel (a) showed that, the scaled inverse chi-squared kernel has good performance when f has a shoulder near zero. On the other hand, the beta prime kernel estimator and BS-PE kernel estimator are unsuitable for such a data set.

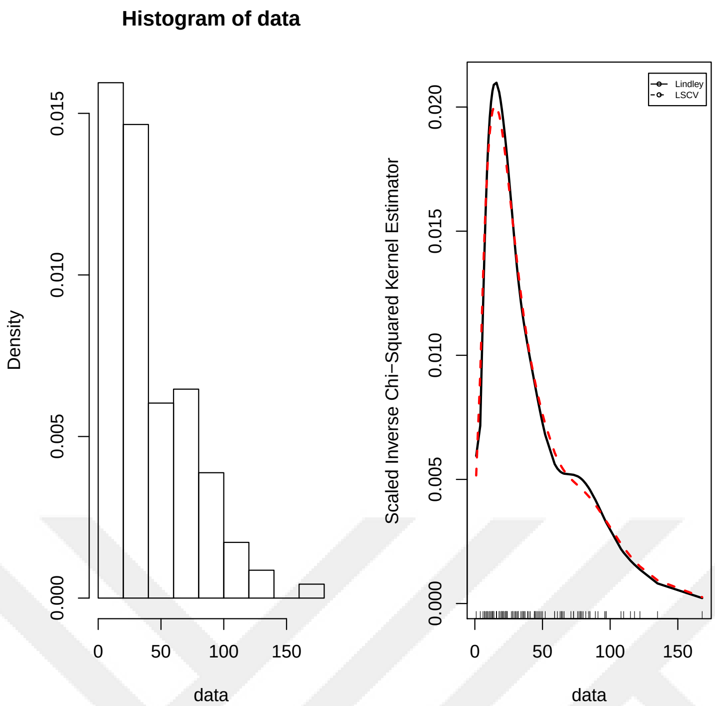

Figure 6.2 : SI-Chi Squared kernel estimator for ozone data.

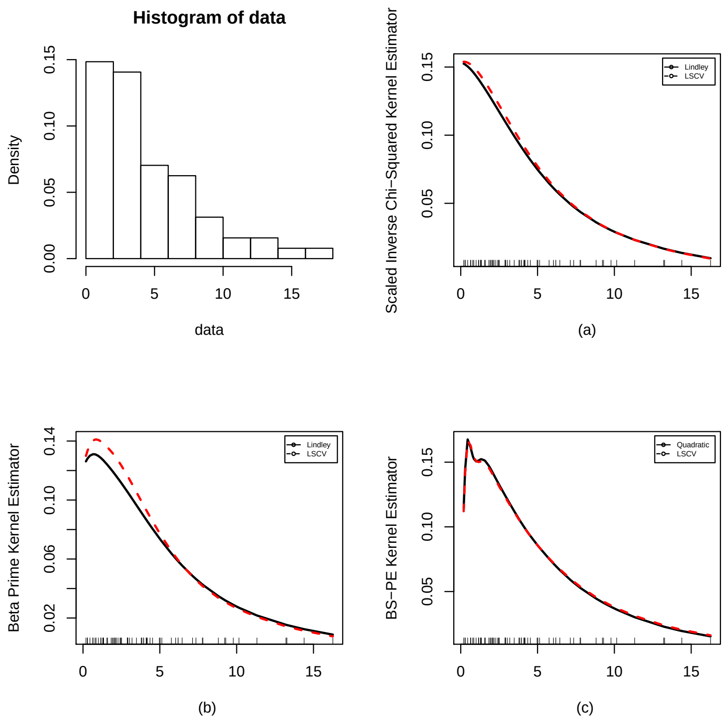

Figure 6.3 : Density estimates for tuna data.

7. CONCLUSIONS AND RECOMMENDATIONS

In this dissertation the density estimation problem is studied using the delta sequence, the orthogonal series and the asymmetric kernel methods. First, delta sequence method is considered. The convergency rate of the MSE of the estimator is obtained for the densities defined by the second order finite differences. Then, it is shown that, the convergency rate is faster than the convergency rate of MSE obtained by using the first order finite differences. Moreover, by writing the conditions on the density in terms of the second order modulus of continuity type majorants, the general assumption of second order differentiability is weakened. Secondly, the orthogonal series method namely Hermite polynomials is considered by using delta sequences. In this chapter, for the densities which have rth derivatives, the convergency rate of the MISE of estimators of densities with infinite support by using delta sequences which are based on the hermite functions is obtained. Then, convergency rate of the MSE and MISE of the estimator for the densities having compact support is obtained. The contribution of this work is improving the results of former publications about the rate of convergence of estimators based on Hermite series. Aforementioned, orthogonal series density estimate may take on negative values. So, the positivity of the Hermite series estimators may be a challenging work.

Furthermore, the asymmetric kernel method is studied and the boundary bias problem are the main interest. A new kernel estimator is proposed by using the asymmetric beta prime distribution function as kernel. It is shown that, beta prime estimator is free of boundary bias problem, has variable shape (so its support matches the support of the density to be estimated) and has the optimal rate of convergence of the MSE and MISE. Simulation studies indicate that, the beta prime kernel estimator has good finite sample properties and generally outperforms the kernel estimators proposed before. Moreover, for the heavy tailed data adaptive Bayesian bandwidth selection method is used with Lindley approximation. Lindley approximation has not been used before for the asymmetric kernel estimators. Then, comparisons are made in terms of average

ISE's obtained by using adaptive Bayesian Lindley method and obtained by using LSCV method via simulation studies. The adaptive Bayesian Lindley method exhibits better results than the LSCV method. Also, the average ISE of beta prime estimators with bandwidths obtained from adaptive Bayesian Lindley method outperforms the average ISE of BS-PE estimator despite the fact that it is obtained from the exact expression for the adaptive bandwidth. Moreover, real data applications illustrate that beta prime estimators with bandwidths obtained from both adaptive Bayesian Lindley and LSCV methods have good smoothing quality and captures the modes and bumps of the models, successfully. So, the proposed model may be useful for the air quality and hydrological data applications.

Finally, in the last chapter, scaled inverse chi-squared kernel estimator is proposed as a new asymmetric estimator. Similar to the existing estimators, the proposed kernel solves the boundary bias problem. For numerical purposes, the MSE comparison is made with some asymmetric kernel estimators. Moreover, adaptive Bayesian bandwidth selection method with Lindley approximation is used for the scaled inverse chi squared kernel estimator. Then, it is shown that, the performance of average ISE is better when using the bandwidths obtained from the adaptive Bayesian bandwidth selection method with Lindley approximation method than the bandwidths obtained from global LSCV method. Real data examples demonstrate that both adaptive Bayesian method and LSCV has good smoothing quality and capture the modes and unimodality of the models successfully. Even though the average ISE performance of the scaled inverse chi-squared estimator is not as good as the beta prime kernel estimator, Tuna data example illustrates that the scaled inverse chi-squared estimator is capable to reproduce the shoulder near zero. As a result, it can be used as an alternative to beta prime kernel estimator for this kind of data sets. The density estimation with dependent data using asymmetric scaled inverse chi-squared kernel can be studied as a further study.

REFERENCES

[1] **Rosenblatt, M.** (1956). Remarks on Some Non-parametric Estimates of a Density Functions, *Ann. Math. Statist.*, *27*(3) 832–837.

[2] **Parzen, E.** (1962). On the Estimation of Probability Density and Mode, *Ann. Math. Statist.*, *33*(3), 1065–1076.

[3] **Walter, G. and Blum, J.** (1979). Probability Density Estimation Using Delta Sequence, *Ann. Statist.*, *7*(2), 328–340.

[4] **Hall, P.** (1981). On Trigonometric Series Estimates of Densities, *Ann. Statist.*, *9*(3), 683–685.

[5] **Susarla, V. and Walter, G.** (1981). Estimation of a Multivariate Density Function Using Delta Sequences, *Ann. Statist.*, *9*(2), 347–355.

[6] **Schwartz, S.C.** (1967). Estimation of a Probability Density by an Orthogonal Series, *Ann. Math. Statist.*, *38*(4), 1261–1265.

[7] **Walter, G.** (1977). Properties of Hermite Series Estimations of Probability Densities, *Ann. Statist.*, *5*(6), 1258–1264.

[8] **Greblicki, W. and Pawlak, M.** (1984). Hermite Series Estimates of a Probability Density and Its Derivatives, *Journal of Multivariate Analysis*, *15*(2), 174–182.

[9] **Letellier, J.** (1997). Rates of Convergence for an Estimator of a Density Function Based on Jacobi Polynomials, *Communication in Statistics- Theory and Methods*, *26*(1), 197–220.

[10] **Nadar, M.** (2011). Local Convergence Rate of Mean Squared Error in Density Estimation, *Communication in Statistics- Theory and Methods*, *40*(1), 176–185.

[11] **Schuster, E.** (1985). Incorporating Support Constraints into Nonparametric Estimation of Densities, *Communication in Statistics- Theory and Methods*, *14*(5), 1123–1136.

[12] **Müller, H.G.** (1991). Smooth Optimum Kernel Estimators Near Endpoints, *Biometrika*, *78*(3), 521–530.

[13] **Müller, H.G.** (1993). On the Boundary Kernel Method for Non-Parametric Curve Estimation Near Endpoints, *Scand. J. Statist.*, *20*(4), 313–328.

[14] **Hall, P. and Wehrly, T.** (1991). A Geometrical Method for Removing Edge Effects from Kernel-type Nonparametric Regression Estimators, *J. Amer. Statist. Assoc.*, *86*(145), 665–672.

[15] **Lejeune, M. and Sarda, P.** (1992). Smooth Estimators of Distribution and Density Functions, *Comput. Statist. Data Anal.*, *14*(4), 457–471.

[16] **Marron, J. and Ruppert, D.** (1994). Transformations to Reduce Boundary Bias in Kernel Density Estimation, *J. R. Statist. Soc. Ser. B*, *56*(4), 653–671.

[17] **Cowling, A. and Hall, P.** (1996). On Pseudodata Methods for Removing Boundary Effects in Kernel Density Estimation, *J. R. Statist. Soc. Ser. B*, *58*(3), 551–563.

[18] **Chen, S.X.** (1999). Beta Kernel Estimators for Density Functions, *Comput. Statist. Data Anal.*, *31*(2), 131–145.

[19] **Chen, S.X.** (2000). Probability Density Function Estimation Using Gamma Kernels, *Ann. Inst. Statist. Math*, *52*(3), 471–480.

[20] **Jin, X. and Kawczak, J.** (2003). Birnbaum-Saunders and Lognormal Kernel Estimators for Modelling Durations in High Frequency Financial Data, *Ann. Econ. Fin.*, *4*(1), 103–124.

[21] **Scaillet, O.** (2004). Density Estimation Using Inverse and Reciprocal Inverse Gaussian Kernels, *J. Nonparametr. Statist.*, *16*(1), 217–226.

[22] **Marchant, C.**, **Bertin, K.**, **Leiva, V. and Saulo, H.** (2013). Generalized Birnbaum-Saunders Kernel Density Estimators and an Analysis of Financial Data, *Computational Statistics and Data Analysis*, *63*(1), 1–15.

[23] **Saulo, H.**, **Leiva, V.**, **Ziegelmann, F. and Marchant, C.** (2013). A Nonparametric Method for Estimating Asymmetric Densities Based on Skewed Birnbaum-Saunders Distributions Applied to Environmental Data, *Stochastic Environ. Res. Risk Assess.*, *27*(6), 1479–1491.

[24] **Igarashi, G. and Kakizawa, Y.** (2014). Re-formulation of Inverse Gaussian, Reciprocal Inverse Gaussian, and Birnbaum-Saunders Kernel Estimators, *Statistics and Probability Letters*, *84*(1), 235–246.

[25] **Mousa, A.M.**, **Hassan, M.K. and Fathi, A.** (2016). A New Nonparametric Estimator for Pdf Based on Inverse Gamma Distribution, *Communication in Statistics- Theory and Methods*, *45*(23), 7002–7010.

[26] **Kakizawa, Y. and Igarashi, G.** (2017). Inverse Gamma Kernel Density Estimation for Nonnegative Data, *Journal of the Korean Statistical Society*, *46*(2), 194–207.

[27] **Igarashi, G.** (2016). Weighted Log-Normal Kernel Density Estimation, *Communication in Statistics- Theory and Methods*, *45*(22), 6670–6687.

[28] **Brewer, M.J.** (2000). A Bayesian Model for Local Smoothing in Kernel Density Estimation, *Stat. Comput.*, *10*(4), 299–309.

[29] **Ziane, Y.**, **Adjabi, S. and Zougab, N.** (2015). Adaptive Bayesian Bandwidth Selection in Asymmetric Kernel Density Estimation for Nonnegative Heavy-Tailed Data, *Journal of Applied Statistics*, *42*(8), 1645–1658.

[30] **Belaid, N.**, **Adjabi, S.**, **Kokonendji, C. and Zougab, N.** (2018). Bayesian Adaptive Bandwidth Selector for Multivariate Discrete Kernel Estimator, *Communication in Statistics- Theory and Methods*, *47*(12), 2988–3001.

[31] **Blum, J. and Susarla, V.** (1980). Maximal Deviation Theory of Density and Faliure Rate Function Estimates Based on Censored Data, *Krishnaiah, P.R. (ed.) Mult. Analysis*, *5*(1), 213–222.

[32] **Padgett, W.J. and McNichols, D.T.** (1984). Nonparametric Density Estimation from Censored Data, *Communication in Statistics- Theory and Methods*, *13*(13), 1581–1611.

[33] **Kuruwita, C.N.**, **Kulasekera, K.B. and Padgett, W.J.** (2010). Density Estimation Using Asymmetric Kernels and Bayes Bandwidths with Censored Data, *Journal of Statistical Planning and Inference*, *140*(7), 1765–1774.

[34] **Bouezmarni, T.**, **El Gouch, A. and Mesfioui, M.** (2011). Gamma Kernel Estimators for Density and Hazard Rate of Right-Censored Data, *Journal of Probability and Statistics*, *2011*, Article ID = 937574.

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

[36] **Timan, A.** (1963). *Theory of Approximation of Functions of a Real Variable*, Oxford, England: Pergammon Press.

[37] **Samko, S. and Yakubov, A.Y.** (1985). Zygmund Estimate for Moduli of Continuity of Fractional Order of the Conjugate Function, In Russian, *Izv. Vyssh. Uchebn. Zaved. Mat.*, *29*(12), 49–53.

[38] **Stein, E. and Weiss, G.** (1971). *Singular Integrals and Differentiability Properties of Functions*, Princeton, NJ: Princeton University Press.

[39] **Szegö, G.** (1939). *Orthogonal Polynomials*, American Mathematical Society.

[40] **Zougab, N.**, **Adjabi, S. and Kokonendji, C.C.** (2014). Bayesian Estimation of Adaptive Bandwidth Matrices in Multivariate Kernel Density Estimation, *Computational Statistics and Data Analysis*, *75*(1), 28–38.

[41] **Lindley, D.V.** (1980). Approximate Bayesian Method, *Trabajos de Estadística e Investigación Operativa*, *31*(1), 223–245.

[42] **Lima, M.S. and Atuncar, G.S.** (2011). A Bayesian Method to Estimate the Optimal Bandwidth for Multivariate Kernel Estimator, *Journal of Nonparametric Statistics*, *23*(1), 137–148.

[43] **Nadarajah, S.** (2008). A Truncated Inverted Beta Distribution with Application to Air Pollution Data, *Stochastic Environmental Research and Risk Assessment*, *22*(2), 285–289.

[44] **Chen, S.X.** (1996). Empirical likelihood confidence intervals for nonparametric density estimation, *Biometrika*, *83*(2), 329–341.

CURRICULUM VITAE

Name Surname: Elif ERÇELİK

Place and Date of Birth: Istanbul, 28.09.1988

E-Mail 1: ercelikel@itu.edu.tr

E-Mail 2: ercelikeliff@icloud.com

EDUCATION:

- **B.Sc.:** 2011, Gebze Institute of Technology, Department of Mathematics.
- **M.Sc.:** 2013, Gebze Institute of Technology, Department of Mathematics.

PROFESSIONAL EXPERIENCE:

- 2013-2019 Research Assistant at Istanbul Technical University, Department of Mathematical Engineering.

PUBLICATIONS, PRESENTATIONS ON THE THESIS:

- Nadar M., **Erçelik E.**, 2017. Local convergency rate of MSE in density estimation using the second-order modulus of smoothness, *Communications in Statistics - Theory and Methods*, 46(7):3164 - 3173.
- **Erçelik E.**, Nadar M., 2018. Nonparametric density estimation based on beta prime kernel, *Communications in Statistics - Theory and Methods*, https://doi.org/10.1080/03610926.2018.1538458
- **Erçelik E.**, Nadar M., 2018. A New Kernel Estimator Based on Scaled Inverse Chi-Squared Pdf., *Communications Series A1: Mathematics and Statistics*, submitted.
- **Erçelik E.** 2019. Convergency Rate of Mean Integrated Square Error in Density Estimation Based on Hermite Polynomials, *TWMS Journal of Applied and Engineering Mathematics*, submitted.
- **Erçelik E.**, Nadar M., Rates of Convergency for an Estimator of a Density Function Based on Hermite Polynomials, *International Congress in Honour of Prof. R. P. Agarval*, 23.06.2014-26.06.2014,Uludağ University, Bursa, Turkey.
- **Erçelik E.**, Nadar M., Convergency Rate of MISE of the Density Estimator by Using Integral Modulus of Continuity Type Majorant, *9th International Statistics Congress, ISTKON*, 28.10.2015-01.11.2015, Antalya, Turkey.

- **Erçelik E.**, Nadar M., Probability Density Function Estimation Using Beta Prime Kernel, *Xth International Statistics Days Conference*, 07.10.2016-09.10.2016, Giresun, Turkey.

- **Erçelik E.**, Nadar M., Probability Density Function Estimation Using Scaled Inverse Chi-Squared Kernel, *11th International Statistics Days Conference*, 3.10.2018-7.10.2018, Muğla, Turkey.