On the Kernel Estimation of the Conditional Median

حول تقدير النواة للوسيط الشرطي

Ammar A. Al Attal

Supervised by

Associate Prof. Raid B. Salha
Prof. of Math

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Mathematics

April/2016
On the Kernel Estimation of the Conditional Median

I understand the nature of plagiarism, and I am aware of the University’s policy on this.

The work provided in this thesis, unless otherwise referenced, is the researcher's own work, and has not been submitted by others elsewhere for any other degree or qualification.

Student's name:  عمر عامر العتال
Signature:  
Date:  23/4/2016
نتيجة الحكم على أطروحة ماجستير

بناءً على موافقة شئون البحث العلمي والدراسات العليا بالجامعة الإسلامية بغزة، على تشكيل لجنة الحكم على أطروحة الباحث/عمر سالم العتال لنيل درجة الماجستير في كلية العلوم، قسم الرياضيات، موضوعها:

(حول تقدير النواة للوسطي الشرطي)

(On the Kernel Estimation of the Conditional Median)

وبعد المناقشة العلمية التي تمت اليوم السبت 16 رجب 1437هـ الموافق 23/04/2016م، الساعة الثانية عشرة ظهراً بمبنى القدس، اجتمعت لجنة الحكم على الأطروحة والموافقة على:

د. رائد بشير صالحة

د. محمد إسماعيل نشوان

النائبين

وقد وافق على منح الباحث/عمر سالم العتال درجة الماجستير في كلية العلوم، قسم الرياضيات، وبهذا يعلن إنه يحظى بالترقية إلى درجة الدكتوراه في الرياضيات.

و_author

نائب الرئيس لشؤون البحث العلمي والدراسات العليا

أ.د. عبدالرؤوف علي المناعمة
الملخص

دالة التوزيع الاحتمالي الشرطي تلعب دوراً مهماً في الإحصاء، فهي تصف العلاقة بين متغيرين، المتغير التابع $Y$، والمتغير المستقل $X$. في هذه الرسالة درسنا تقدير النواة لدالة التوزيع الشرطية عندما تكون مجهولة و درسنا بالإضافة إلى ذلك تقدير النواة للوسيط المشروط.

في هذه الدراسة، لقد تم دراسة دالة التوزيع الشرطي باستخدام مقدرين هما ندارايا و المقدر ثنائي النواة $NW$ و المقدر ثنائي النواة $DK$، وقد تم المقارنة بينهما حسب خواص المقدرات، حيث تم دراسة خواص التقارب كالاتساق، وعدم التحيز والتوزيع الطبيعي لتقدير دالة التوزيع الاحتمالي الشرطي والوسيط الشرطي، وأخيراً قارنا بين المقدرين باستخدام ثلاث مجموعات من بيانات المحاكاة ومجموعة من بيانات حقيقية.

نتائج المقارنة أثبت أن المقدر ثنائي النواة كان أفضل من مقدر ندارايا واتسن.
To my family...
# Table of Contents

ACKNOWLEDGEMENTS iv

ABSTRACT v

List of Abbreviations vi

List of Symbols vii

INTRODUCTION 1

1 Basics and Concepts 2
   1.1 Preliminaries 2
   1.2 The Kernel estimator 9
   1.3 Properties of the Kernels 11
   1.4 Properties of the Kernels Estimator 13
      1.4.1 Bias and Variance of the kernel density estimation 13
      1.4.2 The MSE and MISE Criteria 16
   1.5 Asymptotic normality of the kernel density estimator 20

2 Nadaraya-Watson Estimator 23
   2.1 Estimating the Conditional Probability distribution function 24
   2.2 The Nadaraya-Watson Estimator 26
   2.3 Asymptotic Properties 29
Acknowledgments

First of all, I am awarding my great thanks for the Almighty Allah who all the time helps me and grants me the power and courage to finish this study and give me the success in my live.

I would like to introduce my sincere thanks to my supervisor Dr. Raid Salha for all the interesting discussions I had with him.

I am grateful to the Islamic University in Gaza for offering me the opportunity to get the Master Degree of Mathematics, and my thanks to all Professor who teaching me in mathematics department. I would like to express my deep thanks and appreciation to my family, especially parents for their encouragement.

Also, I wish to thank my colleagues and my friends who provided suggestion in the study. Finally, I pray to allah to accept this work.
ABSTRACT

The conditional probability distribution function (cdf) plays a basic role in statistics. It describes the relationship between the dependent variable $Y$ and the independent variable $X$. Sometimes the cdf is unknown, which makes the estimation of cdf necessary. In this context, the cdf function of quantiles is very important, especially the conditional median distribution function, when the conditional mean is not actually describing the relationship between the dependent variable $Y$ and the independent variable $X$.

In this thesis, we estimate the conditional median distribution function using two types of kernel estimation, the Nadaraya-Watson kernel estimator (NW) and the Double Kernel estimator (DK). For both estimators, we investigate the main asymptotic properties, consistency, unbiasedness, and normality.

Finally, a comparison of two estimators using three simulated data and one real data, indicates that, the Double Kernel estimator is better than the Nadaraya-Watson estimator.
# List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMISE</td>
<td>Asymptotic MISE.</td>
</tr>
<tr>
<td>a.s.</td>
<td>Almost surely.</td>
</tr>
<tr>
<td>Bias</td>
<td>bias.</td>
</tr>
<tr>
<td>P</td>
<td>Probability.</td>
</tr>
<tr>
<td>cdf</td>
<td>Cumulative distribution function.</td>
</tr>
<tr>
<td>Cov</td>
<td>Covariance.</td>
</tr>
<tr>
<td>CLT</td>
<td>Central Limit Theorem.</td>
</tr>
<tr>
<td>i.i.d.</td>
<td>Independent and identically distributed.</td>
</tr>
<tr>
<td>ISE</td>
<td>Integrated Squared Error</td>
</tr>
<tr>
<td>MSE</td>
<td>Mean Square Error.</td>
</tr>
<tr>
<td>MISE</td>
<td>Mean integral square error.</td>
</tr>
<tr>
<td>NW</td>
<td>Nadaraya-Watson.</td>
</tr>
<tr>
<td>KDE</td>
<td>Kernel density estimation.</td>
</tr>
<tr>
<td>KCDF</td>
<td>Kernel conditional density function.</td>
</tr>
<tr>
<td>pdf</td>
<td>Probability density function.</td>
</tr>
<tr>
<td>Var</td>
<td>Variance.</td>
</tr>
<tr>
<td>w.r.t.</td>
<td>With respect to</td>
</tr>
</tbody>
</table>
# List of Symbols

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E(Y \mid X = x)$</td>
<td>Regression mean $Y$ given $X = x$ (realization of $E(Y \mid X)$)</td>
</tr>
<tr>
<td>$\sigma^2(x)$</td>
<td>Conditional variance of $Y$ given $X = x$ (realization of $Var(Y \mid X)$)</td>
</tr>
<tr>
<td>$U[0, 1]$</td>
<td>Uniform distribution on $[0, 1]$</td>
</tr>
<tr>
<td>$N(0, 1)$</td>
<td>Standard normal or Gaussian distribution</td>
</tr>
<tr>
<td>$N(\mu, \sigma^2)$</td>
<td>Normal distribution with mean $\mu$ and variance $\sigma^2$</td>
</tr>
<tr>
<td>$\mathbb{R}$</td>
<td>real numbers</td>
</tr>
<tr>
<td>$f(x)$</td>
<td>Density function</td>
</tr>
<tr>
<td>$\hat{f}(x)$</td>
<td>Estimated density function</td>
</tr>
<tr>
<td>$K$</td>
<td>Kernel Function (univariate)</td>
</tr>
<tr>
<td>$K_h$</td>
<td>Scaled Kernel Function, i.e. $K_h(x) = K(x/h)/h$</td>
</tr>
<tr>
<td>$\mu_2(K)$</td>
<td>Second Moment of $K$, i.e. $\int_{-\infty}^{\infty} z^2 K(z) dz$</td>
</tr>
<tr>
<td>$I$</td>
<td>Indicator function</td>
</tr>
<tr>
<td>$X$</td>
<td>Univariate random variable, $x \in \mathbb{R}$</td>
</tr>
<tr>
<td>$h$</td>
<td>The bandwidth</td>
</tr>
<tr>
<td>$h_{opt}$</td>
<td>The optimal bandwidth</td>
</tr>
<tr>
<td>$</td>
<td>\cdot</td>
</tr>
<tr>
<td>$o(.)$</td>
<td>Small oh</td>
</tr>
<tr>
<td>$O(.)$</td>
<td>Big oh</td>
</tr>
<tr>
<td>$\nabla$</td>
<td>Gradient</td>
</tr>
<tr>
<td>Symbol</td>
<td>Description</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>$m(y</td>
<td>x)$</td>
</tr>
<tr>
<td>$f(y</td>
<td>x)$</td>
</tr>
<tr>
<td>$F(y</td>
<td>x)$</td>
</tr>
<tr>
<td>$\hat{m}_{NW}(y</td>
<td>x)$</td>
</tr>
<tr>
<td>$\hat{m}_{DK}(y</td>
<td>x)$</td>
</tr>
<tr>
<td>$\hat{F}_{NW}(y</td>
<td>x)$</td>
</tr>
<tr>
<td>$\hat{F}_{DK}(y</td>
<td>x)$</td>
</tr>
<tr>
<td>$\tau_i(x)$</td>
<td>Probability weight coefficients</td>
</tr>
<tr>
<td>$R^2_{y,\hat{y}}$</td>
<td>The correlation coefficients</td>
</tr>
<tr>
<td>$X_n \xrightarrow{P} X$</td>
<td>Convergence in Probability</td>
</tr>
<tr>
<td>$X_n \xrightarrow{a.s.} X$</td>
<td>Almost sure convergence</td>
</tr>
<tr>
<td>$X_n \xrightarrow{D} X$</td>
<td>Convergence in Distribution</td>
</tr>
<tr>
<td>$X_n \xrightarrow{M.S} X$</td>
<td>Convergence in mean square</td>
</tr>
<tr>
<td>$O_p(\bullet)$</td>
<td>$U = O_p(V)$ iff for all $\epsilon &gt; 0$ exists $c &gt; 0$ such that $P(U \mid V &gt; c) &lt; \epsilon$</td>
</tr>
</tbody>
</table>
List of Figures

1.1 Kernel density estimation based on 7 points (Fan and Yao (2003)). . . 11
4.1 Graph of the NW and the DK estimators for simulation 1 . . . . . . . 57
4.2 Graph of the NW and the DK estimators for simulation 2 . . . . . . . 59
4.3 Graph of the NW and the DK estimators for simulation 3 . . . . . . . 61
4.4 Graph of the NW and the DK estimators for Realdata1 . . . . . . . 62
4.5 Graph of the NW and the DK estimators for Realdata2 . . . . . . . 62
List of Tables

1.1 Some classical kernel functions ..................................................... 12
4.1 The Comparison Between The NW and The RNW estimators ........ 54
4.2 The MSE for the NW and the DK estimators (Simulation 1) .... 56
4.3 The $R^2_{y,y}$ for the NW and the DK estimators (Simulation1) ... 56
4.4 The MSE for the NW and the DK estimators (Simulation 2) ...... 58
4.5 The $R^2_{y,y}$ for the NW and the DK estimators (Simulation2) ... 58
4.6 The MSE for the NW and the DK estimators (Simulation 3) ...... 60
4.7 The $R^2_{y,y}$ for the NW and the DK estimators (Simulation3) ... 60
INTRODUCTION

The philosophy os statistic inference is to model the observed data of various phenomena, it is helpfull to know the pdf of these data, which is not available in many times, which makes the estimation methods of the pdf is necessary. It is worth mentioning that, there are two methods of estimationm the parametric and non-parametric methods, the first method is suitable when the data distribution is know or some all parameter of the pdf is available, the MLE and MM are popular techniques to estimate the parameter of the pdf. On the other hand if the data distribution is unknown, the non-parametric methods is applicable. Hence the estimation process is very important in many aspects of real life, specially in the econometric field studing the relationship between the dep. R.V. and indep. R.V. which make the conditional distribution of the median or quantile is very important, specially when the conditional mean is un available or is not actually significantly describe this relationship. The aim of this study is to estimate the cdf of the conditional median, NW and DK methods are used to estimate the conditional distribution function and the conditional median. Moreover we hold comparison between the two estimators.

Chapter one is devoted to introduce some concepts and definitions that are needed in the next chapter, it introduce also the concepts of the kernel estimator, and its properties, biasness, MSE and MISE.

In the second chapter, we introduce the NW estimator of the conditional distribution function, in order to estimate the conditional median $m_{NW}(x)$. then study the asymptotic behavior of the $\hat{m}_{NW}(x)$, the NW estimator of the conditional median $m(x)$. The asymptotic normality and consistently properties of the $\hat{m}_{NW}(x)$ are investigated . In Chapter 3, the DK estimator of the conditional distribution function, the $\hat{m}_{DK}(x)$, the DK estimator of the conditional median $m(x)$ are introduced and the asymptotic consistency and normality of the $\hat{m}_{DK}(x)$ are investigated.

The fourth chapter contains a comparison between the NW and the DK estimators using three simulated data, and an application using real data.
Chapter 1

Basics and Concepts

Introduction

This chapter contains some basic definitions and facts that we will need in this thesis. We will introduce the kernel estimation notion of the probability density function (pdf), also we introduce the main properties the kernel, the bias, variance, mean squared error (MSE) and the mean integral squared error (MISE).

1.1 Preliminaries

In this section, we will introduce some basic definitions and theorems that will be helpful in the remanning of this thesis. The basics definitions and theorems in this section were taken from Hogg, Mckean, and Craig (2005), Silverman (1986) and Freund. J. (1992).

Definition 1.1.1. (Random Variable)

Consider a random experiment with a sample space $S$. A function $X$, which assigns to each element $c \in S$ one and only one number $X(c) = x$, is called a random variable. The space or range of $X$ is the set of real numbers $D$ such that
\( D = \{x : x = X(c); \ c \in S\}, \)

\( D \) will generally be a countable set or an interval of real numbers.

**Definition 1.1.2.** If \( D \) is a countable set, then \( X \) is called a discrete random variable. If \( D \) is an interval, then \( X \) is called a continuous random variable.

**Definition 1.1.3.** (Probability Distribution)

If \( X \) is a discrete random variable, then the function given by \( f(x) = P(X = x) \) for each \( x \) within the range of \( X \) is called the **probability distribution** of \( X \).

**Definition 1.1.4.** (Probability Density Function)

A non-negative function with values \( f(x) \), defined over the set of all real numbers, is called a **probability density function** of the continuous random variable \( X \) if and only if

\[
P(a \leq X \leq b) = \int_a^b f(x)dx,
\]

for any real constants \( a \) and \( b \) with \( a \leq b \).

**Definition 1.1.5.** (Cumulative Distribution Function)

If \( X \) is a continuous random variable and the value of its probability density at \( t \) is \( f(t) \), then the function given by

\[
F(x) = P(X \leq x) = \int_{-\infty}^x f(t)dt \quad \text{for} \quad -\infty < x < \infty.
\]

For each \( x \), \( F(x) \) is the area under the density curve to the left of \( x \) is called the **distribution function**, or the **cumulative distribution**, of \( X \).

**Definition 1.1.6.** The **support** of a continuous random variable \( X \) consists of all points \( x \) such that \( f_X(x) > 0 \).

**Definition 1.1.7.** (Independence)

Let the random variables \( X_1 \) and \( X_2 \) have the joint pdf \( f_{X_1,X_2}(x_1,x_2) \) and the marginal pdfs \( f_{X_1}(x_1) \) and \( f_{X_2}(x_2) \) respectively. The random variables \( X_1 \) and \( X_2 \) are said to be **independent** if, and only if, \( f_{X_1,X_2}(x_1,x_2) \equiv f_{X_1}(x_1)f_{X_2}(x_2) \).

Random variables that are not independent are said to be **dependent**.
Definition 1.1.8. (a Lipschitz function)
Let $A \subseteq \mathbb{R}$, $f : A \rightarrow \mathbb{R}$. If there exist a constant $K > 0$, such that
\[|f(x) - f(u)| \leq K|x - u|\]
for all $x, u \in A$, then $f$ is said to be a **Lipschitz function** on $A$.

Definition 1.1.9. (Uniformly continuous)
Let $A \subseteq \mathbb{R}$, $f : A \rightarrow \mathbb{R}$, we say that $f$ is **Uniformly continuous** on $A$ if for each $\epsilon > 0$ there is a $\delta(\epsilon) > 0$ such that if $x, u \in A$ are a number satisfying $|x - u| \leq \delta(\epsilon)$, then $|f(x) - f(u)| \leq \epsilon$.

Definition 1.1.10. (Bounded function)
A function $f : A \rightarrow \mathbb{R}$ is said to be **bounded on $A$**, if there exists a constant $M > 0$, such that $|f(x)| \leq M$ for all $x \in A$.

Definition 1.1.11. If $A$ is any set, we define the Indicator function $I_A$ of the set $A$ to be the function given by
\[ I_A = \begin{cases} 1 & : x \in A \\ 0 & : x \notin A \end{cases} \]

Definition 1.1.12. (Converge in Probability) Let $\{X_n\}$ be a sequence of random variables and let $X$ be a random variable defined on a sample space. We say $X_n$ **converges in probability** to $X$ if for all $\epsilon > 0$, we have
\[ \lim_{n \rightarrow \infty} P(|X_n - X| \geq \epsilon) = 0. \]
or equivalently
\[ \lim_{n \rightarrow \infty} P(|X_n - X| < \epsilon) = 1. \]
If so, we write $X_n \xrightarrow{P} X$.

Definition 1.1.13. (Converge in Distribution)
Let $\{X_n\}$ be a sequence of random variables and let $X$ be a random variable. Let $F_{X_n}$ and $F_X$ be, respectively, the cdfs of the $X_n$ and $X$. Let $C(F_X)$ denote the set of all points where $F_X$ is continuous. We say that $X_n$ **converges in distribution** to $X$ if
\[
\lim_{n \to \infty} F_{X_n}(x) = F_X(x), \text{ for all } x \in C(F_X).
\]

We denote this convergence by \( X_n \xrightarrow{D} X \).

**Definition 1.1.14. (Converge with Probability 1)**

Let \( \{X_n\} \) be a sequence of random variables and let \( X \) be a random variable defined on a sample space. We say that \( X_n \) **converge almost surly** to \( X \) \( (X_n \xrightarrow{a.s.} X) \) or **converge with probability 1** to \( X \) if and only if

\[
P(\{w : X_n(w) \to X(w) \text{ as } n \to \infty\}) = 1.
\]

or equivalently, for all \( \epsilon > 0 \), there exist \( N \in \mathbb{N} \)

\[
P(|X_n - X| < \epsilon, \ n \geq N) = 1.
\]

**Definition 1.1.15.**

1. A function \( f \) is of order less than \( g \) as \( x \to \infty \) if

\[
\lim_{x \to \infty} \frac{f(x)}{g(x)} = 0
\]

and we indicate this by writing \( f = o(g) \) ("\( f \) is little oh \( g \)).

2. Let \( f(x) \) and \( g(x) \) be a positive for \( x \) sufficiently large. Then \( f \) is of at the most order of \( g \) as \( x \to \infty \) if there is a positive integer \( M \) for which

\[
\frac{f(x)}{g(x)} \leq M
\]

for \( x \) sufficiently large. We indicate this by writing \( f = O(g) \) ("\( f \) is big oh of \( g \)).

**Definition 1.1.16.** Given two sequences \( \{a_n\} \) and \( \{b_n\} \) such that \( b_n \geq 0 \) for all \( n \).

We write

\[
a_n = O(b_n) \text{ if there exists a constant } M > 0 \text{ such that } |a_n| \leq Mb_n \text{ for all } n.
\]

We write
\[ a_n = O(b_n) \quad \text{as} \quad n \to \infty \quad \text{if} \quad \lim_{n \to \infty} \frac{a_n}{b_n} = 0. \]

**Definition 1.1.17.** We say that \{a_n\} is asymptotically equivalent to \{b_n\}, or simply \{a_n\} is asymptotic to \{b_n\}, and we write

\[ a_n \sim b_n \quad \text{if and only if} \quad \lim_{n \to \infty} \frac{a_n}{b_n} = 1. \]

**Definition 1.1.18 (Estimator).**

An estimator is any statistic from the sample data which is used to give information about an unknown parameter in the population. For example, the sample mean is an estimator of the population mean. Estimators of population parameters are sometimes distinguished from the true value by using the symbol 'hat'. For example, the normal distribution has two parameters, the mean \( \mu \) and the standard deviation \( \sigma \). There estimators are denoted by \( \hat{\mu} \) and \( \hat{\sigma} \).

**Definition 1.1.19. (Unbiased estimator and Asymptotic Unbiasedness)**

Let \( X \) be a random variable with pdf \( f(\theta) \). Let \( X_1, \ldots, X_n \) be a random sample from the distribution of \( X \) and let \( \hat{\theta} \) denotes an estimator of \( \theta \). We say \( \hat{\theta} \) is an unbiased estimator of \( \theta \), if \( E(\hat{\theta}) = \theta \).

If \( \hat{\theta} \) is not unbiased, we say that \( \hat{\theta} \) is a biased estimator of \( \theta \). We say \( \hat{\theta} \) is asymptotic Unbiased estimator of \( \theta \) if its expected value converges to \( \theta \) as \( n \to \infty \).

\[ \lim_{n \to \infty} E(\hat{\theta}) = \theta \]

If this is not true, then \( \hat{\theta} \) is asymptotically biased.

**Example 1.1.1.** Let \( X_1, \ldots, X_n \) be a random sample distribution with mean \( \mu \) and variance \( \sigma^2 \), then the sample mean \( \bar{X} \) is unbiased estimator for \( \mu \).

\[
E(\bar{X}) = E\left( \frac{\sum_{i=1}^{n} X_i}{n} \right) = \frac{1}{n} E\left( \sum_{i=1}^{n} X_i \right),
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} E(X_i) = \frac{1}{n} \sum_{i=1}^{n} \mu = \frac{1}{n}(n\mu) = \mu.
\]

So, \( \bar{X} \) is unbiased estimator for \( \mu \).
**Example 1.1.2.** Let $X_1, \ldots, X_n$ be a random sample from a normal distribution with mean $\mu$ and variance $\sigma^2$, then sample variance $S^2$ is biased estimator for $\sigma^2$. we know $\sigma^2 = \frac{\sum_{i=1}^{n} (X_i - \mu)^2}{n}$ and $S^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{x})^2}{n-1}$.

Then,

$$E(S^2) = E\left(\frac{\sum_{i=1}^{n} (X_i - \bar{x})^2}{n-1}\right)$$

$$= E\left(\frac{\sum_{i=1}^{n} (X_i - \bar{x})^2}{n-1}\frac{n}{n}\right)$$

$$= E\left(\frac{\sum_{i=1}^{n} (X_i - \bar{x})^2}{n-1}\frac{n}{n-1}\right)$$

$$= E\left(\sigma^2\frac{n}{n-1}\right)$$

$$= \left(\frac{n}{n-1}\right)E(\sigma^2)$$

$$= \left(\frac{n}{n-1}\right)\sigma^2$$

So, $S^2$ is biased estimator for $\sigma^2$.

But

$$\lim_{n \to \infty} E(S^2) = \lim_{n \to \infty} \left(\frac{n}{n-1}\sigma^2\right) = \sigma^2 \lim_{n \to \infty} \left(\frac{n}{n-1}\right) = \sigma^2.$$ 

So, $S^2$ is an asymptotically unbiased estimator for $\sigma^2$.

**Definition 1.1.20. (Relative Efficiency of Estimators )**

If $\hat{\theta}_1$ and $\hat{\theta}_2$ are both unbiased estimators of parameter $\theta$. We say that $\hat{\theta}_1$ is relatively more efficient if $Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$, we use the ratio

$$\frac{Var(\hat{\theta}_1)}{Var(\hat{\theta}_2)},$$

as a measure of the relative efficiency of $\hat{\theta}_1$ w.r.t $\hat{\theta}_2$.

**Definition 1.1.21. (Consistent Estimators )**

The estimator $\hat{\theta}$ of a parameter $\theta$ is said to be consistent estimator if and only if for each $c > 0$
\[
\lim_{n \to \infty} P(|\hat{\theta} - \theta| < c) = 1.
\]

If this is not true, then \( \hat{\theta} \) is inconsistent.

**Theorem 1.1.1.** If \( \hat{\theta} \) is an unbiased estimator of \( \theta \) and \( \text{Var}(\hat{\theta}) \to 0 \) as \( n \to \infty \), then \( \hat{\theta} \) is consistent estimator of \( \theta \).

**Theorem 1.1.2. (Taylor Theorem)**

Suppose that \( f \) is a real-valued function defined on \( R \) and let \( x \in R \). Assume that \( f \) has \( p \) continuous derivatives in an interval \( (x - \delta, x + \delta) \) for some \( \delta > 0 \) and the \( (p + 1) \text{th} \) derivative of \( f \) exists. Then for any sequence \( (\alpha_n) \) converging to zero, we have

\[
f(x + \alpha_n) = \sum_{j=0}^{p} \frac{\alpha_n^j}{j!} f^{(j)}(x) + o(\alpha_n^p)\]
1.2 The Kernel estimator

The univariate kernel density estimation (KDE) is a non-parametric way to estimate the pdf \( f(x) \) of a random variable \( X \). It is a fundamental data smoothing problem where inferences about the population are made, based on a finite data sample. These techniques are widely used in various inference procedures such as signal processing, data mining, and econometrics. For more details see Silverman (1986), Wand and Jones (1995), and Alexandre (2009).

**The stages evolution of the estimator**

The first and the oldest is the histogram estimator.

The histogram estimator is defined by

\[
\hat{f}(x) = \frac{\text{Number of observations in the same bin as } x}{n h}
\]

and then from the definition of a PDF we get the Naive estimator.

The Naive estimator is defined by

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

If we replace the weight function by the kernel function then we get the kernel estimator.

The kernel estimator of a probability density function is,

\[
\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{\infty} K \left( \frac{x-x_i}{h} \right),
\]

\( h \) is called the bandwidth and \( K(\cdot) \) is kernel function which satisfies

\[
\int_{-\infty}^{\infty} K(x) dx = 1.
\]

**Definition 1.2.1. (Kernel Estimator of a Probability Density Function)**

Suppose that \( X_1, \ldots, X_n \) is a random sample of data from an unknown continuous distribution with pdf \( f(x) \) and cumulative distribution function (cdf) \( F(x) \), the
kernel estimator of a probability density function is defined as

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K \left( \frac{x - X_i}{h} \right),$$  \hspace{1cm} (1.1)$$

where the bandwidth $h = h_n$ is a sequence of positive numbers that converging to zero and $K(\cdot)$ is the kernel function considers to be both symmetric and satisfies

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

Kernels are always chosen to integrate to one, but there can be asymptotic advantages to kernels that are negative in places. The density estimates derived using such kernels can fail to be probability densities, because they can be negative for some values of $x$. Typically, $K$ is chosen to be a symmetric probability density function. There is a large body of literature on choosing $K$ and $h$ well, where "well" means that the estimate converges asymptotically as rapidly as possible in some suitable norm on probability density functions.

Note that Equation (1.1), can be written as

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^{n} K_{h}(x - X_i).$$  \hspace{1cm} (1.2)$$

where $K_{h}(x) = K(x/h)/h$.

The kernel estimator can be considered as a sum of bumps placed at the observations. The kernel function $K$ determines the shape of the bumps where the bandwidth $h$ determines their width.

From Figure (1.1), we have

1. The shape of bump is defined the kernel function.
2. The shape of the bump is determined by a bandwidth $h$.

That is, the value of the kernel estimate at the point $x$ is the average of the $n$ kernel ordinates at this point.
1.3 Properties of the Kernels

In this section, we will discuss some fundamental properties of the kernels and give some examples of common the kernel functions.

Properties of the Kernels:
We consider some properties of the kernels:
1) The kernel is a piecewise continuous function, symmetric around zero, and integrating to one, i.e.

\[ K(x) = K(-x), \quad \int_{-\infty}^{\infty} K(x)dx = 1. \quad (1.3) \]

2) The kernel function need not have bounded support and in most applications \( K \) is a positive probability density function.

**Definition 1.3.1.** A kernel function \( K \) is said to be of order \( p \), if its first nonzero
moment is $\mu_p \neq 0$, i.e. if

$$
\mu_j(K) = 0, \quad j = 1, 2, \ldots, p - 1; \quad \mu_p(K) \neq 0;
$$

where

$$
\mu_j(K) = \int_{-\infty}^{\infty} y^j K(y) dy.
$$

(1.4)

Some examples of the kernel functions and their formula are given in Table 1.1, where $I$ is the indicator function.

<table>
<thead>
<tr>
<th>Equation</th>
<th>Kernel</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K(u) = \frac{1}{2} I(</td>
<td>u</td>
</tr>
<tr>
<td>$K(u) = (1 -</td>
<td>u</td>
</tr>
<tr>
<td>$K(u) = \frac{3}{4}(1 - u^2)I(</td>
<td>u</td>
</tr>
<tr>
<td>$K(u) = \frac{15}{16}(1 - u^2)^2I(</td>
<td>u</td>
</tr>
<tr>
<td>$K(u) = \frac{1}{\sqrt{2\pi}} \exp(-u^2/2)$</td>
<td>The Gaussian kernel</td>
</tr>
<tr>
<td>$K(u) = \frac{n}{\pi} \cos(\frac{u}{2})I(</td>
<td>u</td>
</tr>
</tbody>
</table>
1.4 Properties of the Kernels Estimator

In this section, we will discuss some important properties of the kernel estimator.

To establish the properties of the kernel density estimator \( \hat{f}(x) \) of the pdf \( f(x) \), we consider the following conditions:

i) The unknown density function \( f(x) \) has continuous second derivative \( f''(x) \).

ii) The bandwidth \( h = h_n \) satisfies

\[
\lim_{n \to \infty} h = 0, \quad \text{and} \quad \lim_{n \to \infty} nh = \infty \quad (1.5)
\]

iii) The kernel \( K \) is a bounded pdf of order 2 and symmetric about the origin.

\[
\int_{-\infty}^{\infty} zK(z)dz = 0 \quad \text{and} \quad \int_{-\infty}^{\infty} z^2K(z)dz \neq 0 < \infty.
\]

1.4.1 Bias and Variance of the kernel density estimation

In the light of the previous properties, the bias and the variance of the kernel density estimation function will be derived.

**Definition 1.4.1. (The Bias Estimator)**

The Bias of an estimator \( \hat{f}(x) \) of a density \( f(x) \) is the difference between the expected value of \( \hat{f}(x) \) and \( f(x) \). That is

\[
Bias(\hat{f}(x)) = E(\hat{f}(x)) - f(x).
\]

**Lemma 1.4.1.** Let \( X \) be a random variable having density function \( f(x) \), then the bias of \( \hat{f}(x) \) can be expressed as

\[
Bias[\hat{f}(x)] = \frac{1}{2} h^2 f^{(2)}(x) \mu_2(K) + o(h^2), \quad (1.6)
\]

where \( \mu_2(K) = \int_{-\infty}^{\infty} z^2K(z)dz \).

**Proof:** Note that

\[
E[\hat{f}(x)] = \int K_h(x - y)f(y)dy = \int \frac{1}{h} K(\frac{x - y}{h})f(y)dy.
\]
let \( z = \frac{x - y}{h} \), then we get
\[
E[\hat{f}(x)] = \int_{-\infty}^{\infty} K(z) f(x - zh) \, dz,
\]
since \( f(x) \) is continuous derivative of order 2.
Then we can expand \( f(x - zh) \) in Taylor series as follows:
\[
f(x - zh) = \sum_{j=0}^{2} \frac{(-zh)^j}{j!} f^{(j)}(x) + o(-zh)^2
= f(x) + (-zh)f^{(1)}(x) + \frac{z^2h^2}{2} f^{(2)}(x) + o(z^2h^2)
= f(x) - zh f^{(1)} + \frac{1}{2} z^2 h^2 f^{(2)}(x) + o(h^2).
\]
Therefore,
\[
E[\hat{f}(x)] = \int_{-\infty}^{\infty} K(z) \{ f(x) - zh f^{(1)}(x) + \frac{z^2h^2}{2} f^{(2)}(x) + o(h^2) \} \, dz
= \int_{-\infty}^{\infty} K(z) f(x) \, dz - hf^{(1)}(x) \int_{-\infty}^{\infty} z K(z) \, dz + \frac{h^2}{2} f^{(2)}(x) \int_{-\infty}^{\infty} z^2 K(z) \, dz + o(h^2)
= f(x) - zf^{(1)} + \frac{1}{2} z^2 h^2 f^{(2)}(x) \int_{-\infty}^{\infty} z^2 K(z) \, dz + o(h^2).
\]
so, we have
\[
E[\hat{f}(x) - f(x)] = \frac{1}{2} h^2 f^{(2)}(x) \int_{-\infty}^{\infty} z^2 K(z) \, dz + o(h^2).
\]
\[
Bias[\hat{f}(x)] = \frac{1}{2} h^2 f^{(2)}(x) \int_{-\infty}^{\infty} z^2 K(z) \, dz + o(h^2).
\]
By assumption, we have the result.

**Lemma 1.4.2.** Let \( X \) be a random variable having density function \( f(x) \), then
\[
Var\{\hat{f}(x)\} = (nh)^{-1} R(K) f(x) + o\{(nh)^{-1}\}, \tag{1.7}
\]
where \( R(K) = \int_{-\infty}^{\infty} K^2(z) \, dz \).
Proof:

\[
\text{Var}\{\hat{f}(x)\} = \text{Var}\left[ \frac{1}{nh} \sum_{i=1}^{n} K\left( \frac{x - X_i}{h} \right) \right] \\
= \frac{1}{nh^2} \text{Var}\left[ K\left( \frac{x - X_i}{h} \right) \right] \\
= \frac{1}{nh^2} \left[ E\left[ K^2\left( \frac{x - X_i}{h} \right) \right] - \left( E\left[ K\left( \frac{x - X_i}{h} \right) \right] \right)^2 \right] \\
= \frac{1}{nh^2} \left[ E\left[ K^2\left( \frac{x - y}{h} \right) \right] - \left( E\left[ K\left( \frac{x - y}{h} \right) \right] \right)^2 \right] \\
= \frac{1}{nh^2} \left[ \left( \int_{-\infty}^{\infty} K^2\left( \frac{x - y}{h} \right) f(y)dy \right) - \left( \int_{-\infty}^{\infty} K\left( \frac{x - y}{h} \right) f(y)dy \right)^2 \right]
\]

let \( z = \frac{x - y}{h} \), then we get

\[
\text{Var}\{\hat{f}(x)\} = \frac{1}{nh^2} \left[ \left( h \int_{-\infty}^{\infty} K^2(z) f(x - zh)dz \right) - \left( h \int_{-\infty}^{\infty} K(z) f(x - zh)dz \right)^2 \right]
\]

Using Taylor series for \( f(x - zh) \) as before, we have

\[
\text{Var}\{\hat{f}(x)\} = \frac{1}{nh^2} \left[ \left( h \int_{-\infty}^{\infty} \{ f(x) - hzf^{(1)}(x) \} K^2(z)dz \right) - \o(h^2) \right]
\]

\[
= \frac{1}{nh} \left[ \left( \int_{-\infty}^{\infty} \{ f(x) - hzf^{(1)}(x) \} K^2(z)dz \right) - \o(h) \right]
\]

\[
= (nh)^{-1} f(x) \int_{-\infty}^{\infty} K^2(z)dz + o\{(nh)^{-1}\}
\]

By assumption, the result holds.

According the above lemmas, the following properties is hold :
1. The bias is of order \( h^2 \), which implies that \( \hat{f}(x) \) is an asymptotically unbiased estimator.
2. The bias is large, whenever the absolute value of the second derivative \( |f^{(2)}(x)| \) is large. This occurs for several densities at peaks where the bias is negative, and valleys, where the bias is positive.
3. The variance is of order \( (nh)^{-1} \), which means that the variance converges to zero by Condition (ii)(page 14).

Parzen (1962), studied an other statistical properties of the kernel estimators. In
addition to the above, \( \hat{f}(x) \) is a consistent estimate of \( f(x) \), also the sequence of estimates \( \hat{f}(x) \) is asymptotically normally distributed. Moreover, he proved that if the probability density function \( f(x) \) is uniformly continuous (see Chon, D.(1980)) such that \( \lim_{n \to \infty} nh^2 = \infty \), then \( \hat{f}(x) \) tends uniformly continuously (in probability) to \( f(x) \), if

\[
\lim_{n \to \infty} P( \sup_{-\infty < x < \infty} |\hat{f}(x) - f(x)| < \epsilon) = 1, \quad \forall \epsilon > 0.
\] (1.8)

1.4.2 The MSE and MISE Criteria

The important role played by the kernel density estimator makes us concerned with its performance, its efficiency and accuracy in estimating the true density. We study two types of the error criteria, the MSE and the MISE.

**Definition 1.4.2.** The mean squared error (MSE) is used to measure the error when estimating the density function at a single point. It is defined by

\[
MSE\{\hat{f}(x)\} = \mathbb{E}\{(\hat{f}(x) - f(x))^2\}
\] (1.9)

MSE measures is the average squared difference between the density estimator and the true density. In general, any function of the absolute distance \( |\hat{f}(x) - f(x)| \) (often called metric) would serve as a measurement of the goodness of an estimator. But MSE metric has at least two advantages over other metrics. First it is tractable analytically. Second it has an interesting decomposition into the variance and squared the bias provided \( f(x) \) is not random.

**Lemma 1.4.3.** The MSE is a sum of the squared bias and the variance at \( x \), as follow:

\[
MSE\{\hat{f}(x)\} = Var\{\hat{f}(x)\} + bias^2(\hat{f}(x)).
\] (1.10)
Proof:

\[ \text{MSE}\{\hat{f}(x)\} = E\{f(x) - \hat{f}(x)\}^2 \]
\[ = E\{f^2(x) - 2f(x)\hat{f}(x) + \hat{f}^2(x)\} \]
\[ = E\{f^2(x)\} - 2f(x)E\{\hat{f}(x)\} + E\{\hat{f}^2(x)\} \]
\[ = f^2(x) - 2f(x)E\{\hat{f}(x)\} + Var\{\hat{f}(x)\} + \{E\hat{f}(x)\}^2 \]
\[ = Var\{\hat{f}(x)\} + (E\{\hat{f}(x)\} - f(x))^2 \]
\[ = Var\{\hat{f}(x)\} + \text{Bias}^2(\hat{f}(x)) \]

The second type of criteria measures the error when estimating the density over the whole real line. The most well known of this type the mean integral squared error (MISE) which was proposed by Rosenblatt (1956).

**Definition 1.4.3.** An error criterion that measures the distance between \( \hat{f}(x) \) and \( f(x) \) is the integrated squared error (ISE) given by

\[ \text{ISE}\{\hat{f}(x)\} = \int_{-\infty}^{\infty} (\hat{f}(x) - f(x))^2 dx \] (1.11)

Note that the ISE is not appropriate if we deal with all data sets, so we prefer to analyze the expected value of this random quantity, the ISE.

**Definition 1.4.4.** The expected value of the ISE is called the mean integrated squared error (MISE) is given by

\[ \text{MISE}\{\hat{f}(x)\} = E(\text{ISE}\{\hat{f}(x)\}) = E\int_{-\infty}^{\infty} (\hat{f}(x) - f(x))^2 dx \] (1.12)

By changing the order of integration, we have

\[ \text{MISE}\{\hat{f}(x)\} = \int_{-\infty}^{\infty} \text{MSE}(\hat{f}(x)) dx \] (1.13)
\[ = \int_{-\infty}^{\infty} \{E\hat{f}(x) - f(x)\}^2 dx + \int_{-\infty}^{\infty} Var\{\hat{f}(x)\} dx \] (1.14)

**Theorem 1.4.1.**

\[ \text{MISE}\{\hat{f}(x)\} = \text{AMISE}(\hat{f}(x)) + o\{h^4 + (nh)^{-1}\} \] (1.15)
where,
\[
AMISE[\hat{f}(x)] = \frac{1}{4} h^4 \mu_2^2(K) \int_{-\infty}^{\infty} (f''(x))^2 dx + (nh)^{-1} R(K),
\] (1.16)
is called the asymptotic mean integral squared error of \( \hat{f}(x) \), and \( R(K) = \int_{-\infty}^{\infty} K^2(x) dx \).

**Proof:**
From equations (1.6) and (1.7) and applying Equation (1.10), we get
\[
MSE(\hat{f}(x)) = Var\{\hat{f}(x)\} + Bias^2(\hat{f}(x))
\]
\[
= (nh)^{-1} R(K) f(x) + o((nh)^{-1}) + \frac{1}{4} h^4 \mu_2^2(K) (f^{(2)}(x))^2 + o(h^4) + h^2 \mu_2^2(K) f^{(2)}(x)
\]
\[
= (nh)^{-1} R(K) f(x) + \frac{1}{4} h^4 \mu_2^2(K) (f^{(2)}(x))^2 + \{o(nh)^{-1} + h^4\}.
\]

Integrating this expression yields:
\[
MISE = (nh)^{-1} R(K) + \frac{1}{4} h^4 \mu_2^2(K) \int_{-\infty}^{\infty} (f^{(2)}(x))^2 dx + \{o(nh)^{-1} + h^4\}
\]
hence,
\[
MISE\{\hat{f}(x)\} = AMISE\{\hat{f}(x)\} + o\{h^4 + (nh)^{-1}\}
\]
Notice that the integral squared bias is asymptotically proportional to \( h^4 \), so to reduce \( AMISE \{\hat{f}(x)\} \), we must minimize the bandwidth \( h \), but increases the integral variance since it is proportional to \( (nh)^{-1} \). Therefore, as \( n \) increases, \( h \) should vary in such a way that each of the components of the MISE becomes small. This is known as the variance-bias trade-off. The trade-off between bias and variance in the bandwidth distributions seems to be an intrinsic part of the performance of data-based bandwidth selectors. Less bias seems to entail more variance, and at some cost in bias, much less variance can be obtained. See Wand and Jones (1995).

**Corollary 1.4.1.** The AMISE-optimal bandwidth, \( h_{opt} \), has a closed form
\[
h_{opt} = \left[ \frac{R(K)}{n \mu_2^2(K) \int_{-\infty}^{\infty} (f^{(2)}(x))^2 dx} \right]^{\frac{1}{2}}.
\] (1.17)
**Proof:**

By differentiating Equation (1.16) with respect to \( h \) and setting the derivative equal to zero we can find the optimal bandwidth

\[
\frac{d}{dh}\{AMISE \hat{f}(x)\} = -(nh^2)^{-1}R(K) + h^3 \mu_2^2(K) \int_{-\infty}^{\infty} (f^{(2)}(x))^2 dx = 0
\]

\[
h^5 \mu_2^2(K) \int_{-\infty}^{\infty} (f^{(2)}(x))^2 dx = n^{-1} R(K)
\]

\[
h_{opt} = \left\{ \frac{R(K)}{n \mu_2^2(K) \int_{-\infty}^{\infty} (f^{(2)}(x))^2 dx} \right\}^{\frac{1}{5}}
\]

Therefore if we substitute Equation (1.17) into Equation (1.16), we obtain the smallest value of AMISE (since the second derivative is grater than zero) for estimating \( f \) using the kernel \( K \).

\[
AMISE\{\hat{f}(x)\} = (nh)^{-1}\{R(K) + \frac{n}{4} h^5 \mu_2^2(K) \int_{-\infty}^{\infty} (f^{(2)}(x))^2 dx\}
\]

\[
= \frac{5}{4} \frac{R(K) \{\mu_2^2(K) \int_{-\infty}^{\infty} (f^{(2)}(x))^2 dx\}^{\frac{1}{5}}}{n^{\frac{4}{5}} (R(K))^{\frac{1}{5}}}
\]

\[
= \frac{5}{4} n^{-\frac{4}{5}} (R(K))^{\frac{1}{5}} \left( \mu_2^2(K) \int_{-\infty}^{\infty} (f^{(2)}(x))^2 dx \right)^{\frac{1}{5}}
\]

take the infimum over \( h > 0 \), we get

\[
\inf_{h>0} AMISE\{\hat{f}\} = \frac{5}{4} \{\mu_2^2(K) R^4(K) \int_{-\infty}^{\infty} (f^{(2)}(x))^2 dx\}^{\frac{1}{5}} n^{-\frac{4}{5}}.
\]

(1.18)

Notice that from Equation (1.17), note that the optimal bandwidth depends on the unknown density being estimated, so we can not use Equation (1.17) directly to find the optimal bandwidth \( h_{opt} \). Also, from Equation (1.17), we can conclude the following useful conclusions:

1. The optimal bandwidth will converge to zero as the sample size increases, but at very slow rate.
2. The optimal bandwidth is inversely proportional to \( (\int_{-\infty}^{\infty} (f^{(2)}(x))^2 dx)^{\frac{1}{5}} \). Since \( \int_{-\infty}^{\infty} f^{(2)}(x) dx \) measures the curvature of \( f(x) \), this means that for \( f(x) \) with little curvature, the optimal bandwidth will be large. Conversely, if \( f(x) \) has a large curvature, the optimal bandwidth will be small.
1.5 Asymptotic normality of the kernel density estimator

In this section, we will derive the asymptotic normality of the kernel estimator \( \hat{f}(x) \) of the pdf \( f(x) \). First, we will state some preliminaries results that will be used to get the asymptotic normality. In the following, we will present some theorem that will help in this thesis. The theorems in this section was taken from Pranab (1993).

**Theorem 1.5.1. (Central Limit Theorem (CLT))**

If the distribution of the i.i.d. sample \( X_1, \ldots, X_n \) is such that \( X_i \) has finite expectation and variance, i.e. \( E[X_i] < \infty \) and \( \sigma^2 = Var(X_i) < \infty \), then converges in distribution to normal distribution with zero mean and variance \( \sigma^2 \)

\[
\sqrt{n} (\bar{X} - E(X_i)) \xrightarrow{D} N(0, \sigma^2).
\]

**Theorem 1.5.2. (Lyapounov CLT)**

Suppose that for each \( n \), \( v_{n1}, \ldots, v_{nn} \) are independent. Assume that \( E v_{ni} = 0 \) and \( \sigma^2_{ni} = E v^2_{ni} \) and define \( s^2_n = \sum_{i=1}^n \sigma^2_{ni} \). Suppose further that for some \( \delta > 0 \) the following condition holds:

\[
\lim_{n \to \infty} \frac{1}{s^2_n} \sum_{i=1}^n \frac{1}{s^2_n} E|v_{ni}|^{2+\delta} = 0. \tag{1.19}
\]

then,

\[
\sum_{i=1}^n v_{ni}/s_n \xrightarrow{D} N(0, 1).
\]

**Theorem 1.5.3.** Suppose that for each \( n \), \( v_{n1}, \ldots, v_{nn} \) are i.i.d., \( Ev_{ni} = 0 \) and \( \sigma^2_{ni} = Ev^2_{ni} < \infty \) and \( \lim_{n \to \infty} \frac{E|v_{ni}|^{2+\delta}}{n^\delta} = 0 \) for some \( \delta > 0 \), then

\[
\frac{1}{n^\delta} \sum_{i=1}^n v_{ni} \xrightarrow{D} N(0, \lim_{n \to \infty} Ev^2_{ni}).
\]

**Theorem 1.5.4. (The normality of the kernel density estimator)**

Suppose that \( nh \to \infty \) and \( \sqrt{nh^5} \to 0 \), then we have

\[
\sqrt{nh}[\hat{f}(x) - f(x)] \xrightarrow{D} N\left(0, f(x) \int_{-\infty}^{\infty} K^2(u)du\right)
\]
Proof:

\[ \sqrt{nh}[\hat{f}(x) - f(x)] = \sqrt{nh}\left( \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) - f(x) \right) \]

\[ = \sqrt{nh}\left( \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) - E\left[ \frac{1}{h} K\left(\frac{x - X_i}{h}\right) \right] \right) \]

\[ + \sqrt{nh}\left( \frac{1}{nh} \sum_{i=1}^{n} E[K\left(\frac{x - X_i}{h}\right)] - f(x) \right) \]

\[ = \sqrt{n}\left( \frac{1}{\sqrt{h}} K\left(\frac{x - X_i}{h}\right) - \frac{1}{\sqrt{h}} E\left[ K\left(\frac{x - X_i}{h}\right) \right] \right) \quad (1) \]

\[ + \sqrt{nh}\left( \frac{1}{h} E[K\left(\frac{x - X_i}{h}\right)] - f(x) \right). \quad (2) \]

Analyze term (1)

\[ \sqrt{n}\left( \frac{1}{\sqrt{h}} K\left(\frac{x - X_i}{h}\right) - \frac{1}{\sqrt{h}} E\left[ K\left(\frac{x - X_i}{h}\right) \right] \right) \]

Now, apply the CLT Theorem for this part

\[ \text{Term (1)} \sim \mathcal{N}\left(0, \text{Var}\left( \frac{1}{\sqrt{h}} K\left(\frac{x - X_i}{h}\right) \right) \right) \]

\[ \text{Var}\left( \frac{1}{\sqrt{h}} K\left(\frac{x - X_i}{h}\right) \right) = \frac{1}{h} E\left( \left. K^2\left(\frac{x - X_i}{h}\right) \right| \right) - \frac{1}{h} \left( E\left( K\left(\frac{x - X_i}{h}\right) \right) \right)^2 \]

\[ = \frac{1}{h} \int_{-\infty}^{\infty} K^2\left(\frac{x - X_i}{h}\right) f(X_i) dX_i - h E\left( \frac{1}{h} K\left(\frac{x - X_i}{h}\right) \right)^2 \]

\[ = \int_{-\infty}^{\infty} K^2(u) f(x - hu) du - h\{ f(x) + o(1) \} \]

\[ = \int_{-\infty}^{\infty} K^2(u) f(x + h(u)) du - h\{ f(x) + o(1) \} \]

\[ = f(x) \int_{-\infty}^{\infty} K^2(u) du. \]

\[ \text{Term (1)} \sim \mathcal{N}\left(0, f(x) \int_{-\infty}^{\infty} K^2(u) du \right) \]
Analyze term (2)

\[
\frac{1}{h} E[K\left(\frac{x - X_i}{h}\right)] = \frac{1}{h} \int_{-\infty}^{\infty} K\left(\frac{x - X_i}{h}\right) f(X_i) dX_i, \quad \text{let } u = \frac{x - X_i}{h},
\]

\[
= \int_{-\infty}^{\infty} K(s) f(x - sh) ds, \quad \text{using Taylor series expansion for } f(x - hs)
\]

\[
= \int_{-\infty}^{\infty} K(s) \left[ f(x) - shf^{(1)}(x) + \frac{s^2h^2}{2} f^{(2)}(x) + o(h^2) \right] ds,
\]

\[
= f(x) \int_{-\infty}^{\infty} K(s) ds - h f^{(1)}(x) \int_{-\infty}^{\infty} s K(s) ds + \frac{h^2}{2} f^{(2)}(x) \int_{-\infty}^{\infty} s^2 K(s) ds + o(h^2)
\]

since \( \int_{-\infty}^{\infty} K(s) ds = 1, \int_{-\infty}^{\infty} s K(s) ds = 0, \) and \( |f^{(2)}(x)| \leq M, \) then

\[
\frac{1}{h} E[K\left(\frac{x - X_i}{h}\right)] = f(x) \int_{-\infty}^{\infty} K(s) ds - h f^{(1)}(x) \int_{-\infty}^{\infty} s K(s) ds + \frac{h^2}{2} f^{(2)}(x) \int_{-\infty}^{\infty} s^2 K(s) ds + o(h^2)]
\]

\[
= f(x) + \frac{h^2}{2} f^{(2)}(x) \int_{-\infty}^{\infty} s^2 K(s) ds + o(h^2).
\]

so,

\[
\frac{1}{h} E[K\left(\frac{x - X_i}{h}\right)] - f(x) = o(h^2)
\]

This implies that

\[
\sqrt{nh} \left( \frac{1}{h} E[K\left(\frac{x - X_i}{h}\right)] - f(x) \right) = o(\sqrt{nh}.h^2) = o(\sqrt{nh^5}).
\]

Since \( \sqrt{nh^5} \to 0, \) so this is completed proof.
Chapter 2

Nadaraya-Watson Estimator

Introduction

Conditional distribution functions underlie many popular statistical object of interest. They are rarely modeled directly in parametric setting and have perhaps received even less attention in kernel setting. Nevertheless, as will be seen, they are extremely useful for range of tasks, whether directly estimation the conditional distribution function (see Cameron and Trivedi (1998)), or perhaps molding conditional quantiles. The conditional median depends directly on the conditional distribution function. Indeed, estimating the conditional distribution is actually much more informative, since it allows us not only to recalculate the expected value $E(Y \mid X)$ and the variance $\sigma^2(Y \mid X)$, but also to provide the general shape of the conditional distribution.

In this context several nonparametric methods can be applicable for estimating the conditional distribution function based on data $(X_1, Y_1), \ldots, (X_n, Y_n)$. A class of the kernel-type estimators is called the Nadaraya-Watson estimator which is one of the most widely known and used for estimating the conditional distribution function. Conditional distribution estimation was introduced by Rosenblatt (1956). A bias
correction was proposed by Hyndman et al. (1996). Fan et al. (1996) proposed a direct estimator on Local polynomial estimation. The Nadaraya-Watson (NW) estimator is created by the two researchers Watson (1964) and Nadaraya (1964). In this chapter, we will present the NW estimator in order to estimate the conditional density function \( f(y \mid x) \) and conditional distribution function \( F(y \mid x) \) which is a bois to derive the NW estimator of conditional median function \( \hat{m}_{NW}(x) \). In Section 2.1, we will present kernel estimation of the conditional cdf. In Section 2.2, we will introduce the NW estimator of the conditional distribution function \( \hat{F}_{NW}(y \mid x) \) and use it to introduce \( \hat{m}_{NW}(x) \) based on data \( \{(x_i, y_i)\}_{i=1}^n \). In Section 2.3, the asymptotic properties of the \( \hat{m}_{NW}(x) \), the asymptotic expressions for the bias, the variance and the MSE of the \( \hat{m}_{NW}(x) \) will be discussed, and we will investigate the asymptotic normality of the \( \hat{m}_{NW}(x) \).

### 2.1 Estimating the Conditional Probability distribution function

Assume that \( (X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n) \) are i.i.d. sample from a real-valued random variable \((X, Y)\).

**Definition 2.1.1** (The conditional median).

The conditional median can be written as,

\[
m(x) = \inf\{y \in \mathbb{R} : F(y \mid x) \geq 0.5\}
\]

**Definition 2.1.2** (Conditional probability distribution function).

If \( f(x, y) \) is the joint pdf of the random variables \( X \) and \( Y \) at \( (x, y) \), and \( g(x) \) is the marginal pdf of \( X \) at \( x \), the conditional pdf of \( Y \) given \( X = x \) is given by

\[
f(y \mid x) = \frac{f(x, y)}{g(x)}, \quad g(x) > 0,
\]

(2.1)
for each \( y \) within the range of \( Y \), and then,

\[
F(y|x) = \int_{-\infty}^{y} f(u|x)du
\]  

(2.2)

Now, we introduce the basic equation of the kernel conditional distribution estimation (KCDF). The Kernel function \( K(u) \) is assumed to be the kernel function, and \( h_n = h \) is a sequence of positive numbers converging to zero. Standard kernel estimators of \( f(x,y) \), \( g(x) \), \( f(y|x) \) and \( F(y|x) \) are

\[
\hat{f}(x,y) = (nh^2)^{-1} \sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)K\left(\frac{y-Y_i}{h}\right)
\]

(2.3)

\[
\hat{g}(x) = (nh)^{-1} \sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)
\]

(2.4)

and

\[
\hat{f}(y|x) = \frac{\hat{f}(x,y)}{\hat{g}(x)}
\]

\[
= \frac{(nh^2)^{-1} \sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)K\left(\frac{y-Y_i}{h}\right)}{(nh)^{-1} \sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)}
\]

\[
= \frac{1}{h} \sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)K\left(\frac{y-Y_i}{h}\right)
\]

(2.5)

Now by using the estimator of the conditional cdf, we introduce the estimator for the conditional median as follows,

\[
\hat{m}(x) = \inf \{ y \in \mathbb{R} : \hat{F}(y \mid x) \geq 0.5 \}
\]
2.2 The Nadaraya-Watson Estimator

Let \((X_1, Y_1), ..., (X_n, Y_n)\) be real valued independent random variables with a common cdf \(F(X,Y)\). Also assume that \(X\) admits the marginal pdf \(g\). The most common method for studying the relationship between two variables \(X\) and \(Y\) is to estimate the conditional cdf of \(Y\) given \(X = x\), and the conditional median \(m(x) = \inf[y \in \mathbb{R} : F(y \mid x) \geq 0.5]\).

Definition 2.2.1.

Suppose that we are given \(n\) observations of \((X, Y)\) denoted by \((X_1, Y_1), ..., (X_n, Y_n)\). The Nadaraya-Watson estimator of the conditional distribution function \(F(y \mid x)\) is defined as

\[
\hat{F}_{NW}(y \mid x) = \frac{\sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right) I(Y_i \leq y)}{\sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right)},
\]

(2.6)

where \(I\) is the indicator function, \(K\) is a kernel function, and \(h_n\) is sequence of positive number converging to zero and it is called bandwidth.

Definition 2.2.2. The Nadaraya-Watson estimator of the conditional median \(m(x)\) is defined as

\[
\hat{m}_{NW}(x) = \inf[y \in \mathbb{R} : \hat{F}_{NW}(y \mid x) \geq 0.5]
\]
**Assumptions**

We will consider the following assumptions:

**Assumption (A1)** The process \( \{(X_i, Y_i)\}_{i=1}^{n} \subseteq \mathbb{R} \times \mathbb{R} \) are i.i.d. sample from a real-valued random variable \((X, Y)\)

**Assumption (A2)** For fixed \( y \) and \( x \), \( 0 < F(y \mid x) < 1 \), \( F^{(2,0)}(y \mid x) = F''(y \mid x) \) exist in a neighborhood of \( x \).

**Assumption (A3)** The kernel function \( K \) is probability density function satisfies the following:

i. \( K \) is symmetric probability density function;

ii. \( K \) is Lipschitz continuous.

**Assumption (A4)** The bandwidth \( \{h_n\} \) satisfies the following:

i. \( \lim_{n \to \infty} h_n = 0; \)

ii. \( \lim_{n \to \infty} nh_n = \infty; \)

iii. \( \lim_{n \to \infty} nh_n^5 = 0. \)

Now, we will derive for independent \( \{Y_i\} \) the expectation and the variance of the
Now, the variance is given as follows:

\begin{align*}
E[\hat{F}_{NW}(y \mid x)] &= E\left[\frac{\sum_{i=1}^{n} I_{(Y_i \leq y)} K\left(\frac{x-X_i}{h_n}\right)}{\sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right)}\right] \\
&= \frac{1}{\sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right)} \sum_{i=1}^{n} I_{(Y_i \leq y)} K\left(\frac{x-X_i}{h_n}\right) f(y \mid X_i) dy \\
&= \frac{\sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right) \int_{-\infty}^{y} I_{(Y_i \leq y)} f(y \mid X_i) dy}{\sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right)} \\
&= \frac{\sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right) \int_{-\infty}^{y} f(t \mid X_i) dt}{\sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right)} \\
&= \frac{\sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right) F(y \mid X_i)}{\sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right)} \\
&= \frac{\sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right) F(y \mid X_i)}{\sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right)} \\
&= \frac{\sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right) F(y \mid X_i)}{\sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right)}.
\end{align*}

Now, the variance is given as follows:

\begin{align*}
Var[\hat{F}_{NW}(y \mid x)] &= E(\hat{F}_{NW}^2(y \mid x)) - [E(\hat{F}_{NW}(y \mid x))]^2 \\
&= E\left[\frac{\sum_{i=1}^{n} I_{(Y_i \leq y)} K\left(\frac{x-X_i}{h_n}\right) + \sum_{1 \leq i < j \leq n} I_{(Y_i \leq y)} I_{(Y_j \leq y)} K^2\left(\frac{x-X_i}{h_n}\right)}{\left[\sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right)\right]^2}\right] \\
&- \left[\frac{\sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right)}{\sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right)} \cdot F(y \mid X_i)\right]^2 \\
&= \frac{E[\sum_{i=1}^{n} I_{(Y_i \leq y)}^2 K^2\left(\frac{x-X_i}{h_n}\right)]}{\left[\sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right)\right]^2} - \sum_{i=1}^{n} \frac{K\left(\frac{x-X_i}{h_n}\right)}{\left[\sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right)\right]^2} \cdot F^2(y \mid X_i) \\
&= \frac{\int_{-\infty}^{y} \sum_{i=1}^{n} I_{(Y_i \leq y)} K^2\left(\frac{x-X_i}{h_n}\right) f(y \mid X_i) dy}{\left[\sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right)\right]^2} - \sum_{i=1}^{n} \frac{K^2\left(\frac{x-X_i}{h_n}\right)}{\left[\sum_{i=1}^{n} K^2\left(\frac{x-X_i}{h_n}\right)\right]^2} \cdot F^2(y \mid X_i).
\end{align*}
\[
\sum_{i=1}^{n} K^2 \left( \frac{x - X_i}{h_n} \right) \int_{-\infty}^{\infty} I(Y_i \leq y) f(y \mid X_i) - \sum_{i=1}^{n} \frac{K^2 \left( \frac{x - X_i}{h_n} \right)}{\left[ \sum_{i=1}^{n} K \left( \frac{x - X_i}{h_n} \right) \right]^2} F^2(y \mid X_i)
\]

\[
= \sum_{i=1}^{n} \frac{K^2 \left( \frac{x - X_i}{h_n} \right) F(y \mid X_i)}{\left[ \sum_{i=1}^{n} K \left( \frac{x - X_i}{h_n} \right) \right]^2} - \sum_{i=1}^{n} \frac{K^2 \left( \frac{x - X_i}{h_n} \right)}{\left[ \sum_{i=1}^{n} K \left( \frac{x - X_i}{h_n} \right) \right]^2} F^2(y \mid X_i)
\]

\[
= \sum_{i=1}^{n} \frac{K^2 \left( \frac{x - X_i}{h_n} \right) F(y \mid X_i)}{\left[ \sum_{i=1}^{n} K \left( \frac{x - X_i}{h_n} \right) \right]^2} - \sum_{i=1}^{n} \frac{K^2 \left( \frac{x - X_i}{h_n} \right)}{\left[ \sum_{i=1}^{n} K \left( \frac{x - X_i}{h_n} \right) \right]^2} F^2(y \mid X_i)
\]

\[
= \sum_{i=1}^{n} \frac{K^2 \left( \frac{x - X_i}{h_n} \right)}{\left[ \sum_{i=1}^{n} K \left( \frac{x - X_i}{h_n} \right) \right]^2} [F(y \mid X_i) - F^2(y \mid X_i)]
\]

Lemma 2.2.1. (Integral approximation of the sum over the kernel function).

With A3(i)Lipschist-continuity A3(ii) and the mean value theorem of integral it follows

let \( U_i = \frac{x - X_i}{h_n} \)

i. \( \lim_{n \to \infty} \sum_{i=1}^{n} \frac{1}{nh} K(U_i) = \int K(u) du \)

ii. \( \lim_{n \to \infty} \sum_{i=1}^{n} \frac{1}{nh} K^2(U_i) = \int K^2(u) du \)

iii. \( \lim_{n \to \infty} \sum_{i=1}^{n} \frac{1}{nh} U_i K(U_i) = \int uK(u) du \)


2.3 Asymptotic Properties

In this section, we will study the asymptotic properties of the \( \hat{m}_{NW}(x) \). We calculate the bias and the variance of the \( \hat{m}_{NW}(x) \). Then we will investigate the MSE of the \( \hat{m}_{NW}(x) \). For more details see Fan(1992) and Geenens (2013).

Theorem 2.3.1. Suppose that \( g(x) > 0 \) and \( f(\hat{m}_{NW}(x_i) \mid x_i) > 0 \), \( i=1,2 \), then under the assumptions A1-A4, the following holds

\[
\sqrt{nh_n} [\hat{m}_{NW}(x_1) - m_{NW}(x_1), \hat{m}_{NW}(x_2) - m_{NW}(x_2)] \xrightarrow{d} N(0, \mathbf{D}). \quad (2.7)
\]
where $\mathbf{0} = (0, 0)^T$ and $D$ is diagonal covariance matrix with the $(i,i)$ the element,

$$d_{ii} = \frac{0.25}{g(x_i) f^2(m(x_i) \mid x_i)} \int_{-\infty}^{\infty} K^2(u) du, \; i=1,2.$$  

**Proof:** The following notation will be considered.

Let

$$S_{nr} = F(y_r \mid x) - E\hat{F}(y_r \mid x), \; r=1,2$$

$$W_{ni}(x) = \frac{K(\frac{x-X_i}{\sigma_n})}{\sum_{j=1}^{n} K(\frac{X_j-X_i}{\sigma_n})}, \; i = 1, 2, ..., n$$

and

$$Z_{ri} = W_{ni}(x)I(Y_i \leq y_r) - EW_{ni}(x)I(Y_i \leq y_r), \; i=1,2,...,n \; , r=1,2.$$  

Therefore,

$$S_{nr} = \sum_{i=1}^{n} Z_{ri}$$

Set that

$$S_n = \sum_{i=1}^{2} c_r S_{nr}$$

Where $c_r$ are constant, $r=1,2$.

The proof of Theorem will be established through the following lemmas.

**Lemma 2.3.1.** Under the Assumptions A1-A4, and as $n \to \infty$, the following holds

(i) $\hat{f}(y \mid x) = \hat{F}^{(0,1)}(y \mid x) \xrightarrow{P} f(y \mid x) = F^{(0,1)}(y \mid x)$

(ii) $\hat{m}_{NW}(x) \xrightarrow{P} m(x)$

*Proof.* This lemma follows from the proofs of theorems 1, 2 and 3 in Aberger (1997). 

\[\square\]
**Definition 2.3.1. (The Covariance)**

The covariance between two random variables $X$ and $Y$ is defined as,

$$
\text{Cov}(X,Y) = E[(X - \mu_x)(Y - \mu_y)] = E[XY] - E[X]E[Y].
$$

**Lemma 2.3.2.** Under the Assumptions A1,A3,A4(i),(ii), and as $n \to \infty$, the following is true

$$
\text{Var}(S_{nr}) \to \frac{F(y_r | x)[1 - F(y_r | x)] \int_{-\infty}^{\infty} K^2(u)du}{nh_ng(x)}.
$$

*Proof.*

\[\text{Var}(S_{nr}) = \text{Cov}(S_{nr}, S_{nr}) = \text{Cov}(\hat{F}(Y_r | x), \hat{F}(Y_r | x)).\]
\[
= \sum_{i=1}^{n} W_{ni}(x)W_{nj}(x)\text{Cov}(I(Y_i \leq y_r), I(Y_j \leq y_r)).
\]
\[
= [\sum_{i=1}^{n} W_{ni}(x)]^2\text{Cov}(I(Y_i \leq y_r), I(Y_i \leq y_r)).
\]
\[
= \frac{\sum_{i=1}^{n} K^2(\frac{x - X_i}{h})\text{Cov}(I(Y_i \leq y_r), I(Y_i \leq y_r))}{[\sum_{i=1}^{n} K^2(\frac{x - X_i}{h})]^2} \to \frac{F(Y_r | x)[1 - F(Y_r | x)]}{nh_ng(x)} \int_{-\infty}^{\infty} K^2(u)du,
\]

by Theorem 1 in Abberger (1997).

**Lemma 2.3.3.** Under the Assumptions of lemma 2.3.2 and as $n \to \infty$, the following is true

$$
\text{Var}(S_n) \to \sigma^2
$$

31
where
\[
\sigma^2 = \sum_{r=1}^{2} c_r^2 F(Y_r \mid x)[1 - F(Y_r \mid x)] \int_{\infty}^{-\infty} K^2(u)du.
\]

Proof. The proof follows directly from lemma 2.3.2.

Proof of theorem 2.3.1.

Now, by expanding \( \hat{F}(\hat{m}_{NW}(x) \mid x) \) around \( m(x) \), \( i = 1, 2 \), we get
\[
F(m(x) \mid x) = \hat{F}(\hat{m}_{NW}(x) \mid x) = \hat{F}(m(x) \mid x) + (\hat{m}_{NW}(x_i) - m(x)) \hat{F}^{(0,1)}(m^*(x)) \mid x).
\]
This implies that
\[
\hat{m}_{NW}(x) - m(x) = \frac{F(m(x) \mid x) - \hat{F}(m(x) \mid x)}{\hat{F}^{(0,1)}(m^*(x) \mid x)},
\]
where \( m^*(x) \) is some random point between \( \hat{m}_{NW}(x) \) and \( m(x) \).

The proof of the theorem follows from lemma 2.3.2 and lemma 2.3.1(i),(ii).

Now we want to approximate the mean square error as in the next theorem.

Theorem 2.3.2.

Let \( \{Y_i\} \) be independent and let A2, A3(i,ii) and A3(i,ii) be satisfied.

Then it holds for \( n \to \infty \) and \( x \in (h_n, 1 - h_n) \)
\[
MSE(\hat{F}(y \mid x)) \approx \left[ \frac{h_n^2}{2} F^{(2)}(y \mid x) \int u^2 K(u)du \right]^2
+ \frac{1}{nh_n} F(y \mid x) - F^2(y \mid x) \int K^2(u)du.
\]

Proof. Since \( MSE(\hat{F}(y \mid x)) = (E[\hat{F}(y \mid x)] - F(y \mid x))^2 + Var(\hat{F}(y \mid x)). \)

Let \( U_i = \frac{x - X_i}{h_n} \) and \( x \in (h_n, 1 - h_n) \), then
\[
E[\hat{F}(y \mid x)] - F(y \mid x) = \frac{1}{2} h_n^2 \mu_2(K) F^{(2)}(y \mid x) + o(h_n^2),
\]
where
\[
\mu_2(K) = \int u^2 K(u)du.
\]
Then

\[ E[\hat{F}(y \mid x)] = F(y \mid x) + \frac{h_n^2 \sum_i U_i^2 K(U_i)}{2 \sum_i K(U_i)} F^{(2)}(y \mid x) + o(h_n^2). \]

Now we want to find \( \text{Var}(\hat{F}(y \mid x)) \).

Taylor expansion yields the variance

\[ F(y \mid x - h_n U_i) = F(y \mid x) - h_n U_i F^{(1)}(y \mid x) + h_n^2 U_i^2 F^{(2)}(y \mid x) + o(h_n^2), \]

\[ F^2(y \mid x - h_n U_i) = F^2(y \mid x) - 2 h_n U_i F(y \mid x) F^{(1)}(y \mid x) + h_n^2 U_i^2 F^{(1)}(y \mid x) + o(h_n^2) \]

\[ + h_n^2 U_i^2 F(y \mid x) F^{(2)}(y \mid x), \]

let the condition A2(ii) hold and let \( A = 1/\sum_i K(U_i)^2 \). Then

\[ \text{Var}[\hat{F}(y \mid x)] = \frac{\sum_{i=1}^n K^2(U_i)}{\sum_{i=1}^n [K(U_i)]^2} [F(y \mid X_i) - F^2(y \mid X_i)] \]

\[ = A \sum_{i=1}^n K^2(U_i) [F(y \mid X_i) - F^2(y \mid X_i)] \]

\[ = A \sum_{i=1}^n K^2(U_i) [F(y \mid x - h_n U_i) + h_n U_i F^{(1)}(y \mid x) - h_n^2 U_i^2 F^{(2)}(y \mid x) + o(h_n^2)] \]

\[ - F^2(y \mid x - h_n U_i) - 2 h_n U_i F(y \mid x) F'(y \mid x) + h_n^2 U_i^2 F'(y \mid x) \]

\[ + h_n^2 U_i^2 F(y \mid x) F^{(2)}(y \mid x) + o(h_n^2)] \]

\[ = A [F(y \mid x) - F^2(y \mid x)] \sum_i K^2(U_i) \]

\[ + A h_n^2 [F^{(2)}(y \mid x) + F^{(1)}(y \mid x) + F(y \mid x) F^{(2)}(y \mid x)] \sum_i U_i^2 K^2(U_i) \]

\[ + A o(h_n^2) \sum_i K^2(U_i). \]

With integral approximation it holds

\[ \text{Var}[\hat{F}(y \mid x)] \approx \frac{1}{nh_n} [F(y \mid x) - F^2(y \mid x)] \int K^2(u)du. \]

So the proof of this theorem is completed.
Thus the bias of \( \hat{F}(y \mid x) \) depends on the smoothness of the underlying condition distribution function by \( F^{(2)}(y \mid x) \). It is now possible to give a formal assessment about the asymptotic mean squared error.

Observe that the mean squared error depends on second derivative of the conditional distribution and the difference between \( [F(y \mid x) - F^2(y \mid x)] \). This means that the variance of the estimator is highest in the middle of the distributions. (Since the maximum of \( [F(y \mid x) - F^2(y \mid x)] \) is 1/4 and happens when \( F(y \mid x) = 1/2 \).) From the last theorem it follows that the kernel estimator is consistent. Next, the asymptotic normality of the \( (nh_n)^{1/2} (\hat{F}(y \mid x) - E[\hat{F}(y \mid x)]) \) and \( (nh_n)^{1/2} (\hat{F}(y \mid x) - F(y \mid x)) \) is shown.

**Theorem 2.3.3.**

Assuming the conditions of last theorem are satisfied. Then it holds for \( n \to \infty \),

\[
(nh_n)^{1/2} (\hat{F}(y \mid x) - E[\hat{F}(y \mid x)]) \xrightarrow{d} N(0, [F(y \mid x) - F^2(y \mid x)] \int K^2(u)du).
\]

**Proof.**

To prove this theorem, we use Liapunov condition.

Let

\[
Q_{n,i}(x) = \frac{K(z_{n,i} - X_i)}{\sum_{i=1}^{n} K(z_{n,i} - X_i)} [I_{Y_i \leq y} - F(y \mid X_i)] \sqrt{Var[\hat{F}(y \mid x)]}
\]

Therefore

\[
\sum_{i=1}^{n} Q_{n,i}(x) = \sum_{i=1}^{n} \frac{K(z_{n,i} - X_i)}{\sum_{i=1}^{n} K(z_{n,i} - X_i)} [I_{Y_i \leq y} - F(y \mid X_i)] \frac{1}{\sqrt{Var[\hat{F}(y \mid x)]}}
\]

That is

\[
\sum_{i=1}^{n} Q_{n,i}(x) = \sum_{i=1}^{n} \frac{K(z_{n,i} - X_i)}{\sum_{i=1}^{n} K(z_{n,i} - X_i)} [I_{Y_i \leq y}] - \sum_{i=1}^{n} \frac{K(z_{n,i} - X_i)}{\sum_{i=1}^{n} K(z_{n,i} - X_i)} F(y \mid X_i) \frac{1}{\sqrt{Var[\hat{F}(y \mid x)]}}.
\]
This means that
\[
\frac{\hat{F}(y \mid x) - E[\hat{F}(y \mid x)]}{\sqrt{\text{Var}[\hat{F}(y \mid x)]}} = \sum_{i=1}^{n} Q_{n,i}(x)
\]

The Lipaunovs condition
\[
\lim_{n \to \infty} \sum_{i=1}^{\infty} E |Q_{n,i}(x)|^3 = \lim_{n \to \infty} \frac{\sum_{i=1}^{n} E \left| \frac{K(x - X_i)}{\sum_i K(x - X_i)} [I_{Y_i \leq y} - F(y \mid X_i)] \right|^3}{(\text{Var}[\hat{F}(y \mid x)])^{\frac{3}{2}}},
\]
is satisfied.

With the integral approximation it holds for the numerator,
\[
\sum_{i=1}^{n} E \left| \frac{K(x - X_i)}{\sum_i K(x - X_i)} [I_{Y_i \leq y} - F(y \mid X_i)] \right|^3
\]
\[
= \sum_{i=1}^{n} E \left| \frac{K(x - X_i)}{\sum_i K(x - X_i)} \right|^3 \left[ I_{Y_i \leq y} - F(y \mid X_i) \right]^3
\]
\[
\leq \sum_{i=1}^{n} \left| \frac{K(x - X_i)}{\sum_i K(x - X_i)} \right|^3 = \sum_i K^3(x - X_i)
\]
\[
= \frac{nh_n \int K^3(x - X_i) du}{[nh_n \int K(x - X_i) du]^3}
\]
\[
= o(nh_n) = o\left(\frac{1}{(n^2h_n^2)}\right),
\]

For the variance of \(\hat{F}(y \mid x)\) follows from the last theorem
\[
\text{Var}[\hat{F}(y \mid x)] = o\left(\frac{1}{(nh_n)}\right)
\]

Thus it holds
\[
\text{Var}[\hat{F}(y \mid x)]^{\frac{3}{2}} = o\left(\frac{1}{(n^\frac{3}{2}h_n^2)}\right)
\]

35
It follows for Liapnovs condition \( \lim_{n \to \infty} \sum_i E \mid Q_{i,n}(x) \mid^3 \leq \frac{O(\frac{1}{n h_n^3})}{O(\frac{1}{n^2 h_n^2})} = O(\frac{1}{n^2 h_n^2}) = o(1) \)

From Lipunovs condition and the variance of \( \hat{F}(y \mid x) \) from the last theorem, it follows asymptotic normality

\[
\frac{\hat{F}(y \mid x) - E[\hat{F}(y \mid x)]}{\sqrt{[\hat{F}(y \mid x)]}} \overset{d}{\to} N(0, 1).
\]

Therefore,

\[
(nh_n)^{\frac{1}{2}}(\hat{F}(y \mid x) - F(y \mid x)) = N(0, \hat{F}(y \mid x) - F^2(y \mid x) \int K^2(u)du).
\]

**Corollary 2.3.1.**

Let the condition of the last theorem be satisfied and let \( nh_n^5 \to 0 \), for \( n \to \infty \).

Then it follows

\[
(nh_n)^{\frac{1}{2}}(\hat{F}(y \mid x) - F(y \mid x)) = N(0, \hat{F}(y \mid x) - F^2(y \mid x) \int K^2(u)du).
\]

**Proof.**

The last theorem gives the asymptotic normality \( (nh_n)^{\frac{1}{2}}(\hat{F}(y \mid x) - E[\hat{F}(y \mid x)]) \).

That we replace \( E[\hat{F}(y \mid x)] \) by \( F(y \mid x) \) to get \( (nh_n)^{\frac{1}{2}}(\hat{F}(y \mid x) - F(y \mid x)) \) converge to 0, as \( n \to \infty \)

\[
E[\hat{F}(y \mid x)] - F(y \mid x) = \frac{h_n^2}{2} \sum_i U_i K(U_i) F^{(2,0)}(y \mid x) + o(h_n^2)
\]

That is,

\[
(nh_n)^{\frac{1}{2}}E[\hat{F}(y \mid x)] - F(y \mid x) = (nh_n)^{\frac{1}{2}}O(h_n^2)
\]

\[
= O(nh_n^5)^{\frac{1}{2}}
\]

with \( nh_n^5 \to 0 \) for \( n \to \infty \), it follows the asymptotic normality \( (nh_n)^{\frac{1}{2}}(\hat{F}(y \mid x) - F(y \mid x)) \).
Above theorems deal with the estimator of the conditional distribution. Now the behavior of the estimator of the conditional median is analyzed. So assumed that 
\[ \hat{F}_x(\hat{m}(x)) = F_x(m(x)) = 0.5 \] is a unique and \{Y_i\} independent.

Now, let
\[
H_{n, \frac{1}{2}}(\theta(x)) = \sum_i \frac{1}{\sum_i K(\frac{x - X_i}{h_n})} K(\frac{x - X_i}{h_n}) \left[ \frac{1}{2} - I(Y_i \leq \theta(x)) \right]
\]
\[
= \sum_i H_{i, \frac{1}{2}}(\theta(x)).
\]

Using the central limit theorem,
\[
\frac{H_{n, \frac{1}{2}}(\theta(x)) - E[H_{n, \frac{1}{2}}(\theta(x))]}{\sqrt{\text{Var}[H_{n, \frac{1}{2}}(\theta(x))]} \to N(0, 1), \quad n \to \infty}
\]

With \(H_{n, \frac{1}{2}}(\theta(x))\) the mean squared error of \(\hat{m}_{NW}(x)\) can be calculated.

**Theorem 2.3.4.**

Let the condition of theorem (2.3.3) be satisfied and let \(\hat{F}_x(\hat{m}_{NW}(x)) = F_x(m(x)) = \frac{1}{2}\) be unique. Then it holds
\[
\text{MSE}[\hat{m}_{NW}(x)] = \left[ \frac{1}{2} h_n^2 \frac{F^{(2,0)}(m(x) \mid x)}{f(m(x) \mid x)} \int u^2 K(u)du \right]^2
\]
\[
+ \frac{1}{nh_n} \frac{1}{4} f^2(m(x)) \int K^2(u)du
\]

**Proof.**

By the Taylor expansion of the conditional distribution function of theorem (2.3.3) and \(\theta(x) = \hat{m}_{NW}(x)\) follows
\[
E(H_{n, \frac{1}{2}}(\hat{m}_{NW}(x))) \approx f(m(x) \mid x)[\hat{m}_{NW}(x) - m(x)]
\]
\[
+ \frac{1}{2} h_n^2 F^{(2,0)}(m(x)) \sum U_i^2 K(U_i) \sum_i K(U_i)
\]

and with integral approximation holds
\[ E(H_{n, \frac{1}{2}}(\hat{m}_{NW}(x))) \approx f(m(x) \mid x)[\hat{m}_{NW}(x) - m(x)] + \frac{1}{2} h_n^2 F^{(2,0)}(m(x)) \int u^2 K(u) du. \]

Now,
\[
Var[H_{n, \frac{1}{2}}(\hat{m}_{NW}(x))] = \frac{1}{\sum_i U_i^2 K(U_i)^2} \sum_i K^2(U_i)[F(\hat{m}_{NW}(x) \mid x - h_n U_i) - F^2(\hat{m}_{NW}(x) \mid x - h_n U_i)] \\
\approx \frac{1}{4[\sum_i K(U_i)^2]} \sum_i U_i^2 K^2(U_i) \\
\approx \frac{1}{4nh_n} \int K^2(u) du.
\]

\[ nh_n H_{n, \frac{1}{2}}(\hat{m}_{NW}(x)) \text{ is bounded random variable and } \sum_i Var(nh_n H_{i, \frac{1}{2}}(\hat{m}_{NW}(x))) \to \infty \text{ for } n \to \infty. \]

From this the asymptotic normality, it follows.
\[
\frac{nh_n[H_{n, \frac{1}{2}}(\hat{m}_{NW}(x)) - E[H_{n, \frac{1}{2}}(\hat{m}_{NW}(x))]]}{nh_n \sqrt{Var[H_{n, \frac{1}{2}}(\hat{m}_{NW}(x))]} \to N(0, 1), \quad n \to \infty.
\]

Since \( \hat{F}(\hat{m}_{NW}(x) \mid x) = \frac{1}{2}, H_{n, \frac{1}{2}}(\hat{m}_{NW}(x)) = 0. \) This implies for \( n \to \infty \)
\[
\frac{f(m(x) \mid x)[\hat{m}_{NW}(x) - m(x)] + \frac{1}{2} h_n^2 F^{(2,0)}m(x) \int u^2 K(u) du}{\sqrt{\frac{1}{4nh_n} \int K^2(u) du}} \to N(0, 1).
\]

From that bias and variance of \( \hat{m}_{NW}(x) \) can be calculated.

From the method of proof of theorem (2.3.4), asymptotic normality can be satisfied.

**Corollary 2.3.2.**

Let the condition of theorem (2.3.4) be satisfied and let \( nh_n^5 \to 0, \) for \( n \to \infty. \)

Then it holds
\[
(nh_n)^{\frac{3}{2}}(\hat{m}_{NW}(x) - m(x)) \overset{d}{\to} N(\frac{1}{2}(nh_n^5)^{\frac{1}{2}} f^{(2,0)}(m(x) \mid x) \int u^2 K(u) du, \frac{1}{4f(m(x) \mid x) \int K^2(u) du}^2)
\]
Proof. :

Since,
\[
f(m(x) | x)[\hat{m}_{NW}(x) - m(x)] + \frac{1}{2} h_n^2 F^{(2,0)}(m(x)) \int u^2 K(u) du \xrightarrow{d} N(0, 1).
\]

Then
\[
(nh_n)^{\frac{3}{2}} E[(\hat{m}_{NW}(x) - m(x))] \rightarrow \frac{1}{2} (nh_n)^{\frac{3}{2}} F^{(2,0)}(m(x)) \int u^2 K(u) du.
\]

Now, we have
\[
Var]\left[ \frac{f(m(x) | x)[\hat{m}_{NW}(x) - m(x)] + \frac{1}{2} h_n^2 F^{(2,0)}(m(x)) \int u^2 K(u) du}{\sqrt{\frac{1}{4nh_n} \int K^2(u) du}} \right] \rightarrow 1.
\]

This implies that
\[
(nh_n)Var[(\hat{m}_{NW}(x) - m(x))] \rightarrow \frac{1}{4f^2(m(x) | x)} \int K^2(u) du.
\]

From above we have,
\[
(nh_n)^{\frac{3}{2}} (\hat{m}_{NW}(x) - m(x)) \xrightarrow{d} N\left( \frac{1}{2} (nh_n)^{\frac{3}{2}} F^{(2,0)}(m(x) | x) \int u^2 K(u) du, \frac{1}{4f^2(m(x) | x)} \int K^2(u) du \right).
\]

Since \( nh_n^5 \rightarrow 0 \). Then the proof is completed. \( \square \)

Through the studying the NW estimator, the large bias and boundary effects is the most defect of NW. Hence, in the next chapter, we will study another estimator, which is know as a double kernel (DK) estimator.
Chapter 3

The Double Kernel Estimator

Introduction

Let \((X, Y)\) be a two dimensional random variable with a joint distribution function \(F(x, y)\). The large bias and boundary effects are considered to be the most important defect in the case of the NW estimator. The NW estimator was treated and modified in order to obtain some more refinement estimator, which is called the double kernel estimator (DK), see Fan, J., T. C. Hu and Y.K.Troung,(1994). This chapter studies the double kernel estimation of the conditional median of \(Y\) for a given value of \(X\) based on a random sample from the above distribution. In this chapter, the joint asymptotic normality of the conditional median estimated at a finite number of distinct points is established under some regularity conditions.

The aim of this chapter is to summarize the Double Kernel estimator (DK) and its aspects, discussion the properties of the \(\hat{m}_{DK}(x)\), the DK estimator of the conditional median function \(m(x)\) and studying the asymptotic normality of the \(\hat{m}_{DK}(x)\). This chapter consists of three sections. In Section 3.1, we will introduce the DK estimator of the conditional distribution function \(\hat{F}_{DK}(y \mid x)\) and use it to introduce \(\hat{m}_{DK}(x)\) based on data \(\{(x_i, y_i)\}_{i=1}^n\). In Section 3.2, some preliminaries theorems and
lemmas that will help us to study the asymptotic properties of the \( \hat{m}_{DK}(x) \). Finally, in Section 3.3, we will investigate the asymptotic normality of the \( \hat{m}_{DK}(x) \).

### 3.1 The Double Kernel Estimator

If \( f(x, y) \) is the joint pdf of the random variables \( X \) and \( Y \) at \( (x, y) \), and \( g(x) \) is the marginal pdf of \( X \) at \( x \), the conditional pdf of \( Y \) given \( X = x \) is given by

\[
f(y|x) = \frac{f(x, y)}{g(x)}, \quad g(x) > 0,
\]

for each \( y \) within the range of \( Y \), and then

\[
F(y|x) = \int_{-\infty}^{y} f(u|x) du
\]

Now, we introduce the basic equation of the kernel conditional distribution function estimation (KCDF). The Kernel function \( K(u) \) is assumed to be the kernel function, \( h_n = h \) is a sequence of positive numbers converging to zero.

Standard kernel estimators of \( f(x, y) \), \( g(x) \), \( f(y|x) \) are

\[
\hat{f}(x, y) = (nh^2)^{-1} \sum_{i=1}^{n} K\left( \frac{x - X_i}{h} \right) K\left( \frac{y - Y_i}{h} \right)
\]

\[
\hat{g}(x) = (nh)^{-1} \sum_{i=1}^{n} K\left( \frac{x - X_i}{h} \right)
\]

and

\[
\hat{f}(y|x) = \frac{\hat{f}(x, y)}{\hat{g}(x)}
\]

\[
= \frac{(nh^2)^{-1} \sum_{i=1}^{n} K\left( \frac{x - X_i}{h} \right) K\left( \frac{y - Y_i}{h} \right)}{(nh)^{-1} \sum_{i=1}^{n} K\left( \frac{x - X_i}{h} \right)}
\]

\[
= \frac{\sum_{i=1}^{n} K\left( \frac{x - X_i}{h} \right) K\left( \frac{y - Y_i}{h} \right)}{h \sum_{i=1}^{n} K\left( \frac{x - X_i}{h} \right)}.
\]
Definition 3.1.1. The double kernel estimator of the conditional distribution function $F(x, y)$ is defined as:

$$
\hat{F}_{DK}(y \mid x) = \int_{-\infty}^{y} \hat{f}(u \mid x) du = \frac{B_n(x, y)}{\hat{g}(x)}
$$

where,

$$
B_n(x, y) = \frac{1}{nh_n} \sum_{i=1}^{n} K \left( \frac{x-X_i}{h_n} \right) \hat{K} \left( \frac{y-Y_i}{h_n} \right), \quad \hat{K}(y) = \int_{-\infty}^{y} K(u) du.
$$

and $K$ is a probability density function, $h_n$ is a sequence of positive numbers converging to zero.

Definition 3.1.2. The double kernel estimator of the conditional median $m(x)$ is defined as:

$$
\hat{m}_{DK}(x) = \inf\{y \in \mathbb{R} : \hat{F}_{DK}(y \mid x) \geq 0.5\}.
$$

3.2 Preliminaries

In this section, some regularity conditions, notations and basic theorems that will be used in proving the lemmas and the main theorem. Also, four lemmas which are necessary to prove the main theorem in after section.

First, we shall assume the following conditions:

1. $F^{(i,j)}(x, y) = \frac{\partial^{i+j} F(x, y)}{\partial x^i \partial y^j}$ exist and are bounded for $(i, j) = (1, 2), (2, 0), (2, 1), (3, 0)$.

2. $g^{(i)}(x) = \int_{-\infty}^{\infty} \frac{\partial^i f(x, y)}{\partial x^i} dy$ exist for $i = 1, 2$.

3. Both $h(x) = \int_{-\infty}^{\infty} | \frac{\partial f(x, y)}{\partial x} | dy$ and $g^{(2)}(x, y)$ are bounded.

4. The conditional population median $m(x)$ are unique and defined by

$$
F(m(x) \mid x) = \frac{F^{(1,0)}(x, m(x))}{g(x)} = 0.5.
$$
5. $K(u)$ is a function of bounded variation.

6. $\int_{-\infty}^{\infty} uK(u)du = 0$.

7. $\int_{-\infty}^{\infty} u^2K(u)du < \infty$.

8. $h_n = n^{-\delta}, \frac{1}{5} < \delta \leq \frac{1}{4}$.

Now, some notations, which are needed in the remaining of this chapter, are introduced.

Define for, $i = 1, 2$. For, $j = 1, 2, \ldots, n$, the following:

$$U_{nj}^*(x_i) = \frac{1}{h_n} K \left[ \frac{x_i - X_j}{h_n} \right],$$

$$V_{nj}^*(x_i) = \frac{1}{h_n} \hat{K} \left[ \frac{m(x_i) - Y_j}{h_n} \right] K \left[ \frac{x_i - X_j}{h_n} \right],$$

$$U_{nj}(x_i) = h_n^{\frac{1}{2}} \left[ U_{nj}^*(x_i) - E\{U_{nj}^*(x_i)\} \right],$$

$$V_{nj}(x_i) = h_n^{\frac{1}{2}} \left[ V_{nj}^*(x_i) - E\{V_{nj}^*(x_i)\} \right],$$

$$U_n(x_i) = \sum_{j=1}^{n} U_{nj}(x_i),$$

$$V_n(x_i) = \sum_{j=1}^{n} V_{nj}(x_i),$$

$$W_{nj} = \begin{bmatrix} U_{nj}(x_1) \\ U_{nj}(x_2) \\ V_{nj}(x_1) \\ V_{nj}(x_2) \end{bmatrix}, \quad n^\frac{1}{2} Z_n = \begin{bmatrix} U_n(x_1) \\ U_n(x_2) \\ V_n(x_1) \\ V_n(x_2) \end{bmatrix},$$

$$w(x_i) = F^{(1.0)}(x_i, m(x_i)), \quad n^\frac{1}{2} Z_n^* = h_n^{\frac{1}{2}} \begin{bmatrix} \sum_{j=1}^{n} [U_{nj}^*(x_1) - g(x_1)] \\ \sum_{j=1}^{n} [U_{nj}^*(x_2) - g(x_2)] \\ \sum_{j=1}^{n} [V_{nj}^*(x_1) - w(x_1)] \\ \sum_{j=1}^{n} [V_{nj}^*(x_2) - w(x_2)] \end{bmatrix}.$$ 

Now, we state Bochner theorem, Liapnouv’s theorem, Cramer- Wold theorem and Slutsky theorem, since they play an important role in this chapter.

**Theorem 3.2.1.** (Bochner Theorem).

Suppose $K(y)$ is a Borel function satisfying the following conditions:
1. \( \text{Sup}_{-\infty < y < \infty} |K(y)| < \infty \).

2. \( \int_{-\infty}^{\infty} |K(y)| \, dy < \infty \).

3. \( \lim_{y \to \infty} |yK(y)| \, dy = 0 \).

Let \( g(y) \) satisfying \( \int_{-\infty}^{\infty} |g(y)| \, dy < \infty \), and let \( h_n \) be a sequence of positive constants satisfying \( \lim_{n \to \infty} h_n = 0 \).

Define \( g_n(x) = \frac{1}{h_n} \int_{-\infty}^{\infty} K\left(\frac{y}{h_n}\right)g(x-y) \, dy \). Then at every point \( x \) of continuity of \( g(\cdot) \),

\[
\lim_{n \to \infty} g_n(x) = g(x) \int_{-\infty}^{\infty} K(y) \, dy.
\]

**Proof.** This theorem is Theorem 1A in Parzen (1962).

---

**Theorem 3.2.2.** (Liapnou’s Theorem).

Let \( X_k, k \geq 1 \), be independent random variables such that \( EX_k = \mu_k \) and \( VarX_k = \sigma_k^2 \), and for some \( 0 < \delta \leq 1 \), \( \nu_{2+\delta}(k) = E |X_k - \mu_k|^{2+\delta} < \infty, k \geq 1 \).

Also, let \( T_n = \sum_{k=1}^{n} X_k \), \( \xi_n = ET_n = \sum_{k=1}^{n} \mu_k \), \( s_n^2 = VarT_n = \sum_{k=1}^{n} \sigma_k^2 \), \( Z_n = (T_n - \xi_n)/s_n \), and \( \rho_n = s_n^{-(2+\delta)} \sum_{k=1}^{n} \nu_{2+\delta}^{(k)} \). Then if \( \lim_{n \to \infty} \rho_n = 0 \), we have \( Z_n \xrightarrow{D} N(0,1) \).

**Proof.** This theorem is Theorem 3.3.2 in Pranab (1993, p.108).

---

**Theorem 3.2.3.** (Cramer-Wold Theorem).

Let \( X, X_1, X_2, \ldots \) be a random vectors in \( \mathbb{R}^p \); then \( X_n \xrightarrow{D} X \), if and only if, for every fixed \( c \in \mathbb{R}^p \), we have \( c^TX_n \xrightarrow{D} c^TX \).

**Proof.** This theorem is theorem 3.2.4 in Pranab (1993, p.106).

---

**Theorem 3.2.4.** (Slutsky Theorem).

Let \( \{X_n\} \) and \( \{Y_n\} \) be sequences of random \( p- \) vectors such that \( X_n \xrightarrow{D} X \) and \( Y_n \xrightarrow{P} 0 \). Also let \( \{W_n\} \) be sequence of random \( \{w \times p\} \) matrices such that \( tr\{(W_n - W)^T(W_n - W)\} \xrightarrow{P} 0 \), where \( W \) is a nonstochastic matrix. Then
(i) $X_n \pm Y_n \xrightarrow{D} X$.

(ii) $W_nX_n \xrightarrow{D} WX$.

**Proof.** This theorem is Theorem 3.4.3 in Pranab (1993, p.130).

**Theorem 3.2.5.** Let $\{T_n\}$ be a sequence of random of $p-$ vectors such that $\sqrt{n}(T_n - \theta) \rightarrow N(0, \Sigma)$ and consider a vector- valued function $g : \mathbb{R}^p \rightarrow \mathbb{R}^q$ such that $G(\theta) = \partial / \partial x^T g(x) \big|_{x=\theta}$ exist. Then $\sqrt{n}\{g(T_n) - g(\theta)\} \xrightarrow{D} N(0, G(\theta)\Sigma G(\theta)^T)$.

**Proof.** This theorem is Theorem 3.4.6 in Pranab (1993, p.136).

To prove the main theorem, the following four lemmas are needed:

**Lemma 3.2.1.** Under the conditions 1,2,3,5,6,7 and 8 the following hold:

i. $\lim_{n \rightarrow \infty} E\{U_{nj}^2(x_i)\} = g(x_i) \int_{-\infty}^{\infty} K^2(u) du, i = 1, 2.$

ii. $\lim_{n \rightarrow \infty} E\{V_{nj}^2(x_i)\} = w(x_i) \int_{-\infty}^{\infty} K^2(u) du, i = 1, 2.$

iii. $\lim_{n \rightarrow \infty} E\{U_{nj}(x_i)V_{nj}(x_s)\} = w(x_i) \int_{-\infty}^{\infty} K^2(u) du, i = s = 1, 2.$

iv. $\lim_{n \rightarrow \infty} E\{U_{nj}(x_i)V_{nj}(x_s)\} = 0, i \neq s, i = 1, 2, s = 1, 2.$

v. $\lim_{n \rightarrow \infty} E\{V_{nj}(x_i)V_{nj}(x_s)\} = 0, i \neq s, i = s = 1, 2.$

**Proof.** The proof of this lemma is obtained by using Bochner theorem.

\[ i.EU_{nj}^2(x_i) = h_n \frac{1}{h_n} \int_{-\infty}^{\infty} K^2(u) \left( \frac{x_i - u}{h_n} \right) g(u) du - \left( \frac{1}{h_n} \int_{-\infty}^{\infty} K \left( \frac{x_i - u}{h_n} \right) g(u) du \right)^2 \]

\[ \lim_{n \rightarrow \infty} \{EU_{nj}^2(x_i)\} = \lim_{n \rightarrow \infty} \frac{1}{h_n} \int_{-\infty}^{\infty} K^2(\frac{x_i - u}{h_n}) g(u) du - \lim_{n \rightarrow \infty} h_n \frac{1}{h_n} \int_{-\infty}^{\infty} K \left( \frac{x_i - u}{h_n} \right) g(u) du^2 \]

\[ = g(x_i) \int_{-\infty}^{\infty} K^2(u) du - 0 = g(x_i) \int_{-\infty}^{\infty} K^2(u) du, \]
By an application of Bochner theorem.

ii. Using the definition of $V_{nj}(x_i)$ and Bochner Theorem, we obtain,

$$\lim_{n \to \infty} \left\{ EV_{nj}^2(x_i) \right\} = \lim_{n \to \infty} \frac{1}{h_n^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{K}^2 \left( \frac{m(x_i) - v}{h_n} \right) K^2 \left( \frac{x_i - u}{h_n} \right) f(u, v) dudv$$

$$= \lim_{n \to \infty} \frac{1}{h_n^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{K}^2 \left( \frac{m(x_i) - v}{h_n} \right) f(v | x_i - h_n u) dv K^2(u) g(x_i - h_n u) du$$

$$= g(x_i) \int_{-\infty}^{\infty} K^2(u) du. \lim_{n \to \infty} \frac{1}{h_n^2} \int_{-\infty}^{\infty} \hat{K}^2 \left( \frac{m(x_i) - v}{h_n} \right) f(v | x_i - h_n u) dv$$

$$= g(x_i) \int_{-\infty}^{\infty} K^2(u) du. \int_{-\infty}^{\infty} f(v | x_i) dv$$

$$= g(x_i) F(m(x_i) | x_i) \int_{-\infty}^{\infty} K^2(u) du.$$

iii. and iv. Similarly as in part(ii), we have,

$$\lim_{n \to \infty} h_n E\{U_{nj}(x_i)V_{nj}(x_i)\} = \lim_{n \to \infty} \frac{1}{h_n^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{K}^2 \left( \frac{m(x_i) - v}{h_n} \right) K^2 \left( \frac{x_i - u}{h_n} \right) f(u, v) dudv$$

$$= g(x_i) F(m(x_i) | x_i) \int_{-\infty}^{\infty} K^2(u) du.$$

On the other hand,

$$\lim_{n \to \infty} h_n E\{U_{nj}(x_i)V_{nj}(x_i)\} = \lim_{n \to \infty} \frac{1}{h_n^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{K}^2 \left( \frac{m(x_i) - v}{h_n} \right) K \left( \frac{x_i - u}{h_n} \right) K \left( \frac{x_i - u}{h_n} \right) f(u, v) dudv$$

$$= 0,$$

Since

$$\lim_{n \to \infty} \frac{1}{h_n} \int_{-\infty}^{\infty} K \left( \frac{x_i - u}{h_n} \right) K \left( \frac{x_i - u}{h_n} \right) g(u) du = 0.$$

v. Similarly as in part(iii).
Lemma 3.2.2. Under the conditions 1,2,3,5,6,7 and 8, \( Z_n \) as \( n \to \infty \) converge in distribution to a four dimensional random variable with the zero mean vector and a covariance matrix, \( A \), where,

\[
A = \int_{-\infty}^{\infty} K^2(u) du. \begin{bmatrix}
g(x_1) & 0 & w(x_1) & 0 \\
0 & g(x_2) & 0 & w(x_2)
w(x_1) & 0 & w(x_1) & 0 \\
0 & w(x_2) & 0 & w(x_2)
\end{bmatrix}
\]

Proof.

Let \( \mathbf{c} = (c_1, c_2, c_3, c_4)^T, \mathbf{c} \not= 0 \). For \( j = 1, 2, ..., n \) define,

\[
\sigma_{nj}^2 = \text{Var}\{c^TW_{nj}\}, \quad \rho_{nj}^3 = E\left|n^{-\frac{1}{2}}\{c^TW_{nj}\}\right|^3.
\]

\[
\sigma_n^2 = \frac{1}{n} \sum_{j=1}^{n} \sigma_{nj}^2, \quad \rho_n^3 = \sum_{j=1}^{n} \rho_{nj}^3.
\]

Note that \( A \) is a positive definite matrix whenever \( g(x_i) > 0 \) and \( w(x_i) > 0, i = 1, 2 \).

Using lemma 3.2.6, we obtain for \( j = 1, 2, ..., n \)

\[
\lim_{n \to \infty} \sigma_{nj}^2 = \lim_{n \to \infty} \text{Var}\{c_1U_{nj}(x_1) + c_2U_{nj}(x_2) + c_3V_{nj}(x_1) + c_4V_{nj}(x_2)\}
\]

\[
= \int_{-\infty}^{\infty} K^2(u) du\{c_1^2g(x_1) + 2c_1c_3w(x_1) + c_2^2g(x_2) + 2c_2c_4w(x_2) + c_3^2w(x_1) + c_4^2w(x_2)\}
\]

\[
= (c_1, c_2, c_3, c_4) \int_{-\infty}^{\infty} K^2(u) du. \begin{bmatrix}
g(x_1) & 0 & w(x_1) & 0 \\
0 & g(x_2) & 0 & w(x_2)
w(x_1) & 0 & w(x_1) & 0 \\
0 & w(x_2) & 0 & w(x_2)
\end{bmatrix} \begin{bmatrix}
c_1 \\
c_2 \\
c_3 \\
c_4
\end{bmatrix} = \mathbf{c}^T \mathbf{A} \mathbf{c} > 0,
\]

Since \( \mathbf{c}^T \mathbf{A} \mathbf{c} \) is quadratic from associated with positive definite matrix \( \mathbf{A} \).

Therefore by the definition of \( \sigma_n^2 \) the following holds:

\[
\lim_{n \to \infty} \sigma_n^2 = \mathbf{c}^T \mathbf{A} \mathbf{c} > 0.
\]

Now, using computations similar to those in lemma 3.2.6 implies that:

\[
E\left|U_{n1}(x_i)\right|^3 = O(h_n^{-\frac{1}{2}}), \quad E\left|V_{n1}(x_i)\right|^3 = O(h_n^{-\frac{1}{2}}), i = 1, 2.
\]

47
This implies that:

\[
\rho_n^3 = E \left| n^{-\frac{3}{2}} c^T W_{nj} \right|^3 \leq \left| c^T \right|^3 n^{-\frac{3}{2}} E \left| W_{nj} \right|^3 \leq 4^\frac{3}{2} \left| c^T \right|^3 n^{-\frac{3}{2}} \max \{ E \left| U_{n1}(x_i) \right|^3, E \left| V_{n1}(x_i) \right|^3, i = 1, 2. \}
\]

Therefore, from the definition of \( \rho_n^3 \), the following holds:

\[
\rho_n^3 \leq 4^\frac{3}{2} n \left| c^T \right|^3 n^{-\frac{3}{2}} \max \{ E \left| U_{n1}(x_i) \right|^3, E \left| V_{n1}(x_i) \right|^3, i = 1, 2. \} = O((nh_n)^{-\frac{1}{2}}).
\]

Now, an application of condition (8), implies that:

\[
\lim_{n \to \infty} \rho_n^3 = 0.
\]

Thus a combination of above equation and \( \lim_{n \to \infty} \sigma_n^2 = c^T A c > 0 \), implies that

\[
\lim_{n \to \infty} \frac{\rho_n^3}{\sigma_n^2} = 0.
\]

Next, an application of Liapnouv’s theorem, implies that \( c^T Z_n = n^{-\frac{1}{2}} \sum_{j=1}^{n} c^T W_{nj} \) converges in distribution to univariate normal random variable with zero mean and variance \( c^T A c \).

Let \( Z \) be a four dimensional random variable with zero mean and covariance matrix \( A \). Then \( c^T Z \) is a univariate normal random variable with zero mean and variance \( c^T A c \).

From above we have \( c^T Z_n \xrightarrow{D} c^T Z \) in distribution. Now, an application of the Cramer-Wold theorem implies that \( Z_n \xrightarrow{D} Z \) in distribution. This completes the proof of the lemma.

Now, the definition of the notation \( o_p(1) \) is given, since it is needed in Lemma 3.2.9 and Lemma 3.2.10.

**Definition 3.2.1.** For a sequence \( \{X_n\} \) of random variable, if for every \( \eta > 0, \varepsilon > 0 \), there exist a positive integer \( n(\varepsilon, \eta) \), such that

\[
p\{ | X_n | > \eta \} < \varepsilon, \quad n \geq n(\varepsilon, \eta),
\]

48
then we say that \( X_n = O_p(1) \). In other words, \( X_n = O_p(1) \) is equivalent to saying that \( X_n \to 0 \) in probability. The definition extends directly to the vector case by adapting the Euclidean norm. Here, we write \( X_n = o_p(1) \), see Pranab (1993, pp.37).

**Lemma 3.2.3.** Under conditions 1,2,3,5,6,7 and 8, \( Z_n^* \) as \( n \to \infty \) converges in distribution to a four dimensional normal random variable with zero mean vector and a covariance matrix, \( A \).

**Proof.** let:

\[
C_n = \begin{bmatrix}
E\{U_{n1}^*(x_2)\} - g(x_1) \\
E\{U_{n2}^*(x_1)\} - g(x_2) \\
E\{V_{n1}^*(x_1)\} - w(x_1) \\
E\{V_{n2}^*(x_2)\} - w(x_2)
\end{bmatrix}
\]

By Taylor expansion and conditions 6 and 7, we get that:

\[
E\{U_{n1}^*(x_1)\} - g(x_1) = \frac{1}{h_n} \int_{-\infty}^{\infty} K\left(\frac{x_1 - u}{h_n}\right)g(u)du - g(x_1)
\]

\[= \int_{-\infty}^{\infty} K(u)g(x_1 - h_n u)du - g(x_1)
\]

\[= \int_{-\infty}^{\infty} K(u)\{g(x_1) - h_n u g^{(1)}(x_1) + \frac{h_n^2 u^2}{2} g^{(2)}(x_1) + o(h_n^2)\}du - g(x_1)
\]

\[= \frac{h_n^2}{2} g^{(2)}(x_1) \int_{-\infty}^{\infty} u^2 K(u)du \leq C h_n^2 = O(h_n^2).
\]

Similarly, for the other.

From the above and condition 8, the following holds:

\[ (nh_n)^{\frac{1}{2}} C_n = (nh_n)^{\frac{1}{2}} \begin{bmatrix}
O(h_n^2) \\
O(h_n^2) \\
O(h_n^2) \\
O(h_n^2)
\end{bmatrix} = \begin{bmatrix}
(nh_n^2)^{\frac{1}{2}} \\
(nh_n^2)^{\frac{1}{2}} \\
(nh_n^2)^{\frac{1}{2}} \\
(nh_n^2)^{\frac{1}{2}}
\end{bmatrix} \to \begin{bmatrix}
0 \\
0 \\
0 \\
0
\end{bmatrix}.
\]

Now, from the definition of \( Z_n^* \) and \( Z_n \), we obtain

\[ Z_n^* = Z_n + (Z_n^* - Z_n) = Z_n + (nh_n)^{\frac{1}{2}} C_n.
\]

This implies that \( Z_n^* - Z_n = o_p(1) \).
Lemma 3.2.4. Under the conditions, 1-5, 7, and 8 if \( g(x_i) > 0 \) then as \( n \to \infty \)

\[
\hat{f}(\xi_i \mid x_i) = f(m(x_i) \mid x_i) + o_p(1), \quad i = 1, 2.
\]

Proof. This lemma is lemma 6 in Sammanta (1989).

3.3 Asymptotic Properties

In this section, we will study some asymptotic properties of the \( \hat{m}_{DK}(x) \), the two next theorems are the main theorems in this section.

Theorem 3.3.1. Under the condition 1-8, if \( g(x_i) > 0 \) and \( f(x_i, m(x_i)) > 0 \), \( i = 1, 2 \)
then \( (nh_n)^{-1/2}(\hat{m}_{DK}(x_1), \hat{m}_{DK}(x_2))^T \) is asymptotically normally distributed with mean vector \( (m(x_1), m(x_2))^T \) and a diagonal covariance matrix

\[
B = \int_{-\infty}^{\infty} K^2(u)du \begin{bmatrix}
\frac{1}{4g(x_1)f^2(m(x_1) \mid x_1)} & 0 \\
0 & \frac{1}{4g(x_2)f^2(m(x_2) \mid x_2)}
\end{bmatrix}
\]

Proof. Let \( H : R^4 \to R^2 \), defined by:

\[
H(y) = \begin{pmatrix}
y_3 \\
y_4
\end{pmatrix}^T, \quad y = (y_1, y_2, y_3, y_4).
\]

Then the matrix of partial derivatives of \( H \) is given by:

\[
\frac{\partial H(y)}{\partial y} = \begin{bmatrix}
-\frac{y_3}{y_1^2} & 0 & \frac{1}{y_1} & 0 \\
0 & -\frac{y_3}{y_2^2} & 0 & \frac{1}{y_2}
\end{bmatrix}.
\]

Let \( \theta = (g(x_1), g(x_2), w(x_1), w(x_2))^T, \quad T_n = (T_{n1}, T_{n2}, T_{n3}, T_{n4})^T \),
where

\[
T_{ni} = \begin{cases}
\frac{1}{n} \sum_{j=1}^n U_{nj}^*(x_i), \quad i = 1, 2; \\
\frac{1}{n} \sum_{j=1}^n V_{nj}^*(x_i), \quad i = 3, 4.
\end{cases}
\]
This implies that:

\[ H(\theta) = \left( \frac{w(x_1)}{g(x_1)}, \frac{w(x_2)}{g(x_2)} \right)^T. \]

Let \( D \) denotes the matrix of partial derivatives of \( H \), evaluated at \( \theta \). Then we obtain that:

\[
D = \begin{bmatrix}
-\frac{w(x_1)}{g^2(x_1)} & 0 & \frac{1}{g(x_1)} & 0 \\
0 & -\frac{w(x_2)}{g^2(x_2)} & 0 & \frac{1}{g(x_2)} \\
\end{bmatrix}.
\]

Now, \( Z^*_n \) can be written as:

\[ Z^*_n = \left( nh_n, \frac{1}{\sqrt{n}} \right)^T (T_n - \theta) \]

Therefore by application of lemma 3.2.9 and Theorem 3.2.4, we conclude that:

\[
(nh_n)^{\frac{1}{2}} [H(T_n) - H(\theta)] = (nh_n)^{\frac{1}{2}} \left[ \hat{F}(m(x_1) \mid x_1) - F(m(x_1) \mid x_1) \right] = X_n \xrightarrow{D} \mathbf{X},
\]

Where \( \mathbf{X} \) is bivariate normal random variable with zero mean vector and a covariance matrix, \( \mathbf{DAD}^T \),

where:

\[
\mathbf{DAD}^T = \begin{bmatrix}
-\frac{w(x_1)}{g^2(x_1)} & 0 & \frac{1}{g(x_1)} & 0 \\
0 & -\frac{w(x_2)}{g^2(x_2)} & 0 & \frac{1}{g(x_2)} \\
\end{bmatrix}.
\]

\[
\int_{-\infty}^{\infty} K^2(u)du \begin{bmatrix}
g(x_1) & 0 & w(x_1) & 0 \\
0 & g(x_2) & 0 & w(x_2) \\
w(x_1) & 0 & w(x_1) & 0 \\
0 & w(x_2) & 0 & w(x_2) \\
\end{bmatrix} \begin{bmatrix}
-\frac{w(x_1)}{g^2(x_1)} & 0 \\
0 & -\frac{w(x_2)}{g^2(x_2)} \\
\frac{1}{g(x_1)} & 0 \\
0 & \frac{1}{g(x_2)} \\
\end{bmatrix}.
\]

\[
= \int_{-\infty}^{\infty} K^2(u)du \begin{bmatrix}
\frac{1}{g(x_1)} & \frac{w(x_1)}{g(x_1)} \left( 1 - \frac{w(x_1)}{g(x_1)} \right) & 0 \\
0 & \frac{1}{g(x_2)} & \frac{w(x_2)}{g(x_2)} \left( 1 - \frac{w(x_2)}{g(x_2)} \right) \\
\end{bmatrix}.
\]

\[
= \int_{-\infty}^{\infty} K^2(u)du \begin{bmatrix}
0.25 & 0 \\
0 & 0.25 \\
\end{bmatrix}.
\]
since by condition 4, we have:
\[
\frac{w(x_i)}{g(x_i)} = \frac{F^{(1,0)}(x_i, m(x_i))}{g(x_i)} = 0.5, \quad i = 1, 2.
\]
Now using Taylor expansion of \( \hat{F}(\hat{m}_{DK}(x_i) \mid x_i) \) around \( m(x_i), i = 1, 2 \), the following holds,
\[
F(m(x_i) \mid x_i) \approx \hat{F}(\hat{m}_{DK}(x_i) \mid x_i) \approx \hat{F}(m(x_i) \mid x_i) + (\hat{m}_{DK}(x_i - m(x_i))) \hat{f}(\xi \mid x_i),
\]
where \( \xi_i \) is some random point between \( \hat{m}_{DK}(x_i) \) and \( m(x_i), i = 1, 2 \).
\[
\hat{m}_{DK}(x_i) - m(x_i) \approx \frac{F(m(x_i) \mid x_i) - \hat{F}(m(x_i) \mid x_i)}{\hat{f}(\xi \mid x_i)}, \quad i = 1, 2.
\]
Therefore,
\[
(nh_n)^{1/2}(\hat{m}_{DK}(x_i) - m(x_i)) \approx (nh_n)^{1/2}\left\{ \frac{\hat{F}(m(x_i) \mid x_i) - \hat{F}(\hat{m}_{DK}(x_i) \mid x_i)}{\hat{f}(\xi \mid x_i)} \right\}.
\]
Above equation implies that:
\[
(nh_n)^{1/2} \left[ \begin{array}{c}
\hat{m}_{DK}(x_1) - m(x_1) \\
\hat{m}_{DK}(x_2) - m(x_2)
\end{array} \right] \approx (nh_n)^{1/2} \left[ \begin{array}{c}
\frac{\hat{F}(m(x_1) \mid x_1) - \hat{F}(\hat{m}_{DK}(x_1) \mid x_1)}{\hat{f}(\xi_1 \mid x_1)} \\
\frac{\hat{F}(m(x_2) \mid x_2) - \hat{F}(\hat{m}_{DK}(x_2) \mid x_2)}{\hat{f}(\xi_2 \mid x_2)}
\end{array} \right] = W_nX_n,
\]
where,
\[
W_n = \left[ \begin{array}{cc}
\frac{1}{\hat{f}(\xi_1 \mid x_1)} & 0 \\
0 & \frac{1}{\hat{f}(\xi_2 \mid x_2)}
\end{array} \right].
\]
Now let:
\[
W = \left[ \begin{array}{cc}
\frac{1}{\hat{f}(\xi_1 \mid x_1)} & 0 \\
0 & \frac{1}{\hat{f}(\xi_2 \mid x_2)}
\end{array} \right].
\]
Then by lemma 3.2.10 we get that \( \text{tr}\{(W_n - W)^T(W_n - W)\} \overset{P}{\longrightarrow} 0 \)
Next, an application of Slutsky theorem implies that:
\[
(nh_n)^{1/2} \left[ \begin{array}{c}
\hat{m}_{DK}(x_1) - m(x_1) \\
\hat{m}_{DK}(x_2) - m(x_2)
\end{array} \right] = (W_nX_n) \overset{D}{\longrightarrow} WX,
\]
52
which completes the proof of the theorem.

**Theorem 3.3.2.** Under conditions 1,5,7 and 8, if \( g(x) > 0 \) then,

\[
\lim_{n \to \infty} \hat{m}_{DK}(x) = m(x), i = 1, 2,
\]

with probability one.

**Proof.**

Since \( \hat{m}_{DK}(x) \) is unique. Then for every \( \epsilon > 0 \) there exist an \( \delta = \eta(\epsilon) > 0 \) defined by

\[
\eta(\epsilon) = \min \{ F(m(x) + \epsilon \mid x) - F(m(x) \mid x), F(m(x) \mid x) - F(m(x) - \epsilon \mid x) \}
\]

Such that \( |\hat{m}_{DK}(x) - m(x)| > \epsilon \) implies that \( |\hat{F}(\hat{m}_{DK}(x) \mid x) - F(m(x) \mid x)| > \eta(\epsilon) \).

Expanding \( \hat{F}(\hat{m}_{DK}(x) \mid x) \) around \( m(x) \) to get

\[
F(m(x)) = \frac{1}{2} = \hat{F}(\hat{m}_{DK}(x) \mid x) = \hat{F}(m(x) \mid x) + (\hat{m}_{DK}(x) - m(x)) \hat{f}(m \mid x).
\]

Where \( m \) is some random point between \( \hat{m}_{DK}(x) \) and \( m(x) \).

Hence

\[
\hat{m}_{DK}(x) - m(x) = \frac{F(m(x) \mid x) - \hat{F}(m(x) \mid x)}{\hat{f}(m \mid x)},
\]

and so

\[
(nh_n)^{\frac{1}{2}}(\hat{m}_{DK}(x) - m(x)) = \frac{-(nh_n)^{\frac{1}{2}} \hat{F}(m(x) \mid x) - F(m(x) \mid x)}{\hat{f}(m \mid x)}.
\]

Since \( \lim_{n \to \infty} \hat{F}(m(x) \mid x) = F(m(x) \mid x) \), then we have the result.
Chapter 4

Comparison and Applications

In this chapter we compare the NW and DK estimators using three simulated data and one application of real data.

According to the second and third chapter, the NW estimator suffers from large bias and boundary effects, while the DK estimator combines the bias reduction and has no boundary affects, at the same time, it preserves the NW properties see the following table.

<table>
<thead>
<tr>
<th>NW estimator</th>
<th>DK estimator</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{F}_{NW}(y</td>
<td>x) = \sum_{i=1}^{n} \frac{K\left(\frac{x-X_i}{h_n}\right)I\left(Y_i \leq y\right)}{\sum_{i=1}^{n} K\left(\frac{X_i}{h_n}\right)} )</td>
</tr>
<tr>
<td>( \hat{m}<em>{NW}(x) = \inf{y \in \mathbb{R} : \hat{F}</em>{NW}(y \mid x) \geq 0.5} )</td>
<td>( \hat{m}<em>{DK}(x) = \inf{y \in \mathbb{R} : \hat{F}</em>{DK}(y \mid x) \geq 0.5} )</td>
</tr>
</tbody>
</table>

Table 4.1: The Comparison Between The NW and The RNW estimators

where,

\( B_n(x,y) = \frac{1}{n h_n} \sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right)\hat{K}\left(\frac{Y_i}{h_n}\right), \quad \hat{K}(y) = \int_{-\infty}^{y} K(u) du. \) and \( K \) is a probability density function, \( h_n \) is a sequence of positive numbers converging to zero.

And,

\[
\hat{g}(x) = (nh)^{-1} \sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right) \quad (4.1)
\]
4.1 Simulation Studies.

In this section, the performance of the NW and the DK estimators of the conditional median will be tested using three simulated data. The performance of the two estimators have been tested using the MSE and the correlation coefficients.

Simulation study 1

Sample of sizes 100, 200 and 400 are simulated from the model

\[ y = x \sin(2\pi x) + e, \]

where

\[ x \sim U[0, 1], \quad e \sim N(0, 0.1). \]

The fixed bandwidth \( h_n \) was computed. The NW and the DK estimators were computed using the Exponential and Gaussian kernel functions.

We computed the MSE and the correlation coefficients between the predicted values \( \hat{y} \), and the actual values \( y \), \( R^2_{y,\hat{y}} \) for the DK and the NW estimators, where

\[ MSE = \frac{SSE}{n} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2, \]

\[ R^2_{y,\hat{y}} = 1 - \frac{SSE}{SSTO} = 1 - \frac{1}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}, \]

where, \( \bar{y} \) denotes the mean of actual values.

The results are listed in Table (4.2) and Table (4.3) respectively.

Figure (4.1) present the first simulated data, the perfect curve, the NW estimator, and the DK estimator.
Table 4.2: The MSE for the NW and the DK estimators (Simulation 1)

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>NW</th>
<th>DK</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.003137917</td>
<td>0.003041035</td>
</tr>
<tr>
<td>200</td>
<td>0.001710991</td>
<td>0.002028771</td>
</tr>
<tr>
<td>400</td>
<td>0.001165013</td>
<td>0.001052282</td>
</tr>
</tbody>
</table>

Table 4.3: The $R^2_{\hat{y}, \hat{y}}$ for the NW and the DK estimators (Simulation 1)

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>NW</th>
<th>DK</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.9762615</td>
<td>0.9769944</td>
</tr>
<tr>
<td>200</td>
<td>0.9871621</td>
<td>0.9847778</td>
</tr>
<tr>
<td>400</td>
<td>0.9911704</td>
<td>0.9920248</td>
</tr>
</tbody>
</table>

From Table (4.2) and Table (4.3), for all sample sizes, the MSE and the $R^2_{\hat{y}, \hat{y}}$ values of the DK estimator has the best performance than the NW estimator.
Simulation study 2

Sample of sizes 100, 200 and 400 are simulated from the model

\[ y = \sin(2\pi(1 - x)^2) + xe, \]

where

\[ x \sim U[0,1], \ e \sim N(0,0.1). \]

The fixed bandwidth \( h_n \) was computed. The NW and the DK estimators were computed using the Exponential kernel functions.

We computed the MSE and the \( R^2_{y:y} \) of the DK and the NW estimators. The results are listed in Table (4.4) and Table (4.5) respectively.
<table>
<thead>
<tr>
<th>Sample Size</th>
<th>NW</th>
<th>DK</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.02564404</td>
<td>0.02080606</td>
</tr>
<tr>
<td>200</td>
<td>0.03012957</td>
<td>0.02074708</td>
</tr>
<tr>
<td>400</td>
<td>0.01555065</td>
<td>0.01007544</td>
</tr>
</tbody>
</table>

**Table 4.4:** The MSE for the NW and the DK estimators (Simulation 2)

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>NW</th>
<th>DK</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.9313106</td>
<td>0.9442695</td>
</tr>
<tr>
<td>200</td>
<td>0.920108</td>
<td>0.9449867</td>
</tr>
<tr>
<td>400</td>
<td>0.9582464</td>
<td>0.9729474</td>
</tr>
</tbody>
</table>

**Table 4.5:** The $R^2_{\hat{y}, \hat{y}}$ for the NW and the DK estimators (Simulation 2)

From Table (4.4) and Table (4.5), for all sample sizes, the MSE and the $R^2_{\hat{y}, \hat{y}}$ values of the DK estimator has the best performance than the NW estimator.

In Figure (4.2), we note that the three functions (DK estimator, NW estimator and Perfect Curve) are far apart at the boundary and approximately convergent in the otherwise.
Simulation study 3

Sample of sizes 100, 200 and 400 are simulated from the model

\[ y = \frac{1}{2} x^2 + e, \]

where

\[ x \sim U[-1, 1], \quad e \sim N(0, 0.25). \]

The fixed bandwidth \( h_n \) was computed. The NW and the DK estimators were computed using the Exponential kernel functions.

We computed the MSE and the \( R^2_{y,y} \) of the DK and the NW estimators. The results are listed in Table (4.6) and Table (4.7) respectively.
Figure (4.3) the third simulated data, the perfect curve, the NW estimator, and the DK estimator.

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>NW</th>
<th>DK</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.007542803</td>
<td>0.005265237</td>
</tr>
<tr>
<td>200</td>
<td>0.00717857</td>
<td>0.005718339</td>
</tr>
<tr>
<td>400</td>
<td>0.005030053</td>
<td>0.005035178</td>
</tr>
</tbody>
</table>

Table 4.6: The MSE for the NW and the DK estimators (Simulation 3)

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>NW</th>
<th>DK</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.9659354</td>
<td>0.9762213</td>
</tr>
<tr>
<td>200</td>
<td>0.9663612</td>
<td>0.9732039</td>
</tr>
<tr>
<td>400</td>
<td>0.9774927</td>
<td>0.9774697</td>
</tr>
</tbody>
</table>

Table 4.7: The $R^2_{y,\hat{y}}$ for the NW and the DK estimators (Simulation 3)

From Table (4.5) and Table (4.6), for all sample sizes, the MSE and the $R^2_{y,\hat{y}}$ values of the DK estimator has the best performance than the NW estimator.
4.2 Real Data Studies

A real data was conducted to compare the performances of the two estimators, a data frame with 111 observations (rows), and 4 variables (columns), taken from an environmental study that measured the four variables ozone, solar radiation, temperature, and wind speed for 111 consecutive days.

- ozone: surface concentration of ozone in New York, in parts per million.
- temperature: observed temperature, in degrees Fahrenheit.
- wind: wind speed, in miles per hour.


Our main study is talk a bout the relation between the temperature and ozone using the two estimators, in the first figure the $x$-axis represent the temperature and the $y$-axis represent ozone, in the second figure the $x$-axis represent ozone the and the $y$-axis represent temperature.
4.3 Discussion and Conclusion

In this thesis, we have studied two kernel estimators of the conditional distribution function and we have used them to estimate the conditional median function for the NW and the
DK estimators.
The NW estimator suffer from some disadvantages properties especially near the boundary points. To solve this problems a new version of the NW estimator has been proposed, which is called the DK estimator.

From this study, we have concluded that the asymptotic properties and the performance of the DK estimator is better than the NW estimator.

At the end of this thesis, we suggest some ideas to improve the estimators of the conditional median which have been discussed in this thesis.

Some of these ideas concern with the bandwidth. We can replace the constant bandwidth by a locally weighted bandwidth or by a variable bandwidth, which will improve the results of the estimators. For more details, see Salha (2009a,b).

Also another method to improve the estimator can be obtain by using adaptive kernel estimator. For more details see Salha (2009a, 2009b, 2016)
Bibliography


Pranab K.S.


