Kernel Estimation of the Conditional Probability Density Function and Some of Its Aspects

Thesis
Submitted for the degree
of
DOCTOR OF PHILOSOPHY
in
Pure Mathematics

Mathematical Statistics
by

Hazem Ismail A. El Shekh Ahmed
Mathematics Department
University College for Women
Ain Shams University

Supervised
By

Later: Prof.Dr. Abd El Azim A. Anis
Professor of Mathematical statistics
University College for Women
Ain Shams University

Prof.Dr. Gamal A. F. Ismail
Professor of Applied Mathematics
University College for Women
Ain Shams University

Dr. Raid Salha
Department Mathematics
Faculty of Science
Islamic University of Gaza

2010

AIN SHAMS UNIVERSITY
UNIVERSITY COLLEGE FOR WOMEN
MATHEMATICS DEPARTEMENT

Ph.D. Thesis

(Pure Mathematics)

Mathematical Statistics

Title of Thesis:

Kernel Estimation of the Conditional Probability Density Function and Some of Its Aspects

Thesis Supervisors

Later: Prof.Dr. Abd El Azim A. Anis
Professor of Mathematical statistics
University College for Women
Ain Shams University

Prof.Dr. Gamal A. F. Ismail
Professor of Applied Mathematics
University College for Women
Ain Shams University

Dr. Raid Salha
Department Mathematics
Faculty of Science
Islamic University of Gaza

الإهداء

إلى روح والدي الطاهرة الذي أحب لغة القران الكريم و علمها

إلى والدتي الغالية أطال الله في عمرها

إلى زوجتي و أبنائي الأحباء

إلي إخواني و أخواتي الأعزاء

أقدم ثمرة جهدي

ACKNOWLEDGMENT

Firstly, *gratitude and thanks to ALLAH.*

I would like to express my great thanks to my great sincere supervisor Prof. Dr. Abd AlAzem Anees for his efforts in life and also after he died to complete this study, and for that I must express my sincere gratitude and appreciation to that great man. I also thank my supervisor Prof. Dr. Gamal A. F. Ismail for her effort in this study.

Also my best great thanks to my second supervisor Dr. Raid Salha, Associate Prof. Of mathematics department at Islamic university of Gaza for providing me with rich ideas throughout this study.

I don't forget Prof Dr. Samia S. Al Azab, the head of department of mathematics, Women's University College, Ain Shams University for her continuous encouragement and constructive guidance throughout this work.

In our life we may recognize persons, proud of them and difficult to forget, to M. Alice my best great gratitude for her efforts in reviewing the language of the thesis.

Also I would like to thank all my kindly professors, for their great efforts in reviewing this thesis.

Finally, I would like to thank the stuff members of mathematics department, Women's University College, Ain Shams University, Cairo Egypt.

CONTENTS

Abstract III

Summary V

Chapter 1 Preliminaries 1

1.1. Histograms 4

1.2. Kernel Estimation 10

1.2.1 Naïve Estimator 11

1.2.2 Kernel Density Estimation 13

1.3. Properties of the Kernels 15

1.4. Properties of the Kernel Density Estimators 16

1.4.1 Bias and Variance of the Kernel Density Estimation 18

1.4.2 MSE and MISE Criteria 20

1.5. Optimal Density Kernel Estimation 23

1.5.1 Optimal Bandwidth 24

1.5.2 Optimal Kernel 29

1.6. Density Derivatives Estimation 32

Chapter 2 On the Kernel Estimation of the Conditional Mode 35

2.1. Basic Concepts 38

2.2. Kernel Estimation of a Conditional PDF 40

2.3. Nadaraya-Watson (NW) Estimator 42

2.4. Re-weighed Nadaraya-Watson (RNW) Estimator 45

2.5. Analysis of the Kernel Estimation of the Conditional Mode 50

2.5.1 Conditions 53

2.5.2 Main Results 54

2.5.3 Applications 61

Chapter 3 Local Variable Kernel Estimation of the Conditional Mode 65

3.1.	The Local Variable Kernel Method	**67**
3.2.	Presentation and Proof of the Modified Local Kernel Estimation	**70**
3.3.	Applications	**82**
	3.3.1. Real Life Data	**83**
	3.3.2. Simulation Studies	**84**

Chapter 4 A Comparison of Reweighted Nadaraya-Watson and Local Kernel Mode Estimators **87**

4.1.	Basic Properties and Formulas	**88**
4.2.	Theoretical Comparison	**91**
4.3.	Practical Comparison	**92**
	4.3.1 Real Life Data	**92**
	4.3.2 Simulation Study	**93**

Chapter 5 Conclusion and Future Work **95**

REFERENCES **98**

Published Papers **105**

ARABIC SUMMARY

Abstract

Statistical problems involving the nonparametric kernel estimation of density and conditional functions are widely encountered.

This dissertation concerns itself with nonparametric models that deal with mode as a one of central tendency measures.

We consider the problem of estimating the conditional mode kernel estimation under regularity conditions, so as to obtain modifications for some formal estimators of the conditional mode. We improve such estimators by specifying some optimal probability weighted coefficients $\tau_i(x)$ and using them in maximizing the Re-Weighted Nadaraya-Watson (RNW) kernel estimator of the conditional density function.

Additionally we improve the estimator of the mode of a conditional probability density function, by using some called local variable kernel density estimation (lvkmes), $\hat{f}(y|x)$, which depends on different bandwidths $h(x)$, for each point x, where $f(y|x)$ is estimated. The two obtained conditional mode estimators are tested as a strongly consistent and asymptotically normally distributed. Moreover we test the performance of the Reweighted Nadaraya-Watson conditional mode estimator by using the correlation

coefficient between the predicated and actual values which gives a strong result at 0.9089. We go on to evaluate the mean square errors (MSE) for different conditional mode kernel estimators to get the smallest value among these errors, which was 0.0199 for the local variable kernel mode density estimator.

We conclude by testing the goodness of performance of the proposed new estimators by using both real life data and two simulation studies. In addition, the two conditional mode estimators, Reweighted Nadaraya-Watson, $\hat{M}_{RNW}(x)$ and local variable kernel, $\hat{M}_{L}(x)$ are compared theoretically and practically. The comparison indicates that $\hat{M}_{L}(x)$ is better than $\hat{M}_{RNW}(x)$, in relation to the strong effect of the bandwidth h(x) as a smooth parameter in the local variable conditional mode kernel estimator.

Key words: Kernel estimation, conditional distribution, conditional mode, Reweighted Nadaraya Watson estimator, large sample theory, Variant bandwidths, identically distributed, constant kernel density estimation (ckmes), bounded variation, asymptotic properties, strongly consistent.

SUMMARY

Summary

This thesis focuses on presenting the kernel estimation of the conditional probability density function and some of its aspects.

We try to propose a modified estimator of the conditional mode in two ways, one by maximizing the Reweighted Nadaraya-Watson (RNW) kernel estimator, and the other by using the local variable kernel density estimation (ivkmes).

Finally we make a comparison between the two modified estimators to indicate the better one theoretically and practically.

The thesis divides into four chapters with contents as the following:

Chapter 1 gives a brief introduction to the thesis topic throughout six sections, we begin with the histogram as a banner of the kernel estimator, and focus through the two basic effected parameters, the bin width, and the shifted of the bins of the data in the histogram. Then we present the second step towards the kernel estimation using Naïve estimator, Naïve estimator seemed to be as a refinement estimator of the histogram. The basic properties of the kernel function $K(u)$ were considered, and later we present the kernel estimator of the probability density function with its error calculations.

In Chapter 2, more work will be done to obtain modifications for some known estimator of the conditional mode. Five sections contains the ideas of that approach using some called Nadaraya-

Watson estimation, then we derive the kernel conditional mode density estimation deals with Nadaraya-Watson and Reweighted Nadaraya-Watson estimators in order to obtain a refine formula of such estimator.

The asymptotic normality and consistency properties of the proposed estimator are established and later its efficiency was examined by two applications of simulated and real life data.

We present, in Chapter 3 another idea to improve the results of conditional mode kernel density estimation throughout three sections, using local variable kernel estimation of the conditional mode. We introduce a general description of the method and derive the formula of such conditional mode estimation in order to get good results. This formula was tested in the end of the chapter using two applications of simulated and real life data.

Chapter 4, consists of three sections; and describes the basic properties and formulas of the investigated estimators with literature reviews. We give a theoretical comparison of the mean squared errors (MSE's) of the two estimators, and we briefly indicate the results of the comparisons of practical performance based on data applications and simulations. And finally we discuss these results through the conclusion summary.

Finally, concluding remarks and directions for future research are outlined in chapter 5

Chapter 1: Preliminaries

Preliminaries

Kernel estimation refers to a general class of techniques for non–parametric estimation of functions. In comparison to parametric estimators where the estimator has a fixed functional form (structure) and the parameters of this function are the only information we need to store, non-parametric estimators have no fixed structure and depend upon all the data points to reach an estimate. Nonparametric methods are statistical techniques that do not require a researcher to specify functional forms for objects being estimated. The methods we survey are known as kernel methods. Such methods are becoming increasingly popular for applied data analysis; they are often used- in situations involving large data sets for which the number of variables involved is also large. Such estimators have no fixed structure and can utilize all the data points to reach an estimate.

Nonparametric methods have attracted a great deal of attention from statisticians in the past few decades, as evidenced by articles, researches and texts written by statisticians including Prakasa Rao (1983), Devroye et al. (1996), Silverman (1986), Scott (1992), Bickel et al. (1993), Wand and Jones (1995), Fan and Gijbels (1992), Simonoff (1996), Azzaline and Bowman (1997), Hart (1997), Efromovich (1999), Eubank (1999), Ruppert et al. (2003), Hardle et al. (2004), and Fan and Yao (2005). The first published paper in kernel estimation was written by Rosenblatt (1956), he was interested in some aspects of the estimation of the density function of a univeriate probability distribution.

There are certain standard notations it would be convenient to mention in the presentation to be undertaken here. If, we have a random sample $X_1, X_2, ..., X_n$ from a continuous univeriate distribution with a probability density function (pdf), that function – which we will be try to estimate, will be denoted here by f. The probability density function may be considered to be one of the most important concepts in statistics. Specifying the probability density function f gives a natural description of the distribution of X, and allows probabilities associated with X to be found from the relation:

$$P(a < x < b) = \int_a^b f(x)\, dx \qquad \text{for all } a < b$$

The symbol $\hat{f}$ will be used to denote the estimator of the probability density function f.

In addition and without lost of generality we will integrate the functions over the real line R, using the symbol $\int f(.)$.

The present chapter will classified into six sections as follows:

In Section 1.1, some basic concepts are introduced about the histogram, which is considered to be one of the oldest and most popular density estimator. The smooth of the data will be focused through the two basic parameters, the bin width, and the shift of the bins of the data in the histogram.

In Section 1.2, the Naïve estimator- a special case of the usual kernel estimator- is introduced as refinement formula of the histogram, This section will recognize and discuss the probability and wide use of this method of estimating.

In Section 1.3, we turn to present the basic properties of the kernel estimation K as a symmetric and continuous function with zero values at both extremes.

In Section 1.4, the bias and variance of the kernel estimator is derived for identical independent distributed random variables (iid). Different types of errors of what will be presented with some related calculations.

In Section 1.5, the optimal case will be discussed for both the kernel function and bandwidth of the kernel estimator.

Finally in Section 1.6, the kernel density derivative estimation is presented.

Understanding of these concepts allows us to proceed, in the chapters that follow, to see their role in certain aspects of mode estimation.

1.1 Histograms

In this section, we present the first step towards the kernel estimations and its aspects, and cover some important definitions in this subject. Our topic is the histogram, one of the basic concepts in the kernel estimation.

The histogram is the oldest and most popular tool for graphical display of a univeriate set of data. It is the simplest non-parametric