On the Inverse Gaussian Kernel Estimator of the Hazard Rate Function

May 16, 2016

The Islamic University of Gaza
Deanery of Higher Studies
Faculty of Science
Department of Mathematics

On the Inverse Gaussian Kernel Estimator of the Hazard Rate Function

حول مقدّر نواة معكوس جاوس لدالة معدل المخاطرة

Presented by
Noha Khaled ALjayeh

Supervised by
Associate Prof. Dr. Raid B. Salha

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master in Mathematics

1437 - 2016

نموذج رقم (1)

إقـــــــرار

أنا الموقع أدناه مقدم الرسالة التي تحمل العنوان:

حول مقدّر نواة معكوس جاوس لدالة معدل المخاطرة

On the Inverse Gaussian Kernel Estimator of the Hazard Rate Function

أقر بأن ما اشتملت عليه هذه الرسالة إنما هو نتاج جهدي الخاص، باستثناء ما تمت الإشارة إليه حيثما ورد، وإن هذه الرسالة ككل أو أي جزء منها لم يقدم من قبل لنيل درجة أو لقب علمي أو بحثي لدى أي مؤسسة تعليمية أو بحثية أخرى.

DECLARATION

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

Student's name: **Noha Khaled Aljayeh**	اسم الطالب: **نهى خالد الجايح**
Signature:	التوقيع: نهى
Date: 19/3/2016	التاريخ: 2016/3/19

بسم الله الرحمن الرحيم

الجامعة الإسلامية – غزة
The Islamic University - Gaza
مكتب نائب الرئيس للبحث العلمي والدراسات العليا   هاتف داخلي 1150

الرقم ج س غ/35/ Ref

التاريخ 2016/03/12 Date

نتيجة الحكم على أطروحة ماجستير

بناءً على موافقة شئون البحث العلمي والدراسات العليا بالجامعة الإسلامية بغزة على تشكيل لجنة الحكم على أطروحة الباحثة/ **نهى خالد ربحي الجايح** لنيل درجة الماجستير في **كلية *العلوم*** قسم **الرياضيات** وموضوعها:

(حول مقدر نواة معكوس جاوس لدالة معدل المخاطرة)

(On the Inverse Gaussian Kernel Estimator of the Hazard Rate Function)

وبعد المناقشة العلنية التي تمت اليوم الاثنين 12 جمادى الآخر 1437هـ، الموافق 2016/03/21م الساعة التاسعة صباحاً بمبنى طيبة، اجتمعت لجنة الحكم على الأطروحة والمكونة من:

د. رائد بشير صالحة	مشرفاً و رئيساً	....
د. بشر مأمون عقيلان	مناقشاً داخلياً	....
أ.د. عبد الله محمد الهبيل	مناقشاً خارجياً	....

وبعد المداولة أوصت اللجنة بمنح الباحثة درجة الماجستير في **كلية *العلوم*/** قسم **الرياضيات**.

واللجنة إذ تمنحها هذه الدرجة فإنها توصيها بتقوى الله ولزوم طاعته وأن تسخر علمها في خدمة دينها ووطنها.

والله ولي التوفيق،،،

نائب الرئيس لشئون البحث العلمي والدراسات العليا

أ.د. عبدالرؤوف علي المناعمة

Contents

Abstract

In this thesis, we consider the nonparametric estimation of the hazard rate function for identically independent data. To solve the problem of bias effect near the zero, when the hazard rate function is estimated by asymmetric kernel, we proposed to estimate it by using the Inverse Gaussian (IG) kernel estimation. The asymptotic mean squared error (AMSE) and the asymptotic normality of the proposed estimator are investigated under criteria conditions. Also, the problem of optimal bandwidth selection has been discussed. The performance of the proposed has been tested by applications using simulated and real data. Then we compared its performance with the performance of the Gaussian kernel. The comparasion indecated that the IG kernel is better than the Gaussian kernel especially near the zero.

الملخص

في هذا البحث ندرس التقدير اللامعملي لدالة معدل المخاطرة لبيانات مستقلة.
ولحل مشكلة التحيز عند تقدير دالة المخاطرة في منطقةالحدود بالقرب من الصفر
فإننا نستخدم تقدير معكوس جاوس.
كما تم في هذا البحث دراسة الخطأ التربيعي و تقارب الخطأ التربيعي للتقدير المقترح.
بالاضافة لذلك تم مناقشة كيفية اختيار اتساع النافذ حيث انه يلعب دورا مهما في تقدير النواة.
وأخيرا، تم اختبار أداء مقدر معكوس جاوس بشكل عملي باستخدام بيانات حقيقية و أخرى محاكاه. ثم المقارنة بين مقدر جاوس ومعكوس جاوس وكان مقدر معكوس جاوس الافضل خصوصا عند منطقة الحدود بالقرب من الصفر.

Acknowledgements

First, I thank God the most for everything and for my success. I would like to thank several people for helping me with this work. First, I would like to thank my supervisor Dr. Raid Salha for his support, patience, encouragement, and long hours without which I could have never accomplished this thesis. I would like to thank the whole group of Department of Mathematics gave me great support helped me in my work. I would like to thank all my colleagues and friends at the Department of Mathematics for encouraging me. I want to express my great gratitude to my entire family for your continuous support and encouragement during my whole life and especially during this work. And special thanks to my father, mother and my husband.

List of Abbreviations

Abbreviation	Description
AMSE	Asymptotic Mean Square Error.
AMISE	Asymptotic mean integrated squared error.
Bias	bias.
cdf	Cumulative distribution function.
Cov	Covariance.
i.i.d.	independent and identically distributed.
IG(μ,λ)	Inverse Gaussian distribution with parameter μ and λ.
ISE	Integral Square Error.
MISE	Mean Integral Square Error.
MIAE	Mean Integral Absolute Error.
MSE	Mean Square Error.
mgf	moment generating function.
o	small oh.
O	big oh.
pdf	probability density function.
pmf	probability mass function.
Var	variance.

List of Symbols

Symbol	Description
$f(x)$	Probability density function.
$F(x)$	Cumulative distribution function.
K_h	Scaled uni-variate kernel function.
$\hat{f}$	kernel estimator for the function f.
P	probability set function.
h	the bandwidth smoothing parameter.
h^*	Optimal bandwidth minimizing $AMSE$.
h^{**}	Optimal bandwidth minimizing $AMISE$ (global).
μ	the mean.
σ^2	the variance.
E	the expectation.
$R(K)$	$\int K^2(x)dx$.
I_A	the indicator function.
$\vert \cdot \vert$	the absolute value function.
S	the sample standard deviation.
R	the set of real numbers.
$\prod$	product.
$N(0,1)$	Standard normal distribution .
$\mu_j(K)$	j-th moment of a kernel K, or Gaussian distribution.
$k(u)$	the characteristic function.
$k(.)$	the kernel function.
$\overset{p}{\hookrightarrow}$	converge in probability.
$\overset{d}{\hookrightarrow}$	converge in distribution.

List of Figures

List of Tables

Preface

A density estimation is a fundamental concept in statistics defined as the construction of an estimate of the density function f from a set of observed data points assumed to be a sample from an unknown probability density function f.

There are two ways for estimating the density function the first is the parametric way which assumes that the data are drawn from a known parametric distribution which depends only on finitely many parameters. The main goal in this approach is to estimate these parameters using the sample data, the normal and gamma distributions are familiar examples of a parametric distribution families . The second one is the nonparametric estimation which assumes that the data does not belong to a known distribution family and then the estimation depend only on the data, the oldest and most widely uses is histograms, naive estimators and kernel estimator,etc.

In this thesis, we will consider the kernel estimation (as a tool in the non parametric method) for the hazard rate function, which is one of the most important ways for representing the life time distribution in the survival analysis. To get this point, we will study the kernel estimator for the probability density function (*pdf*) for independent and identically distributed (*iid*) data. Next, a study for the Gaussian kernel estimator for the hazard rate function will present. After this the main aim of the thesis with a theoretical and practical comparison will discuss. We will follow same way where Chen and Scallite were proposed in [2] and [15] respectively for solving the boundary bias near the zero. For the theoretical comparison we will consider the MSE criteria for both estimators and for the practical comparison we will use simulated and real data to test the performance of the two estimators.

This thesis will consist of the following chapters

Chapter 1. Introduction

This chapter will contain some basic definitions, facts and notations that will be used in the thesis. Also, it will contain an introduction to the kernel estimation and the Inverse Gaussian distribution .

Chapter 2. Kernel estimator of the hazard rate function

we will study the symmetric kernel estimator of the hazard rate function.

Chapter 3. Estimation of the Hazard Rate Function Using the IG Kernel

This chapter is the main chapter of the thesis. we will introduce the Inverse Gaussian (IG) kernel and using it to estimate the pdf, cdf and the hazard Rate function.

Chapter 4. Application

This chapter will contain applications using simulated and real life data to test the performance of the IG estimator. Also, we will compare it to the symmetric Gaussian kernel estimator.

The applications will construct using S-Plus program.

Chapter 1

Introduction

This chapter contains some basic definitions and facts that we need in the remaining of this thesis . In Section 1.1, we present some preliminaries from probability theory and statistics. The idea of the density estimation and some important subjects related to it will be discussed in Section 1.2. In Section 1.3, we present the kernel density estimation and some important subjects related to it. In Section 1.4, we present1 the invers Gaussian distribution and some definitions and facts related to it.

1.1 Preliminaries

Definition 1.1.1. *[1] A random variable X is a function from a sample space into the reals numbers.*

If we have a sample space $S = \{s_1 \cdots s_n\}$ with a probability function P and if we define a random variable X with range $\chi = \{x_1 \cdots x_n\}$. We can define a probability function P_X on χ as follow : we observe that $X = x_i$ if and only if the outcome of the random experiment is an $s_j \in S$ such that $X(s_j) = x_i$ and hence $P_X(X = x_i) = P(\{s_j \in S : X(s_j) = x_i\})$.

For every random variable X, we a associate a function called **cumulative distribution function** , which is defined as follows:

Definition 1.1.2. *[1] The cumulative distribution function or cdf of a random vari-*

able X, denoted by $F_X(x)$ or $F(x)$ is defined by

$$F_X(x) = P_X(X \leq x), \textit{for all } x.$$

A random variable X continuous if $F_X(x)$ is continuous function of x and its discrete if $F_X(x)$ is step function.

Definition 1.1.3. ***probability distribution****[3]*
*If X is a discrete random variable, 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. *[3]*
*A 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 $a \leq b$

Theorem 1.1.1. *[1] A function $f_X(x)$ (or $f(x)$) is a pdf or pmf of a random variable random X if and only if:*

1. *$f_X(x) \geq 0$ for all x.*
2. *$\sum_x f_X(x) = 1$ (pmf) and $\int_{-\infty}^{\infty} f_X(x)dx = 1$(pdf).*

Definition 1.1.5. ***Expectation*** *[4]*
Let X be a random variable. If X is a continuous random variable with pdf $f(x)$ and

$$\int_{-\infty}^{\infty} |x| f(x)\, dx < \infty,$$

*then the **expectation** of X is*

$$E(X) = \int_{-\infty}^{\infty} x f(x)\, dx.$$

*If X is a discrete random variable with pmf $p(x)$ and $\sum_x |x| p(x) < \infty$. The **expectation** of X is*

$$E(X) = \sum_x x p(x)$$

Definition 1.1.6. ***The* m*th* *Moment*** *[4]*

Let X be a random variable. If X is a continuous random variable with $\int_{-\infty}^{\infty} |x|^m f(x)\,dx < \infty$, then ***The* m*th moment*** *of X is*

$$E(X^m) = \int_{-\infty}^{\infty} x^m f(x)\,dx,\ m = 1, 2, 3, \ldots$$

If X has discrete random variable with $\sum_x |x|^m p(x) < \infty$. ***The* m*th moment*** *of X is*

$$E(X^m) = \sum_x x^m p(x),\ m = 1, 2, 3, \ldots$$

Definition 1.1.7. ***Mean****[3]*

Let X be a random variable whose expectation exists the ***mean*** *of X is defined by $\mu = E(X)$.*

Definition 1.1.8. ***Variance****[3]*

Let X be a random variable with finite mean μ and such that $E[(X-\mu)^2]$ is finite. Then the variance of X is defined to be $E[(X-\mu)]^2$. It is usually denoted by σ^2 or by $Var(X)$

$$\sigma^2 = E[(X-\mu)^2] = E[(X^2 - 2\mu X + \mu^2)]$$

and since E is linear operator.

$$\begin{aligned} \sigma^2 &= E(X^2) - 2\mu E(X) + \mu^2 \\ &= E(X^2) - 2\mu^2 + \mu^2 \\ &= E(X^2) - \mu^2 \end{aligned}$$

Definition 1.1.9. ***Independence*** *[4]*

Let the random variables X_1 and X_2 have the joint pdf $f(x_1, x_2)$ and the marginal pdf $f_1(x_1)$ and $f_2(x_2)$ respectively. The random variables X_1 and X_2 are said to be independent if and only if,

$$f(x_1, x_2) = f_1(x_1) f_2(x_2)$$

Random variables that are not independent are said to be ***dependent****.*

Definition 1.1.10. *[1] The random variables $X_1, \cdots, X_n$ are called* ***a random sample*** *of size n from population $f(x)$ if $X_1, \cdots, X_n$ mutually independent random variables and the marginal pdf or pmf of each X_i is the same function $f(x)$. Alternatively, $X_1, \cdots, X_n$ are called* ***independent and identically*** *distributed random variables abbreviated to iid random variables.*

Definition 1.1.11. *[11]*
If A is any set, we define the **Indicator function** *I_A of the set A to be the function given by*

$$I_A = \begin{cases} 1 & if x \in A, \\ 0 & if x \notin A. \end{cases}$$

Definition 1.1.12. *(***Converge in Probability***)[4].*
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

$$\lim_{n \longrightarrow \infty} P\left[|X_n - X| \geq \epsilon\right] = 0, \tag{1.1.1}$$

or equivalently,

$$\lim_{n \longrightarrow \infty} P\left[|X_n - X| < \epsilon\right] = 1. \tag{1.1.2}$$

If so, we write $X_n \xrightarrow{p} X$

Definition 1.1.13. ***Converge in Distribution****[4].*
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 X_n and X. Let $C(F_X)$ denote the set of all points where F_X is continuous. We say that X_n **converge in distribution** *to X if*

$$\lim_{n \longrightarrow \infty} F_{X_n}(x) = F_X(x), \quad \text{for all } x \in C(F_X). \tag{1.1.3}$$

We denote this convergence by

$$X_n \xrightarrow{d} X$$

Theorem 1.1.2. *[4]*

1. *If X_n converge to X with probability 1, then X_n converge to X in probability.*

2. *If X_n converge to X in probability, then X_n converge to X in distribution.*

3. *Let X_n converge to X in probability and let g be a continuous function on R; then $g(X_n)$ converge to $g(X)$ in probability.*

Definition 1.1.14. Characteristic Function*[4].*
The characteristic function of a random variable X with distribution function F, denoted by $k(u)$, is defined be

$$k(u) = \int_{-\infty}^{\infty} e^{-iuy} K(y) dy.$$

Definition 1.1.15. *[1] Let X be a random variable with a cdf F_X. The moment generating function (mgf) of X, denoted by $M_X(t)$, is*

$$M_X(t) = E(e^{tX})$$

provided that the expectation exists.

By taking the natural logarithm of $M_X(t)$, we get the cumulant moment $K_X(t)$.

$$K_X(t) = Log M_X(t)$$

A very important and most famous probability inequality that we will need it is **Chebychev's inequality** were presented in the next theorem.

Theorem 1.1.3. *[1] Let X be a random variable and $g(x)$ be non negative function. Then for any $r > 0$,*

$$P(g(X) \geq r) \leq \frac{E(g(X))}{r} \tag{1.1.4}$$

Definition 1.1.16. ***Order Notation** O **And** o [20].*

Let a_n and b_n each be sequences of real numbers. Then we say that a_n is of order b_n or (a_n is big oh b_n) as $n \to \infty$ and write $a_n = O(b_n)$ as $n \to \infty$, if and only if

$$\limsup_{n\to\infty} |\frac{a_n}{b_n}| < \infty.$$

In other words, $a_n = O(b_n)$ *if* $|\frac{a_n}{b_n}|$ *remains bounded as* $n \to \infty$.
We say that a_n *is of small order* b_n *and write* $a_n = o(b_n)$ *as* $n \to \infty$*, if and only if*

$$\lim_{n\to\infty} |\frac{a_n}{b_n}| = 0.$$

we will use the **Liapounov Theorem** stated in the next theorem.

Theorem 1.1.4. ***Liapounov Theorem*** *[9]*
Let $X_1, X_2, \cdots,$ *be (iid) random variables such that* $E(X_k) = \mu_k$ *and* $Var(X_k) = \sigma_k^2$ *and for some* $0 < \delta \leq 1$, $v_{2+\delta}^k = E(|X_k - \mu_k|^{2+\delta}) < \infty$ *for all* $k \geq 10$*. Also let* $T_n = \sum_{k=1}^n X_k$, $\zeta_n = E(T_n) = \sum_{k=1}^n \mu_k$, $s_n^2 = Var(T_n = \sum_{k=1}^n \sigma_k^2)$, $Z_n = \frac{T_n - \zeta_n}{s_n}$ *and* $\rho_n = s_n^{-(2+\delta)} \sum_{k=1}^n v_{2+\delta}^k$*. Then if* $\lim_{n\to\infty} \rho_n = 0$*, we have* $Z_n \xrightarrow{d} N(0,1)$*.*

Taylor expansion is important mathematical tool for obtaining asymptotic approximations in kernel smoothing and allows us to approximate function values close to a given point in term of higher-order derivatives at that point(provided the derivatives exists).

Theorem 1.1.5. ***Taylor's Theorem*** *[20]*
Suppose that f *is real-valued function defined on* $\mathbb{R}$ *and let* $x \in \mathbb{R}$*. Assume that* f *has* p *continuous derivatives in an interval* $(x - \delta, x + \delta)$ *for some* $\delta > 0$*. Then for any sequence* α_n *converging to zero.*

$$f(x + \alpha_n) = \sum_{j=0}^{p} (\frac{\alpha_n^j}{j!}) f^j(x) + o(\alpha_n^p)$$

1.2 Density Function Estimation

In this section, the concepts of density estimation parametric and non parametric density estimation are introduced.

Density estimation has experienced a wide explosion of interest over the last 40 years. Density estimation has been applied in many fields, including archaeology chemistry, banking, climatology, genetics, economics, hydrology and physiology. For more details See, [12] and [22].

1.2.1 Estimation

The purpose of inferential statistical conclusion of community properties sample drawn from it, when you use the sample data **Statistic** to infer from the community because we don't have all the facts, community urge for practical way we can trust the fact required within a given dependent on the nature of the desired community appreciation transactions **Parameter** trying to access values **numerical** to community through sample data drawn from it at random. Statistical inference is divided into two sections:

- Statistical Estimation.
- Hypothesis testing.

The main purpose of this thesis is the first section, the statistical estimation.
The probability density function is a fundamental concept in statistics. Consider any random variable X that has probability density function f. Specifying the function f gives a natural description of the distribution of X, and allows probabilities associated with X to be found from the relation.

$$P(a < X < b) = \int_a^b f(x)dx,$$

for any real constants a and b with a < b.

Definition 1.2.1. ***Estimator****[3]. An estimator is a rule, often expressed as a formula, that tells how to calculate the value of an estimate based on the measurements contained in a sample.*

Definition 1.2.2. *[3]. Let X be a random variable with pdf with parameter θ. Let X_1,X_2,... ,X_n be a random sample from the distribution of X and let $\hat{\theta}$ denotes an estimator of θ. We say $\hat{\theta}$ is an* **unbiased** *estimator of θ if*

$$E(\hat{\theta}) = \theta.$$

If $\hat{\theta}$ is not unbiased, we say that $\hat{\theta}$ is **a biased** *estimator of θ.*

Definition 1.2.3. *[3]. If $\hat{\theta}$ is an unbiased estimator of θ and*

$$Var(\hat{\theta}) = \frac{1}{nE\left[\left(\frac{\partial lnf(X)}{\partial\theta}\right)\right]^2} \tag{1.2.1}$$

then $\hat{\theta}$ is called a minimum variance unbiased estimator **(efficient)** *of θ.*

Definition 1.2.4. *[3]. The statistic $\hat{\theta}$ is* **a Consistent estimator** *of the parameter θ if and only if for each $c > 0$*

$$\lim_{n\longrightarrow\infty} P(|\hat{\theta} - \theta| < c) = 1. \tag{1.2.2}$$

Theorem 1.2.1. *[3]. If $\hat{\theta}$ is an unbiased estimator of θ and $Var(\hat{\theta})\rightarrow 0, as\ \ n\rightarrow \infty$, then $\hat{\theta}$ is a consistent estimator of θ.*

Definition 1.2.5. *[3]. The statistic $\hat{\theta}$ is* **a sufficient estimator** *of the parameter θ if, and only if for each value of $\hat{\theta}$ the conditional probability distribution or density of the random sample X_1,X_2, ... ,X_n given $\hat{\theta} = \theta$ is independent of θ.*

1.2.2 Density Estimator

Suppose, now, that we have a set of observed data points assumed to be a sample from an unknown probability density function. Density estimation is the construction of an estimate of the density function from the observed data.

The two main aims are to explain how to estimate a density from a given data set and to explore how density estimates can be used, both in their own right and as an ingredient of other statistical procedures.

One approach to density estimation is **parametric**. Assume that the data are drawn from one of a known parametric family of distributions, for example the normal distribution with mean μ and variance σ^2 . The density f underlying the data could then be estimated by finding estimates of μ and σ^2 from the data and substituting these estimates into the formula for the normal density.

Another approach to a density estimation is **a nonparametric** .

1.2.3 Parametric Estimation

Parametric statistics is a branch of statistics that assume that the data has come from a type of probability distribution and makes inferences about the parameters of the distribution. Most well known elementary statistical methods are Parametric. Parametric formula are often simpler to write down and faster to compute .

The parametric approach for estimating $f(x)$ is to assume that $f(x)$ is a member of some parametric family of distributions, e.g. $N(\mu, \sigma^2)$, and then to estimate the parameters of the assumed distribution from the data. For example, fitting a normal distribution leads to the estimator

$$\hat{f}(x) = \tfrac{1}{\sqrt{2\pi}\hat{\sigma}} exp\left(\frac{(x-\hat{\mu})^2}{2\hat{\sigma}^2}\right), \quad x \in \mathbf{R},$$

where,

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i, \text{ and } \hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{\mu})^2.$$

This approach has advantages as long as the distributional assumption is correct, or at least is not seriously wrong. It is easy to apply and it yields (relatively) stable estimates. The main disadvantage of the parametric approach is lack of flexibility. Each parametric family of distributions imposes restrictions on the shapes that $f(x)$ can have. For example the density function of the normal distribution is symmetrical and bell-shaped, and therefore is unsuitable for representing skewed densities or bimodal densities.

Methods of finding parametric Estimator :-

Here, we will introduce two main methods of parametric estimation, the method of moments and the method of maximum likelihood function.

(1) The Method of Moments

In statistics, the method of moments is a method of estimation of population parameters such as mean, variance, median, etc. (which need not be moments), by equating sample moments with unobservable population moments and then solving those equations for the quantities to be estimated .

(2) The Method of Maximum Likelihood

The maximum likelihood method which depends on finding the value of the unknown parameter θ that maximize the joint distribution $f(x_1, x_2, ..., x_n; \theta)$.

Definition 1.2.6. *If $x_1, x_2, ..., x_n$ are the values of the random sample from a population with the parameter θ , the likelihood function of the sample is given by*

$$\boldsymbol{L(\theta) = f(x_1, x_2, ..., x_n; \theta) = \prod f(x_i/\theta)}$$

as $x_1, x_2, ..., x_n$ are independent random the Maximum likelihood Method.
The Maximum likelihood Method for finding an estimator of θ, consist of finding the estimator θ which make the function $L(\hat{\theta})$ is maximum . That is to find θ by finding

$$\frac{\partial LnL(\theta)}{\partial\theta} = 0.$$

Example 1.2.1. *If $x_1, x_2, .., x_n$ are the values of a random sample of size n, from the Bernoulli population.*

$$f(x) = \theta^x(1-\theta)^{1-x} \ , \ x = 0, 1 \ , \ 0 < \theta < 1$$

$$\begin{aligned}
\mathrm{L}(\theta) &= \prod \theta^x(1-\theta)^{1-x} = \theta^{\sum x_i}(1-\theta)^{n-\sum x_i} \\
LnL(\theta) &= \sum x_i Ln(\theta) + (n - \sum x_i)Ln(1-\theta) \\
\frac{\partial LnL(\theta)}{\partial\theta} &= \frac{\sum x_i}{\theta} - \frac{n - \sum x_i}{1-\theta} = 0 \\
\frac{\sum x_i}{\theta} &= \frac{n - \sum x_i}{1-\theta} \\
\sum x_i - \theta \sum x_i &= n\theta - \theta \sum x_i \\
\sum x_i &= n\theta \\
\hat{\theta} &= \frac{\sum x_i}{n}
\end{aligned}$$

1.2.4 Non Parametric Estimation

Nonparametric density estimation extracts information about the underlying structure of a data set when no appropriate parametric model is available. It is an

important data analytic tool which provides a very effective way of showing structure in a set of data at the beginning of its analysis.
For obtaining a nonparametric estimation of a probability density function there are many methods. Three of them are the following methods:

- Histogram
- The naive estimator
- Kernel density estimation

Non Parametric Methods

(1) Histogram :

The oldest and most widely used density estimator is the histogram. The idea of the nonparametric approach is to avoid restrictive assumptions about the form of $f(x)$ and to estimate this directly from the data. A well known nonparametric estimator of the *pdf* is the histogram. It has the advantage of simplicity but it also has disadvantages, such as lack of continuity. Secondly, in terms of various mathematical measures of accuracy there exist alternative nonparametric estimators that are superior to histograms.

To construct a histogram one needs to select a left bound, or starting point, x_0, and the bin width, h. The bins are of the form $[x_0 + mh; x_0 + (m+1)h]$, for positive and negative integers m. The estimator of $f(x)$ is then given by

$$\hat{f}(x) = \frac{1}{nh}(\text{number of } X_i \text{ in same bin as } x). \tag{1.2.3}$$

More generally one can also use bins of different widths, in which case

$$\hat{f}(x) = \frac{1}{nh}\frac{(\text{number of } X_i \text{ in same bin as } x)}{(Width\ \ of\ \ bin\ \ containing\ \ x)}. \tag{1.2.4}$$

The choice of bins, especially the bin widths, has a substantial effect on the shape and other properties of $\hat{f}(x)$. For more detalies see [6]

(2)The naive estimator

A generalization of the histogram method, is the naive estimator. From the definition of a probability density function, if the random variable X has density f, then

$$f(x) = \lim_{h \longrightarrow 0} \frac{1}{2h} P(x - h < X \leq x + h). \tag{1.2.5}$$

where,

$$\begin{aligned} \lim_{\mathrm{h} \longrightarrow 0} (\frac{1}{2h} P(x - h < X \leq x + h)) &= \lim_{h \longrightarrow 0} \frac{1}{2h} (F_X(x + h) - F_X(x - h)) \\ &= F'_X(x) = f(x). \end{aligned}$$

For any given h, we can estimate $P(x - h < X < x + h)$ by the proportion of the sample falling in the interval $(x - h,\ x + h)$. Thus a natural estimator $\hat{f}$ of the density function is given by choosing a small number h and setting

$$\hat{f}(x) = \frac{1}{2nh} (\text{number of } X_1, ..., X_n \text{ falling in } (x - h,\ x + h)). \tag{1.2.6}$$

This estimator is called the naive estimator.

To express the estimator more transparently, define the weight function w by

$$w(x) = \begin{cases} \frac{1}{2} & : |x| < 1 \\ 0 & : \text{other wise.} \end{cases} \tag{1.2.7}$$

Using this notation, we can express the naive estimator as

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h} w\left(\frac{x - X_i}{h}\right). \tag{1.2.8}$$

where X_i are the data samples.

In this form, it is easy to see that the naive estimator places a box of width $2h$ and height $(2hn)^{-1}$ at each data point and sums up the contributions.
This interpretation is useful in deriving the kernel estimator, which we discuss in the next section. For more detalies see [4]

1.3 Kernel Density Estimation

A generalization of the naive density estimation is the Kernel Density Estimation .

1.3.1 Kernel Estimator

From the definition of the pdf, $f(x)$, of a random variable, X, one has that

$$P(x-h<X<x+h)=\int_{x-h}^{x+h} f(t)dt \approx 2hf(x) \tag{1.3.1}$$

and hence

$$f(x) \approx \frac{1}{2h} P(x-h<X<x+h) \tag{1.3.2}$$

Let $X_1, X_2, ..., X_n$ be i.i.d. random variables with distribution function $F(x)=\int_{-\infty}^{x} f(y)dy$ with probability density function $f(x)$. The sample distribution function $\hat{f}(x)$ at a point x is defined as

$\hat{f}(x)=\frac{1}{n}$ (number of observations $x_1, x_2, ..., x_n$ falling in $(-\infty,\ x\ ]$). It is natural to take $\hat{f}(x)$ as an estimate of $f(x)$ at a given point x, where h is chosen as a positive number and can be written as

$$\begin{aligned}
\hat{f}(x) &= \frac{1}{2nh}(number\ of\ observations\ falling\ in\ the\ interval[x-h, x+h]) \\
&= \frac{1}{2nh}\sum_{i=1}^{n} I(|X_i-x| \leq h) \\
&= \frac{1}{nh}\sum_{i=1}^{n} \frac{1}{2} I(\frac{|X_i-x|}{h} \leq 1) \\
&= \frac{1}{nh}\sum_{i=1}^{n} w(\frac{X_i-x}{h})
\end{aligned}$$

where, $w(\frac{X_i-x}{h})=\frac{1}{2}I(|\frac{X_i-x}{h}| \leq 1)=\begin{cases} \frac{1}{2}, & -1 \leq \frac{X_i-x}{h} \leq 1, \\ 0 & \text{otherwise.} \end{cases}$

Definition 1.3.1. *We consider the function that centered at the estimation point used to weight nearby data points as a weight function and will call it the kernel* *function and denoted by $K(.)$ which defined as*

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K(\frac{x - X_i}{h}) \tag{1.3.3}$$

Note that Equation (1.3.3) can be written as

$$\hat{f}(x) = \tfrac{1}{n} \textstyle\sum_{i=1}^{n} K_h(x - X_i)$$

where, $K_h(x) = \frac{K(\frac{x}{h})}{h}$

Figure 1.1 shows kernel density estimation with different bandwidths h, constructed using seven points with kernel chosen to be $N(0,1)$ density ($f_G(x)$), i.e. $K(x) = f_G(x)$, where $f_G(x) = \frac{1}{2\sqrt{\pi}} e^{-\frac{x^2}{2}}$.
From this we have

$$K_h(x) = \tfrac{1}{h} K(\tfrac{x}{h}) = \tfrac{1}{h} f_G(\tfrac{x}{h})$$

hence $K_h(x)$ have $N(0, h^2)$ distribution. So h determine the spread of the kernel.

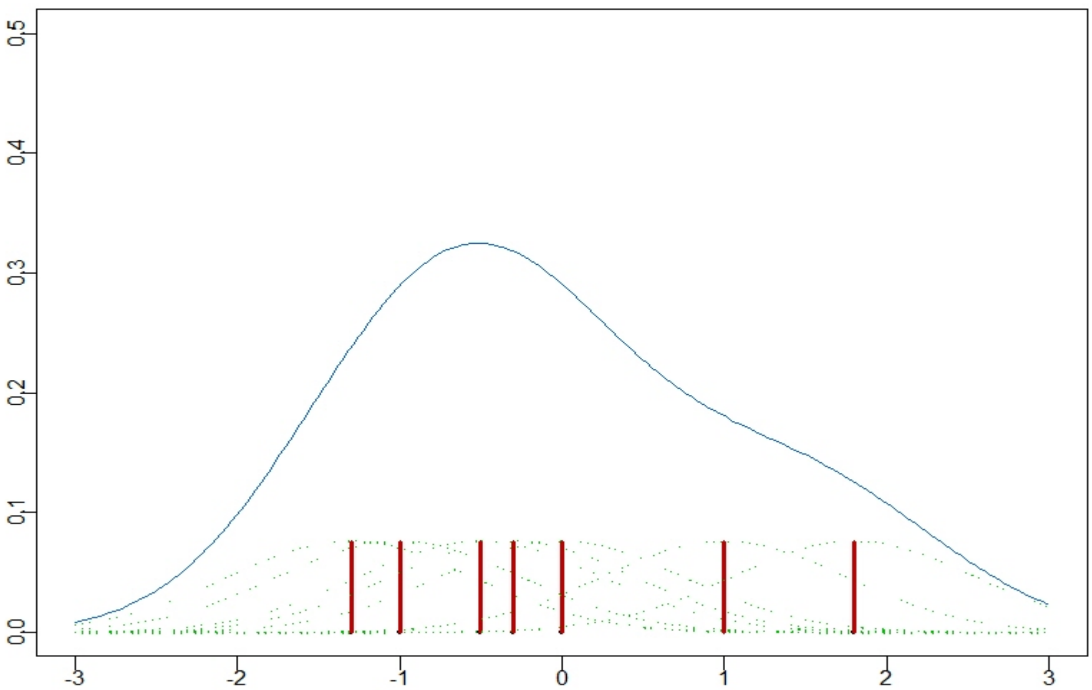

Figure 1.1: Kernel density estimation based on 7 points.

From Figure 1.1, we have:

(1) The shape of the bump is defined the kernel function.

(2) The spread of the bump is determined by a bandwidth h, that is analogous to the bandwidth of a histogram.

That is the value of the kernel estimate at the point x is the average of the n kernel ordinates at this point.

1.3.2 The Properties of the Kernel Estimator

There are various ways to quantify the accuracy of a density estimator. We will study two types of the error criteria, the mean squared error (MSE) and the mean integrated squared error ($MISE$), also we discuss the asymptotic normality and the consistency of the kernel density estimator.

Definition 1.3.2. ***Biasedness and Unbiasedness*** *[6]*

If $\hat{f}(\theta)$ is an estimator of θ and assume Θ be a set of parameters then the bias of an estimator is defined to be expected value (assuming it exists) of its sampling error, that is,

$$\text{Bias}(\hat{f}(\theta)) = E(\hat{f}(\theta)) - f(\theta).$$

If $\text{Bias}(\hat{f}(\theta)) = 0$ *for all $\theta \in \Theta$, then $E(\hat{f}(\theta)) = \theta$ and the estimator $\hat{f}(\theta)$ is defined to be an unbiased estimator of θ.*

Definition 1.3.3. *The mean squared error (MSE) is used to measure the error when estimating the density function at a single point. It is defined by*

$$MSE(\hat{f}(x)) = E(\hat{f}(x) - f(x))^2 \tag{1.3.4}$$

From its definition, the MSE measures the average squared difference between the density estimator and the true density.

$$\begin{aligned}
\text{MSE}(\hat{f}(x)) &= E(\hat{f}(x) - f(x))^2 \\
&= E(\hat{f}(x))^2 - 2E\hat{f}(x)f(x) + E(f(x))^2 \\
&= [E(\hat{f}(x))^2 - (E\hat{f}(x))^2] + [(E\hat{f}(x))^2 - 2E\hat{f}(x)f(x) + f(x)^2] \\
&= Var(\hat{f}(x)) + (E\hat{f}(x) - f(x))^2.
\end{aligned}$$

Definition 1.3.4. *An error criterion that measures the distance between* $\hat{f}(x)$ *and* $f(x)$ *is the integrated squared error* (ISE) *given by*

$$ISE\hat{f}(x) = \int_{-\infty}^{\infty} [\hat{f}(x) - f(x)]^2 dx \tag{1.3.5}$$

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 integrated squared error.

Theorem 1.3.1. *Let* X *be a random variable having a density* f*; then*

$$\begin{aligned} \text{MISE}(\hat{f}(x)) &= \frac{1}{n}[\int_{-\infty}^{\infty} K^2(x-y)f(y)dy - [\int_{-\infty}^{\infty} K(x-y)f(y)dy]^2] \\ &+ [\int_{-\infty}^{\infty} K(x-y)f(y)dy]^2 - f(x)]^2 \end{aligned} \tag{1.3.6}$$

Proof: See [8]

Now to compute the MISE of $\hat{f}(x)$, we use variance and bias of $\hat{f}(x)$.

Definition 1.3.5. *The expected value of* ISE *is called the mean integrated squared error* $(MISE)$ *is given by*

$$MISE(\hat{f}(x)) = E(ISE\hat{f}(x)) = E\int_{-\infty}^{\infty} [\hat{f}(x) - f(x)]^2 dx \tag{1.3.7}$$

By changing the order of integration, we have

$$\begin{aligned} \text{MISE}(\hat{f}(x)) &= \int_{-\infty}^{\infty} MSE(\hat{f}(x))dx \\ &= \int_{-\infty}^{\infty} Var(\hat{f}(x))dx + \int_{-\infty}^{\infty} [(E\hat{f}(x) - f(x))^2]dx \end{aligned}$$

Theorem 1.3.2. *The MISE of an estimator* $\hat{f}(x)$ *of a density* $f(x)$ *is given by*

$$\begin{aligned} \text{MISE}(\hat{f}(x)) &= \frac{1}{n}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K^2(x-y)f(y)dydx \\ &+ (1-\frac{1}{n})\int_{-\infty}^{\infty}[\int_{-\infty}^{\infty} K(x-y)f(y)dy]^2 dx] \\ &- 2\int_{-\infty}^{\infty}[\int_{-\infty}^{\infty} K(x-y)f(y)dy]f(x)dx \\ &+ \int_{-\infty}^{\infty} f^2(x)dx \end{aligned}$$

Proof: See [8]

Definition 1.3.6. ***Asymptotic Unbiasedness****[6]*

An estimator $\hat{\theta}_i$ *of* θ_i *is said to be asymptotically unbiased if*

$$E(\hat{\theta}_i) \to \theta_i \text{ as } i \to \infty$$

for all possible values of θ_i *,* $\hat{\Theta}$ *is asymptotically unbiased estimator of* Θ *if* $\hat{\theta}_i$ *is asymptotically unbiased for* $i = 1, ..., k$.

Theorem 1.3.3. *Let* X *be a random variable having a density* f *; then the bias of* $\hat{f}(x)$ *can be expressed as*

$$E(\hat{f}(x)) - f(x) = \frac{1}{2}h^2\mu_2(K)f''(x) + o(h^2), \tag{1.3.8}$$

where,

$\int_{-\infty}^{\infty} K(z)dz = 1, \int_{-\infty}^{\infty} zK(z)dz = 0,$

$\int_{-\infty}^{\infty} z^2K(z)dz < \infty, \mu_2(K) = \int_{-\infty}^{\infty} z^2K(z)dz$

Proof: See [23]

Theorem 1.3.4. *Let* X *be a random variable having a density* f *; then*

$$Var(\hat{f}(x)) = \frac{1}{nh}R(K)f(x) + o(\frac{1}{nh}), \tag{1.3.9}$$

where, $R(K) = \int_{-\infty}^{\infty} K^2(x)dx$

Proof: See [23]

Corollary 1.3.1. *The Mean-Squared Error of* $\hat{f}(x)$ *is given by :*

$$MSE(\hat{f}(x)) = \frac{\mu_r^2(K)}{(r!)^2}f^{(r)}(x)^2h^{2r} + \frac{f(x)R(K)}{nh} \tag{1.3.10}$$

where, $\mu_j(K) = \int_{-\infty}^{\infty} u^jK(u)du$*,* r *is the order of the kernel and* $R(K) = \int_{-\infty}^{\infty} K^2(u)du$.

Proof: By the definition of the mean squared error we have :

$$
\begin{aligned}
MSE(\hat{f}(x)) &= E(\hat{f}(x) - f(x))^2 \\
&= Bias(\hat{f}(x))^2 + Var(\hat{f}(x)) \\
&\cong (\frac{1}{r!} f^r(x) h^r \mu_r(k))^2 + \frac{f(x)R(K)}{nh} \\
&= \frac{\mu_r(k)^2}{(r!)^2} f^r(x)^2 h^{2r} + \frac{f(x)R(K)}{nh}
\end{aligned}
$$

Remark 1.3.1. *Since our approximation for the MSE is based on asymptotic expansions, this is also called the asymptotic mean squared-error $AMSE$, which mean that*

$$
AMSE(\hat{f}) = MSE(\hat{f}) = \frac{\mu_r(k)^2}{(r!)^2} f^r(x)^2 h^{2r} + \frac{f(x)R(K)}{nh} \tag{1.3.11}
$$

The next remark is very important and we will use it in Chapter 3, especially for the conditions under which the results of the chapter hold.

Remark 1.3.2. *In Equation (1.2.10), the first term (squared bias) is increasing in h and the second term (the variance) is decreasing in nh and hence to make the $MSE(\hat{f}(x))$ to decline as $n \to \infty$ we have to make both of these terms small, which meaning that as $n \to \infty$ we must have $h \to 0$and $nh \to \infty$. That is, the bandwidth h must decrease, but not at a rate faster than sample size n.*

Theorem 1.3.5. *The MISE of an estimator $\hat{f}(x)$ of the unknown density f is given by*

$$
MISE(\hat{f}(x)) = AMISE(\hat{f}(x)) + o\{h^4 + (nh)^{-1}\} \tag{1.3.12}
$$

where AMISE is the asymptotic mean integral squared error of $\hat{f}(x)$ given by

$$
AMISE(\hat{f}(x)) = \frac{1}{4} h^4 \mu_2^2(K) R(f'') + (nh)^{-1} R(K). \tag{1.3.13}
$$

Proof:

Now to compute the MISE of $\hat{f}(x)$, we use variance and bias of $\hat{f}(x)$

$$
\begin{aligned}
\text{MISE } (\hat{f}(x)) &= \int_{-\infty}^{\infty} [Var(\hat{f}(x))] dx + \int_{-\infty}^{\infty} [E(E\hat{f}(x) - f(x))^2] dx \\
&= n^{-1} h_n^{-1} \int_{-\infty}^{\infty} K^2(t) dx + \int_{-\infty}^{\infty} \text{bias}_{h_n}^2(x) dx \\
&= n^{-1} h_n^{-1} \int_{-\infty}^{\infty} K^2(t) dx + \frac{1}{4} h_n^4 \mu_2(K) \int_{-\infty}^{\infty} f''^2(t) dt
\end{aligned}
$$

Notice that the integral squared bias is asymptotically proportional to h^4, so to reduce this quantity one needs to take h to be small. On the other hand, taking a small h 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.

Remark 1.3.3. *A kernel is higher-order kernel if $r > 2$, such kernels will have negative parts and are not probability densities. In our thesis we will consider that the kernels are of the second order $r = 2$ and the* ***assumptions(C)*** *that we will need are summarized below :*

1. *The unknown density function $f(x)$ has continuous second derivative $f^{(2)}(x)$.*
2. *The bandwidth $h = h_n$ is a sequence of positive numbers and satisfies $h \to 0$ and $nh \to \infty$ as $n \to \infty$ (see Remark 1.3.2).*
3. *The kernel K is a bounded probability density function of order 2 and symmetric about the zero.*

Under the assumptions in Remark 1.3.1, we have the following results :

1. The bias of $\hat{f}(x)$ is given by :
$$Bias(\hat{f}(x)) = \frac{1}{2!}f^{(2)}(x)h^2\mu_2(K) + o(h^2) \tag{1.3.14}$$
which means that the Bias is of order $o(h^2)$, which implies that $\hat{f}(x)$ is asymptotically unbiased estimator since assumption C_2.
2. The bias is large, whenever the absolute value of the second derivative $|f^{(2)}(x)|$ is large.
3. The variance of $\hat{f}(x)$ is given by :
$$Var(\hat{f}(x)) = \frac{f(x)R(K)}{nh} + o(\frac{1}{n}) \tag{1.3.15}$$
which meas that The variance is of order nh, hence the variance converges to zero by assumption C_2.

4. The asymptotic mean squared error is given by :

$$AMSE(\hat{f}) = MSE(\hat{f}) = \frac{\mu_2(k)^2}{4} f^{(2)}(x)^2 h^4 + \frac{f(x)R(K)}{nh} \tag{1.3.16}$$

5. The asymptotic mean integrated squared error is given by :

$$AMISE(\hat{f}) = \frac{\mu_2(k)^2}{4} R(f^{(2)}(x))h^4 + \frac{R(K)}{nh} \tag{1.3.17}$$

The next table present some of common second order kernels with $R(K)$ and $\mu_2(K)$ already evaluated.

Kernel	**Equation**	$R(K)$	$\mu_2(K)$
Normal	$K_{NW}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$	$\frac{1}{2\sqrt{\pi}}$	1
Epanechnikov	$K_1(x) = \frac{3}{4}(1-x^2)I_{(\lvert x\rvert \leq 1)}$	$\frac{3}{5}$	$\frac{1}{5}$
Gaussian	$K_G(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$	$\frac{1}{2\sqrt{\pi}}$	1

Table 1.1: Common Second Order Kernels

Corollary 1.3.2. *Under the assumptions C_1,C_2 and C_3 we have :*

$$\hat{f}(x) \xrightarrow{p} f(x)$$

Proof: Using Chebychevs inequality and Equation 1.3.16, we have for $\epsilon > 0$:

$$\begin{aligned} P\left(\left|\hat{f}(x) - f(x)\right| > \epsilon\right) &\leq \frac{E\left(\hat{f}(x) - f(x)\right)^2}{\epsilon^2} \\ &= \frac{\mu_2(k)^2}{4\epsilon^2} f^{(2)}(x)^2 h^4 + \frac{f(x)R(K)}{nh\epsilon^2} \\ &\to 0, \qquad \text{as } n \to \infty \end{aligned}$$

since $h \to 0$ and $nh \to \infty$, the first and second term vanishes respectively.

The next theorem present the asymptotic normality of the kernel density estimator.

Theorem 1.3.6. *: Under the assumptions C_1,C_2 and C_3 with additional condition $(nh^5)^{\frac{1}{2}} \to 0$ as $n \to \infty$, we have :*

$$(nh)^{\frac{1}{2}} \left(\hat{f}(x) - f(x)\right) \xrightarrow{d} N\left(0, f(x)R(K)\right)$$

Proof: See [7].

Finally, we present the kernel estimator for the cdf $\hat{F}(x)$.

Definition 1.3.7. *The kernel estimator of the cdf is defined by :*

$$\hat{F}(x) = \int_{-\infty}^{x} \hat{f}(u)du = \frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{x} K\left(\frac{u - X_i}{h}\right) du. \tag{1.3.18}$$

Remark 1.3.4. *: By Corollary 1.3.1, Definition 1.3.7 and under the assumptions C_1,C_2 and C_3 we have $\hat{F}(x) \xrightarrow{p} F(x)$.*

1.3.3 Optimal Bandwidth

The problem of bandwidth selection is very important in density estimation. Choice of the appropriate bandwidth is critical to the performance of most nonparametric density estimators. When the bandwidth is very small, the estimate will be very close to the original data. The estimate will be almost unbiased , but it will have large variation under repeated sampling. If the bandwidth is very large, the estimate will be very smooth, lying close to the mean of all the data. Such an estimate will have small variance, but it will be highly biased.

If we differentiate the expression of the $AMSE$ in Equation (1.3.16) with respect to h and setting it equal to zero, we get :

$$4h^3\frac{\mu_2(K)^2 f^{(2)}(x)^2}{4} = \frac{f(x)R(K)}{n}h^{-2} \tag{1.3.19}$$

Multiply Equation (1.3.19) by h^2 both sides to get :

$$h^5\left(\mu_2(K)^2 f^{(2)}(x)^2\right) = \frac{f(x)R(K)}{n} \tag{1.3.20}$$

Next, solving Equation (1.3.20) for h, we get the optimal bandwidth h^*:

$$h^* = \left(\frac{f(x)R(K)}{\mu_2(K)^2 f^{(2)}(x)^2}\right)^{\frac{1}{5}} n^{-\frac{1}{5}} \tag{1.3.21}$$

Note that the optimal bandwidth h^* is proportional to $n^{-\frac{1}{5}}$, and we say that the optimal bandwidth is of order $o(n^{-\frac{1}{5}})$.

Now to find the optimal $AMSE$ ($AMSE^*$), we substitute h^* in Equation (1.3.16):

$$\begin{aligned} AMSE^* &= \left(\frac{f^{(2)}(x)\mu_2(K)}{2!}\right)^2 \left(\frac{(2!)^2 f(x) R(K)}{4\mu_2(K)^2 f^{(2)}(x)^2}\right)^{\frac{4}{5}} n^{-\frac{4}{5}} \\ &+ f(x)R(K) \left(\frac{(2!)^2 f(x) R(K)}{4\mu_2(K)^2 f^{(2)}(x)^2}\right)^{-\frac{1}{5}} n^{-\frac{4}{5}} \\ &= \left(\frac{f^{(2)}(x)\mu_2(K)}{2!}\right)^2 \left(\frac{(2!)^2 f(x) R(K)}{4\mu_2(K)^2 f^{(2)}(x)^2}\right)^{\frac{4}{5}} n^{-\frac{4}{5}} \\ &+ 4\left(\frac{f^{(2)}(x)\mu_2(K)}{2!}\right)^2 \left(\frac{(2!)^2 f(x) R(K)}{4\mu_2(K)^2 f^{(2)}(x)^2}\right)^{\frac{4}{5}} n^{-\frac{4}{5}} \\ &= 5\left(\frac{f^{(2)}(x)\mu_2(K)}{2!}\right)^2 \left(\frac{f(x) R(K)}{\mu_2(K)^2 f^{(2)}(x)^2}\right)^{\frac{4}{5}} n^{-\frac{4}{5}} \qquad (1.3.22) \end{aligned}$$

Now for the **global properties** we take the integration of the formula of h^* in

Equation 1.3.19 and using the facts that $\int_{-\infty}^{\infty} f(x)dx = 1$ and $\int_{-\infty}^{\infty} f^{(2)^2}(x)dx = R(f^{(2)})$, we get the optimal bandwidth h^{**}:

$$h^{**} = \left(\frac{R(K)}{n\mu_2(K)^2 R(f^{(2)})}\right)^{\frac{1}{5}} \qquad (1.3.23)$$

Now to find the optimal $AMISE$ ($AMISE^{**}$), we integrate the formula of the $AMSE^*$ in Equation 1.3.22 :

$$AMISE^{**}(\hat{f}(x)) = \frac{5}{4}\left(R(f^{(2)})\mu_2(K)^2 R(K)^4\right)^{\frac{1}{5}} n^{-\frac{4}{5}} \qquad (1.3.24)$$

Remark 1.3.5. *Note that the $AMISE^{**}(\hat{f}(x))$ in Equation 1.3.22, is of order $n^{-\frac{4}{5}}$, which is the best obtainable rate of convergence for the class of the second order kernels as [20] stated.*

Remark 1.3.6. *The rule of thumb were [17] used replaces the unknown pdf f in Equation 1.3.21 by a reference distribution function having variance equal to the sample variance. An illustration given in Chapter 2 and later in Chapter 3 we follow the same way.*

1.4 Inverse Gaussian Distribution

Tweedie [18] who was first study and apply this to a certain class of distribution when he noted the inverse relationship between the cumulant generating functions of these distributions and those of Gaussian distributions. Wald [19] derived the same class of distributions. The following definition describe the canonical form of the two parameter $IG(\mu, \lambda)$ inverse Gaussian distribution.

Definition 1.4.1. *[5] A random variable X have an inverse Gaussian distribution if $f_X(x|\mu, \lambda)$ is the density defined as :*

$$f_X(x|\mu, \lambda) = \frac{\sqrt{\lambda}}{\sqrt{2\pi x^3}} exp\left(-\frac{\lambda}{2\mu}\left(\frac{x}{\mu} - 2 + \frac{\mu}{x}\right)\right), x > 0, \mu > 0, \lambda > 0. \tag{1.4.1}$$

We will denote this distribution by $IG(\mu, \lambda)$.

Remark 1.4.1. *By the previous definition, if X is distributed as $IG(\mu, \lambda)$, then aX ($a > 0$)distributed as $IG(a\mu, a\lambda)$.*

Remark 1.4.2. *The following theorem show that using definition 1.4.1 we can prove that the $IG(\mu, \lambda)$ inverse Gaussian distribution can be written in exponential family form.*

Theorem 1.4.1. *The inverse Gaussian distribution is a two-parameter exponential family.*

Proof: The pdf of the IG distribution can be written as: for all $x > 0, \mu > 0, \lambda > 0$

we have,

$$\begin{aligned} f_X(x|\mu, \lambda) &= \frac{1}{\sqrt{x^3}}\sqrt{\frac{\lambda}{2\pi}} exp\left(\frac{\lambda}{\mu}\right) exp\left(-x\frac{\lambda}{2\mu^2} - \frac{\lambda}{2x}\right) \\ &= \frac{1}{\sqrt{x^3}}\sqrt{\frac{\lambda}{2\pi}} exp\left(\frac{\lambda}{\mu}\right) exp\left(-\frac{\lambda}{2\mu^2}x - \frac{\lambda}{2}\frac{1}{x}\right) \end{aligned}$$

Setting :

1. $h(x) = \frac{1}{\sqrt{x^3}}, \forall x > 0$,
2. $c(\mu, \lambda) = \sqrt{\frac{\lambda}{2\pi}} exp\left(\frac{\lambda}{\mu}\right)$,
3. $w_1(\mu, \lambda) = -\frac{\lambda}{2\mu^2}$,
4. $w_2(\mu, \lambda) = -\frac{\lambda}{2}$ and

5. $t_1(x) = x$ and $t_2(x) = \frac{1}{x}$.

Hence, the pdf of the IG distribution becomes :

$$f_X(x|\mu, \lambda) = h(x)c(\mu, \lambda)exp\left(w_1(\mu, \lambda)t_1(x) + w_2(\mu, \lambda)t_2(x)\right)$$

Therefore, the IG distribution is in two parameters exponential family form.

Figure 1.2, Shows the pdf of IG for different μ and λ.

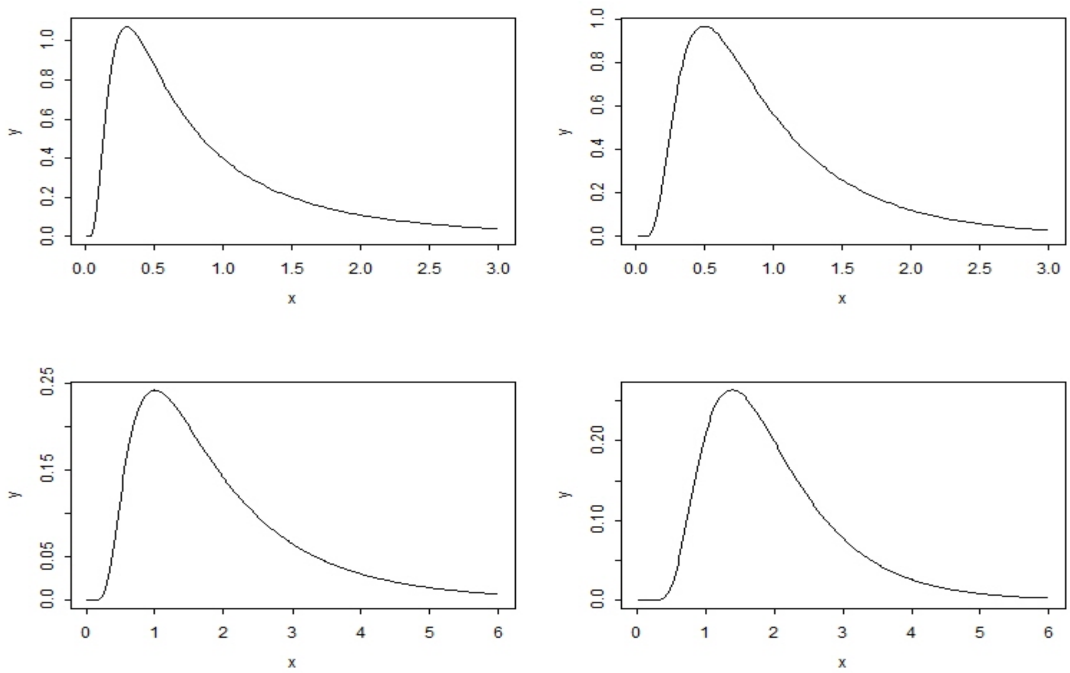

Figure 1.2: IG distribution with different μ and λ.

The Moments of the IG distribution

Next we give in detail the **moments** of the IG distribution.

Recall that the moment generating function for any random variable X is defined as

$$M_X(t) = E(e^{tX})$$

and the cumulant moment is defined as

$$K_X(t) = logM_X(t)$$

Theorem 1.4.2. *Let X be a random variable have an $IG(\mu, \lambda)$ distribution, then the moment generating function of X is given by :*

$$M_X(t) = exp\left(\frac{\lambda}{\mu}\left(1 - \left(1 - \frac{2t\mu^2}{\lambda}\right)^{1/2}\right)\right), \mu > 0, \lambda > 0 \qquad (1.4.2)$$

Proof : Using the definition of moments generating function we have:

$$\begin{aligned}
M_X(t) &= E\left(e^{tX}\right) \\
&= \int_0^\infty e^{tx} f_X(x|\mu, \lambda) dx \\
&= \int_0^\infty e^{tx} \frac{\sqrt{\lambda}}{\sqrt{2\pi x^3}} exp\left(-\frac{\lambda}{2\mu}\left(\frac{x}{\mu} - 2 + \frac{\mu}{x}\right)\right) dx \\
&= \int_0^\infty \frac{\sqrt{\lambda}}{\sqrt{2\pi x^3}} exp\left(-\frac{x\lambda}{2\mu^2} - \frac{\lambda}{2x} + \frac{\lambda}{\mu} + tx\right) dx \\
&= \int_0^\infty \frac{\sqrt{\lambda}}{\sqrt{2\pi x^3}} exp\left(-\lambda x(\frac{\lambda - 2t\mu^2}{2\lambda\mu^2}) - \frac{\lambda}{2x} + \frac{\lambda}{\mu}\right) dx \\
&= exp\left(\frac{\lambda}{\mu} - \frac{\lambda}{\mu_t}\right) \int_0^\infty \frac{\sqrt{\lambda}}{\sqrt{2\pi x^3}} exp\left(\frac{-\lambda x}{2\mu_t^2} - \frac{\lambda}{2x} + \frac{\lambda}{\mu_t}\right) dx.
\end{aligned}$$

where

$$\mu_t = \left(\frac{\lambda\mu^2}{\lambda - 2t\mu^2}\right)^{\frac{1}{2}}$$

and assuming that $t < \frac{\lambda}{2\mu^2}$, hence we have :

$$\begin{aligned}
M_X(t) &= exp\left(\frac{\lambda}{\mu} - \frac{\lambda}{\mu_t}\right) \\
&= exp\left(\frac{\lambda}{\mu}\left(1 - \left(1 - \frac{2t\mu^2}{\lambda}\right)^{1/2}\right)\right).
\end{aligned}$$

From Theorem 1.4.2, we can conclude the following results:

Corollary 1.4.1. *Let X be a random variable have an $IG(\mu, \lambda)$ distribution, then the cumulant moment function of X is given by :*

$$K_X(t) = \frac{\lambda}{\mu}\left(1 - \left(1 - \frac{2t\mu^2}{\lambda}\right)^{1/2}\right), \mu > 0, \lambda > 0 \qquad (1.4.3)$$

Proof : Taking the logarithm both sides in equation (1.4.2), we get the cumulant moment

$$K_X(t) = \frac{\lambda}{\mu}\left(1 - \left(1 - \frac{2t\mu^2}{\lambda}\right)^{1/2}\right).$$

Corollary 1.4.2. *Let X be a random variable have an $IG(\mu, \lambda)$distribution, then the mean of X is given by :*

$$E(X) = \mu \tag{1.4.4}$$

Proof: Using the cumulant moment we can get the mean,

$$K'_X(0) = E(X)$$

so taking the first derivative both sides in equation (1.4.3) we have :

$$K'_X(t) = \frac{\lambda}{\mu}\left(\left(\frac{1}{2}\left(1 - \frac{2t\mu^2}{\lambda}\right)^{-1/2}\right)\frac{-2\mu^2}{\lambda}\right)$$

putting $t = 0$,we have :

$$E(X) = K'_X(0) = \mu$$

Corollary 1.4.3. *Let X be a random variable have an $IG(\mu, \lambda)$ distribution, then the variance of X is given by :*

$$Var(X) = \frac{\mu^3}{\lambda} \tag{1.4.5}$$

Proof : To find the variance we use the cumulant moment:

$$K''_X(0) = Var(X)$$

so taking the second derivative both sides in equation (1.4.3) we get

$$K''_X(t) = \frac{-\mu}{2}\left(1 - 2t\frac{\mu^2}{\lambda}\right)^{-3/2}\frac{-2\mu^2}{\lambda}$$

putting $t = 0$ we get the variance :

$$\begin{aligned} Var(X) &= K''_X(0) \\ &= \frac{\mu^3}{\lambda} \end{aligned}$$

Chapter 2

Kernel Estimator of the Hazard Rate Function

In this chapter we will discuss some methods used for smoothing the hazard rate function to specify the appropriate method which we consider in our thesis especially in Chapter 3. The Chapter is divided in three sections. Section one presents the definition of the hazard rate function and we discuss some of its properties. Section two proposes the kernel estimator for the hazard rate function and we discuss its properties and we present the Gaussian kernel estimator as an example and in order to compare it with the IG kernel estimator in Chapter 4.

2.1 Hazard Rate Function

Survival analysis is a branch of statistics that deals with analysis of time duration until one or more events happen, such as death in biological organisms and failure in mechanical systems. This topic is called reliability theory or reliability analysis in engineering, duration analysis or duration modeling in economics, and event history analysis in sociology.

Theorem 2.1.1. *The hazard rate function $r(.)$ is defined as*

$$r(x) = \lim_{\Delta x \longrightarrow 0} \frac{P(X \leq x + \Delta x \mid X > x)}{\Delta x}, x > 0 \tag{2.1.1}$$

and it can be written as

$$r(x) = \frac{f(x)}{1 - F(x)} = \frac{f(x)}{S(x)} \tag{2.1.2}$$

where $f(.)$ and $F(.)$ are the density and distribution function of a continuous random variable X respectively, and S(x) is the Survival function.

Proof :

Let $A := X > x$ (meaning the life time greater than x) and $B := X > x + y$ this implies that $A \cap B := X > x + y$, hence $P(A) = P(X > x) = S(x) = 1 - F(x)$ and $P(A \cap B) = P(X > x + y) = S(x + y)$, then we have :

$$P(X > x + y|X > x) = \frac{S(x + y)}{S(x)} \tag{2.1.3}$$

now since $P(X > x + y|X > x) = P(X - x > y|X > x)$, letting $Y = X - x$ and by Equation (2.1.3) we have :

$$S(Y|X > x) = \frac{S(x + y)}{S(x)} \tag{2.1.4}$$

taking the complement of Equation (2.1.4) we get the conditional distribution function :

$$\begin{aligned} F(y|X > x) &= 1 - S(y|X > x) \\ &= 1 - \frac{S(x + y)}{S(x)} \\ &= \frac{S(x) - S(x + y)}{S(x)} \\ &= \frac{F(x + y) - F(x)}{1 - F(x)} \end{aligned} \tag{2.1.5}$$

taking the derivative of Equation (2.1.5) we get the conditional failure density :

$$\begin{aligned} f(y|X > x) &= \frac{d}{dy}(F(y|X > x)) \\ &= \frac{d}{dy}(\frac{F(x + y) - F(x)}{1 - F(x)}) \\ &= \frac{f(x + y)}{1 - F(x)} \\ &= \frac{f(x + y)}{S(x)} \end{aligned} \tag{2.1.6}$$

now For small Δ, we have:

$$\frac{P(x < X < x + \Delta | X > x)}{\Delta} \approx f(\Delta | X > x)$$
$$= \frac{f(x + \Delta)}{S(x)}$$

letting $\Delta \to 0$ we get the hazard rate as :

$$\begin{aligned}
r(x) &= \lim_{\Delta \to 0} \frac{P(x < X < x + \Delta | X > x)}{\Delta} \\
&= \lim_{\Delta \to 0} f(\Delta | X > x) \\
&= \lim_{\Delta \to 0} \frac{f(x + \Delta)}{S(x)} \\
&= \frac{f(x)}{S(x)} \\
&= \frac{f(x)}{1 - F(x)} \qquad (2.1.7)
\end{aligned}$$

Example 2.1.1. *If we suppose that $f(x)$ is an exponential distribution, then we have : $f(x) = \alpha e^{-\alpha x}$ and the cdf is given by $F(x) = 1 - e^{-\alpha x}$ which imply that $S(x) = 1 - F(x) = e^{-\alpha x}$. Hence the hazard rate function is given by*

$$\begin{aligned}
r(x) &= \frac{f(x)}{1 - F(x)} \\
&= \frac{\alpha e^{-\alpha x}}{e^{-\alpha x}} \\
&= \alpha.
\end{aligned}$$

Therefore, an exponential failure density corresponds to a constant hazard function.

2.2 Estimation the Hazard Rate Function

Several methods for estimation the hazard rate function have been studied. The non parametrically method has the advantage of flexibility because no formal assumptions are made about the mechanism that generates the sample order than the randomness, such as the kernel estimation as we discussed in section 1.2. In this section, we will discuss this estimator and we discuss its properties under the assumptions we assumed in Chapter 1 and we present the definition of the boundary effect which hold due to the symmetric kernels.

Definition 2.2.1. *If $X_1, X_2, .., X_n$ is a random sample distributed as X , then Watson and leadbetter [21] proposed the following estimator for $r(.)$*

$$\hat{r}(x) = \frac{\hat{f}(x)}{1 - \hat{F}(x)} \tag{2.2.1}$$

where $\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{x-X_i}{h})$, *and* $\hat{F}(x) = \frac{1}{nh}\sum_{i=1}^{n}\int_{-\infty}^{x} K(\frac{u-X_i}{h})du$

where K is a bounded and symmetric kernel, integrating to one.

Conditions

The following conditions will be used in the sequel:

C1 *Suppose that the kernel function K satisfies the following:*
(i) K is asymmetric density function.
(ii) $\lim_{y\longrightarrow\infty} |y|K(y) = 0$
(iii) $\int_{-\infty}^{\infty} K^2(u)du < \infty$
(iv) $\int_{-\infty}^{\infty} uK(u)du = 0$
(v) $\int_{-\infty}^{\infty} u^2K(u)du < \infty$

C2 *Suppose that the bandwidth h satisfies the following:*
(i) $h \to 0$.
(ii) $nh \to \infty$.
(iii) $nh^5 \to 0$.

C3 *f'' exists and integrable.*

Next, we discuss properties of the estimator given in Equation 2.2.2. The mean, variance and bias of the kernel estimator for the hazard rate function will discuss, the MSE and $AMSE$ will present and the asymptotic normality will investigate.

Remark 2.2.1. *By Remark 1.3.2 , we have $\hat{F}(x) \xrightarrow{p} F(x)$. Hence, since $\hat{S}(x) = 1 - \hat{F}(x)$ we get that $\hat{S}(x) \xrightarrow{p} S(x)$.*

Theorem 2.2.1. *Under the assumptions C_1,C_2 and C_3 we have :*

$$Bias(\hat{r}(x)) = \frac{f^{(2)}(x)}{2S(x)}h^2\mu_2(K) + o(h^2) \tag{2.2.2}$$

and

$$Var(\hat{r}(x)) = \frac{r(x)}{S(x)nh}R(K) + o(\frac{1}{nh}) \tag{2.2.3}$$

Proof: *For the first part, we find first the mean, by Remark 2.2.1 we have:*

$$\hat{S}(x) \xrightarrow{p} S(x)$$

hence,

$$E(\hat{r}(x)) = \frac{E(\hat{f}(x))}{S(x)} \tag{2.2.4}$$

Now, by proof of Theorem 1.3.3 (with $r = 2$*), we have*

$$E(\hat{f}(x)) = f(x) + \frac{1}{2} f^{(2)}(x) h^2 \mu_2(K) + o(h^2) \tag{2.2.5}$$

using Equation 2.2.5 in Equation 2.2.4 we have :

$$E(\hat{r}(x)) = \frac{f(x) + \frac{1}{2} f^{(2)}(x) h^2 \mu_2(K)}{S(x)} + o(h^2) \tag{2.2.6}$$

then , the bias is given by :

$$\begin{aligned} Bias(\hat{r}(x)) &= E(\hat{r}(x)) - r(x) \\ &= \frac{\frac{1}{2} f^{(2)}(x) h^2 \mu_2(K)}{S(x)} + o(h^2) \end{aligned}$$

For the second part, we have

$$\begin{aligned} Var(\hat{r}(x)) &= \frac{Var(\hat{f}(x))}{S(x)^2} \\ &= \frac{1}{S(x)^2} \frac{f(x)R(K)}{nh} + o(\frac{1}{nh}), \qquad \textit{(by Equation 1.3.9)} \\ &= \frac{r(x)R(K)}{S(x)nh} + o(\frac{1}{nh}) \end{aligned}$$

Remark 2.2.2. *From Equations 2.2.2 and 2.2.3 we can see that the bias increasing in* h^2 *and the variance decreasing in* nh*, hence under the assumptions we have* $Bias(\hat{r}(x)) \to 0$ *and* $Var(\hat{r}(x)) \to 0$ *as* $n \to \infty$*.*

Corollary 2.2.1. *The mean squared error of the kernel estimator for the hazard rate function is given by :*

$$MSE(\hat{r}(x)) = \left(\frac{f^{(2)}(x)}{2S(x)} h^2 \mu_2(K) + o(h^2) \right)^2 + \frac{r(x)R(K)}{S(x)nh} + o(\frac{1}{nh}) \tag{2.2.7}$$

Proof: *By Equations 2.2.2 and 2.2.3 and the definition of the mean squared error.*

Corollary 2.2.2. *The asymptotic mean squared error of the kernel estimator for the hazard rate function is given by :*

$$AMSE(\hat{r}(x)) = \left(\frac{f^{(2)}(x)}{2S(x)} h^2 \mu_2(K)) \right)^2 + \frac{r(x)R(K)}{S(x)nh} \tag{2.2.8}$$

Proof: *By letting* $o(h^2) \to 0$ *and* $o(\frac{1}{nh}) \to 0$ *in Equation 2.2.7.*

Remark 2.2.3. *Under the assumptions* C_1,C_2 *and* C_3 *and Equation 2.2.8, we have* $AMSE(\hat{r}(x)) \to 0$ *as* $n \to \infty$.

2.2.1 The Asymptotic Normality of the Kernel Estimator $\hat{r}(x)$

Theorem 2.2.2. *Under the assumptions* C_1,C_2 *and* C_3 *with additional condition* $(nh^5)^{\frac{1}{2}} \to 0$ *as* $n \to \infty$*, we have :*

$$\sqrt{nh}\left(\hat{r}(x) - r(x)\right) \xrightarrow{d} N\left(0, \frac{r(x)}{S(x)}R(K)\right) \tag{2.2.9}$$

Proof: *We have :*

$$\begin{aligned}
\sqrt{nh}\left(\hat{r}(x) - r(x)\right) &= \sqrt{nh}\left(\frac{\hat{f}(x)}{\hat{S}(x)} - \frac{f(x)}{S(x)}\right) \\
&= \sqrt{nh}\left(\frac{\hat{f}(x)}{\hat{S}(x)} - \frac{f(x)}{\hat{S}(x)} - \frac{f(x)}{S(x)} + \frac{f(x)}{\hat{S}(x)}\right) \\
&= \frac{\sqrt{nh}}{\hat{S}(x)}\left(\hat{f}(x) - f(x) - \frac{f(x)\hat{S}(x)}{S(x)} + f(x)\right) \\
&= \frac{\sqrt{nh}}{\hat{S}(x)}\left(\hat{f}(x) - f(x)\right) + \frac{\sqrt{nh}f(x)}{S(x)\hat{S}(x)}\left(\hat{S}(x) - S(x)\right) \\
&\xrightarrow{d} \frac{1}{S(x)}N\left(0, f(x)R(K)\right) \\
&\xrightarrow{d} N\left(0, \frac{r(x)}{S(x)}R(K)\right)
\end{aligned}$$

because Remark 2.2.1 (the second term vanishes) and Theorem 1.2.1 complete the proof.

Bandwidth selection:

In order to find the optimal bandwidth we will consider the same way discussed in Chapter 1, so first we take the derivative of the Equation 2.2.7 with respect to h *and Equating it to zero we get :*

$$4h^3\frac{f''(x)^2\mu_2(K)^2}{S^2(x)} = h^{-2}\frac{r(x)R(K)}{S(x)n} \tag{2.2.10}$$

Multiplying Equation 2.2.10 by h^2 and solving for h we get the optimal bandwidth that minimizes AMSE:

$$h^* = \left(\frac{f(x)R(K)}{f''(x)^2\mu_2(K)^2}\right)^{\frac{1}{5}} n^{-\frac{1}{5}} \tag{2.2.11}$$

Note that the optimal bandwidth is of order $o(n^{-\frac{1}{5}})$. Substituting the result 2.2.10 in Equation 2.2.7 and after some simplification we have :

$$\begin{aligned} AMSE^*(\hat{r}(x)) &= \left(\frac{f''(x)^{\frac{1}{5}}}{2S(x)}\mu_2(K)^{\frac{1}{5}}f(x)^{\frac{2}{5}}R(K)^{\frac{2}{5}}\right)^2 n^{-\frac{4}{5}} + \left(\frac{f''(x)^{\frac{1}{5}}}{S(x)}\mu_2(K)^{\frac{1}{5}}f(x)^{\frac{2}{5}}R(K)^{\frac{2}{5}}\right)^2 n^{-\frac{4}{5}} \\ &= \frac{5}{4}\left(\frac{f''(x)^{\frac{1}{5}}}{S(x)}\mu_2(K)^{\frac{1}{5}}f(x)^{\frac{2}{5}}R(K)^{\frac{2}{5}}\right)^2 n^{-\frac{4}{5}} \end{aligned} \tag{2.2.12}$$

2.2.2 The Gaussian Kernel Estimator for The Hazard Rate Function

Next we discuss the Gaussian kernel estimator for the hazard rate function as an example of the estimation the hazard rate function using symmetric kernels. The mean, variance and bias will be investigated and we will find the AMSE in order to make a comparison with the IG kernel estimator which we present in Chapter 4.

Definition 2.2.2. *The Gaussian kernel $K_G(x)$ is defined by :*

$$K_G(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}, for\ all\ x \in \Re. \tag{2.2.13}$$

Hence, by Definition 1.3.1, the Gaussian kernel estimator of the pdf is given as in the next definition.

Definition 2.2.3. *The Gaussian kernel estimator of the pdf is given by :*

$$\hat{f}_G(x) = \frac{1}{nh}\sum_{i=1}^{n} K_G\left(\frac{x-X_i}{h}\right) \tag{2.2.14}$$

where $K_G(x)$ as in definition 2.2.2 .

Next, the G kernel estimator for the cdf will be presented.

Definition 2.2.4. *The Gaussian kernel estimator of the pdf is given by :*

$$\hat{F}_G(x) = \frac{1}{nh}\sum_{i=1}^{n}\int_0^x K_G\left(\frac{u-X_i}{h}\right) du \tag{2.2.15}$$

Now by using $\hat{S}_G(x) = 1 - \hat{F}_G(x)$, we present the G kernel estimator for the hazard rate function.

Definition 2.2.5. *The G kernel estimator for the hazard rate function is given by :*

$$\hat{r}_G(x) = \frac{\hat{f}_G(x)}{1 - \hat{F}_G(x)} = \frac{\hat{f}_G(x)}{\hat{S}_G(x)} \tag{2.2.16}$$

where, $\hat{f}_G(x)$ and $\hat{F}_G(x)$ as in definitions 2.2.3 and 2.2.4 respectively.

Now we give the properties of $\hat{r}_G(x)$.

Theorem 2.2.3. *Under the assumptions C_1,C_2 and C_3 we have,*

$$Bias(\hat{r}_G(x)) = \frac{f''(x)h^2}{2S(x)} + o(h^2) \tag{2.2.17}$$

and

$$Var(\hat{r}_G(x)) = \frac{f(x)}{2nh\sqrt{\pi}S^2(x)} + o(\frac{1}{nh}) \tag{2.2.18}$$

Proof: *To prove the first part, we use Equation 2.2.2 with substituting the value of ($\mu_2(K_G) = 1$). Similarly for the second part we use Equation 2.2.3 with substituting the value of ($R(K_G) = \frac{1}{2\sqrt{\pi}}$).*

The next corollaries present the $MSE(\hat{r}_G(x))$ and $AMSE(\hat{r}_G(x))$.

Corollary 2.2.3. *The mean squared error of the Gaussian kernel estimator of the hazard rate function is given by :*

$$MSE(\hat{r}_G(x)) = \left(\frac{f''(x)h^2}{2S(x)} + o(h^2)\right)^2 + \frac{r(x)}{2nh\sqrt{\pi}S(x)} + o(\frac{1}{nh}) \tag{2.2.19}$$

Proof: *Using Equations 2.2.12 and 2.2.13 and by the definition of MSE directly we get the results.*

Corollary 2.2.4. *The asymptotic mean squared error of the Gaussian kernel estimator of the hazard rate function is given by :*

$$AMSE(\hat{r}_G(x)) = \left(\frac{f''(x)h^2}{2S(x)}\right)^2 + \frac{r(x)}{2nh\sqrt{\pi}S(x)} \tag{2.2.20}$$

Proof: *By letting $o(h^2) \to 0$ and $o(\frac{1}{nh}) \to 0$ in Equation 2.2.14 we get the result.*

The asymptotic normality of the kernel estimator $\hat{r}_G(x)$ presented in the next theorem.

Theorem 2.2.4. *Under the assumptions C_1,C_2 and C_3 with additional condition $(nh^5)^{\frac{1}{2}} \to 0$ as $n \to \infty$, we have :*

$$\sqrt{nh}\,(\hat{r}_G(x) - r(x)) \xrightarrow{d} N\left(0, \frac{r(x)}{2\sqrt{\pi}S(x)}\right) \tag{2.2.21}$$

Proof: *By Theorem 2.2.2 with substituting the value of ($R(K_G) = \frac{1}{2\sqrt{\pi}}$).*

Corollary 2.2.5. *The optimal bandwidth of the Gaussian kernel estimator for the hazard rate function is given by:*

$$h^* = \left(\frac{f(x)}{2\sqrt{\pi}f''(x)^2}\right)^{\frac{1}{5}} n^{-\frac{1}{5}} \tag{2.2.22}$$

Proof: *By Equation 2.2.11 with substituting the values of $R(K_G)$ and $\mu_2(K_G)$.*

Corollary 2.2.6. *The optimal mean squared error of the Gaussian kernel estimator for the hazard rate function is given by:*

$$AMSE^*(\hat{r}_G(x)) = \frac{5}{4}\left(f''(x)^{\frac{1}{5}}f(x)^{\frac{2}{5}}(2\sqrt{\pi})^{\frac{2}{5}}S(x)\right)^2 n^{-\frac{4}{5}} \tag{2.2.23}$$

Proof: *By Equation 2.2.12 with substituting the values of $R(K_G)$ and $\mu(K_G)$. Regarding the global properties the optimal bandwidths and mean integrated squared errors can be derived by taking the the integration of Equations 2.2.22 and 2.2.23 and using that $\int_0^\infty f(x)dx = 1$ we have :*

$$h^{**} = \left(\frac{1}{2\sqrt{\pi}\int_0^\infty f''(x)^2 dx}\right)^{\frac{1}{5}} n^{-\frac{1}{5}} \tag{2.2.24}$$

and

$$AMISE^{**}(\hat{r}_G(x)) = \frac{5}{4}\left(\frac{\int_0^\infty f''(x)^{\frac{1}{5}}dx}{(2\sqrt{\pi})^{\frac{2}{5}}\int_0^\infty S(x)dx}\right)^2 n^{-\frac{4}{5}} \tag{2.2.25}$$

2.2.3 Practical Optimal Bandwidth:

Now we consider the rule of the thumb were [17] used to find the practical optimal bandwidth(see Remark 1.3.1).

Example 2.2.1. *If we consider the standard normal distribution as reference distribution (with variance σ^2) and setting f_G be the standard normal density in Equation*

2.2.24, we have:

$$\begin{aligned} R(f'') &= \int_0^\infty (f''(x))^2 dx \\ &= \sigma^{-5} \int_0^\infty (f_G^{(2)}(x))^2 dx \\ &= \frac{3}{8}\pi^{-\frac{1}{2}}\sigma^{-5}, \end{aligned}$$

then, the rule of thumb gives :

$$\begin{aligned} h^{**} &= (4\pi)^{-\frac{1}{10}}(\frac{3}{8}\pi^{-\frac{1}{2}})^{-\frac{1}{5}}\sigma n^{-\frac{1}{5}} \\ &= 1.06\sigma n^{-\frac{1}{5}}. \end{aligned}$$

Note that h^{**} depends on estimating σ from the data , hence by taking $\hat{\sigma}^2 = \frac{\sum(x_i-\bar{x})}{n-1}$, which means that :

$$h^{**} = 1.06\hat{\sigma}n^{-\frac{1}{5}}. \qquad (2.2.26)$$

Boundary Effect

Note that the support of the hazard rate function is in the non-negative part of the real line $[0,\infty)$, so when the estimation is based on symmetric kernels its will be under the **boundary effect** (called **a boundary bias** problem) near the zero, its causes that the estimator of the hazard rate function will take values outside the support.

To solve this problem, Chen replaced the symmetric kernels by asymmetric Gamma kernel which never assigns weight outside the support. Scaillet used this idea and proposed two new classes of density estimators, rely on the use of inverse Gaussian IG and the RIG kernels in place of the Gamma kernel. In [14], the estimation of the hazard rate function using the IG kernel has been considered. In Chapter 3, we will consider the nonparametric estimation of the hazard rate function for (iid) data using the IG kernel based.

Chapter 3

Estimation of the Hazard Rate Function Using the IG Kernel

In this chapter, we will study the inverse Gaussian(IG) kernel estimator that can be used as a non-parametrically estimation for the hazard rate function. In Chapter 1, Section 4. we studied the inverse Gaussian distribution and we discussed some of its properties, such as the flexible shape and its support in the non-negative part of the real line. The IG kernel estimator is free of boundary bias. In parallel its achieves the optimal rate of convergence for the mean integrated squared error ($MISE$) within the class of non-negative kernel density estimators, (See Scaillet2004) [15]. In Section 3.1 we discuss the inverse Gaussian kernel estimator of the pdf and cdf for independent and identically distributed data, Section 3.2 will contain a study for the (IG) kernel estimator for the hazard rate function.

3.1 The IG Kernel Estimator

In this section we state first the conditions under which the results of this chapter will be proved. Also we present the IG kernel estimator of the pdf and the cdf for independent and identically distributed data, then we will derive the asymptotic normality and the strong consistency of the proposed estimator.

Conditions

1. Let $X_1, X_2, \cdots, X_n$ be a random sample from a distribution with an unknown probability density function f defined on $[0, \infty)$, such that f is twice continuously differentiable, and $\int_0^\infty (x^3 f''(x))^2 dx < \infty$.

2. h is a smoothing parameter satisfying $h + \frac{1}{nh} \to 0$, and $nh^{\frac{5}{2}} \to 0$, as $n \to \infty$.

Under the previous two conditions the inverse Gaussian kernel $K_{IG}(u)$ and the (IG) kernel estimators of the pdf and the cdf will state in the following definitions.

Definition 3.1.1. *[15] The inverse Gaussian (IG) kernel is defined by :*

$$K_{IG(x,\frac{1}{h})}(u) = \frac{1}{\sqrt{2\pi h u^3}} exp\left(-\frac{1}{2hx}\left(\frac{u}{x} - 2 + \frac{x}{u}\right)\right), u > 0 \quad (3.1.1)$$

where $h + \frac{1}{nh} \to 0$ as $n \to \infty$.

Using this kernel, the inverse Gaussian (IG) pdf kernel estimator which proposed by Scaillet [15] is defined as follows :

Definition 3.1.2. *[15] The inverse Gaussian (IG) kernel estimator of the pdf is defined by :*

$$\hat{f}_{IG}(x) = \frac{1}{n}\sum_{i=1}^{n} K_{IG(x,\frac{1}{h})}(X_i) \quad (3.1.2)$$

Definition 3.1.3. *The inverse Gaussian (IG)kernel estimator of the cdf is defined by :*

$$\hat{F}_{IG}(x) = \int_0^x \hat{f}_{IG}(u)du = \frac{1}{n}\sum_{i=1}^{n}\int_0^x K_{IG(u,\frac{1}{h})}(X_i)du. \quad (3.1.3)$$

The Properties of the IG Kernel Estimator

Next, we discuss the properties of $\hat{f}_{IG}(x)$ and the estimator of the cdf $\hat{F}_{IG}(x)$. Firstly, the variance and the bias of the inverse Gaussian kernel estimator of the pdf $\hat{f}_{IG}(x)$ will be investigated in the following theorem.

Theorem 3.1.1. *[15] Under conditions C_1 and C_2, we have :*

$$Bias(\hat{f}_{IG}(x)) = \frac{1}{2}x^3 f''(x)h + o(h). \quad (3.1.4)$$

and

$$Var(\hat{f}_{IG}(x)) = \frac{1}{2n\sqrt{\pi h}} x^{-\frac{3}{2}} f(x) + o(n^{-1}h^{-\frac{1}{2}}) \quad (3.1.5)$$

Proof:

$$E(\hat{f}_{IG}(x)) = \int_0^\infty K_{IG}(x, \frac{1}{h})(u) f(u) du$$
$$= E(f(\zeta_x)), \quad (3.1.6)$$

where, ζ_x follows a $IG(x, \frac{1}{h})$ distribution. where, $\mu_x = E(\zeta_x) = x$ and $V_x = Var(\zeta_x) = x^3 h$. Using this and since f is twice continuously differentiable(by condition 2), we can expand $f(\zeta_x)$ about μ_x using Taylor series (Theorem 1.1.5) as follows:

$$f(\zeta_x) = f(\mu_x) + f'(\mu_x)(\zeta_x - \mu_x) + \frac{f''(\mu_x)}{2!}(\zeta_x - \mu_x)^2 + o(h)$$

Next we have,

$$E(f(\zeta_x)) = E(f(\mu_x)) + E(f'(\mu_x))E((\zeta_x - \mu_x)) + E(\frac{f''(\mu_x)}{2!})E((\zeta_x - \mu_x)^2) + o(h)$$
$$= f(x) + f'(x)(\mu_x - \mu_x) + \frac{f''(x)}{2!} V_x + o(h)$$
$$= f(x) + \frac{x^3}{2} f''(x) h + o(h). \quad (3.1.7)$$

Hence,

$$Bias(\hat{f}_{IG}(x)) = E(\hat{f}_{IG}(x)) - f(x)$$
$$= E(f(\zeta_x)) - f(x), \quad \text{(using Equation (3.1.6))}$$
$$= f(x) + \frac{x^3}{2} f''(x) h + o(h) - f(x), \quad \text{(using Equation (3.1.7))}$$

Therefore,

$$Bias(\hat{f}_{IG}(x)) = \frac{1}{2} x^3 f''(x) h + o(h).$$

Next, we prove the second part of the theorem, note that:

$$Var(\hat{f}_{IG}(x)) = \frac{1}{n} Var(K_{IG}(x, \frac{1}{h})(X_i))$$
$$= \frac{1}{n} E(K_{IG}(x, \frac{1}{h})(X_i)^2) + o(\frac{1}{h}). \quad (3.1.8)$$

Equation (3.1.7) can be proved in the same way as in Theorem 1.2.2.

$$
\begin{aligned}
E\left(K_{IG}(x,\frac{1}{h})(X_i)^2\right) &= \int_0^{\infty} \frac{1}{2\pi h u^3} exp(-\frac{1}{hx}(\frac{u}{x}-2+\frac{x}{u}))f(u)du \\
&= \frac{1}{\sqrt{4\pi h}} \int_0^{\infty} \frac{1}{\sqrt{\pi \frac{h}{2} u2}} u^{-\frac{3}{2}} exp(-\frac{1}{hx}(\frac{u}{x}-2+\frac{x}{u}))f(u)du \\
&= \frac{1}{\sqrt{4\pi h}} \int_0^{\infty} u^{-\frac{3}{2}} f(u) \frac{\sqrt{\frac{2}{h}}}{\sqrt{2\pi u}} exp(-\frac{1}{2hx}(\frac{u}{x}-2+\frac{x}{u}))du \\
&= \frac{1}{\sqrt{4\pi h}} E\left(\eta_x^{-\frac{3}{2}} f(\eta_x)\right),
\end{aligned}
\tag{3.1.9}
$$

where, η_x follows a $IG\left(x, \frac{2}{h}\right)$.

Since $\mu_x = E(\eta_x) = x$ and $V_x = Var(\eta_x) = \frac{x^3 h}{2}$, and we obtain by Taylor series :

$$
\begin{aligned}
E(\eta_x^{-\frac{3}{2}} f(\eta_x)) &= \mu_x^{-\frac{3}{2}} f(\mu_x) + \frac{1}{2}(\frac{15}{4} x^{-\frac{7}{2}} f(x) - x^{-\frac{5}{2}} f'(x) + x^{-\frac{3}{2}} f''(x))V_x + o(h) \\
&= \mu_x^{-\frac{3}{2}} f(\mu_x) + \frac{1}{4}(\frac{15}{4} x^{-\frac{1}{2}} f(x) - 3x^{-\frac{1}{2}} f'(x) + x^{\frac{3}{2}} f''(x))h + o(h) \\
&= \mu_x^{-\frac{3}{2}} f(\mu_x) + o(h) \\
&= x^{-\frac{1}{2}} f(x) + o(h)
\end{aligned}
\tag{3.1.10}
$$

Now, using the Equations (3.1.10),(3.1.9) in (3.1.8) we get :

$$
Var(\hat{f}_{IG}(x)) = \frac{1}{2n\sqrt{\pi h}} x^{-\frac{3}{2}} f(x) + o(n^{-1}h^{-\frac{1}{2}})
$$

From Theorem 3.1.1 and under the condition 1 and 2, we conclude that $\hat{f}_{IG}(x)$ appears to have the following **asymptotic properties**:

1. The IG kernel estimator $\hat{f}_{IG}(x)$ is free **boundary bias**, because its bias $Bias(\hat{f}_{IG}(x))$ is of order $o(h)$ in the interior of $[0, \infty)$ and near zero.

2. Since the expression of the $Bias(\hat{f}_{IG}(x))$ depends on $x^3 f''(x)$, then as $\int_0^{\infty} (x^3 f''(x))^2 < \infty$ (by C_1), the terms $x^3 f''(x) \to 0$ and hence $x^3 f''(x) \to 0$ as $x \to \infty$. So the $Bias(\hat{f}_{IG}(x))$ is smaller as x increases.

3. By Equation 3.1.4, we deduce that the $Bias(\hat{f}_{IG}(x))$ increase in h, and another look to Equation 3.1.5, we can deduce that the $Var(\hat{f}_{IG}(x))$ is decrease in $nh^{\frac{1}{2}}$. Hence by assuming that as $n \to \infty$ we must have $h \to 0$ and $nh^{\frac{1}{2}} \to 0$ were hold by C_2 we can establish the strong consistency of the estimator.

Asymptotic Normality of the Estimator $\hat{f}_{IG}(x)$:

Theorem 3.1.2. *Under the conditions C_1 and C_2 the following holds :*

$$\sqrt{nh^{\frac{1}{2}}}(\hat{f}_{IG}(x) - f(x)) \xrightarrow{d} N(0, \tfrac{1}{2\sqrt{\pi}}x^{-\frac{3}{2}}f(x)).$$

Proof: Let $V_{ni} = K_{IG}(x, \frac{1}{h})(X_i)$, $i = 1, 2, \cdots, n$, then by Definition 3.1.2, $\hat{f}_{IG}(x)$ can be written as : $\hat{f}_{IG}(x) = \frac{1}{n}\sum_{i=1}^{n} V_{ni}$, where V_{ni}, $i = 1, 2, \cdots, n$ are (iid) random variables.

Next, we show that Liapounove condition (Theorem 1.1.4) is satisfied, that is for some $\delta > 0$,

$$\lim_{n\to\infty} \frac{E|V_n - E(V_n)|^{2+\delta}}{n^{\frac{\delta}{2}}\sigma^{2+\delta}(V_n)} = 0. \tag{3.1.11}$$

Assume that η_x follows a $IG(x, \frac{2+\delta}{h})$ distribution. where, $\mu_x = E(\eta_x) = x$ and $T_x = Var(\eta_x) = \frac{x^3}{2+\delta}$.

Hence,

$$\begin{aligned}
E|V_n|^{2+\delta} &= E\left[\left(\frac{1}{\sqrt{2\pi hy}}\right)^{2+\delta} exp\left(\left(-\frac{(2+\delta)(x-h)}{2h}\right)\left(\frac{y}{x-h} - 2 + \frac{x-h}{y}\right)\right)\right] \\
&= \frac{1}{(2\pi h)^{\left(1+\frac{\delta}{2}\right)}}\int_0^\infty y^{-\left(1+\frac{\delta}{2}\right)}\left[exp\left(\left(-\frac{(2+\delta)(x-h)}{2h}\right)\left(\frac{y}{x-h} - 2 + \frac{x-h}{y}\right)\right)\right] f(y)dy \\
&= \frac{\sqrt{2\pi h}}{\sqrt{2+\delta}\,(2\pi h)^{\left(1+\frac{\delta}{2}\right)}}\int_0^\infty y^{-\frac{1}{2}(1+\delta)} K_{IG}\left(x, \frac{2+\delta}{h}\right)(y) f(y)dy \\
&= \frac{\sqrt{2\pi h}}{\sqrt{2+\delta}\,(2\pi h)^{\left(1+\frac{\delta}{2}\right)}} E\left(\eta_x^{-\frac{3}{2}(1+\delta)} f(\eta_x)\right).
\end{aligned} \tag{3.1.12}$$

Using Taylor's series we expand $\eta_x^{-\frac{3}{2}(1+\delta)} f(\eta_x)$ about μ_x as :

$$\begin{aligned}
\eta_x^{-\frac{3}{2}(1+\delta)} f(\eta_x) &= \mu_x^{-\frac{3}{2}(1+\delta)} f(\mu_x) + (\mu_x^{-\frac{3}{2}(1+\delta)} f'(\mu_x) - \frac{3}{2}\mu_x^{-\frac{5}{2}(1+\delta)} f(\mu_x))(\eta_x - \mu_x) \\
&\quad + (\mu_x^{-\frac{3}{2}(1+\delta)} f''(\mu_x) - \frac{3}{2}\mu_x^{-\frac{5}{2}(1+\delta)} f'(\mu_x) - \frac{3}{2}\mu_x^{-\frac{5}{2}(1+\delta)} f(\mu_x) \\
&\quad + \frac{15}{4}\mu_x^{-\frac{7}{2}(1+\delta)} f(\mu_x))(\eta_x - \mu_x)^2 + o(h)
\end{aligned}$$

Hence,

$$\begin{aligned}
E(\eta_x^{-\frac{3}{2}(1+\delta)}f(\eta_x)) &= x^{-\frac{3}{2}(1+\delta)}f(x) + \frac{1}{2}(x^{-\frac{3}{2}(1+\delta)}f''(x) - \frac{3}{2}x^{-\frac{5}{2}(1+\delta)}f' \\
&\quad - \frac{3}{2}x^{-\frac{5}{2}(1+\delta)}f') + \frac{15}{4}x^{-\frac{7}{2}(1+\delta)}f(x))T_x o(h) \\
&= x^{-\frac{3}{2}(1+\delta)}f(x) + \frac{1}{2(2+\delta)}(x^{-\frac{3}{2}(1+\delta)}f''(x) - \frac{3}{2}x^{-\frac{5}{2}(1+\delta)}f' \\
&\quad - \frac{3}{2}x^{-\frac{5}{2}(1+\delta)}f') + \frac{15}{4}x^{-\frac{7}{2}(1+\delta)}f(x))x^3h + o(h) \\
&= x^{-\frac{3}{2}(1+\delta)}f(x) + o(h).
\end{aligned}$$

subtituing in equation (3.1.12) this implies that :

$$E|V_n|^{2+\delta} = \frac{1}{\sqrt{2+\delta}(2\pi h)^{\frac{1+\delta}{2}}}x^{-\frac{3}{2}(1+\delta)}f(x) + o(h^{-\frac{(1+\delta)}{2}})$$

Now, substitute $\delta = 0$ we get :

$$Var(V_n) = \frac{1}{2\sqrt{\pi}}h^{-\frac{1}{2}}x^{-\frac{3}{2}}f(x) + o(h^{-\frac{1}{2}}).$$

using this we have :

$$\begin{aligned}
\frac{E|V_n - E(V_n)|^{2+\delta}}{n^{\frac{\delta}{2}}\sigma^{2+\delta}(V_n)} &\leq \frac{E|V_n|^{2+\delta}}{n^{\frac{\delta}{2}}\left(\frac{1}{2\sqrt{\pi}}h^{-\frac{1}{2}}x^{-\frac{3}{2}}f(x)\right)^{\frac{2+\delta}{2}}} \\
&\to \frac{\frac{1}{\sqrt{2+\delta}(2\pi h)^{\left(\frac{1+\delta}{2}\right)}}x^{-\frac{3}{2}(1+\delta)}f(x)}{n^{\frac{\delta}{2}}\left(\frac{1}{2\sqrt{\pi}}h^{-\frac{1}{2}}x^{-\frac{3}{2}}f(x)\right)^{\frac{2+\delta}{2}}} \\
&= \frac{\frac{1}{\sqrt{2+\delta}(2\pi h)^{\left(\frac{1+\delta}{2}\right)}}x^{-\frac{3}{2}(1+\delta)}f(x)}{n^{\frac{\delta}{2}}h^{\frac{\delta}{4}}\left(\frac{1}{2\sqrt{\pi}}x^{-\frac{3}{2}}f(x)\right)^{\frac{2+\delta}{2}}} \\
&= \frac{\frac{1}{\sqrt{2+\delta}(2\pi)^{\left(\frac{1+\delta}{2}\right)}}x^{-\frac{3}{2}(1+\delta)}f(x)}{(nh^{\frac{1}{2}})^{\frac{\delta}{2}}\left(\frac{1}{2\sqrt{\pi}}x^{-\frac{3}{2}}f(x)\right)^{\frac{2+\delta}{2}}} \\
&\to 0
\end{aligned}$$

The last term vanishes as $n \to \infty$, since C2 implies that $h \to 0$ and $nh \to \infty$, then $h^{\frac{1}{2}}$ goes to zero slower than h and this implies that $nh^{\frac{1}{2}} \to \infty$. On the other hand, the remaining components of the last term are bounded by condition C1. Since under the same condition $nh^{\frac{1}{2}} \to \infty$ we have $Var\left(\hat{f}_{IG}(x)\right) \to 0$. Therefore by Liapounove theorem (1.1.4) we have :

$$\frac{\hat{f}_{IG}(x) - f(x)}{\sqrt{Var\left(\hat{f}_{IG}(x)\right)}} \xrightarrow{d} N(0,1). \tag{3.1.13}$$

Substituting the expression of $Var\left(\hat{f}_{IG}(x)\right)$, we get :

$$\frac{\hat{f}_{IG}(x) - f(x)}{\sqrt{\frac{1}{2n\sqrt{\pi h}} x^{-\frac{3}{2}} f(x)}} \xrightarrow{d} N(0,1). \tag{3.1.14}$$

Which implies that :

$$\sqrt{nh^{\frac{1}{2}}}\left(\hat{f}_{IG}(x) - f(x)\right) \xrightarrow{d} N(0, \frac{1}{2\sqrt{\pi}} x^{-\frac{3}{2}} f(x)).$$

Next we will show that the error in estimating the cumulative density function vanishes with probability.

Lemma 3.1.1. *Under the conditions* 1 *and* 2 *the following holds :*

$$\sqrt{nh^{\frac{1}{2}}} |\hat{F}_{IG}(x) - F(x)| \xrightarrow{p} 0$$

Proof: Using the definition of $\hat{F}(x)$, we have :

$$\begin{aligned} E(\hat{F}_{IG}(x)) &= \int_0^\infty \int_0^x K_{IG}(u, \frac{1}{h})(y) du f(y) dy \\ &= \int_0^x \int_0^\infty K_{IG}(u, \frac{1}{h})(y) f(y) dy du \\ &= \int_0^x E(f(\zeta_u)) du \end{aligned} \tag{3.1.15}$$

where, ζ_u follows a $IG(u, \frac{1}{h})$ distribution. Using Taylor's series we expand $f(\zeta_u)$ about the mean of ζ_u ($\mu_u = E(\zeta_u) = u$) as :

$$f(\zeta_u) = f(\mu_u) + f'(\mu_u)(\zeta_u - \mu_u) + \frac{1}{2} f''(\mu_u)(\zeta_u - \mu_u)^2 + o(h)$$

Hence we have :

$$\begin{aligned} E(f(\zeta_u)) &= f(u) + \frac{1}{2} f''(u) V_u + o(h), \quad (\text{where } V_u = Var(\zeta_u) = (u-h)h + 2h^2) \\ &= f(u) + \frac{1}{2} f''(u) hu + o(h). \end{aligned}$$

Using the result in (3.1.15) we get :

$$\begin{aligned} E(\hat{F}_{IG}(x)) &= \int_0^x E(f(\zeta_u)) du \\ &= \int_0^x (f(u) + \frac{1}{2} f''(u^3) hu) du + o(h) \\ &= F(x) + \frac{h}{2} (\int_0^x u^3 f''(u) du) + o(h) \\ &= F(x) + o(h). \end{aligned}$$

Which implies the following result :

$$\sqrt{nh^{\frac{1}{2}}}|E(\hat{F}_{IG}(x)) - F(x)| = o((nh^{\frac{5}{2}})^{\frac{1}{2}}) \to 0. \tag{3.1.16}$$

Now, $\hat{F}_{IG}(x)$ can be written in the following form :

$$\begin{aligned}\hat{F}_{IG}(x) &= \frac{1}{n}\sum_{i=1}^{n}\int_0^x K_{IG}(u, \frac{1}{h})(X_i)du \\ &= \frac{1}{n}\sum_{i=1}^{n} W_i(x)\end{aligned}$$

where $W_i(x) = \int_0^x K_{IG}(u, \frac{1}{h})(X_i)du$. Let $\epsilon > 0$, $\delta > 0$ be given, using Chebychev's inequality (1.1.3) we have:

$$\begin{aligned}P((nh^{\frac{1}{2}})^{\frac{1}{2}}|\hat{F}_{IG}(x) - E(F(x))| > \epsilon) &\leq \epsilon^{-2-2\delta}(nh^{\frac{1}{2}})^{1+\delta}E|\frac{1}{n}\sum_{i=1}^{n}(W_i(x) - E(W_i(x)))|^{2+\delta} \\ &= \epsilon^{-2-2\delta}h^{\frac{1+\delta}{2}}n^{-1-\delta}E|\sum_{i=1}^{n}(W_i(x) - E(W_i(x)))|^{2+\delta} \\ &\leq 2^{1+\delta}\epsilon^{-2-2\delta}(n^{-1}h^{\frac{1}{2}})^{1+\delta}\sum_{i=1}^{n}E|W_i(x)|^{2+2\delta} \\ &+ 2^{1+\delta}\epsilon^{-2-2\delta}(n^{-1}h^{\frac{1}{2}})^{1+\delta}\sum_{i=1}^{n}|E(W_i(x))|^{2+2\delta}.\end{aligned}$$

and we have :

$$\begin{aligned}(n^{-1}h^{\frac{1}{2}})^{1+\delta}\sum_{i=1}^{n}E|(W_i(x))|^{2+2\delta} &= (n^{-1}h^{\frac{1}{2}})^{1+\delta}n\int_0^\infty\int_0^x\left(K_{IG}(u, \frac{1}{h})(y)\right)^{2+2\delta}duf(y)dy \\ &= n^{-\delta}h^{\frac{1+\delta}{2}}\int_0^x\int_0^\infty\left(K_{IG}(u, \frac{1}{h})(y)\right)^{2+2\delta}f(y)dydu \\ &= n^{-\delta}h^{\frac{1+\delta}{2}}\int_0^x\frac{1}{\sqrt{2+2\delta}(2)^{\frac{1+2\delta}{2}}}u^{-\frac{3}{2}(1+2\delta)}f(u)du + o(h^{-\frac{1+2\delta}{2}}) \\ &\leq Cn^{-\delta}h^{\frac{1+2\delta}{2}}h^{-\frac{1+2\delta}{2}} \\ &= Cn^{-\delta}h^{-}\frac{\delta}{2} \\ &= C(nh^{\frac{1}{2}})^{-\delta} \\ &\to 0.\end{aligned}$$

This implies that,

$$\sqrt{nh^{\frac{1}{2}}}|\hat{F}_{IG}(x) - E(F(x))| \xrightarrow{p} 0. \tag{3.1.17}$$

Now, using the results(3.1.16),(3.1.17)and by triangle inequality we obtain that :

$$\sqrt{nh^{\frac{1}{2}}}|\hat{F}_{IG}(x) - F(x)| \leq \sqrt{nh^{\frac{1}{2}}}|\hat{F}_{IG}(x) - E(\hat{F}_{IG}(x))| + \sqrt{nh^{\frac{1}{2}}}|E(\hat{F}_{IG}(x)) - F(x)|$$
$$\xrightarrow{p} 0.$$

The proof now is complete.

3.2 The IG Kernel Estimator for the Hazard Rate Function

In this section, we consider the nonparametric estimation of the hazard rate function for (iid) data using the inverse Gaussian kernel IG, the asymptotic normality of the proposed estimator will be derived and we close this section by investigate the selection of the optimal bandwidth.

Firstly, we recall the definition of hazard rate function .

Definition 3.2.1. *The hazard rate function or age-specific failure rate , defined by:*

$$r(x) = \lim_{\Delta \longrightarrow 0} \frac{P(x < X \leq x + \Delta | x \leq X)}{\Delta} \tag{3.2.1}$$

and by the definition of kernel we have :

$$r(x) = \frac{f(x)}{1 - F(x)} \tag{3.2.2}$$

where $f(x)$ is the pdf of the distribution and $F(x)$ is the cdf.

As in Chapter 2, the survivor function $S(x) = 1 - F(x)$, hence the hazard rate function is : $r(x) = \frac{f(x)}{S(x)}$. The next definition state the kernel estimator for the survivor function using Equation 2.2.1 .

Definition 3.2.2. *The kernel estimator for the survivor function is constructed using kernel density estimator in equation (3.1.3), as :*

$$\hat{S}_{IG}(x) = 1 - \hat{F}_{IG}(x) \tag{3.2.3}$$

where,

$$\hat{F}_{IG}(x) = \int_0^x \hat{f}_{IG}(u)du = \frac{1}{n}\sum_{i=1}^{n}\int_0^x K_{IG}(u, \frac{1}{h})(X_i). \tag{3.2.4}$$

By Definition 3.2.2 and using Definition 2.2.1, now we study the smoothed indirect IG kernel estimator for the hazard rate function.

Definition 3.2.3. *The IG kernel estimator for the hazard rate function is given by:*

$$\hat{r}_{IG}(x) = \frac{\hat{f}_{IG}(x)}{\hat{S}_{IG}(x)}, \tag{3.2.5}$$

where, $\hat{f}_{IG}(x)$ and $\hat{S}_{IG}(x)$ as in definition 3.1.2 and 3.2.3 respectively.

3.2.1 The Properties of the IG Kernel Estimator $\hat{r}_{IG}(x)$

By Definition 3.2.3, we present the mean, variance and the bias in the following theorem.

Theorem 3.2.1. *Under the conditions C_1 and C_2 we have :*

$$Bias(\hat{r}_{IG}(x)) = \frac{\frac{1}{2}x^3 f''(x)h}{S(x)} + o(h), \tag{3.2.6}$$

and

$$Var(\hat{r}_{IG}(x)) = \frac{1}{2n\sqrt{\pi h}} x^{-\frac{3}{2}} \frac{r(x)}{S(x)} + o(n^{-1}h^{-\frac{1}{2}}). \tag{3.2.7}$$

Proof: To find the bias we find the mean :

$$\begin{aligned} E(\hat{r}_{IG}(x)) &\approx \frac{E(\hat{f}_{IG}(x))}{E(\hat{S}_{IG}(x))} \\ &= \frac{f(x) + \frac{1}{2}x^3 f''(x)h}{S(x)} + o(h), \qquad \text{(since } \hat{S}_{IG}(x) \xrightarrow{p} S(x)) \\ &= r(x) + \frac{\frac{1}{2}x^3 f''(x)h}{S(x)} + o(h). \end{aligned} \tag{3.2.8}$$

Hence we have :

$$\begin{aligned} Bias(\hat{r}_{IG}(x)) &\approx E(\hat{r}_{IG}(x)) - r(x) \\ &= \frac{\frac{1}{2}x^3 f''(x)h}{S(x)} + o(h). \qquad \text{(using Equation 3.2.8)} \end{aligned}$$

Now we prove the second part of the theorem, Since by definition 3.2.3,

$$\begin{aligned}
Var\left(\hat{r}_{IG}(x)\right) &= Var\left(\frac{\hat{f}_{IG}(x)}{\hat{S}_{IG}(x)}\right) \\
&= Var\left(\frac{\hat{f}_{IG}(x)}{S(x)}\right) \\
&= \frac{1}{S(x)^2} Var\left(\hat{f}_{IG}(x)\right) \\
&= \frac{1}{S(x)^2}\frac{1}{2n\sqrt{\pi h}} x^{-\frac{3}{2}} f(x) + o(n^{-1}h^{-\frac{3}{2}}), \qquad \text{(by Theorem 3.1.1)} \\
&= \frac{1}{2n\sqrt{\pi h}} x^{-\frac{3}{2}} \frac{f(x)}{S(x)^2} + o(n^{-1}h^{-\frac{1}{2}}) \\
&= \frac{1}{2n\sqrt{\pi h}} x^{-\frac{3}{2}} \frac{r(x)}{S(x)} + o(n^{-1}h^{-\frac{1}{2}}).
\end{aligned}$$

Remark 3.2.1. *Form Theorem 3.2.1, we note that by Equation (3.2.6) the* $Bias(\hat{r}_{IG}(x))$ *is increasing in* h *and by Equation (3.2.7) the* $Var\left(\hat{r}_{IG}(x)\right)$ *is decreasing in* $nh^{\frac{1}{2}}$*. And hence under the conditions C1 and C2, we have* $Bias(\hat{r}_{IG}(x)) \to 0$ *and* $Var\left(\hat{r}_{IG}(x)\right) \to 0$ *as* $n \to \infty$ *which give us the asymptotic consistency.*

Next, we analyze the performance of the IG kernel estimator for the hazard rate function $\hat{r}_{IG}(x)$ by considering the mean squared error MSE and asymptotic mean squared error $AMSE$. We will use this to get the global properties and to investigate the optimal bandwidth later. Firstly, using Theorem 3.2.1, we give the mean squared error MSE in the next corollary.

Corollary 3.2.1. *The mean squared error* $MSE(\hat{r}_{IG}(x))$ *is given by :*

$$MSE(\hat{r}_{IG}(x)) = \left(\frac{\frac{1}{2}xf''(x)h}{S(x)} + o(h)\right)^2 + \frac{1}{2n\sqrt{\pi h}} x^{-\frac{3}{2}} \frac{r(x)}{S(x)} + o(n^{-1}h^{-\frac{1}{2}}). \tag{3.2.9}$$

Proof:

$$\begin{aligned}
MSE(\hat{r}_{IG}(x)) &= (Bias(\hat{r}_{IG}(x)))^2 + Var(\hat{r}_{IG}(x)) \\
&= \left(\frac{\frac{1}{2}x^3f''(x)h}{S(x)} + o(h)\right)^2 + \frac{1}{2n\sqrt{\pi h}} x^{-\frac{3}{2}} \frac{r(x)}{S(x)} + o(n^{-1}h^{-\frac{1}{2}}).
\end{aligned}$$

A closer look to Equation 3.2.9, we see that the MSE increases in h (the first term) and decreases in $nh^{\frac{1}{2}}$ (the second term) and hence to make the MSE to decline as $n \to \infty$, we have to make these both terms small and this is hold by condition C2 we assumed before.

Remark 3.2.2. *By the same way in Chapter 1, letting $o(h)$ and $o(n^{-1}h^{-\frac{1}{2}}) \to 0$, the asymptotic mean squared-error (AMSE) is given by :*

$$AMSE(\hat{r}_{IG}(x)) = \left(\frac{\frac{1}{2}x^3 f''(x)h}{S(x)}\right))^2 + \frac{1}{2n\sqrt{\pi h}} x^{-\frac{3}{2}} \frac{r(x)}{S(x)} \tag{3.2.10}$$

Asymptotic Normality of the Estimator $\hat{r}_{IG}(x)$:

Theorem 3.2.2. *Under the conditions C_1 and C_2, the following holds :*

$$\sqrt{nh^{\frac{1}{2}}}(\hat{r}_{IG}(x) - r(x)) \xrightarrow{d} N\left(0, \frac{1}{2\sqrt{\pi}} x^{-\frac{3}{2}} \frac{r(x)}{S(x)}\right).$$

Proof:

$$\begin{aligned}
\sqrt{nh^{\frac{1}{2}}}(\hat{r}_{IG}(x) - r(x)) &= \sqrt{nh^{\frac{1}{2}}}\left(\frac{\hat{f}_{IG}(x)}{\hat{S}_{IG}(x)} - \frac{f(x)}{S(x)}\right) \\
&= \sqrt{nh^{\frac{1}{2}}}\left(\frac{\hat{f}_{IG}(x)}{\hat{S}_{IG}(x)} - \frac{f(x)}{\hat{S}_{IG}(x)} - \frac{f(x)}{S(x)} + \frac{f(x)}{\hat{S}_{IG}(x)}\right) \\
&= \frac{\sqrt{nh^{\frac{1}{2}}}}{\hat{S}_{IG}(x)}\left(\hat{f}_{IG}(x) - f(x) - \frac{f(x)\hat{S}_{IG}(x)}{S(x)} + f(x)\right) \\
&= \frac{\sqrt{nh^{\frac{1}{2}}}}{\hat{S}_{IG}(x)}(\hat{f}_{IG}(x) - f(x)) + \frac{\sqrt{nh^{\frac{1}{2}}}f(x)}{S(x)\hat{S}_{IG}(x)}(\hat{S}_{IG}(x) - S(x))
\end{aligned} \tag{3.2.11}$$

$$\begin{aligned}
&\xrightarrow{d} \frac{1}{S(x)} N\left(0, \frac{1}{2\sqrt{\pi}} x^{-\frac{3}{2}} f(x)\right) + 0 \\
&\xrightarrow{d} N\left(0, \frac{r(x)}{2\sqrt{\pi}S(x)} x^{-\frac{3}{2}} f(x)\right).
\end{aligned} \tag{3.2.12}$$

The result 3.2.12 is by Theorem 3.1.2 and Lemma 3.1.1 , since by Theorem 3.1.2, the first term in equation 3.2.11 is asymptotically distributed and the second term vanishes by Lemma 3.1.1 .

3.2.2 Bandwidth Selection

In Chapter 1 and Chapter 2 we discussed the ways for finding the optimal bandwidth. Here we will use the same way. We will get the global properties $AMISE$. In order to find the optimal $AMSE$ ($AMSE^*$) , we differentiate the $AMSE$ in Equation

3.2.10 with respect to h, then we equating it to zero, to obtain :

$$2h\left(\frac{\frac{1}{2}x^3 f''(x)}{S(x)}\right)^2 - \frac{1}{2}\frac{1}{2n\sqrt{\pi}}h^{-\frac{3}{2}}x^{-\frac{3}{2}}\frac{r(x)}{S(x)} = 0. \qquad (3.2.13)$$

Next, Multiplying Equation (3.2.13) by $h^{\frac{3}{2}}$ both sides, and replacing $r(x)$ by $\frac{f(x)}{S(x)}$ we get :

$$2h^{\frac{5}{2}}\left(\frac{\frac{1}{2}x^3 f''(x)}{S(x)}\right)^2 = \frac{1}{4n\sqrt{\pi}}x^{-\frac{3}{2}}\frac{f(x)}{S(x)^2}. \qquad (3.2.14)$$

Now, solving (3.2.14) for h we have :

$$\begin{aligned} h^{\frac{5}{2}} &= \left(\frac{1}{4n\sqrt{\pi}}x^{-\frac{3}{2}}\frac{f(x)}{S^2(x)}\right)\left(2\frac{S2^(x)}{f''(x)^2x^2}\right) \\ &= \frac{x^{-\frac{5}{2}}}{2n\sqrt{\pi}}\frac{f(x)}{f''(x)^2} \end{aligned}$$

Hence h^* is given by :

$$\begin{aligned} h^* &= \left(\frac{x^{-\frac{5}{2}}}{2n\sqrt{\pi}}\frac{f(x)}{f''(x)^2}\right)^{\frac{2}{5}} \\ &= \left(\frac{1}{2\sqrt{\pi}}\frac{f(x)}{f''(x)^2}\right)^{\frac{2}{5}} x^{-1}n^{-\frac{2}{5}}. \end{aligned} \qquad (3.2.15)$$

Using the result (3.2.15) in (3.2.10) we have :

$$\begin{aligned} AMSE^*(\hat{r}_{IG}) &= \left(\frac{x^3 f''(x)}{2S(x)}x^{-3}n^{-\frac{2}{5}}\left(\frac{f(x)}{2\sqrt{\pi}}\right)^{\frac{2}{5}}\frac{1}{f''(x)^{\frac{4}{5}}}\right)^2 + \frac{1}{2n\sqrt{\pi}}x^{-\frac{1}{2}}\frac{f(x)}{S(x)^2}h^{-\frac{1}{2}} \\ &= \left(\frac{x f''(x)}{2S(x)}x^{-1}n^{-\frac{2}{5}}\left(\frac{f(x)}{2\sqrt{\pi}}\right)^{\frac{2}{5}}\frac{1}{f''(x)^{\frac{4}{5}}}\right)^2 + \frac{1}{2n\sqrt{\pi}}x^{-\frac{1}{2}}\frac{f(x)}{S(x)^2}\left(\frac{f(x)}{2\sqrt{\pi}f''(x)^2}\right)^{-\frac{1}{5}}x^{\frac{1}{2}}n^{\frac{1}{5}} \\ &= \frac{1}{4}\left(\frac{f(x)}{2\sqrt{\pi}}\right)^{\frac{4}{5}}\left(\frac{n^{-\frac{4}{5}}f''(x)^{\frac{2}{5}}}{S(x)^2}\right) + \left(\frac{f(x)}{2\sqrt{\pi}}\right)^{\frac{4}{5}}\left(\frac{n^{-\frac{4}{5}}f''(x)^{\frac{2}{5}}}{S(x)^2}\right) \\ &= \frac{5}{4}\left(\frac{f(x)}{2\sqrt{\pi}}\right)^{\frac{4}{5}}\left(\frac{n^{-\frac{4}{5}}f''(x)^{\frac{2}{5}}}{S(x)^2}\right). \end{aligned} \qquad (3.2.16)$$

Note that the optimal bandwidth h^* is proportional to $n^{-\frac{2}{5}}$, and as in Chapter 1 we stuied that the optimal bandwidth is of order $o(n^{-\frac{2}{5}})$.

Global Properties :

Regarding **global properties** the optimal bandwidths h^{**} and mean integrated squared errors $MISE^{**}$ will be discussed here :

Using Equation (3.2.15) by multiplying the denominator and nominator by x^2 and taking the integration we get :

$$\begin{aligned} h^{**} &= \left(\frac{\int_0^\infty \frac{1}{2\sqrt{\pi}} x^{-\frac{5}{2}} x^2 f(x) dx}{\int_0^\infty x^2 f''(x)^2 dx} \right)^{\frac{2}{5}} n^{-\frac{2}{5}} \\ &= \left(\frac{\int_0^\infty \frac{1}{2\sqrt{\pi}} x^{-\frac{3}{2}} f(x) dx}{\int_0^\infty \left(x^3 f''(x)\right)^2 dx} \right)^{\frac{2}{5}} n^{-\frac{2}{5}}. \end{aligned} \tag{3.2.17}$$

To find $AMISE^{**}$, we multiply Equation(3.2.17) by $\left(\frac{x^{\frac{2}{5}}}{x^{\frac{2}{5}}}\right)$ and taking the integration as follow :

$$AMISE^{**}(\hat{r}_{IG}(x)) = \frac{5}{4} \left(\frac{1}{2\sqrt{\pi}} \int_0^\infty x^{-\frac{3}{2}} f(x) dx \right)^{\frac{4}{5}} \left(\frac{\int_0^\infty (x^3 f''(x)^2)^{\frac{1}{5}} dx}{S(x)^2} \right) n^{-\frac{4}{5}} \tag{3.2.18}$$

By Remark 1.3.3 and Equation 3.2.18, we deduce that the IG kernel estimator for the hazard rate function, achieve the optimal rate of convergence for the $AMISE$ within the class of non-negative kernels (class of second order kernel functions).

Practical Optimal Bandwidth :

In practice, the bandwidth selection can be done by using the same rule which proposed by Scaillet. Scaillet used the same way were Silverman proposed (see Example 2.2.1), but for the log-normal probability density function in the IG case. For this, if lnx follows a normal distribution with parameters μ and σ^2 we have :

$$\int_0^\infty x^{-\frac{1}{2}} f(x) dx = exp(\frac{1}{8}(\sigma^2 - 4\mu)). \tag{3.2.19}$$

and

$$\int_0^\infty (x f''(x))^2 dx = \frac{12 + 4\sigma^2 + \sigma^4}{32\sqrt{\pi}\sigma^5} exp(\frac{1}{4}(9\sigma^2 - 12\mu)). \tag{3.2.20}$$

Hence, using 3.2.19 and 3.2.20 in Equation 3.2.17 the optimal bandwidth is given by :

$$h^{**} = \left(\frac{16\sigma^5 exp(\frac{1}{8}(-17\sigma^2 + 20\mu))}{12 + 4\sigma^2 + \sigma^4} \right)^{\frac{2}{5}} n^{-\frac{2}{5}}. \tag{3.2.21}$$

where, the unknown parameters σ and μ are estimated by the arithmetic mean as follow :

1. $\bar{x} = \frac{1}{n} \sum_{i=1}^n lnx_i$,

2. $\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(lnx_i - \bar{x})^2$.

Conclusion :

The conclusion of this chapter is that by using a **constant bandwidth** h, the IG kernel estimator for the hazard rate function $\hat{r}_{IG}$ appears to have the following asymptotic properties:

1. $\hat{r}_{IG}(x)$ is free **boundary bias**, because its bias $Bias(\hat{r}_{IG}(x))$ is of order $o(h)$ in the interior of $[0, \infty)$ and near zero.
2. $\hat{r}_{IG}(x)$ is mean square consistent for $r(x)$.
3. $\hat{r}_{IG}(x)$ is an asymptotically unbiased estimator of $r(x)$.
4. $\hat{r}_{IG}(x)$ is asymptotically normally distributed.

Finally, we summarize the **comparison** between the two proposed estimators for the hazard rate function ($\hat{r}_{IG}(x)$ and $\hat{r}_G(x)$) by comparing there biases and variances near the zero and comparing the $AMSE^*$ for each estimator.

1. The Bias :

Recall that

$$Bias(\hat{r}_{IG}(x)) = \frac{\frac{1}{2}x^3 f''(x)h}{S(x)} + o(h),$$

and

$$Bias(\hat{r}_G(x)) = \frac{f''(x)h^2}{2S(x)} + o(h^2).$$

we see that the expressions of the $Bias(\hat{r}_{IG}(x))$ and $Bias(\hat{r}_G(x))$ increases in xh and h^2 respectively, and hence near the zero ($x \in (0, h)$) we have $xh < h^2$, which imply that $Bias(\hat{r}_{IG}(x)) < Bias(\hat{r}_G(x))$.

2. The Variance :

Recall that,

$$Var(\hat{r}_{IG}(x)) = \frac{1}{2n\sqrt{\pi h}}x^{-\frac{1}{2}}\frac{r(x)}{S(x)} + o(n^{-1}h^{-\frac{1}{2}}). \qquad (3.2.22)$$

and

$$Var(\hat{r}_G(x)) = \frac{f(x)}{2nh\sqrt{\pi}S^2(x)} + o(\frac{1}{nh}), \qquad (3.2.23)$$

we see that the expressions of the $Var(\hat{r}_{IG}(x))$ and $Var(\hat{r}_G(x))$ decreases in $\sqrt{xh}$ and h respectively, and hence as $x \in (0, h)$ we have $\sqrt{xh} < h$, which imply that $Var(\hat{r}_{IG}(x)) > Var(\hat{r}_G(x))$.

3. The $AMSE^*$:

Finally, for any $x \in [0, \infty)$, the $AMSE^*$ for both estimators is the same

$$AMSE^*(\hat{r}_{IG}(x)) = AMSE^*(\hat{r}_G(x)) = \frac{5}{4}\left(\frac{f(x)}{2\sqrt{\pi}}\right)^{\frac{4}{5}}\left(\frac{n^{-\frac{4}{5}}f''(x)^{\frac{2}{5}}}{S(x)^2}\right),$$

which means that they have the same behavior in practical applications.
The comparison appear clearly in the next chapter which proposed a comparison with applications using real data and simulated data describing the behavior of the two estimators especially near zero.

Chapter 4

Applications

In this Chapter, we test the performance of the IG kernel estimator of the pdf, cdf and the smoothed indirect IG kernel estimator for the hazard rate function using applications to real and simulated data. The chapter is in three sections. Section one contains the first application which deals with simulated data set from the normal distribution, and we estimated the pdf and the hazard rate function by using two different kernels, the normal and the Epanchinkove kernel. In the second application, we used a simulated data for the exponential distribution, and we estimated the pdf and the hazard rate function by using the IG and Gaussian kernel estimators.
Section two deals with real life data set (the suicide data), and we estimated the hazard rate function by using the normal and Epanchinkove kernel estimators, and we estimated the pdf and the hazard rate function by using the IG kernel estimator.
Section three is a conclusion summarizes the main results of the thesis.
For the practical implementation of the IG estimator, we will use the bandwidth is computed using the following equation from Silverman[17]

$$h = 1.06sn^{-\frac{1}{5}}, \tag{4.0.1}$$

where s is the sample standard deviation and n is the sample size, and for the Gaussian estimator we will use the rule of chapter 2. The applications will construct using S-Plus program.

4.1 An Application with Simulated Data

The performance of IG estimator is tested using two simulated data, and we computed for each estimator the Mean Squared Error (MSE) , where

$$MSE = \sum_{i=1}^{n} \frac{(y_{ni} - y_i)^2}{n}$$

where y_i denotes the true value and y_{ni} denotes its predicted value.
In the first application, we used a simulated data for the standard normal distribution, and we estimated the pdf and hazard rate function by using two different kernels, the normal and the Epanchinkove kernel. In the second application, we used a simulated data for the exponential distribution, and we estimated the pdf and the hazard rate function by using the IG and Gaussian kernel estimators.

Application 1 : In the first application, we have simulated data of size 200 from the standard normal distrubution. Then we used the kernel estimator

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

which we have studied in the first chapters. Using two different kernels, the normal and the Epanchinkove kernel respectively.
$K_{NW}(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}, -\infty < x < \infty$
$K_{EP}(x) = \frac{3}{4}(1 - x^2) I_{(|x| \leq 1)}$

Figure 4.1, Figure 4.2 show the Normal kernel and Epanchinkove kernel estimator of the density function of the data.

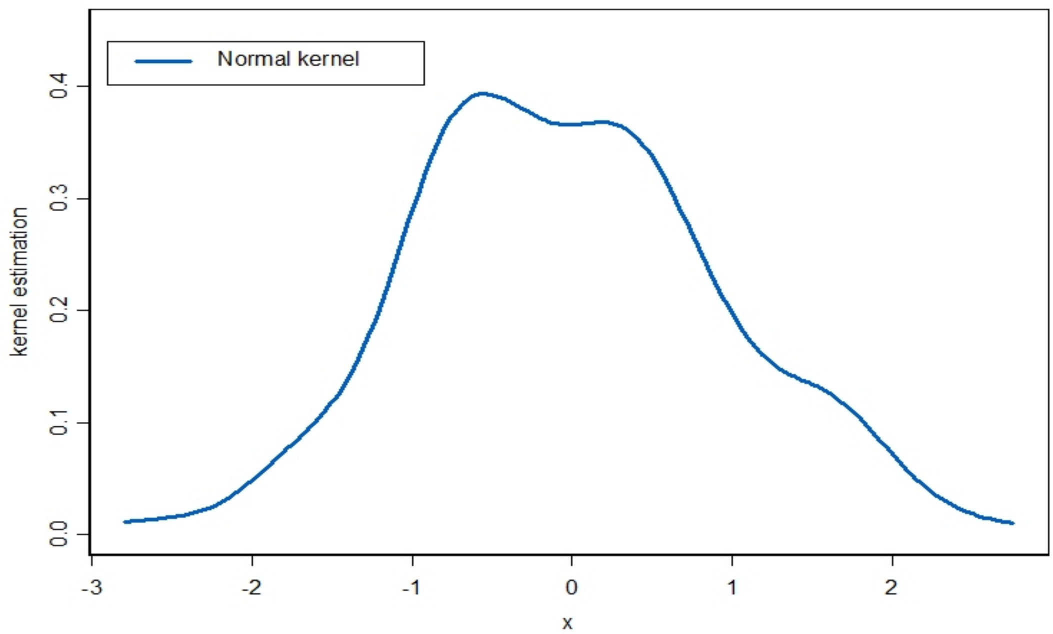

Figure 4.1: The Normal kernel estimator of the pdf

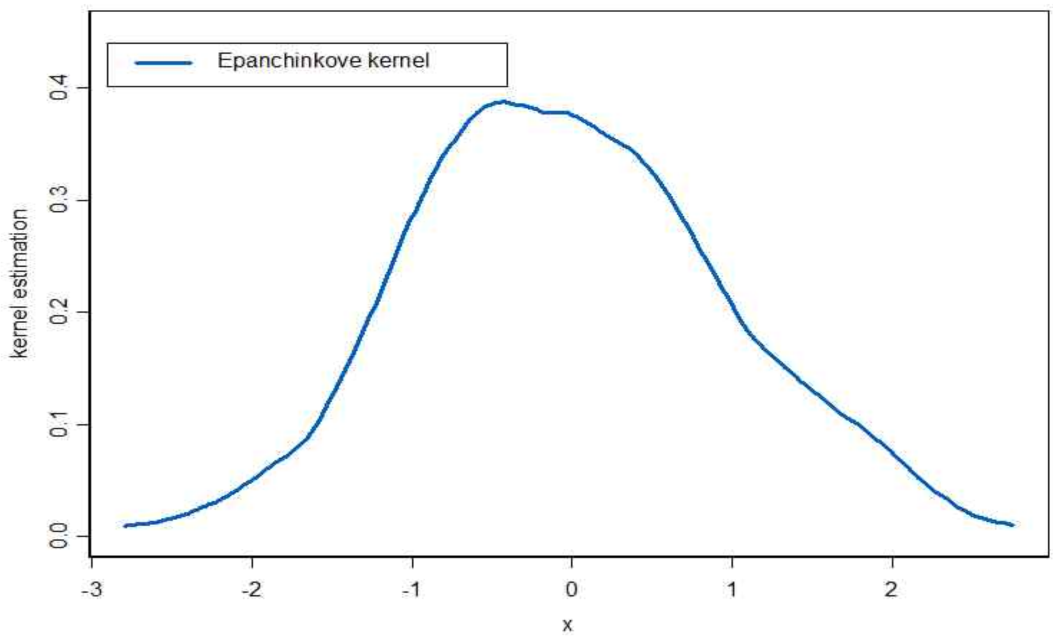

Figure 4.2: The Epanchinkove kernel estimator of the pdf

Figure 4.3 and Figure 4.4 show the kernel estimator of the hazard rate function for the simulated data, which we have studied in chapter 2, Equation 2.1 using the normal kernel and the Epanchinkove kernel respectively.

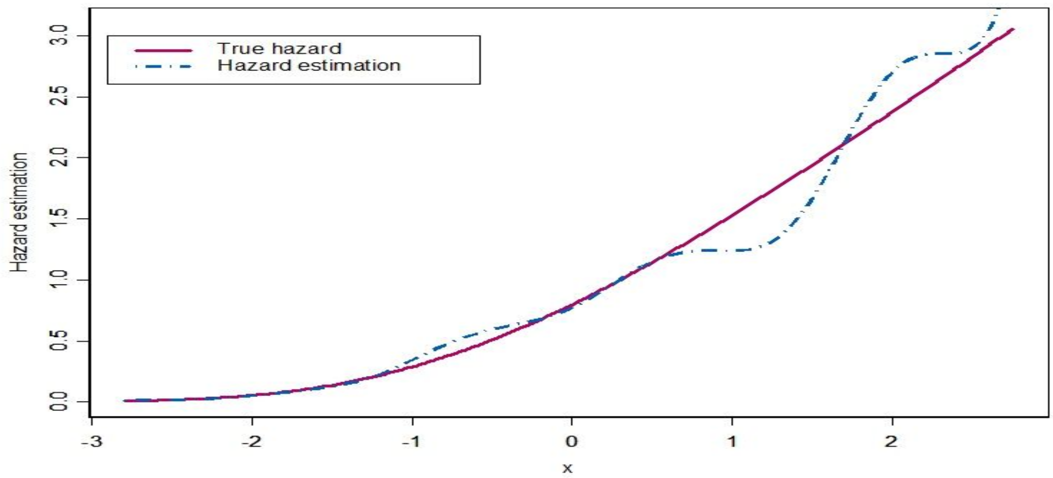

Figure 4.3: The true Hazard Function and hazard estimation of normal kernel.

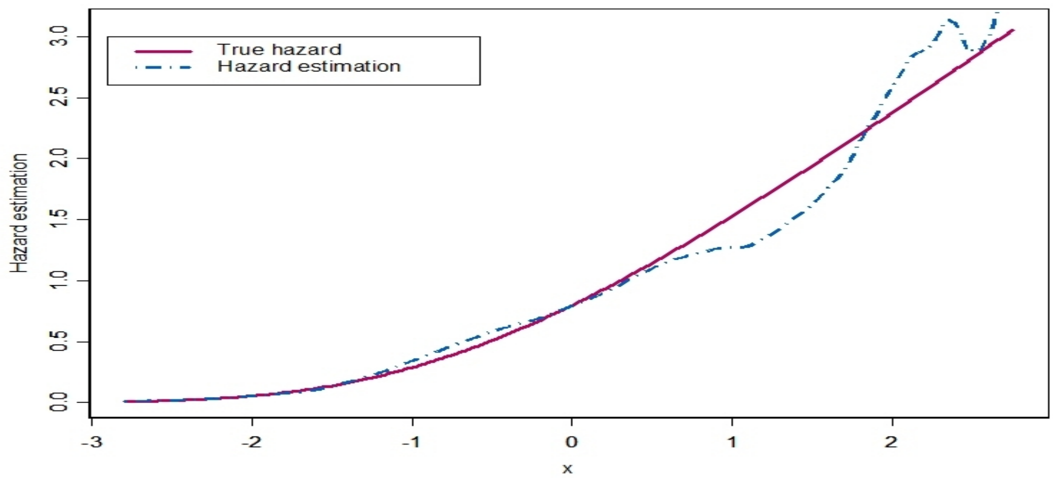

Figure 4.4: The true Hazard Function and hazard estimation of Epanchinkove kernel.

We note of Figure 4.3 and 4.4 , The hazard estimation of the Normal and Epanchinkove kernel estimators are sufficient about zero .

Table 4.1 contains the results for the Normal and Epanchinkove kernel estimators for the simulated data, where we computed the MSE .

Table 4.1: The MSE for Application 1.

	MSE
Normal Krenel	0.008975474
Epanchinkove Kernel	0.001135421

From the table, we show that the Epanchinkove kernel estimator of the hazard gives interesting results for the MSE.

Application 2 : A sample of size 200 from the exponential distribution with pdf $f(x) = e^{-x}$, $x > 0$, is simulated. After that the density function and the hazard rate functions were estimated using the IG and the Gaussian estimators. The IG and Gaussian kernel estimators of the density function for the simulated data of the exponential distribution presented in Figure 4.5.

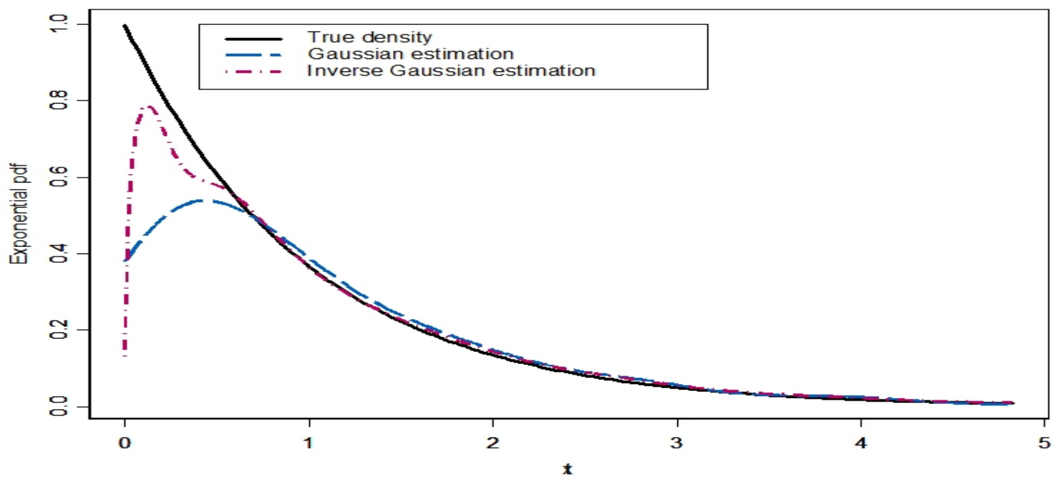

Figure 4.5: The IG and Gaussian kernel estimators of the density function for the simulated data of the exponential distribution

The Figures show that the performance of the IG estimator is better than that of the Gaussian estimator at the boundary near the zero.

The hazard rate function for the exponential distribution

$$h(x) = \frac{f(x)}{1-F(x)} = \frac{e^{-x}}{1-(1-e^{-x})} = 1$$

In Table 4.2, we computed the MSE for both the IG and Gaussian kernel estimators of the hazard rate function for the simulated data of the exponential distribution.

Table 4.2: The MSE for the IG and Gaussian kernel estimators for Application 2.

	MSE
The Gaussian kernel	0.02929727
Inverse Gaussian kernel	0.007413355

Figure 4.6 shows the IG and Gaussian kernel estimators of the hazard rate function for the simulated data of the exponential distribution.

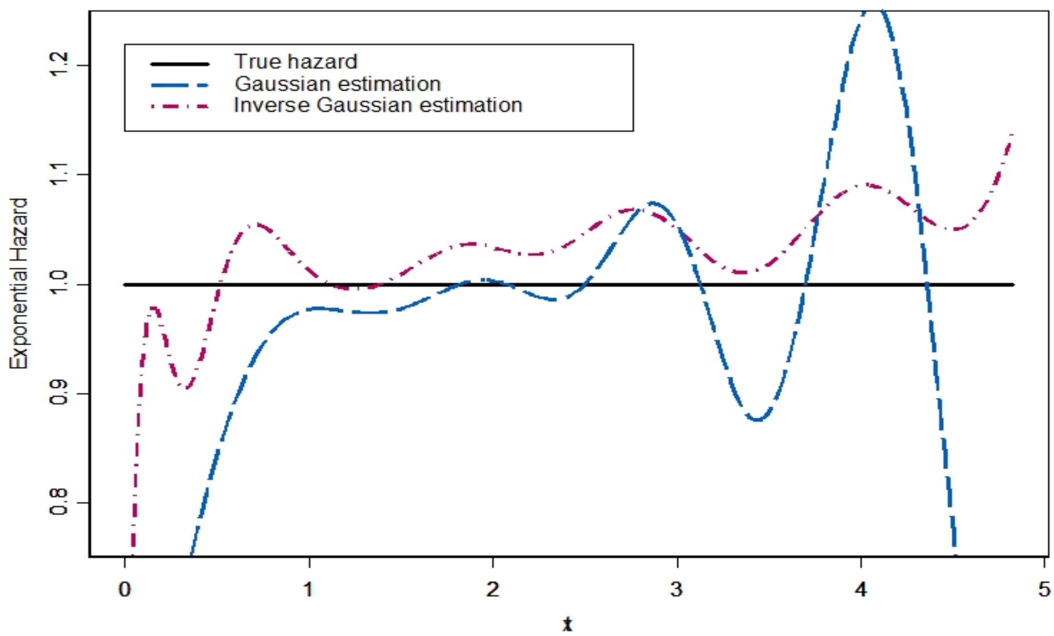

Figure 4.6: The IG and Gaussian kernel estimators of the hazard rate function for the simulated data of the exponential distribution

From Table 4.2 and Figure 4.6 , We note the IG estimator is better than that of the Gaussian estimator at the boundary near the zero.

4.2 An Application with Real Data

In this section, we use the survival time of the lengths of 86 spells of psychiatric treatment undergone by patients used as controls in a study of suicide risks by Silverman (Table 4.1, page 8) [17], to exhibit the practical performance of the IG estimator. The data gives the lengths of the treatment spells (in days) of control patients were hospitalized. The objective is to estimate the hazard rate function which in this case represents the instant potential per unit of time that an individual die within the time interval $(x, x + \Delta)$ given that it was known to be alive up to time x.

Table 4.3: Lengths of treatment spells (in days) of control patients in suicide study (Data from [17]).

Lengths(in days)					
1	25	40	83	123	256
1	27	49	84	126	257
1	27	49	84	129	311
5	30	54	84	134	314
7	30	56	90	144	322
8	31	56	91	147	369
8	31	62	92	153	415
13	32	63	93	163	573
14	34	65	93	167	609
14	35	65	103	175	640
17	36	67	103	228	737
18	37	75	111	231	
21	38	76	112	235	
21	39	79	119	242	
22	39	82	122	256	

We estimated the hazard rate function of the suicide data by using the normal kernel estimator and we presented in Figure 4.7, and it estimated using Epanchinkove kernel is presented in Figure 4.8.

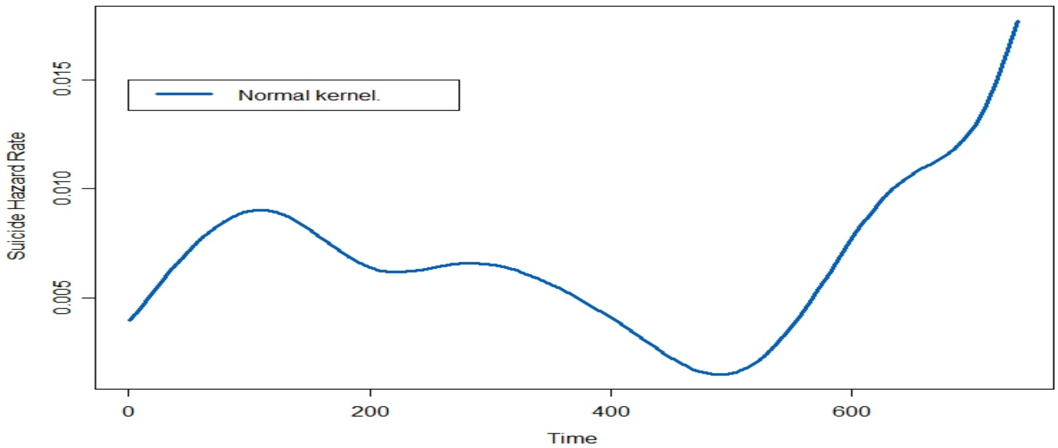

Figure 4.7: The Normal kernel estimatior of the hazard rate function for suicide data

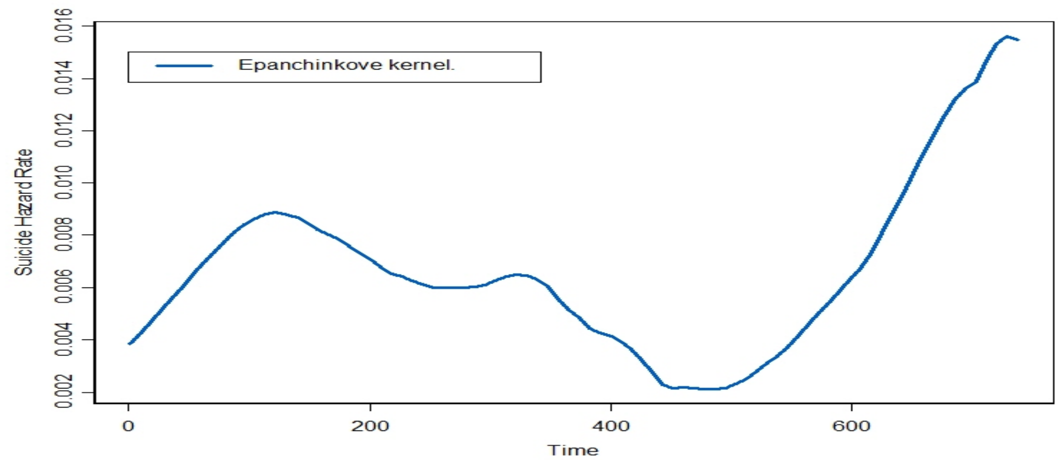

Figure 4.8: The Epanchinkove kernel estimatior of the hazard rate function for suicide data

Then, in Figure 4.9, shows the IG kernel estimators for the pdf of the suicide data, and in Figure 4.10, we presented the IG kernel estimators of the hazard rate function for the suicide data.

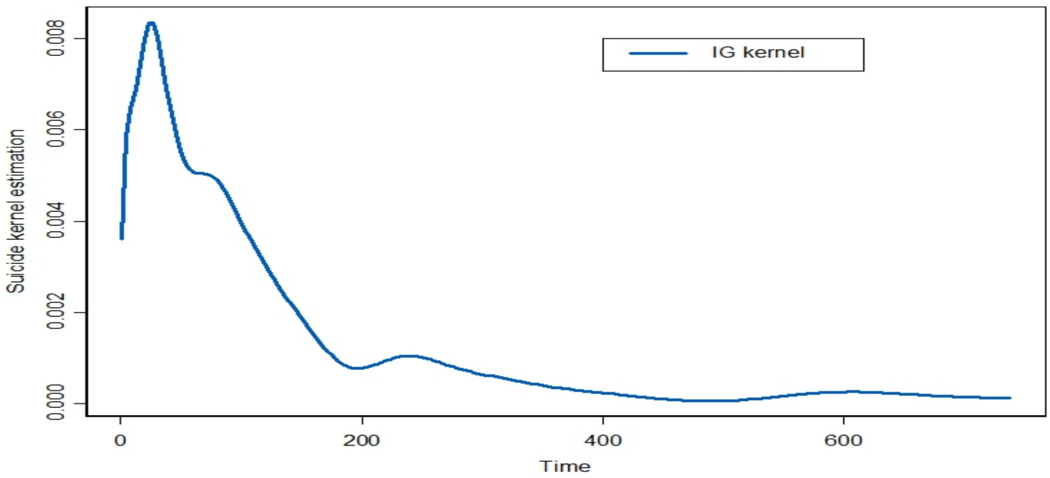

Figure 4.9: The IG kernel estimator for the pdf of suicide data

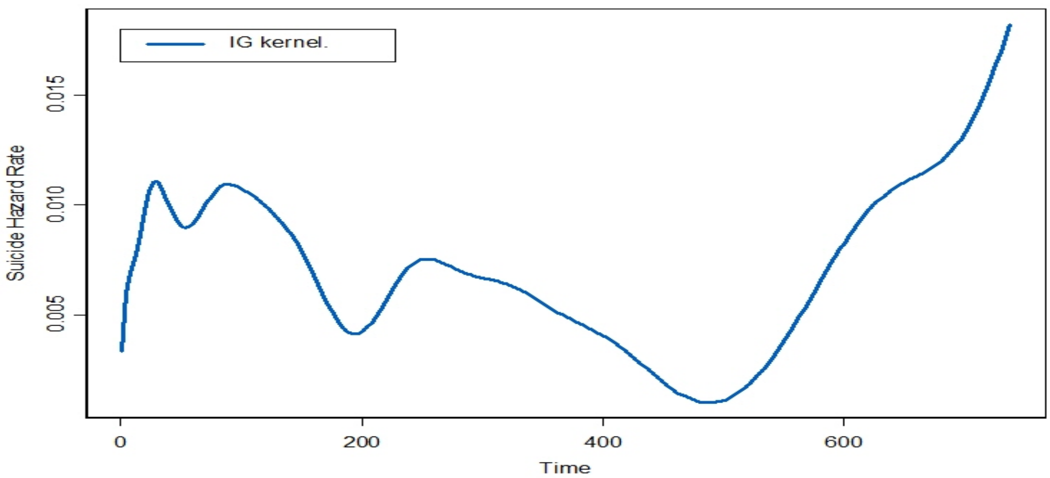

Figure 4.10: The IG kernel estimator of the hazard rate function estimate for the suicide data.

4.3 Summary

Conclusion

In this thesis, we have discussed a new kernel estimator of the hazard rate function for (iid) data based on the IG with non negative support which was proposed by Scaillet in [15]. The proposed estimator overcomes the bias problem when the hazard rate function is estimated at the boundary region near the zero.

The asymptotic normality, the strong consistency and the $AMSE$ of the proposed estimator were obtained. The $AMSE$ of the new estimator is smaller than that of the Gaussian kernel near the zero.

All applications from the simulated and real data show that the performance of the proposed estimator is better than that of the Gaussian kernel estimator at the boundary region near the zero. This is due to weight allocation by the Gaussian kernel outside the density support when smoothing is carried out at the boundary near the zero.

Recommendations

A new estimator can be modified by considering a new bandwidth selection technique that uses **a variable bandwidth** that depends on the points at which the hazard rate function is estimated rather than a constant variable (see [7]).

Bibliography

[1] Casella, G. and Berger, R.(2002). Statistical Inference,USA.

[2] Chen, S. X. (2000). Probability density function estimation using gamma kernels. Ann. Inst. Stat. Math.,52, 471-480.

[3] Freund, J. (1992).Mathematical Statistics , Arizona State University.

[4] Hogg . Mckean . Craig , (2005). Introduction to Mathematical Statistic. University of Iowa, Wester Michigan University, University of Iowa.

[5] Johnson and Kots and Balakrishnan1995, N.L. Johnson, S. Kotz and N. Balakrishnan (1995). univariate distributions .2nd Edition. John Wiley, New York, ISBN.

[6] LE Cam, L.,(1953). On Some Asymptotic Properties of Maximum Likelihood Estimates and Related Baye's Estimates 277-330. Univ. of california Press.

[7] Muller, H. and Wang, J. (1990). Locally adaptive hazard smoothing , Division of Statistics, University of California, Davis, USA .

[8] Nada, G.(2002). On The Kernel Density Estimation, Islamic University of Gaza, Palestine.

[9] Pranab, K.(1993).Large Sample Methods in Statistic "An Introduction With Applications, New York.

[10] Rice. J and Rosenblatt. M (1976). Estimation of the log survivor function and hazard function. Sankhya Series A 38, 60-78.

[11] Royden H.L. (1997). Real Analysis. Stanford University.

[12] Sheather, S.,(2004).Density Estimation, Institute of Mathematical Statistics ,588-597.

[13] Salha, R (2009). Adaptive kernel estimation of the hazard rate function. The Islamic university Gaza Journal, 17(1), 71-81.

[14] Salha, R (2012). Hazard rate function estimation using inverse Gaussian kernel.The Islamic university Gaza Journal, 20(1), 73-84.

[15] Scaillet, O. Scaillet (2004). Density estimation using inverse and reciprocal inverse Gaussian kernels. Nonparametric Statistics, 16, 217-226.

[16] Singpurwalla and M. Y. Wong (1983). Kernel estimation of the failure rate function and density estimation: an analogy. J. Amer. Statist. Assoc. 78(3), 478-481.

[17] Silverman, B. W. (1986). Density estimation for statistics and data analysis. Chapman and Hall, London.

[18] Tweedie M.(1947). functions of a statistical variate with given means with special reference to laplacian destributions , proceedings of the cambridge philosophical socity ,43, 41- 49 .

[19] Wald A. (1947). sequential analysis , wiley , new york.

[20] Wand and Jones, Wand, M. P. and Jones, M. C. (1995). Kernel smoothing, Chapman and Hall.

[21] Watson, G. and Leadbetter, M. (1964). Hazard analysis I, Biometric, 51, 175-184 .

[22] Yang, L. and Marron, J. (1999). Iterated Transformation Kernel Density Estimation . University of North Carolina, Chapel Hill, North Carolina. 1-10.

[23] Zucchini , Walter (2003). Kernel Density Estimation - Smoothing Techniques.