République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université A.MIRA-BEJAIA

Faculté des Sciences Exactes
Département de Mathématiques
Laboratoire de Mathématiques Appliquées

THÈSE

EN VUE DE L'OBTENTION DU DIPLOME DE DOCTORAT

Domaine : Mathématiques et Informatique   Filière : Mathématiques
Spécialité : Probabilités et Statistiques

Présentée par
CHEKKAL Sylia

Thème

Estimation non paramétrique du taux de défaillance par la méthode du noyau

Soutenue le : 27/06/2021   **Devant le Jury composé de :**

Nom et Prénom	Grade		
M^r ADJABI Smail	Prof.	Univ. de Bejaia	Président
M^me LAGHA Karima	MCA	Univ. de Bejaia	Rapporteur
M^r ZOUGAB Nabil	MCA	Univ. de Bejaia	Co-rapporteur
M^me OURBIH Megdouda	Prof.	Centre Universitaire de Tipaza	Examinatrice
M^me BARECHE Aicha	Prof.	Univ. de Bejaia	Examinatrice
M^me BELAIDE Karima	Prof.	Univ. de Bejaia	Examinatrice
M^me ZIANE Yasmina	MCB	Univ. de Bejaia	Invitée

Année Universitaire : 2020/2021

People's Democratic Republic of Algeria
Ministry of Higher Education and Scientific Research
A. MIRA University of Bejaia

جامعة بجاية
Tasdawit n Bgayet
Université de Béjaïa

Faculty of Exact Sciences
Department of Mathematics
Laboratory of Applied Mathematics

THESIS

IN ORDER TO GET A DOCTORAL DEGREE

Domain: Mathematics and Informatics Tracks: Mathematics
Specialty: Probability and Statistics

Presented by
Sylia CHEKKAL

Theme

Nonparametric estimation of hazard rate function using kernel method

Supported on: 27/06/2021 **In front of a jury composed of :**

First and Last names	**Grade**		
Mr. Smail ADJABI	Prof.	Univ. Bejaia.	President
Ms. Karima LAGHA	MCA	Univ. Bejaia	Supervisor
Mr. Nabil ZOUGAB	MCA	Univ. Bejaia	Co-supervisor
Ms. Megdouda OURBIH	Prof.	Tipaza University Center	Examiner
Ms. Aicha BARECHE	Prof.	Univ. Bejaia.	Examiner
Ms. Karima BELAIDE	Prof.	Univ. Bejaia	Examiner
Ms. Yasmina ZIANE	MCB	Univ. Bejaia	Guest

Academid year : 2020/2021

Acknowledgements

I would to express my deepest acknowledgements to the doctors Karima Lagha and Nabil Zougab for proposition of this topic, for confidence and help they have given me throughout these years of my thesis.

I would like to thank also the Professor Smail Adjabi, University of Bejaia, for honoring me by presiding my thesis defense jury, and all members of the jury: Prof. Megdouda Ourbih, Prof. Bareche Aicha, prof. Karima Belaide and the guest Dr. Yasmina Ziane for their careful reading and helpful suggestions which led to a substantial improvement of the original manuscript.

I also want to thank professor Célestin Kokonendji, University of France-Comté in France, for his help and precious advances during my training period. Special thanks go to my familly, Hamza Hamchache and Yasmina Ziane for their help and encouragements, and all my colleagues in laboratory LMA.

Dedications

I dedicate this work

In memory of my dear father who left too early, to my dear mother who is my source of strength and to whom I owe success.

To my dear brothers and sisters.

To my dear sisters-in-law, brothers-in-law, my nephews and nieces.

To all my friends and members of the LMA laboratory.

Scientific production

- **Chekkal, S.**, Lagha, K. and Zougab, N. 2021. Generalized Birnbaum–Saunders kernel for hazard rate function estimation. *Communications in Statistics-Simulation and Computation*. doi: 10.1080/03610918.2021.1887228.

- **Chekkal, S.**, Lagha, K. and Zougab, N. 2019. Hazard rate function estimation using kernel method. the 9th edition of TAMTAM in Tlemcen, Algerie, from 23-27 February 2019.

- **Chekkal, S.**, Lagha, K. and Zougab, N. 2019. Hazard rate function estimation using Generalized Birnbaum-Saunders kernel. The 51st Statistics Days in Nancy, France, from 03-07 June 2019.

- **Chekkal, S.**, Lagha, K. and Zougab, N. 2019. Failure rate estimation using the class of Generalized Birnbaum-Saunders kernels. international congress on mathematical modeling and numerical analysis, Meknès, Morroco, from 16-18 December 2019.

Contents

List of Figures

List of Tables

Introduction

Survival analysis also known as failure time analysis, is one of the most significant advancements of statistics in the last quarter of 20th century. It analysis the expected duration of time until one or more events happen, such as death in biological organism and failure in mechanical system. This topic is called "reliability theory" or "reliability analysis" in engineering.

The term 'reliability' in engineering refers to the probability that a product, or system will perform its designed functions without failures, under a given set of operating conditions for a specific period of time. The primary aim of reliability system is the prevention of these failures that affect the operational capability of the system. Many tools developed in survival analysis, especially in reliability engineering are naturally formulated via the hazard rate (HR) called also failure rate (FR) concept. In actuarial and demographic disciplines, it is usually called "the mortality rate".

The HR function has been subject of several works, particulary in parametric and nonparametric estimation. The main problem with the parametric approach is that existing classical probability distribution families are limited in the face of a multitude of data structures. A wrong assumption concerning the underlying distribution model for the data may lead to misleading interpretations. In situations such these, nonparametric methods may be more suitable. The nonparametric methods impose only mild assumptions, such as smoothness, on the underlying probability distribution and so avoid the risk of specifying the wrong model for the data. There are several nonparametric estimation methods, such, maximum penalized likelihood estimates (de Montricher et al., 1974), orthogonal series estimates (Silverman, 1986), smoothing splines (Gu, 1993), and the one that received the most important attention is the kernel method (Rosenblatt, 1956) known by its simplicity and great flexibility.

Kernel method also known as Parzen-Rosenblatt window method is rooted from the histogram methodology, introduced firstly by Rosenlatt (1956) for the density function estimation and generalized by Parzen (1962), then developed by Nadaraya (1964) and Ferraty and Vieu (2003) for regression function estimation. This method is characterized by two essential parameters: the kernel function K which can be symmetric (classical) or asymmetric, and the smoothing parameter (bandwidth) h. As in all smoothing methods, the inherent issue is the selection of the smoothing parameter, which can be done by using several techniques such as, cross validation, plug-in and bayesian approach. It is well known from the literature that the kernel function has less impact than the bandwidth on the resulting estimate. Despite the fact, the kernel should be properly chosen regarding to the support of the function to be estimated, for instance when the density function of the data have a bounded support, using the classical kernel leads to an estimator with a large bias near the endpoints, called "boundary effect". This is especially the case in survival analysis, since the survival time is assumed to be nonnegative variable. So, near zero, the symmetric kernel estimator of the density and the HR functions underestimates the true ones, and this problem becomes a serious drawback when a large portion of the sampled data are present in the boundary region. In fact, many solutions are proposed to avoid the problem of boundary bias, such as boundary kernel method, see (Jones, 1993; Zhang and Karunamuni, 2000), local linear method (Lejeune and Sarda, 1992; Cheng, 1997; Zhang and Karunamuni, 1998), local renormalization method (Härdle, 1990), pseudo-data method (Cowling and Hall, 1996), reflection method (Cline and Hart, 1991), ect.
The asymmetric kernels have been proposed as a best solution for avoiding these boundary effects. This simple idea is due to Chen (2000) by introducing beta and gamma kernels, Scaillet (2004) by introducing Inverse Gaussian (IG) and Reciprocal Inverse Gaussian (RIG) kernels and Marchant et al. (2013) with Generalized Birnbaum-Sauders (GBS) kernels.

Reliability analysis mostly deals with positive random variables, which are often called lifetimes. The analysis of lifetime data by HR function has received considerable attention, for instance, Bouezmarni et al. (2011) by using gamma kernel in the context of censored data, Salha (2013, 2014) using inverse Gaussian (IG), Erlang and weibull kernels, Altun and Comert (2016) used Weibull-Exponential models to represent the typical L-shaped HR of electronic products, Moriyama and Maesono (2018) proposing a new kernel estimator of HR function, that is based on a modification of $\check{C}$wik and Mielniczuk method and Athayde et al. (2019) have analyzed the failure rate of generalized Birnbaum–Saunders GBS distributions; the change-points and statistical robustness are discussed.

Our work focuses on estimation of HR function using the nonparametric kernel method in the case of complete data, by using the class of GBS kernels. The choice of this class of kernels is motivated by several points. First, the family of GBS kernels includes various special cases such as, BS-classical (BS), BS-power-exponential (BS-PE), BS-Student (BS-t) and BS-laplace (BS-lap) kernels. As second motivation, some applications of GBS kernels method for HR function estimation can be found in various domains, because of its several interesting properties and flexibility. In fact, the GBS distribution contains a wider class of positively skewed densities with nonnegative support that possesses lighter and heavier tails than the BS distribution. Thus, the GBS distribution is essentially flexible in the kurtosis level, see Marchant et al. (2013).

Our study is not restricted to HR function alone; two other important reliability measures related to the HR function are also studied using the class of GBS kernels, as well; reliability function (survival function) and reversed hazard rate function (RHR). These two functions have also attracted considerable attention among researchers, for instance, Brunel et al. (2016) have studied the kernel estimator of reliability function in multiplicative censoring model, Srivastava (2020) has estimated the reliability Function of Log Gompertz Model, Desai et al. (2011) have analyzed the nature of RHR and Veres-Ferrer and Pavia (2014) by studying the relationship between the RHR and elasticity.

We have organized this document in four chapters. In Chapter 1, we present some basic concepts in reliability theory, in particular the general properties of HR, reliability and RHR functions. The Chapter 2 presents the kernel method and gives some clarification about the class of GBS kernel in the case of density function estimation. The Chapter 3 deals with the HR function estimation using kernel method. First we give an overview of some results of kernel estimation of HR function in both cases of complete and censored data, then we introduce our proposed kernel estimator using the class of GBS kernel and we study its asymptotic properties, strong consistence and asymptotic normality. Also the bandwidth parameter h is estimated by two methods: rule of thumb (RT) and unbiased cross validation (UCV). In addition, simulation and application with real data are investigated to test the performance of our proposed estimator. In the Chapter 4, we use the class of GBS kernel in estimation of reliability and RHR functions and establish their asymptotic properties. The bandwidth is selected by RT and UCV methods, then simulation study is investigated to test the performance of the estimators of reliability and RHR functions, and selecting the most appropriate bandwidth method. We finish our document with conclusion and some perspectives.

Chapter 1

Reliability analysis

Introduction

The life distribution is characterized by many useful reliability functions such as survival function (called also reliability function), hazard rate function, reversed hazard rate function, mean residual life, etc. The behaviour of these functions serves to describe the ageing properties of a device, and prevent any eventual failures. These elementary functions and some other basic concepts are illustrated in this introductory chapter. The reader can also refer to Lai and Xie (2006), Marshall and Olkin (2007), Finkelstein (2008) and O'Connor (2011).

Usually in reliability analysis, we deal with positive random variable which represents time to failure of an engineering component. This random variable r.v. is called "lifetime", usually assumed to be continuous. We will restrict ourselves to this case, with density function pdf (called failure density) and cumulative density function cdf.

1.1 Basic reliability concepts

In this section, we summarize the essential elements in reliability analysis, such as failure density function, reliability function, residual life distribution, hazard rate function and reversed hazard rate function.

Let T be a r.v representing lifetime of an item, with pdf f and cdf F.

1.1.1 Failure density function

Generally, the density failure function of the lifetime r.v. T is positively skewed (skewed to the right or steep on the left–hand side). Thus, $f(t)$ has a flat and relatively long right–hand tail, meaning that longer lifetimes are less probable than shorter lifetimes and that the mean life (life expectancy) is greater than the median life.

Especially for a newly born organism or a produced unit, e.g., for a unit starting at age $t = 0$, the probability to fail up to an age $t > 0$, called cumulative distribution function cdf, is given by

$$\mathbb{P}(T \leq t) = \int_0^t f(u)du, \qquad t > 0.$$

Definition 1.1.1. *The failure density function that represents the probability of failure in the interval* $[t, t + dt]$, *with* dt *is small enough, is defined as*

$$\begin{aligned} f(t) = F'(t) &= \lim_{dt \to 0} \frac{F(t+dt) - F(t)}{dt} \\ &= \lim_{dt \to 0} \frac{1}{dt} \mathbb{P}(t < T < t + dt), \quad t > 0. \end{aligned}$$

Therefore, when dt *is sufficiently small*

$$f(t)dt \simeq \mathbb{P}(t < T < t + dt).$$

All probability density functions for the variable 'lifetime' satisfies the two popular conditions: $f(t) > 0, \ \ \forall t > 0$ and $\int_0^\infty f(t)dt = 1$.

The mean time to failure (MTTF), also called average lifetime, expected lifetime, life expectancy, or mean lifetime, is another important descriptor in lifetime analysis. Thus, the mean time to failure of the r.v. T is given by

$$\mu = \mathbb{E}(T) = \int_0^\infty t f(t)dt.$$

1.1.2 Reliability function

The reliability function is considered as a very useful indicator in several fields. Reliability engineers use the reliability function in many types of decision making. In manufacturing, this function provides a tool for setting warranties. In system safety designs, it provides one basis of safety assessment. Thus, accurate estimation of the reliability function is of importance in many industries. The most common method modeling the reliability function is the non-parametric Kaplan–Meier estimator proposed by Kaplan and Meier (1958). Many works are carried out on this function, see for instance, Brunel et al. (2016) and Srivastava (2020).

The reliability of a device is defined as the probability that this device performs its intended function for a given period of time under conditions specified for its operation. When the device does not perform its function satisfactorily, we say that it has failed. When the random variable T represents the lifetime of a device, the observation on T is realized as the time of failure.

Definition 1.1.2. *The reliability function R, also known as the survival function S, is defined as*

$$S(t) = R(t) = \mathbb{P}(T > t) = 1 - \mathbb{P}(T \leq t) = 1 - F(t), \qquad t > 0.$$

It represents the probability that the random event (time of failure) occurs after t.

The reliability function R satisfies the following properties

- $R(t)$ is a decreasing function, with t.
- $\lim_{t\to 0} R(t) = 1$ and $\lim_{t\to\infty} R(t) = 0$,

The Figure 1.1 gives the form cdf and reliability function.

The components and the way in which they are arranged within the system, have a direct effect on the entire system reliability. We consider $T_1, T_2, ...T_n$ lifetimes of n components, and suppose they are independent, with reliability function R_i, $\; i = \overline{1, n}$.
Let T_s and R_s be the lifetime and the reliability function of the system, respectively.

In the case of series system (Figure 1.2), the reliability of this system is always lower than the reliability of any of its components, it fails if any of its elements fails. Then its reliability function is obtained simply as the product of probabilities of each elements.

$$R_s(t) = \mathbb{P}(T_s > t) = \mathbb{P}(\min_i T_i > t) = \prod_{i=1}^{n} \mathbb{P}(T_i > t) = \prod_{i=1}^{n} R_i(t).$$

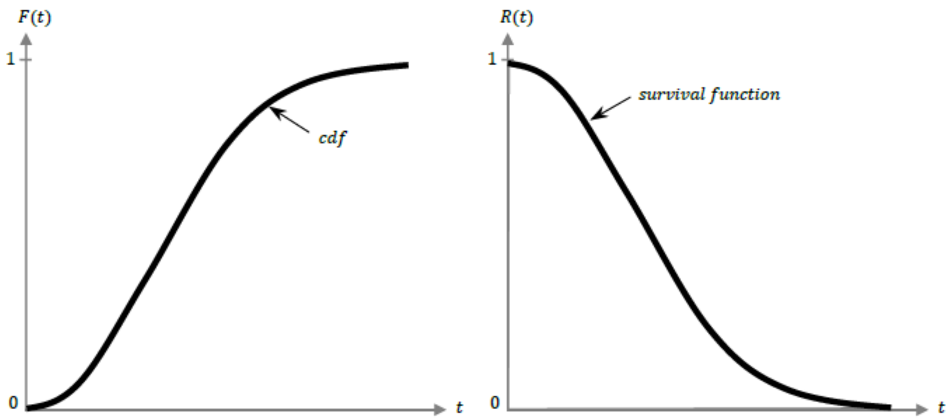

Figure 1.1: Cumulative density function (cdf) and reliability function.

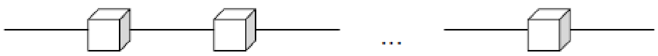

Figure 1.2: Series system model.

For the case of parallel system (Figure 1.3), it is sufficient that one of its components work to make the system work, because it fails only if all its parts fail. Its reliability function is given by

$$R_s(t) = \mathbb{P}(T_s > t) = \mathbb{P}(\max_i T_i > t) = 1 - \mathbb{P}(\max_i T_i \leq t), \quad i = 1..n,$$

and as the components are independent, we write

$$R_s(t) = 1 - \prod_{i=1}^{n}(T_i \leq t) = 1 - \prod_{i=1}^{n}(1 - R_i(t)).$$

1.1.3 Residual life distribution

How much longer will an item of age t live? This question is vital for reliability analysis, survival analysis, actuarial applications and other disciplines. For example, how much time does an average person aged 65 have left to live? The residual life is an important measure in reliability application which summarizes the entire remaining life distribution. Let an item with a lifetime T and a cdf $F(t)$ start operating at $t = 0$. The residual lifetime at time

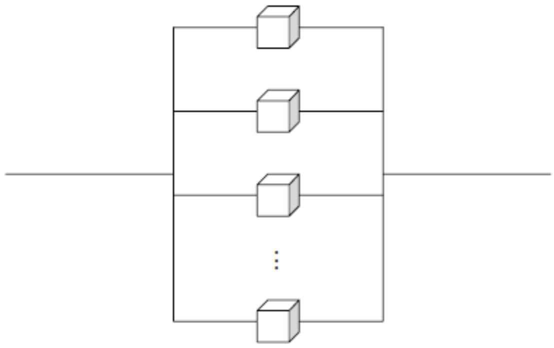

Figure 1.3: Parallel system model.

$t \geq 0$ is the time left up to the failure for a component starting its life at time 0 and it is still alive at time t.

Definition 1.1.3. *Let F be a distribution function such that $F(0) = 0$. The residual life distribution of F at t, noted by F_t is defined for all $t > 0$, $s > 0$ and such that $R(t) > 0$, by*

$$F_t(s) = \mathbb{P}(T \leq t + s \,|_{T>t}) = 1 - \frac{R(t+s)}{R(t)}. \tag{1.1}$$

Clearly, the residual life distribution F_t is a conditional distribution of the remaining life given survival up to time t. This distribution is of considerable practical importance because the remaining life of devices (used cars, etc) or of biological entities (people, for example) is often of interest.

The mean residual life function $m(t)$ is the mean of the residual life distribution F_t as a function of t, and is given by

$$m(t) = \int_0^\infty \frac{R(t+s)}{R(t)} ds, \tag{1.2}$$

for t such that $R(t) > 0$.

Other terms have been used for this function; in the context of actuarial science, it has been called "the average excess claim" or "the mean excess function", see Marshall and Olkin (2007).

1.1.4 Hazard Rate (HR) function

The major notion in survival analysis is the hazard function called also failure rate function, noted by λ. It defines the conditional probability that a component fails in a small time interval, given that it has survived from time zero until the beginning of the time interval.

Definition 1.1.4. *If F is an absolutely continuous cdf with density f, then the hazard rate function λ is defined by*

$$\begin{aligned}\lambda(t) &= \lim_{dt\to 0}\frac{P(t<T\leq t+dt\mid_{T>t}}{dt} = \lim_{dt\to 0}\frac{F(t+dt)-F(t)}{R(t)dt}\\ &= \frac{f(t)}{R(t)}, \qquad t>0.\end{aligned}$$

Note that $R(t) > 0$ and the density function f is well defined for all $t \geq 0$.

The hazard rate is also called failure rate, death rate, force of mortality and intensity function in other disciplines such as survival analysis, actuarial science, demography, extreme value theory and bio-sciences.

The hazard rate measures the propensity to fail depending on the age reached and it thus plays a key role in characterizing the process of aging and in classifying lifetime distributions. For instance, when we want to predict the chance of failure at age t for a newly born or produced unit having $F(t)$ as its cdf we have to use $f(t)$ (failure density), i.e., $f(t)$ is an unconditional predictor for risk to fail at t. But when we know that a unit has survived up to t, we have to use $\lambda(t)$ which is a conditional predictor.

The HR function $\lambda(t)$ satisfies the following properties

$$\lambda(t) \geq 0, \quad \forall t \geq 0, \quad \text{and} \quad \int_0^\infty \lambda(t)dt = \infty.$$

Care should be taken not to confuse the hazard rate with the Rate of Occurrence of Failures (ROCOF). The ROCOF is the probability that a failure (not necessarily the first) occurs in a small time interval. Unlike the hazard rate, the ROCOF is the absolute rate at which system failures occur and is not conditional on survival to time t. The ROCOF is using in measuring the change in the rate of failures for repairable systems. (O'Connor, 2011).

As is well-known, the density probability function of a random variable can be integrated to obtain the cumulative distribution function. Analogously, the hazard rate of a variate can be integrated to obtain the cumulative hazard rate. Specifically, the cumulative hazard rate of an r.v. T is given by

$$\Lambda(t) = \int_0^t \lambda(x)dx, \qquad t > 0,$$

and it satisfies three conditions

- $\Lambda(0) = 0$,
- $\lim_{t\to\infty} \Lambda(t) = \infty$;
- $\Lambda(t)$ is increasing with t.

As the hazard rate can vary over time, then it is useful to find an average value (known as failure rate average) to represent the behavior of this rate in an interval of fixed time, say $[0, t]$. Thus, the failure rate average (FRA) of an r.v. T is given by

$$\text{FRA}(t) = \frac{\Lambda(t)}{t}, \quad t > 0.$$

Relationship among density, distribution, reliability and hazard functions are presented in the Table 1.1.

	$f(t)$	$R(t)$	$\lambda(t)$	$\Lambda(t)$
$f(t) =$	-	$-R'(t)$	$\lambda(t)\exp\{-\int_0^t \lambda(x)dx\}$	$\frac{-d\{\exp[-\Lambda(t)]\}}{dt}$
$R(t) =$	$\int_t^\infty f(x)dx$	-	$\exp\{-\int_0^t \lambda(x)dx\}$	$\exp\{-\Lambda(t)\}$
$\lambda(t) =$	$\frac{f(t)}{1-\int_0^t f(x)dx}$	$-\frac{R'(t)}{R(t)}$	-	$\Lambda'(t)$
$\Lambda(t) =$	$-\ln\int_t^\infty f(x)dx$	$-\ln[R(t)]$	$\int_0^t \lambda(x)dx$	-

Table 1.1: Summary of important functions relationships (O'Connor 2011).

.

The different graphical Shapes of HR function

Hazard rate can take any graphical forms, according to the life time distribution. We distinguish

- Increasing form: The intuitive content of an increasing hazard rate stems from the interpretation of $\lambda(t)dt$ as the conditional probability of failure in the interval $[t, t+dt]$ given survival up to time t. Thus, with an increasing hazard rate, the probability of failure in the next instant of time increases as the device or organism ages. In a very real sense this is a mathematical translation of the intuitive concept of "adverse ageing," but it would be unfair to claim that it is the only mathematical translation of this concept.

- Decreasing form: An item has a decreasing hazard rate if, as it ages, the chance of failure (death) in the next instant of time decreases. This is the opposite of wear-out, and might be called "wear-in." Humans might exhibit a decreasing probability of failing at some particular job as they gain experience and practice. But mixtures may be the most important source of distributions with decreasing hazard rates.

- Bath-Tub form

Definition 1.1.5. *(Marshall and Olkin, 2007) A distribution is said to have a bath-tub hazard rate if for some $0 < a < b$, the hazard rate $\lambda(t)$ is decreasing in t, $0 < t < a$, is constant in the interval $a < t < b$, and is increasing in t, $t > b$.*

The bath-tub curve is the most popular graph in reliability application, and may be broadly classified in three distinct time zones, each one corresponds a distinctive failure mode: infant mortality (wear-in), youth (constant rate) and aging (wear-out) mode, as shown above in Figure 1.4.

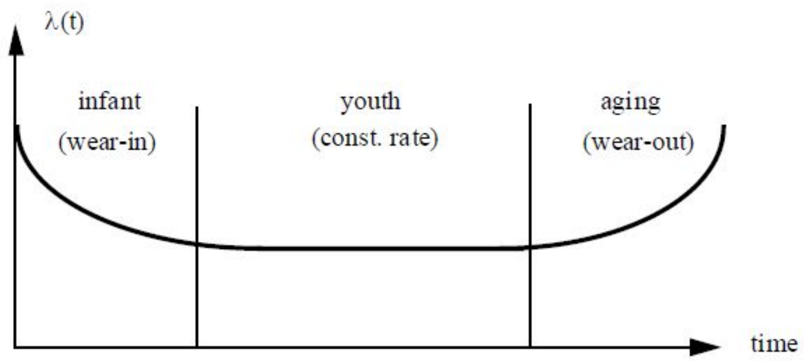

Figure 1.4: Form 01 of bath-tub curve.

The infant mortality or *wear-in* mode is generally short, with a high but decreasing rate such as in the case of human life expectancy and the engineered product, for instance, in Engineering the wear-in mode may be due to defective parts or defects in materials. To correct this situation, one may resort to design improvement, care in materials selection and tightened production quality control.

The youth or *constant rate* mode is exhibited by those product that have survived the wear-in period. The rate is generally the lowest; and in some product it maintains a long and flat behavior.
Random failure can be reduced by improving product design, making it more robust with respect to the service condition to which it is exposed in real life.

The aging or *wear-out* mode is usually due to material fatigue. The wear-out mode is often encountered in mechanical systems with moving parts such as pumps, engines, automobile tires, ect. Onset of rapidly increasing rate requires measures such as increased regularity of inspection, maintenance, replacement, etc. Since in these products the youth period is relatively short while the wear-out period is long, such as depicted in Figure 1.5

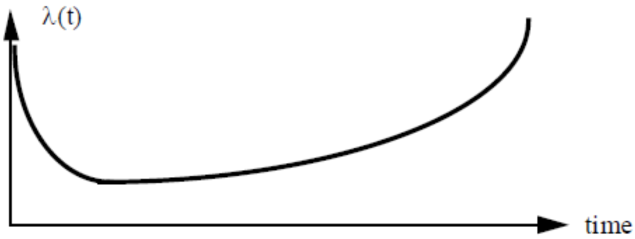

Figure 1.5: Form 02 of bath-tub curve.

- Inverted bathtub (upside-down) form

Definition 1.1.6. *(Marshall and Olkin, 2007) A distribution is said to have an inverted bathtub hazard rate if for some* $0 < a < b$*, the hazard rate* $\lambda(t)$ *is increasing in* t*,* $0 < t < a$*, is constant in the interval* $a < t < b$*, and is decreasing in* t*,* $t > b$*. Alternatively, such hazard rates are said to be unimodal.*

Inverted bathtub hazard rates have not attracted much interest, at least in reliability theory, perhaps because the bathtub hazard rates have been a focus of attention.
The just described origin of bathtub hazard rates for biological organisms has its counterpart for mechanical systems. A new system may suffer from "bugs", that is, from errors of design or of construction. Moreover, the operators of the system may

be initially inexperienced. As the system ages, the potential for bugs or human error diminishes, causing the hazard rate to decrease. But after a while, the effects of aging cause the hazard rate to rise.

Let $T_1, T_2, ..., T_n$ lifetimes of n components, distributed as T, and suppose that the components are independent.
In the case of the series system, the HR is represented as the sum of the HR of the components, that is

$$\lambda(t) = \sum_{i=1}^{n} \lambda_i(t).$$

In the case of parallel system, we use the fact that

$$R(t) = 1 - \prod_{i=1}^{n} \left\{ 1 - \exp\left[-\int_0^t \lambda_i(u) du \right] \right\},$$

and

$$\lambda(t) = -\frac{R'(t)}{R(t)}.$$

We obtain the following formula of HR

$$\lambda(t) = \frac{\sum_{i=1}^{n} \lambda_i(t) \exp\left[-\int_0^t \lambda_i(u) du \right] \prod_{j \neq i} \left\{ 1 - \exp\left[-\int_0^t \lambda_j(u) du \right] \right\}}{1 - \prod_{i=1}^{n} \left\{ 1 - \exp\left[-\int_0^t \lambda_i(u) du \right] \right\}}.$$

1.1.5 Reversed Hazard Rate (RHR) function

The reversed hazard rate RHR function was introduced by Keilson and Sumita (1982), and have attracted considerable interest among researchers, see for instance Chandro and Roy (2001, 2005), and Finkelstein (2002).

Definition 1.1.7. *Let T be a r.v. representing lifetime, with density function f and cumulative distribution function F, the reversed hazard rate of T is defined as*

$$\begin{aligned} \rho(t) &= \lim_{dt \to 0} \frac{\mathbb{P}(t - dt < T \leq t \,|_{T \leq t})}{dt} \\ &= \frac{f(t)}{F(t)}, \quad t > 0. \end{aligned}$$

Thus, $\rho(t)dt$ can be interpreted as an approximate probability of a failure in $(t-dt, t]$, given that the failure had occurred in $]0, t]$. We can establish the relation between the hazard rate function λ and the reversed hazard rate function ρ as

$$\rho(t) = \frac{\lambda(t)}{\exp\left\{\int_0^t \lambda(u)du\right\} - 1}, \quad t > 0.$$

Note that,

$$\lim_{t\to 0} \rho(t) = \infty.$$

In lifetime data analysis, the concepts of reversed hazard rate has potential application when the time elapsed since failure is a quantity of interest in order to predict the actual time of failure, and it is more useful in estimating reliability function when the data are left censored or right truncated.

The RHR function was initially introduced by actuarial research, until now it has mainly been applied to reliability engineering (Desai et al., 2011). So it plays a vital role in the analysis of parallel systems, indeed for identical independently distributed components, the RHR of the system life is proportional to the RHR of each component, and this is not obvious for the HR function. Reliability engineering, however, is not the only field where this tool has proved useful. Reversed hazard can be also employed for analyzing right-truncated and left-censored data. See, for instance Finkelstein (2008) and Desai et al.(2011).

A number of different applications of the RHR function, in the study of lifetime r.v. have been already investigated in the literature. Thus, Andersen et al.(1993) use the RHR in the estimation of the survival function in the presence of left censored observations. Block et al. (1998) characterize some useful properties for k out of n systems in terms of the RHR. Some properties of the waiting time (time elapsed since the failure of an object till the time of observation) with respect to the RHR were studied by Chandra and Roy (2001). In addition, Veres-Ferrer and Pavia (2014) expand the usefulness of RHR in economics.

1.2 Common life distibutions

We use the term "life distributions" to describe the collection of statistical probability distributions that we use in reliability engineering and life data analysis. Naturally any distribution of non-negative random variables could be used to describe durations.

The distributions to be presented here are all continuous and they have appeared more frequently in the literature, such as Marshall and Olkin (2007), Lai and Xie (2006) and O'Connor (2011).

Recall that T is a lifetime r.v. with pdf f and cdf F.

1.2.1 Exponential distribution

The exponential distribution is a fairly simple distribution commonly used in reliability analysis. If the r.v. T follows the exponential distribution, with the parameter λ, $T \sim exp(\lambda)$ then, the corresponding density f, reliability function R, hazard rate λ and reversed hazard rate ρ are defined respectively for $t > 0$, by

$$\begin{aligned} f(t) &= \lambda \exp\{\lambda t\}, \\ R(t) &= \exp\{-\lambda t\}, \\ \lambda(t) &= \lambda \\ \rho(t) &= \frac{\lambda}{\exp\{\lambda t\} - 1}. \end{aligned}$$

where the parameter $\lambda > 0$ acts both as a scale parameter and a frailty parameter. Note that the hazard rate λ is constant, so the exponential distribution is used to model the behavior of items that have a constant failure rate (i.e., items that do not degrade with time or wear out). This is the case of many engineering devices (especially electronic) which have a constant hazard rate ($\lambda > 0$) during the usage period.

The exponential distribution is the only one that possess the memoryless property, in the continuous case, so

$$F(t/x) = F(t), \forall x, t \geq 0,$$

where $F(t/x)$ is a conditional distribution function.

The following propositions is given in Marshall and Olkin (2007).

Proposition 1.2.1. *A distribution has a constant hazard rate if and only if it is an exponential distribution.*

Proposition 1.2.2. *A distribution F has a mean residual life independent of age if and only if it is an exponential distribution.*

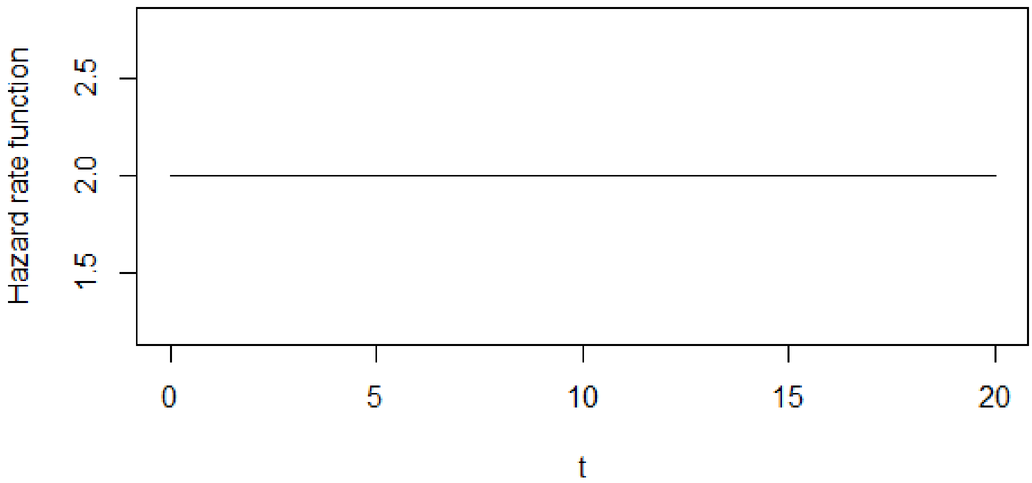

Figure 1.6: HR function of exponential distribution, with $\lambda = 2$.

1.2.2 Gamma distribution

Let the r.v. T follows the gamma distribution with shape and scale parameters $\alpha > 0, \beta > 0$. It is characterized by the density

$$f(t) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} t^{\alpha-1} \exp(-\beta t), \quad t > 0.$$

Where the gamma function is defined in the usual way by

$$\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} \exp(-t) dt.$$

Proposition 1.2.3. *(Marshall and Olkin, 2007)*

The density of the gamma distribution is

- *completely monotone, log convex, and decreasing, for $0 < \alpha < 1$,*
- *log concave and unimodal, for $\alpha \geq 1$, with mode at the point $t = (\alpha - 1)/\beta$.*

The reliability function can be given in closed form only when α is an integer. In that case,

$$R(t) = \sum_{k=0}^{\alpha-1} \exp(-\beta t)^k/(!k), \quad t > 0.$$

When α is a positive integer, the gamma distribution can be called the Erlang distribution. The HR and RHR of the gamma distribution do not take a convenient form. Following Barlow and Proschan (1975, p. 74), we write the survival function as an integral of the density to obtain

$$\frac{1}{\lambda(t)} = \int_t^{\infty} \left(\frac{z}{t}\right)^{\alpha-1} \exp(-\beta(z-t))dz = \int_0^{\infty} \left[1+\frac{u}{t}\right]^{\alpha-1} \exp(-\beta u)du,$$

where the second integral is obtained from the first by the change of variable $u = z - t$. By deriving the expression above, it is easy to see that for all $\beta > 0$

- $\alpha < 1$, $\lambda(t)$ increases with time.
- $\alpha > 1$, $\lambda(t)$ decreases with time.
- $\alpha = 1$, $\lambda(t)$ is constant, (case of exponential distribution),

  and $\lim_{t\to\infty} \lambda(t) = \beta$, for all $\alpha > 0$.

The gamma distribution is flexible in shape and can give good approximations to life data.

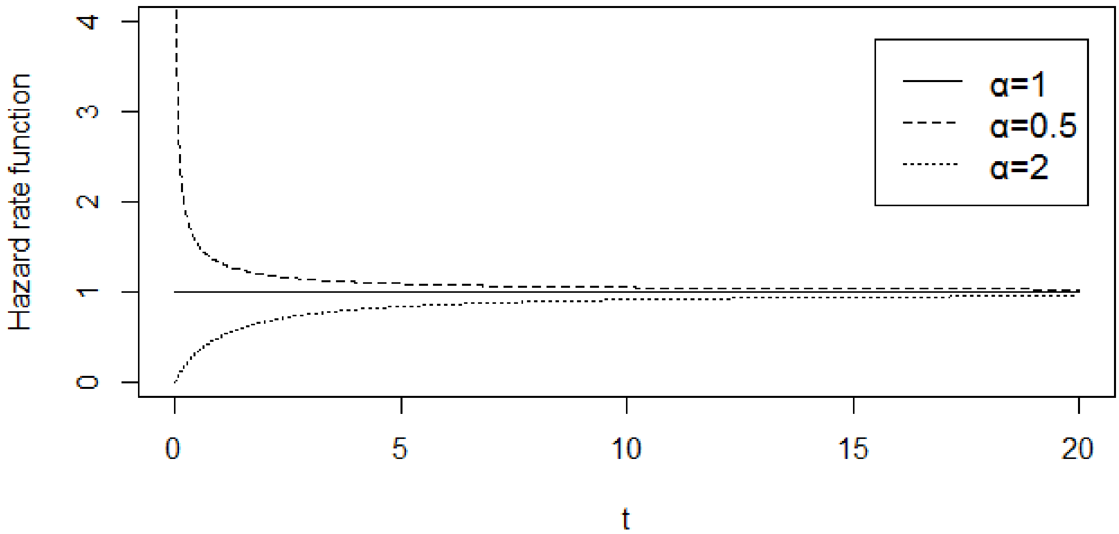

Figure 1.7: HR function of gamma distribution, with $\beta = 1$.

1.2.3 Weibull distribution

The Weibull distribution with parameters η and β denoted by $\mathcal{W}(\eta, \beta)$ can be viewed as a generalization of the exponential distribution, it has wide application in reliability analysis (e.g., engines, mechanical devices) and in human mortality.

Its density f, reliability R, hazard rate λ, and reversed hazard rate ρ, are given respectively for $t > 0$ as

$$\begin{aligned} f(t) &= \frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1} \exp\left[-\left(\frac{t}{\eta}\right)^{\beta}\right], \\ R(t) &= \exp\left(-\frac{t}{\eta}\right)^{\beta}, \\ \lambda(t) &= \frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1}, \\ \rho(t) &= \frac{\frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1} \exp\left[-\left(\frac{t}{\eta}\right)^{\beta}\right]}{1-\exp\left(-\frac{t}{\eta}\right)^{\beta}}. \end{aligned}$$

Where $\eta > 0$ and $\beta > 0$ are shape and scale parameter, respectively. Note that the hazard rate is a function of time, it can be used to model a variety of life behavior. According to the shape parameter β, one can state that for all $\eta > 0$, if (see, O'Connor 2001, page 65)

- $\beta < 1$, $\lambda(t)$ decreases with time, that represent the infant period.
- $\beta > 1$, $\lambda(t)$ increases with time, that represent the wear-out period.
- $\beta = 1$, $\lambda(t)$ is constant, that represents the youth period (case of exponential distribution).
- $1 < \beta < 2$. $\lambda(t)$ increases less as time increases.
- $\beta = 2$, $\lambda(t)$ increases with a linear relationship to time.
- $\beta > 2$, $\lambda(t)$ increases more as time increases.
- $\beta < 3.447798$, the distribution is positively skewed. (Tail to right).
- $\beta \approx 3.447798$, the distribution is approximately symmetrical.
- $\beta > 3.447798$, the distribution is negatively skewed (Tail to left).
- $3 < \beta < 4$, the distribution approximates a normal distribution.
- $\beta > 10$, the distribution approximates a Smallest Extreme Value Distribution.

It follows that, $\lim_{t\to\infty} \lambda(t) = 0$ for $\beta < 1$ and $\lim_{t\to\infty} \lambda(t) = +\infty$ for $\beta > 1$.

The Weibull distribution is by far the most popular life distribution used in reliability engineering. This is due to its variety of shapes and generalization or approximation of many other distributions. Analysis assuming a Weibull distribution already includes the exponential life distribution as a special case.

Some applications where the Weibull distribution has been used are: Acceptance sampling, Warranty analysis, Maintenance and renewal, Strength of material modeling, Wear modeling, Electronic failure modeling, Corrosion modeling, see (O'Connor 2011, p.66) and a detailed list with references to practical examples is contained in (Rinne 2008, p.275).

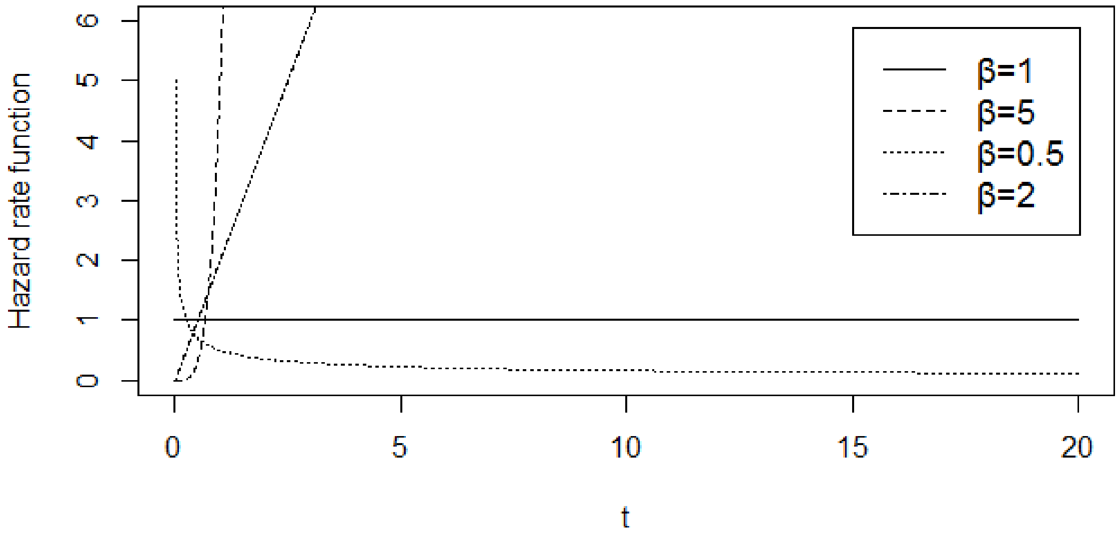

Figure 1.8: HR function of Weibull distribution.

1.2.4 Lognormal distribution

It is well known, the normal distribution is the most used in statistics, however it is not a lifetime distribution as its support is $(-\infty, +\infty)$. Therefore, the lognormal distribution is derived from the normal distribution for positive r.v.

The r.v. T follows the lognormal distribution with parameters m, σ^2, denoted by $\mathcal{LN}(m, \sigma^2)$. If $Y = \ln T$ is normally distributed, $Y \sim N(m, \sigma^2)$, where m and σ^2 are mean and variance of Y respectively. Its density f, reliability function R, hazard rate λ and reversed hazard rate ρ for $t > 0$ are

$$\begin{aligned}
f(t) &= \frac{1}{t\sigma\sqrt{2\pi}} \exp\left[-\frac{(\ln t - m)^2}{2\sigma^2}\right], \\
R(t) &= 1 - \phi\left(\frac{\ln t - m}{\sigma}\right), \\
\lambda(t) &= \frac{\phi\left[\dfrac{\ln(t) - m}{\sigma}\right]}{t\sigma\left\{1 - \phi\left[\dfrac{\ln(t) - m}{\sigma}\right]\right\}}, \\
\rho(t) &= \frac{\dfrac{1}{t\sigma\sqrt{2\pi}} \exp\left[-\dfrac{(\ln t - m)^2}{2\sigma^2}\right]}{\phi\left(\dfrac{\ln t - m}{\sigma}\right)},
\end{aligned}$$

where ϕ denotes the standard normal distribution function. The hazard rate λ is unimodal with slow decrease to zero as $t \to 0$.

The lognormal distribution has been found to accurately model many life distributions and is a popular choice for life distributions. The increasing hazard rate in early life models the weaker subpopulation (burn in) and the remaining decreasing hazard rate describes the main population. In particular this has been applied to some electronic devices and fatigue-fracture data. Its is also considered as a good candidate for modeling the repair time in engineering system. See O'Connor (2011, p. 56).

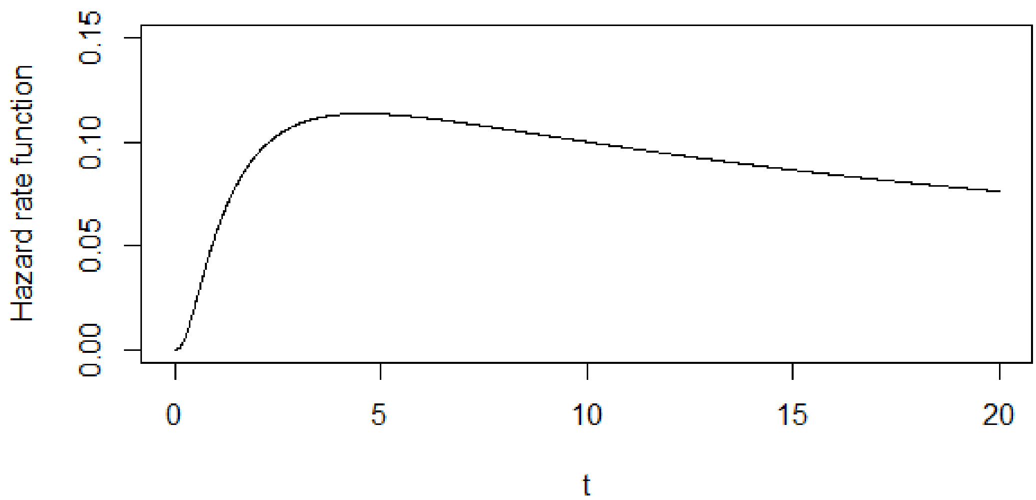

Figure 1.9: HR function of lognormal distribution with $m = 2$ and $\sigma = 1$.

1.2.5 Log-logistic distribution

The log-logistic distribution with parameters α and β is the probability distribution of a r.v whose logarithm has a logistic distribution. It is very useful in a wide variety of applications, especially in the analysis of survival data, and has been quite frequently to analyze positively skewed data.
It is characterized by the density f, reliability function R, hazard rate λ and reversed hazard rate ρ. these are given for $t > 0$ as follows

$$\begin{aligned} f(t) &= \frac{\frac{\beta}{\alpha}\left(\frac{t}{\alpha}\right)^{\beta-1}}{\left[1+\left(\frac{t}{\alpha}\right)^{\beta}\right]^{2}}, \\ R(t) &= \left[1+\left(\frac{t}{\alpha}\right)^{\beta}\right]^{-1}, \\ \lambda(t) &= \frac{\frac{\beta}{\alpha}\left(\frac{t}{\alpha}\right)^{\beta-1}}{1+\left(\frac{t}{\alpha}\right)^{\beta}}, \\ \rho(t) &= \frac{\beta}{t\left[1+\left(\frac{t}{\alpha}\right)^{\beta}\right]}, \end{aligned}$$

where $\alpha > 0$ and $\beta > 0$ are scale and shape parameters, respectively. The shape of the hazard rate function of log-logistic distribution depends on the parameter β; when $\beta > 1$ the hazard function is unimodal and when $\beta \leq 1$ it decreases monotonically. The shape of log-logistic distribution is very similar to those of log-normal distribution. Therefore, often it is very difficult to discriminate between a log-normal and a log-logistic distribution if the sample size is not very large. However, due to the symmetry of the log-logistic distribution, it may be inappropriate for modeling censored survival data, especially for the cases where the hazard rate is skewed or heavily tailed, (Kissell and Poserina 2017).

1.2.6 Birnbaum Saunders distribution (BS)

The Birnbaum-Saunders (BS) family of distributions was proposed to model the length of cracks on surfaces. In fact, it is a two-parameter distribution for a fatigue life with unimodal

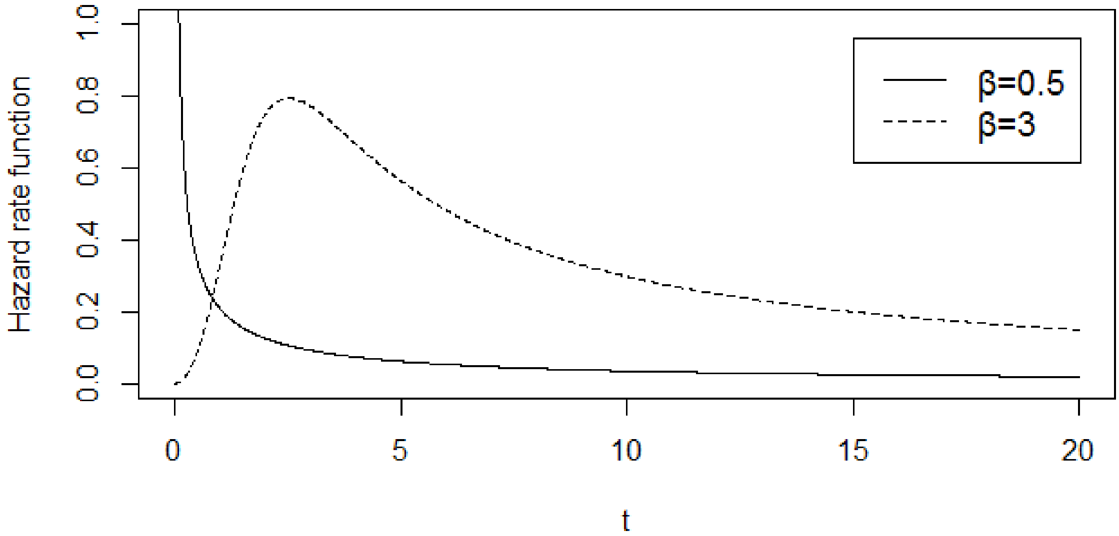

Figure 1.10: HR function of log-logistic distribution.

hazard rate function. Considerable amount of work has been done on this distribution.
If the r.v. T follows BS distribution with parameters α and β, $T \sim BS(\alpha, \beta)$ then, its density f and the reliability R are defined for $t > 0$ as

$$\begin{aligned} f(t) &= \frac{1}{2\sqrt{2\pi}\alpha\beta}\left[\left(\frac{\beta}{t}\right)^{1/2} + \left(\frac{\beta}{t}\right)^{3/2}\right] \exp\left\{\frac{-1}{2\alpha^2}\left[\frac{t}{\beta} + \frac{\beta}{t} - 2\right]\right\}, \\ R(t) &= 1 - \Phi\left\{\frac{1}{\alpha}\left[\left(\frac{t}{\beta}\right)^{1/2} - \left(\frac{\beta}{t}\right)^{1/2}\right]\right\}, \end{aligned}$$

where $\alpha > 0$ is the shape parameter, $\beta > 0$ is the scale parameter, and Φ is the standard normal distribution function. Its HR and RHR functions do not have a closed form. The HR function is always unimodal; it increases from 0 to its maximum value and then decreases to $\frac{1}{2\alpha\beta^2}$, i.e. it is upside-down bathtub shaped; see Kundu et al. (2008) and O'Connor (2011). For comprehensive reviews on various developments concerning the BS distribution, one may refer to Johnson et al. (2005) and Leiva et al. (2008).

1.3 Bathtub life distributions

The class of lifetime distribution having a bathtub shape failure rate function is very important because the lifetime of electronic, electromechanical, and mechanical products are often

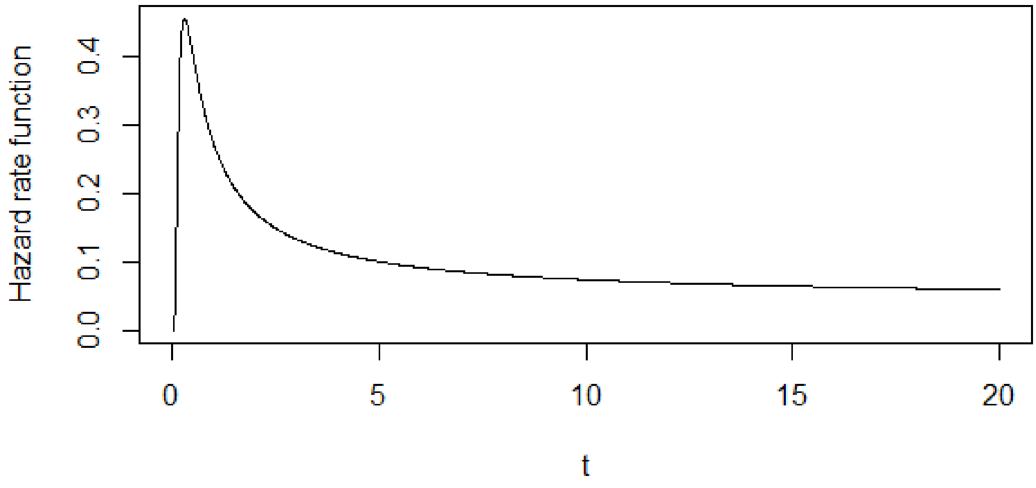

Figure 1.11: HR function of BS distribution, with $\alpha = 2$ et $\beta = 3$.

modeled with this feature. Above some of bathtub life distribution are presented.

1.3.1 Modified Weibull distribution

The modified Weibull distribution with parameters α, β and γ, is derived from the basic Weibull distribution, it is characterized by the density f, the reliability R, hazard rate function λ and reversed hazard rate function ρ, for $t \geq 0$ as

$$\begin{aligned} f(t) &= \alpha(\beta+\gamma t)t^{\beta-1}\exp(\gamma t)\exp[-\alpha t^{\beta}\exp(\gamma t)], \\ R(t) &= \exp[-\alpha t^{\beta}\exp(\gamma t)], \\ \lambda(t) &= \alpha(\beta+\gamma t)t^{\beta-1}\exp(\gamma t), \\ \rho(t) &= \frac{\alpha(\beta+\gamma t)t^{\beta-1}\exp(\gamma t)\exp[-\alpha t^{\beta}\exp(\gamma t)]}{1-\exp[-\alpha t^{\beta}\exp(\gamma t)]}, \end{aligned}$$

where $\alpha > 0$, $\gamma > 0$ are scale parameters and $\beta > 0$ is the shape parameter.

Note that for (See, O'Connor, 2011, p. 81).

- $0 < \beta < 1$ and $\gamma > 0$, the HR function $\lambda(t)$ has a bathtub curve shape.
- $\beta \geq 1$ and $\gamma > 0$, the HR function $\lambda(t)$ is increasing.
- $\gamma = 0$, the HR function $\lambda(t)$ has a same form as a Weibull distribution, with two parameters.

Proposition 1.3.1. *(Xie et al., 2004)*
When the HR function is a bathtub curve ($0 < \beta < 1$ and $\gamma > 0$), then the minimum hazard rate point is given by

$$m = \frac{\sqrt{\beta} - \beta}{\gamma}.$$

1.3.2 Exponentiated Weibull distribution

The exponentiated Weibull distribution is an extension of the Weibull family obtained by adding a second shape parameter, its density f, reliability R, hazard rate function λ and reversed hazard rate ρ are given respectively for $t \geq 0$

$$\begin{aligned}
f(t) &= \frac{\beta\nu t^{\beta-1}}{\alpha^{\beta}}\left[1 - \exp\left\{-\frac{t}{\alpha}\right\}^{\beta}\right]^{\nu-1} \exp\left\{-\left(\frac{t}{\alpha}\right)^{\beta}\right\}, \\
R(t) &= 1 - \left[1 - \exp\left\{-\left(\frac{t}{\alpha}\right)^{\beta}\right\}\right]^{\nu}, \\
\lambda(t) &= \frac{\alpha^{-1}\beta\nu(t/\alpha)^{\beta-1}\left[1 - \exp\left\{-\left(\frac{t}{\alpha}\right)^{\beta}\right\}\right]^{\nu-1} \exp\left\{-\left(\frac{t}{\alpha}\right)^{\beta}\right\}}{1 - \left[1 - \exp\left\{-\left(\frac{t}{\alpha}\right)^{\beta}\right\}\right]^{\nu}}, \\
\rho(t) &= \frac{\frac{\beta\nu t^{\beta-1}}{\alpha^{\beta}}\left[1 - \exp\left\{-\frac{t}{\alpha}\right\}^{\beta}\right]^{\nu-1} \exp\left\{-\left(\frac{t}{\alpha}\right)^{\beta}\right\}}{\left[1 - \exp\left\{-\left(\frac{t}{\alpha}\right)^{\beta}\right\}\right]^{\nu}},
\end{aligned}$$

with α is a scale parameter, β and ν are shape parameters.
Note that for (see, O'Connor, 2011, p. 79).

- $\beta \leq 1$ and $\beta\nu \leq 1$, λ is monotonically decreasing.
- $\beta \geq 1$ and $\beta\nu \geq 1$, λ is monotonically increasing.
- $\beta < 1$ and $\beta\nu > 1$, λ is unimodal.
- $\beta > 1$ and $\beta\nu < 1$, λ is bathtub curve.

This distribution is applied to model failure data and extreme value data.

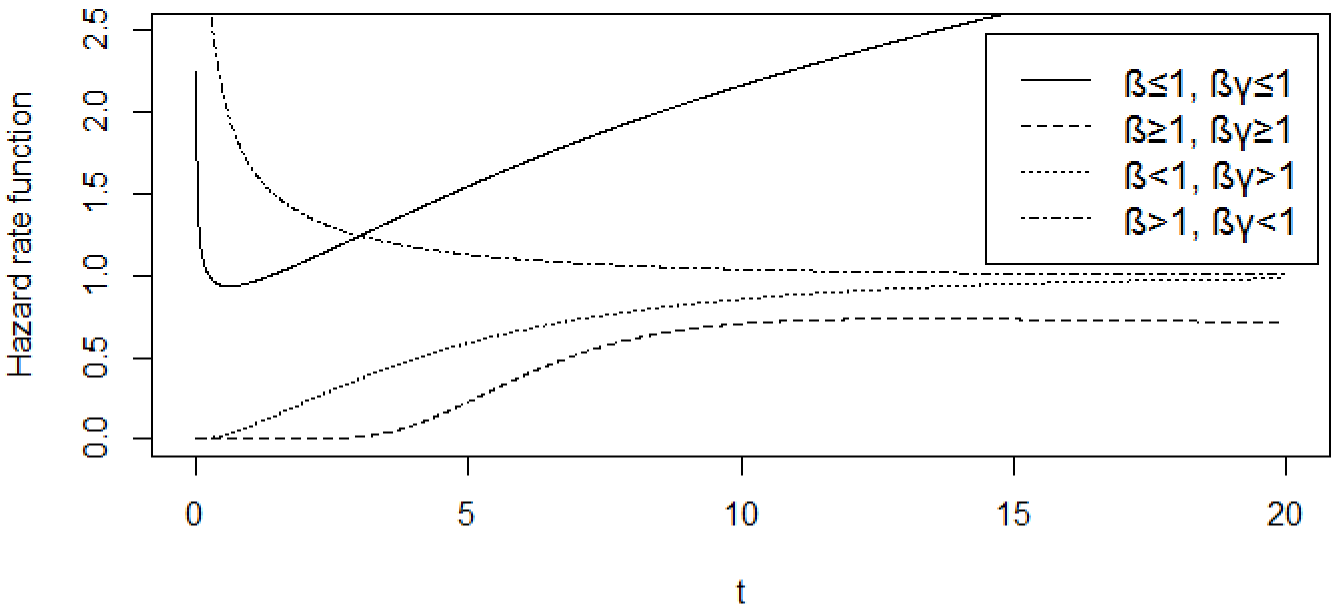

Figure 1.12: HR function of exponentiated Weibull distribution, with $\alpha = 5$.

Chapter 2

Kernel method and pdf estimation

2.1 Introduction

Kernel estimation is the most popular nonparametric method, introduced firstly by Rosenblatt (1956) and Parzen (1962) to estimate un unknown density function pdf f of a univariate random variable r.v. T, in the support $(-\infty, +\infty)$. Let $T_1, T_2, .., T_n$ i.i.d random variables, the kernel estimator $\widetilde{f}_h$ of f is given by

$$\widetilde{f}_h(t) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{T_i - t}{h}\right), \quad \forall t \in \mathbb{R},$$

with t is the target (point where the density is estimated), K is the symmetric kernel and h ($h > 0$) is the smoothing parameter (called also bandwidth) which controls the amount of smoothing of $\widetilde{f}_h$, satisfying $h \to 0$ when $n \to \infty$. The consistency of the estimator $\widetilde{f}_h$ is well documented; see Parzen (1962) or Silverman (1986), for a set of regularity conditions for consistency. However, the estimator above is not appropriate when the density to be estimated f is supported in positive half-line $\mathbb{R}_+$ (nonnegative data), because it causes problem in the boundary, called "boundary effect". The alternative way proposed by Chen (1999, 2000) is to use asymmetric kernel instead of the symmetric one. See, Hirukawa (2018) for more details about this type of kernels.

In the next, we give more details about symmetric and asymmetric kernels and we present the different methods of selection the bandwidth parameter.

2.2 Symmetric (classical) kernel

Let $T_1, T_2, ..., T_n$ a set of independent identically distributed (i.i.d.) r.v. , distributed as T, with pdf f supported in $\mathbb{R}$ and cumulative distribution function cdf F.

The approximation of the derivative f of F at a given t can be written as

$$f(t) = \lim_{h \to 0} \frac{F(t+h) - F(t-h)}{2h}, \quad \forall t \in \mathbb{R}.$$

Then, the estimator $\widetilde{f}_h$ of the pdf f is given by

$$\widetilde{f}_h(t) \simeq \frac{\widetilde{F}(t+h) - \widetilde{F}(t-h)}{2h}, \quad \forall t \in \mathbb{R},$$

where $\widetilde{F}$ is the empirical distribution function estimator of F, with $\widetilde{F}(t) = \frac{1}{n}\sum_{i=1}^{n} 1_{(-\infty,t]}(T_i)$, such that $1_{(-\infty,t]}$ is the indicator function on $(-\infty, t]$.

By conducting some simple developments, we obtain the following

$$\widetilde{f}_h(t) \simeq \frac{1}{2nh}\sum_{i=1}^{n} 1_{[-1 \leq \frac{T_i - t}{h} \leq 1]} \quad \forall t \in \mathbb{R}.$$

This kernel estimator $\widetilde{f}_h(t)$ is introduced by Rosenblatt (1956) with a uniform kernel on $[-1, 1]$ and after any years, Parzen(1962) generalized the above estimator using any probability density function K instead of the uniform kernel function on $[-1, 1]$.

So, the general expression of symmetric kernel density estimator is given by

$$\widetilde{f}_h(t) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{T_i - t}{h}\right) \quad \forall t, \tag{2.1}$$

where h is the bandwidth and K is the symmetric kernel which verify the following conditions

$$\int_{\mathbb{R}} K(u)du = 1, \quad \int_{\mathbb{R}} uK(u)du = 0, \quad \int_{\mathbb{R}} u^2 K(u)du = \sigma_K^2 < \infty. \tag{2.2}$$

The Table 2.1 gives some examples of symmetric kernels, see (e.g, Scott, 1977), and the Figure 2.1 displays their shapes.

Asymptotic properties

In this section, we present some properties of the estimator (2.1); bias, variance, mean squared error (MSE) and integrated mean squared error (MISE). These properties are established under the conditions (2.2) and by supposing that the derivatives f', f'' exist with finite integral on the support $\mathbb{R}$. The bias and the variance are given by Parzen (1962).

Kernel	Density	Support
Epanechnikov	$\frac{3}{4}(1-u^2)$	$[-1,1]$
Uniforme	$1/2$	$[-1,1]$
Triangulaire	$(1-\|u\|)$	$[-1,1]$
Biweight	$\frac{15}{16}(1-u^2)1_{\|}u\|$	$[-1,1]$
Gaussien	$\frac{1}{\sqrt{2\pi}}\exp(-\frac{u^2}{2})$	$\mathbb{R}$

Table 2.1: Some of symmetric kernels.

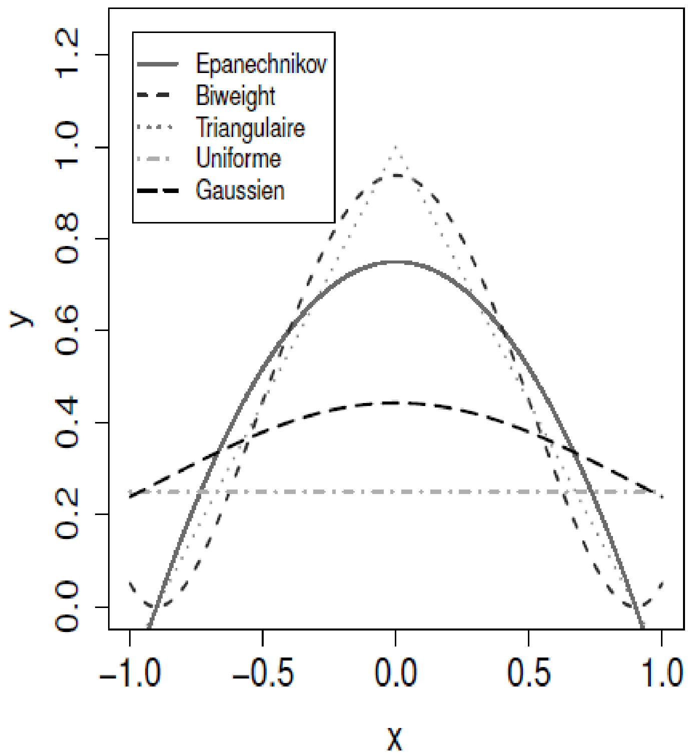

Figure 2.1: Shapes of some symmetric kernels.

Proposition 2.2.1. *(Parzen, 1962)*

For a fixed t in $\mathbb{R}$, the bias and the variance of the estimator $\widetilde{f}_h$ defined in (2.1) are

$$\begin{aligned}\text{Bias}\left[\widetilde{f}_h(t)\right] &= \frac{h^2}{2}f''(t)\int_{\mathbb{R}} z^2K(z)dz + o(h^2),\\ \text{Var}\left[\widetilde{f}_h(t)\right] &= \frac{f(t)}{nh}\int_{\mathbb{R}} K^2(z)dz + o\left(\frac{1}{nh}\right).\end{aligned}$$

In fact, the bias and the variance of $\widetilde{f}_h$ are expressed respectively as

$$\begin{aligned}\text{Bias}\left[\widetilde{f}_h(t)\right] &= \mathbb{E}\left[\widetilde{f}_h(t)\right] - f(t)\\ &= \frac{1}{h}\int_{\mathbb{R}} K\left(\frac{t-u}{h}\right) f(u)du - f(t),\end{aligned}$$

and

$$\begin{aligned}\text{Var}\left[\widetilde{f}_h(t)\right] &= \frac{1}{(nh)^2}\left\{\mathbb{E}\left[\sum_{i=1}^{n} K\left(\frac{T_i-t}{h}\right)\right]^2 - \mathbb{E}^2\left[\sum_{i=1}^{n} K\left(\frac{T_i-t}{h}\right)\right]\right\}\\ &= \frac{1}{nh^2}\left\{\int_{\mathbb{R}} K^2\left(\frac{t-u}{h}\right) f(u)du - \left[\int_{\mathbb{R}} K\left(\frac{t-u}{h}\right) f(u)du\right]^2\right\}.\end{aligned}$$

By conducting the change of variable $u = t - hz$ with $z > 0$, and the Taylor expansion of the function $f(t - hz)$ around t, we fined the results shown in Proposition 2.2.1.

By using the formulas of the bias and the variance given above, we deduce the expressions of MSE and MISE of $\widetilde{f}_h$ as follows

$$\begin{aligned}\text{MSE}[\widetilde{f}_h(t)] &= \text{Var}[\widetilde{f}_h(t)] + \text{Bias}^2[\widetilde{f}_h(t)]\\ &= \frac{1}{nh} f(t)\int_{\mathbb{R}} K^2(z)dz + \frac{h^4}{4}\left[f''(t)\right]^2\left[\int_{\mathbb{R}} z^2 K(z)dz\right]^2 + o(h^4),\end{aligned}$$

and

$$\begin{aligned}\text{MISE}[\widetilde{f}_h] &= \int_{\mathbb{R}} \text{MSE}[\widetilde{f}_h(t)]\\ &= \frac{1}{nh}\int_{\mathbb{R}} K^2(z)dz + \frac{h^4\sigma_K^4}{4}\int_{\mathbb{R}}\left[f''(t)\right]^2 dt + o\left(\frac{1}{nh}\right).\end{aligned}$$

Some convergence results of the symmetric kernel estimator

Different types of convergence results are available, some of them are summarized in the theorems below. Parzen (1962) and Tiago de Oliviera (1963) show the convergence of MSE and MISE in Theorem 2.2.1 and 2.2.2, respectively. Strong and weak consistency are established by Parzen (1962) and Silverman (1986) in 2.2.3 and 2.2.4 for the estimator $\widetilde{f}_h$. The last Theorem 2.2.5 deals with convergence in distribution.

Theorem 2.2.1. *(Parzen, 1962)*

If the density function f is continuous on $\mathbb{R}$ and $\widetilde{f}_h$ its kernel estimator, $h \to 0$, $nh \to \infty$

when $n \to \infty$ and the symmetric kernel K satisfies the following conditions

$$\int_{\mathbb{R}} K(u)du = 1, \quad \sup_{u \in \mathbb{R}} |K(u)| < \infty, \quad \int_{\mathbb{R}} |K(u)|du < \infty, \tag{2.3}$$

then

$$\text{MSE}\left[\widetilde{f}_h(t)\right] \xrightarrow{\mathbb{P}} 0, \quad \forall t \in \mathbb{R}, \quad n \to \infty.$$

Where $\xrightarrow{\mathbb{P}}$ denotes the convergence in probability.

Theorem 2.2.2. *(Tiago de Oliviera, 1963)*
If the density function f is k^{th} power integrable ($\int |f(t)|^k < \infty$), $\widetilde{f}_h$ its kernel estimator, $h \to 0$, $nh \to \infty$ when $n \to \infty$ and the symmetric kernel K satisfies the conditions in (2.3), then

$$\text{MISE}\left[\widetilde{f}_h\right] \xrightarrow{\mathbb{P}} 0, \quad n \to \infty.$$

Theorem 2.2.3. *(Parzen, 1962)*
Let f the density function and $\widetilde{f}_h$ its kernel estimator, if $nh^2 \to \infty$ when $n \to \infty$, the symmetric kernel K satisfies the conditions in (2.3) and the Fourier transform $\mathcal{TF}(t) = \int \exp(-izu)K(u)du$ is absolutely integrable, then

$$\sup_{t \in \mathbb{R}} \left[\widetilde{f}_h(t) - f(t)\right] \xrightarrow{\mathbb{P}} 0, \quad n \to \infty.$$

Theorem 2.2.4. *(Silverman, 1986)*
If the density function f is uniformly continuous and $\widetilde{f}_h$ its kernel estimator. If $h \to 0$ and $\frac{\log n}{nh} \to 0$ when $n \to \infty$ and the symmetric kernel K is positif with bounded variation, then it follows

$$\sup_{t \in \mathbb{R}} \left[\widetilde{f}_h(t) - f(t)\right] \xrightarrow{p.s.} 0, \quad n \to \infty.$$

where $\xrightarrow{p.s.}$ denotes the convergence almost surely.

Theorem 2.2.5. *(Parzen, 1962)*
If the density function f is continuous in $\mathbb{R}$ and $\widetilde{f}_h$ its kernel estimator, $h \to 0$, $nh \to \infty$ when $n \to \infty$ and the symmetric kernel K satisfies the conditions in (2.3), then

$$\frac{\widetilde{f}_h(t) - \mathbb{E}\left[\widetilde{f}_h(t)\right]}{\sqrt{\text{Var}\left[\widetilde{f}_h(t)\right]}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1), \quad \forall t \in \mathbb{R},$$

where $\xrightarrow{\mathcal{L}}$ denotes the convergence in distribution.

Efficiency of symmetric kernels

The more efficient symmetric kernel is the one that minimizes the Asymptotic integrated mean squared error criterion AMISE of the estimator (2.1), given as

$$\begin{aligned}\text{AMISE}[\widetilde{f}_h] &= \text{MISE}[\widetilde{f}_h] - o\left(\frac{1}{nh}\right),\\ &= \frac{1}{nh}\int_{\mathbb{R}} K^2(z)dz + \frac{h^4\sigma^4}{4}\int_{\mathbb{R}} \left[f''(t)\right]^2 dt.\end{aligned}$$

Theorem 2.2.6. *(Epanchnicov, 1969)*
If $\lim_{n\to\infty} h = 0$, $\lim_{n\to\infty} nh = \infty$, $f \in \mathrm{L}^2$, $\int [f''(t)]^2 dt \neq 0$ *and* $\int [f''(t)]^2 dt < \infty$ *the Epanchnikov kernel defined as* $\frac{3}{4}(1-u^2)1_{[-1,1]}$ *is of minimum* AMISE, *with* L^2 *is a set of reel functions* f *such that,* $\int |f(t)|^2 dt < \infty$.

In the case of the Epanchnikov kernel, the minimum value of AMISE is $\frac{3}{5\sqrt{5}}$ and it is declared more efficient comparing to other symmetric kernels, as shown in the theorem above. Thus, the efficiency of the other symmetric kernels can be measured based on the Epanchnikov one, as follow

$$\text{EFF}(K) = \frac{\int K^2(u)du}{\int K_E^2(u)du} = \frac{\int K^2(u)du}{\frac{3}{5\sqrt{5}}}.$$

The Table 2.2 shows the efficiency of some most used symmetric kernels.

Kernel	Efficiency
Epanechnikov	1,0000
Uniforme	1.0758
Triangulaire	1,0143
Biweight	1,0061
Gaussien	1,0513

Table 2.2: Efficiency of some symmetric kernels.

According to the Table 2.2, we note that the efficiency of the symmetric kernels is very closed to that of Epanechnikov one (the values are around 1). That means that the choice of the symmetric kernels has not significant impact on the quality of estimation.

2.3 Asymmetric kernel

It is known from the literature that the symmetric kernels are not suitable when the support of the density f to be estimated is bounded. More specifically, when it lies on the unit interval $[0, 1]$ or the positive half-line $\mathbb{R}_+$ (i.e. when we deal with nonnegative data), then the consistency of the density estimator at the origin no longer holds because the symmetric kernel assigns positive weights outside the support when smoothing is carried out near the origin. So, it causes problem in the boundary called **boundary bias** or **edge effect**.
Several methods have been proposed to remove this bias problem, such as boundary kernel method (Jones 1993; Zhang and Karunamuni, 2000), local linear method (see Lejeune and Sarda, 1992; Cheng, 1997; Zhang and Karunamuni, 1998), local renormalization method (see Härdle, 1990), pseudo-data method (see Cowling and Hall, 1996), reflection method (see Cline and Hart, 1991), and transformation method (see Marron and Ruppert, 1994). For other methods, see Hall and Park (2002) and Karunamuni and Alberts (2006).
An alternative way to remove the aforementioned **boundary bias** is to use kernels of asymmetric distributions with nonnegative support instead of classical kernels, called asymmetric kernels. This type of kernels has been first introduced by Chen (1999, 2000) using Beta and gamma density functions as kernels to estimate densities with support $[0, 1]$ and $[0, \infty)$ respectively. Jin and Kawczak (2003) introduced log-normal and Birnbaum–Saunders (BS) kernels, whereas Scaillet (2004) applied inverse Gaussian (IG) and reciprocal inverse Gaussian kenels and Marchant et al. (2013) introduced a class of Generalized Birnbaum-Saunders kernls, ect.

The asymmetric kernel estimator do not engender **boundary bias**, see the Figure 2.2, and it gave better estimates when data are nonnegative, see for instance, Libengué (2013).

Definition 2.3.1. *For nonnegative i.i.d. r.v. $T_1, T_2, ..., T_n$ distributed as T, with pdf supported in $\mathbb{T}$, $\mathbb{T} \subseteq \mathbb{R}_+$, the density kernel estimator of f, using asymmetric kernel as*

$$\widehat{f}_h(t) = \frac{1}{n} \sum_{i=1}^{n} K_{t,h}(T_i), \qquad t > 0, \tag{2.4}$$

with h is bandwidth parameter and $K_{t,h}$ is the asymmetric kernel which intrinsically depends on the bandwidth h and on the target point t.

Note that the standard kernel density estimator can also be rewritten as in (2.4), with $K_{t,h}(\cdot) = (1/h)K(\cdot - t)/h$.

The function $K_{t,h}(\cdot)$ is said to be an asymmetric kernel if it possesses the following two basic properties, see Hirukawa (2018)

Property 2.3.1. *The kernel function is a pdf with support either on the unit interval* $[0, 1]$ *or on the positive half-line* $\mathbb{R}_+$.

Property 2.3.2. *Both the location and shape parameters in the kernel are functions of the design point* t *where the estimation is made and the smoothing parameter* h.

We cite in the Table 2.3 page 51, some of asymmetric kernels with their statistics properties (where $\mathcal{Z}_{t,h}$ is the r.v. obeying the distribution with pdf K, we give more details about the class of Generalized Birnbaum-Sauders (GBS) kernels in the Section 2.4).

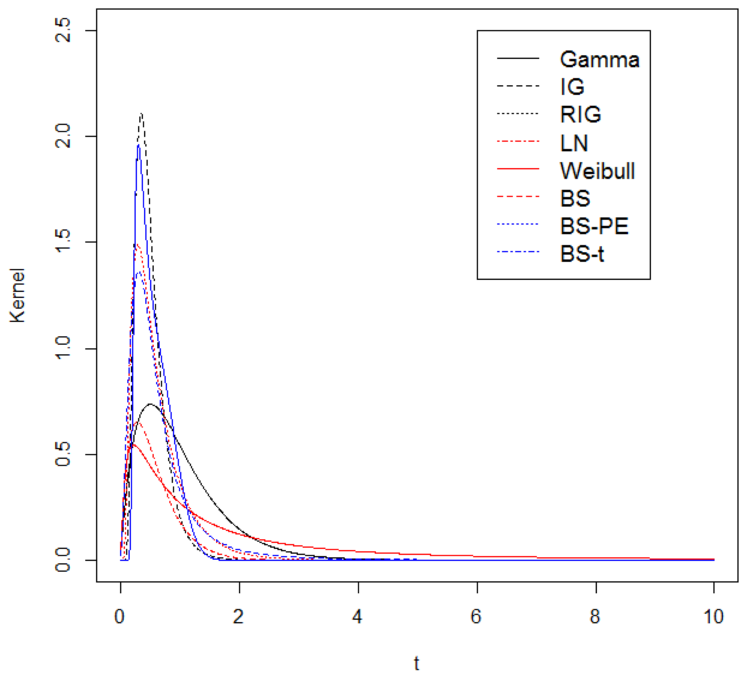

Figure 2.2: Shape of some asymmetric kernels.

Asymptotic properties

In this section, we introduce the bias and the variance of the estimator (2.4) then, we deduce the expressions of MSE and MISE of the estimator $\widehat{f}_h$. We find these results in Zougab (2013).

Proposition 2.3.1. *The bias and the variance of the estimator defined in (2.4) are given as below*

$$\begin{aligned}\text{Bias}\left[\widehat{f}_h(t)\right] &= f\left[\mathbb{E}(\mathcal{Z}_{t,h})\right] + \frac{1}{2}\text{Var}(\mathcal{Z}_{t,h})f''(t) - f(t) + o(h^2).\\ \text{Var}[\widehat{f}_h(t)] &= \frac{1}{n}\int_{\mathbb{R}_+} K^2_{t,h}(u)f(u)du - \frac{1}{n}\left\{\text{Bias}[\widehat{f}(t)] + f(t)\right\}^2.\end{aligned}$$

In fact, the bias of $\widehat{f}_h$ can be written as

$$\begin{aligned}\text{Bias}\left[\widehat{f}_h(t)\right] &= \mathbb{E}[\widehat{f}_h(t)] - f(t)\\ &= \mathbb{E}[K_{t,h}(T)] - f(t)\\ &= \int_{\mathbb{T}} K_{t,h}(u)f(u)du - f(t)\\ &= \mathbb{E}[f(\mathcal{Z}_{t,h})] - f(t),\end{aligned}$$

where $\mathcal{Z}_{t,h}$ is the r.v. obeying the distribution with pdf $K_{t,h}$, and let $\mathbb{E}(\mathcal{Z}_{t,h}) = m$. Suppose that the function to be estimated f admits derivatives of second-order, and we conduct a second-order Taylor expansion of $f(\mathcal{Z}_{t,h})$ around m, as

$$f(\mathcal{Z}_{t,h}) = f(m) + (\mathcal{Z}_{t,h} - m)f'(m) + \frac{1}{2}(\mathcal{Z}_{t,h} - m)^2 f''(m) + o[(\mathcal{Z}_{t,h} - m)^2].$$

By developing $\mathbb{E}[f(\mathcal{Z}_{t,h})]$, we deduce the following expression of $\text{Bias}[\widehat{f}_h(t)]$

$$\begin{aligned}\text{Bias}\left[\widehat{f}_h(t)\right] &= f(m) + \frac{1}{2}E\left[(\mathcal{Z}_{t,h} - m)^2\right] f''(m_{t,h}) - f(t) + o(h^2)\\ &= f\left[\mathbb{E}(\mathcal{Z}_{t,h})\right] + \frac{1}{2}\text{Var}(\mathcal{Z}_{t,h})f''(t) - f(t) + o(h^2).\end{aligned}$$

In other hand, the variance can be expressed as

$$\begin{aligned}\text{Var}[\widehat{f}_h(t)] &= \text{Var}\left[\frac{1}{n}\sum_{i=1}^{n} K_{t,h}(T_i)\right]\\ &= \frac{1}{n}\text{Var}[K_{t,h}(T)]\\ &= \frac{1}{n}\left\{[\mathbb{E}(K^2_{t,h}(T))] - \mathbb{E}^2[K_{t,h}(T)]\right\}\\ &= \frac{1}{n}\int_{\mathbb{T}} K^2_{t,h}(u)f(u)du - \frac{1}{n}\left\{\text{Bias}[\widehat{f}_h(t)] + f(t)\right\}^2.\end{aligned}$$

By using the fact that, $\text{MSE}[\widehat{f}_h(t)] = \text{Var}[\widehat{f}_h(t)] + \text{Bias}^2[\widehat{f}_h(t)]$ and $\text{MISE}[\widehat{f}_h] = \int_{\mathbb{T}} \text{MSE}[\widehat{f}_h(t)]dt$, we deduce the expressions of MSE and MISE of the estimator $\widehat{f}_h$ as shown in the proposition above.

Proposition 2.3.2.

$$\begin{aligned}
\mathrm{MSE}[\widehat{f}_h(t)] &= \left\{ f[\mathbb{E}(\mathcal{Z}_{t,h})] + \frac{1}{2}\mathrm{Var}(\mathcal{K}_{t,h}) f''(t) - f(t) \right\}^2 \\
&\quad + \frac{1}{n}\int_{\mathbb{T}} K_{t,h}^2(u) f(u) du - \frac{1}{n}\left\{ \mathrm{Bias}[\widehat{f}(t)] + f(x) \right\}^2. \\
\mathrm{MISE}[\widehat{f}_h] &= \int_{\mathbb{T}} \left[f[\mathbb{E}(\mathcal{Z}_{t,h})] + \frac{1}{2}\mathrm{Var}(\mathcal{K}_{t,h}) f''(t) - f(t) \right]^2 dt \\
&\quad + \frac{1}{n}\int_{\mathbb{T}} \left\{ \int_{\mathbb{T}} K_{t,h}^2(u) f(u) du - [\mathrm{Bias}(\widehat{f}_h(t)) + f(t)]^2 \right\} dt.
\end{aligned}$$

Some convergence results of the asymmetric kernel estimator

Above, we show the convergence of the two criterion MSE and MISE of the estimator $\widehat{f}_h$ in Proposition 2.3.3 and 2.3.4, see for instance Zougab (2013).

Proposition 2.3.3. *(Zougab, 2013)*
If $\frac{1}{n}\int_{\mathbb{R}_+} K_{t,h}^2(u) f(u) du \to 0$ *and* $h = h(n) \to 0$ *when* $n \to \infty$*, then*

$$\lim_{n\to\infty} \mathrm{MSE}[\widehat{f}_h(t)] = 0 \quad \forall t > 0.$$

Proposition 2.3.4. *(Zougab, 2013)*
If $\frac{1}{n}\int_{\mathbb{R}_+} \left[\int_{\mathbb{R}_+} K_{t,h}^2(u) f(u) du\right] dt \to 0$ *and* $h = h(n) \to 0$ *when* $n \to \infty$*, then*

$$\lim_{n\to\infty} \mathrm{MISE}[\widehat{f}_h] = 0.$$

The uniforms weak and strong consistency of the estimator (2.4) using beta and gamma kernels are studied respectively by Bouezmarni and Rolin (2003) and Bouezmarni and Rombouts (2010), as illustrated in Proposition 2.3.1 and 2.3.2.

Theorem 2.3.1. *(Bouezmarni and Rolin, 2003)*
If f *has support on* $[0,1]$*, and is continuous and bounded on* $[0,1]$*, with* $\widehat{f}_B$ *its beta kernel estimator, and then*

$$\begin{aligned}
\sup_{t\in[0,1]} |\widehat{f}_B(t) - f(t)| &\xrightarrow{P} 0 \qquad \text{if } h + (nh^2)^{-1} \to 0, \text{ as } n \to \infty. \\
\sup_{t\in[0,1]} |\widehat{f}_B(t) - f(t)| &\xrightarrow{a.s} 0 \qquad \text{if } h + \log n/(nh^2) \to 0, \text{ as } n \to \infty.
\end{aligned}$$

Theorem 2.3.2. *(Bouezmarni and Rombouts, 2010)*
If f is supported in $\mathbb{R}_+$, continuous and bounded on a compact interval $I \subset \mathbb{R}_+$, with $\widehat{f}_G$ its gamma kernel estimator, then

$$\sup_{t\in I} |\widehat{f}_G(t) - f(t)| \xrightarrow{P} 0 \qquad \text{if } h + (nh^2)^{-1} \to 0, \text{ as } n \to \infty.$$
$$\sup_{t\in I} |\widehat{f}_G(t) - f(t)| \xrightarrow{a.s} 0 \qquad \text{if } h + \log n/(nh^2) \to 0, \text{ as } n \to \infty.$$

Other convergence results of the estimator (2.4) such as, weak and strong consistencies in the sense of both uniform and L^1 norms are available in Kokonendji and Libengué (2018).

Associated kernel

Kokonendji and Somé (2018) have introduced the *associated kernel* in the multivariate case, to unify the two notions; symmetric and asymmetric kernels.
Above, we give the definition in univariate case.

Definition 2.3.2. *Let $\mathbb{T}(\subseteq \mathbb{R})$ be the support of the pdf to be estimated, $t \in \mathbb{T}$ is a target and h is a bandwidth parameter. A parameterized pdf $K_{t,h}(\cdot)$ on support $\mathbb{S}_{t,h}(\subseteq \mathbb{R})$ is called **associated kernel**, if the following conditions are satisfied*

$$t \in \mathbb{S}_{t,h}, \qquad \mathbb{E}(\mathcal{Z}_{t,h}) = x + A(t,h) \;\; \text{and} \;\; Var(\mathcal{Z}_{x,h}) = B(t,h).$$

Where $\mathcal{Z}_{x,h}$ is a real r.v. with density $K_{t,h}$ and both $A(t,h)$ and $B(t,h)$ tend to 0 when h goes to 0.

2.4 Case of Generalized Birnbaum-Saunders (GBS) kernel

This section describes the class of GBS distribution and introduces the asymmetric associated kernel estimator based on this class of distributions.

Let T be the positive and continuous random variable (r.v.), denoted by

$$T = \beta \left(\frac{\alpha Z}{2} + \left[\left(\frac{\alpha Z}{2} \right)^2 + 1 \right]^{\frac{1}{2}} \right)^2 . \tag{2.5}$$

The r.v. T follows the Generalized Birnbaum-Saunders (GBS) distribution, $T \sim GBS(\alpha, \beta, Z)$, where $\alpha > 0$ is the shape parameter, $\beta > 0$ is the scale parameter and Z is a r.v. with symmetric distribution in $\mathbb{R}$ characterized by a position parameter $\mu = 0$ and scale parameter $\sigma = 1$, $Z \sim S_{\mathbb{R}}(0, 1)$.
The r.v. Z can be expressed in terms of T as $Z = \frac{1}{\alpha}\left(\sqrt{T/\beta} - \sqrt{\beta/T}\right)$, and Z^2 be a r.v. following a generalized chi-squared distribution with one degree of freedom, $Z^2 \sim G\chi^2(1)$.

The density function of the r.v. T is defined as follow

$$f_{\alpha,\beta}(t) = \frac{c_g}{2\alpha}\left(\frac{1}{\sqrt{\beta t}} + \sqrt{\frac{\beta}{t^3}}\right) g\left[\frac{1}{\alpha^2}\left(\frac{t}{\beta} + \frac{\beta}{t} - 2\right)\right], \quad t > 0, \tag{2.6}$$

where g that is, $f_Z(z) = c_g g(z^2)$, $z \in \mathbb{R}$, f_Z is a density function of a r.v. Z and c_g is a normalization constant that is, $c_g \int_{-\infty}^{+\infty} g(z^2)dz = 1$.

The mean and the variance of a r.v. T are given by

$$\mathbb{E}(T) = \frac{\beta}{2}\left(2 + u_1\alpha^2\right) \quad \text{and} \quad \text{Var}(T) = (\beta\alpha)^2\left(u_1 + \frac{\alpha^2}{2}u_2 - \frac{\alpha^2}{4}u_1^2\right),$$

where $u_j = \mathbb{E}[Z^j]$ are the moments of the r.v. Z^j, $j = 1, 2$

For example, if we consider the r.v. Z is normally distributed with $\mu = 0$ and $\sigma = 1$, denoted by $Z \sim N(0, 1)$, then T follows a BS distribution (as particular case of GBS distribution), $T \sim BS(\alpha, \beta)$. Therefore Z^2 follows a chi-square distribution with one degree of freedom, $Z^2 \sim \chi^2(1)$.

In the Table 2.4, page 52, we give some examples of symmetric distributions of the r.v. Z and the values of c_g, g, u_1 and u_2 corresponding to each distribution. The corresponding GBS kernel is presented in the Table 2.5, page 53.

Pdf estimation

Marchant et al. (2013) introduced a class of GBS kernels by substituting in the expression (2.6) the shape and the scale parameters by $\sqrt{h}$ and t respectively, so it is defined as $K_{GBS(t,h)}(y) = f_{\sqrt{h},t}(y)$.

Considering the random variables $T_1, T_2, .., T_n$ with an unknown density function. The estimator of the density function based on the GBS kernel is given in Marchant and al. (2013) as

$$\widehat{f}_{\text{GBS}}(t) = \frac{1}{n}\sum_{i=1}^{n} K_{GBS(t,h)}(T_i), \quad t > 0,$$

where t is the target (point where the density is estimated) and $h > 0$ is the smoothing parameter.

The expressions of the bias and variance for $\widehat{f}_{\text{GBS}}$ are derived by Marchant et al. (2013), under the following conditions

C1. The function f is twice differentiable and its second derivative is continuous and bounded;

C2. The functions $t^{-\frac{1}{2}}f(t)$ and $t^{-\frac{3}{2}}f(t)$ are continuous and bounded;

C3. The bandwidth $h = h(n)$ satisfies $\lim_{n\to\infty} h = 0$ and $\lim_{n\to\infty} nh^{1/2} = \infty$.

The asymptotic bias is given by

$$\text{bias}\left[\widehat{f}_{\text{GBS}}(t)\right] = m^* + o(h), \tag{2.7}$$

where $m^* = hu_1(g)\left[tf'(t) + t^2 f''(t)\right]/2$.

The asymptotic variance is given by

$$\text{Var}\left[\widehat{f}_{\text{GBS}}(t)\right] = \sigma^{*2} + o\left(\frac{1}{nh^{1/2}}\right), \tag{2.8}$$

where $\sigma^{*2} = c_g^2 t^{-1} f(t)/\left(c_{g^2} nh^{1/2}\right)$, $c_g = 1/\int_{-\infty}^{+\infty} g(y^2)dy$ and $c_{g^2} = 1/\int_{-\infty}^{+\infty} g^2(y^2)dy$.

Note that, the GBS distribution contains a wider class of positively skewed densities with nonnegative support that possesses lighter and heavier tails than the BS distribution. Thus, the GBS distribution is essentially flexible in the kurtosis level; see Marchant et al. (2013).

2.5 Construction of asymmetric kernels: mode-dispersion method

The majority of asymmetric kernels that exist in the literature are introduced without revealing the method of their construction such as, gamma kernel of Chen (2000), Inverse Gaussian (IG) and Reciprocal Inverse Gaussian (RIG) kernels of Scaillet (2004), ect.
The associated asymmetric kernel with two parameters (depending on the target t and bandwidth h) is constructed based on the asymmetric density function with two parameters. Libengué (2013) and Libengué and Kokonendji (2017) have proposed a method for constructing associated asymmetric kernel, called " mode-dispersion" method.

Principal of mode-dispersion method

Let $K_{\theta(a,b)}$ be a type of unimodal kernel on the support $\mathbb{S}_{\theta(a,b)}$, with $\theta(a,b)$ is a function of parameters $a > 0$ and $b > 0$. Let $M_{(a,b)}$ and $D_{(a,b)}$ represents the mode and the dispersion parameter respectively of the density kernel $K_{\theta(a,b)}$. For $t > 0$ and $h > 0$, the mode-dispersion method allows the construction of the function $K_{\theta(t,h)}$ by solving in term of a and b the following system S

$$S : \begin{cases} M_{(a,b)} = & \text{t} \\ D_{(a,b)} = & \text{h}, \end{cases}$$

Let $\theta(t,h) = \theta(a(t,h); b(t,h))$ where $a(t,h)$ and $b(t,h)$ are solutions of the system S, for $h > 0$ and $t \in \mathbb{T}$, with $\mathbb{T}$ is the support of the density to be estimated.

The following proposition shows that $K_{\theta(t,h)}$ satisfies the definition of associated kernel given in 2.3.2.

Proposition 2.5.1. *(Libengué and Kokonendji, 2017)*
Let $\mathbb{T}$ be the support of the density f to be estimated. For all $t \in \mathbb{T}$ and $h > 0$, the kernel function constructed by the mode-dispersion method $K_{\theta(t,h)}$ with support $\mathbb{S}_{\theta(a,b)} = \mathbb{S}_{\theta(a(t,h);b(t,h))}$, is such that

$$t \in \mathbb{S}_{\theta(t,h)},$$
$$\mathbb{E}(\mathcal{Z}_{\theta(t,h)}) - t = A_\theta(t,h),$$
$$\mathrm{var}(\mathcal{Z}_{\theta(\mathrm{t,h})}) = \mathrm{B}_\theta(\mathrm{t,h}),$$

where $\mathcal{Z}_{\theta(t,h)}$ is a random variable with pdf $K_{\theta(t,h)}$ and $A_\theta(t,h) \to 0$ and $B_\theta(t,h) \to 0$ when $h \to 0$.

Some examples of construction of kernels

Here, we give some examples of asymmetric kernels to illustrate the mode-dispersion method.

Gamma kernel

The density function of gamma distribution, with shape parameter $a > 0$ and scale parameter $b > 0$, is defined as

$$f_G(a, b; y) = \frac{b^{-a}}{\Gamma(a)} y^{a-1} \exp\left\{-\frac{y}{b}\right\}, \quad y > 0,$$

where Γ is the gamma function defined as

$$\Gamma(a) = \int_0^\infty t^{a-1} \exp(-t) dt.$$

Its mode $M_{(a,b)} = (a-1)b$ and dispersion parameter $D_{(a,b)} = b$.

By resolving the system S, we obtain $a = \frac{t}{h} + 1$ and $b = h$, then gamma kernel can be written as

$$K_{G(\frac{t}{h}+1,h)}(y) = \frac{h^{-(\frac{t}{h}+1)}}{\Gamma(\frac{t}{h}+1)} y^{\frac{t}{h}} \exp\left\{-\frac{y}{h}\right\},$$

with $y > 0$, $t > 0$ and $h > 0$.

Inverse Gamma kernel

The density function of inverse gamma distribution, with shape parameter $a > 0$ and scale parameter $b > 0$, is defined as

$$f_{IG}(a, b; y) = \frac{b^a}{\Gamma(a)} y^{-(a+1)} \exp\left\{-\frac{b}{y}\right\}, \quad y > 0.$$

Its mode $M_{(a,b)} = \frac{b}{a+1}$ and dispersion parameter $D_{(a,b)} = \frac{1}{b}$.

By solving the system S, we obtain $a = \frac{1}{ht} - 1$ and $b = \frac{1}{h}$, then the inverse gamma kernel is expressed as

$$K_{IG(\frac{1}{ht}-1,\frac{1}{h})}(y) = \frac{1}{\Gamma(\frac{1}{ht}-1)} y^{\frac{-1}{ht}} \exp\left\{-\frac{1}{hy}\right\} \left(\frac{1}{h}\right)^{\frac{1}{ht}-1},$$

with $y > 0$, $t > 0$ and $h > 0$.

Lognormal kernel

The density function of lognormal distribution, with the mean μ and standard deviation σ is given as

$$f_{LG}(\mu, \sigma; y) = \frac{1}{\sqrt{2\pi\sigma y}} \exp\left\{-\frac{1}{2\sigma} (\ln(y) - \mu)^2\right\}, \quad y > 0.$$

Its mode $M_{(a,b)} = \exp(\mu - \sigma^2)$ and dispersion parameter $D_{(a,b)} = \sigma$.

By solving the system S, we get $\mu = \ln(t) + h^2$ and $\sigma = h$. Then the lognormal kernel is given by

$$K_{LG(\ln(t)+h^2,h)}(y) = \frac{1}{\sqrt{2\pi h y}} \exp\left\{ \frac{-1}{2h} \left(\ln(y) - \ln(t) + h^2\right)^2 \right\},$$

with $y > 0$, $t > 0$ and $h > 0$.

However, some types of asymmetric kernels do not satisfy the mode-dispersion method, taking example of Birnbaum-Saunders (BS) kernel which do not have an explicit form of the mode.

Note that, the kernel generated from a given distribution may not be unique, regardless of its support on $\mathbb{R}_+$. Rather, it is possible to generate different kernels from the same distribution by changing functional forms of the shape and scale parameters. For example, the gamma kernels proposed by Igarashi and Kakizawa (2014) and Malec and Schienle (2014) and the inverse gamma kernels introduced by Mousa et al. (2016) and Igarashi and Kakizawa (2017), can be obtained via alternative specifications of the shape parameter, see Hirukawa (2018).

2.6 Bandwidth selection methods

As already pointed out, the smoothness of the density kernel estimator depends on the smoothing parameter h, so the selection of an appropriate bandwidth h plays a very important role on the quality of the estimation, see the Figure 2.3. However, when the parameter selected h is not suitable, it engenders an under-smoothing (when h is very small), see the Figure 2.4 or an over-smoothing (when h is very large), see the Figure (2.5) of the estimator.

It exists several methods to select the bandwidth parameter for both symmetric and asymmetric kernel estimator, see for instance, Jones et al. (1996), Zougab (2013) and Ziane (2015). These methods can be classified in two categories: classical approaches (plug-in, cross validation) and Bayesian approach (global, local and adaptative).

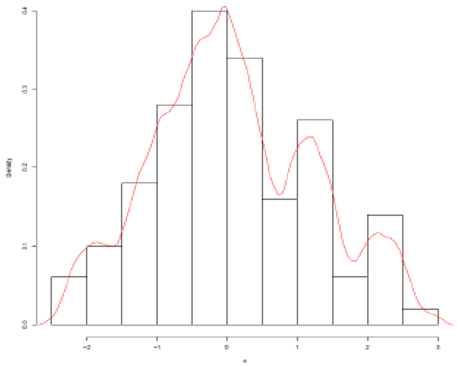

Figure 2.3: Smoothed estimator.

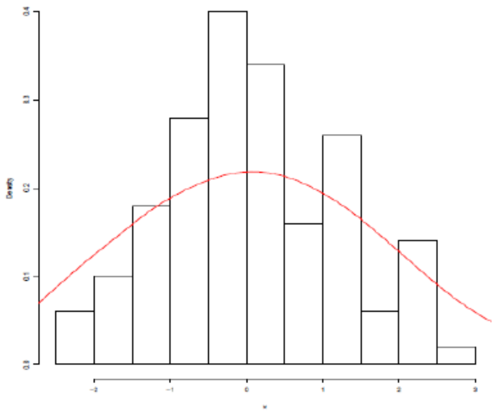

Figure 2.4: Under-smoothed estimator.

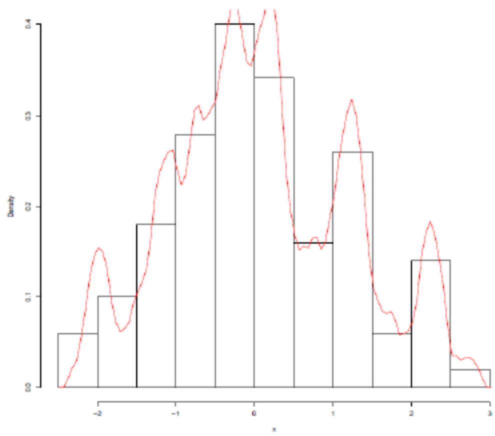

Figure 2.5: Over-smoothed estimator.

2.6.1 Plug-in method

Symmetric kernel case

This method is based on optimization of the following asymptotic mean squared error AMISE

$$\mathrm{AMISE}(\widetilde{\mathrm{f}}) \quad = \quad \frac{1}{nh}\int_{\mathbb{R}} K^2(z)dz + \frac{h^4\sigma_K^4}{4}\int_{\mathbb{R}} \left[f''(t)\right]^2 dt,$$

The minimization on h of the AMISE gives the following optimal bandwidth

$$h_P = \left[\frac{\int_{\mathbb{T}} K^2(y)dy}{\sigma^4 \int_{\mathbb{T}} f''^2(t)dt}\right]^{1/5} n^{-1/5}.$$

The above optimal bandwidth is not easily usable because it depends on the unknown quantity $f''^2(t)$.

Many issues have been proposed to overcome this problem, for instance the rule of thumb method. This method consists in supposing that the unknown function f is normally distributed with mean 0 and variance σ_f^2, where σ_f^2 is estimated using the observations $T_1, T_2, ..., T_n$ by $S_n^2 = \frac{1}{n}\sum_{i=1}^n (T_i - \overline{T})^2$, with $\overline{T} = \frac{1}{n}\sum_{i=1}^n T_i$, so the modified version of the optimal bandwidth is

$$h_{RT} = 1.06 S_n n^{-1/5}.$$

This bandwidth gives a good result when the observations are really normally distributed. Otherwise, this method is not efficient, for more details see Silverman (1986).

In another hand, Scott et al. (1977) introduced iterated plug-in method, Park and Marron (1990) presented modern iterated plug-in and Sheather and Jones (1991) introduced another method called plug-in in three steps, which is considered the most efficient.

The principal of the method of Sheather and Jones (1991) is to replace $\int_{\mathbb{R}} f''^2(t)dt$ by the following estimator

$$\widetilde{R}_a = \int_{\mathbb{R}} \widetilde{f}_h''^2(t)dt = \frac{1}{n^2 a^5}\sum_{i,i\neq j}^{n} L^{(4)}\left(\frac{T_i - T_j}{a}\right),$$

where $L^{(4)}$ is a derivative in forth order of the kernel L and a is a new bandwidth called pilot parameter. This estimator is obtained under sufficient regularity conditions

$$\int_{\mathbb{R}} f''^2(t)dt = \int_{\mathbb{R}} f^{(4)}(t)f(t)dt.$$

The new bandwidth parameter $\widehat{a}$ that minimizes the term $\mathbb{E}\left\{\left[\widetilde{R}_a - \int_{\mathbb{R}} f''^2(t)dt\right]^2\right\}$ is represented as

$$\widehat{a} = \left[\frac{2L^{(4)}(0)}{\sigma^2 \int_{\mathbb{R}} f'''^2(t)dt}\right]^{1/7} n^{-1/7}.$$

It appears again in the term above of $\widehat{a}$ an unknown quantity $\int_{\mathbb{T}} f^{(3)^2}(t)dt$. So the authors proposed the following estimator

$$\widetilde{R}_b = \frac{1}{n^2 b^7} \sum_{i,j=1}^{n} L^{(6)}\left(\frac{T_i - T_j}{b}\right),$$

where $b = 0.912\ \widehat{\lambda}\ n^{-1/9}$ and $\widehat{\lambda}$ is the estimator of λ, that represents the scale parameter of the function f (for example, its interquartile range).

The packages **bw.nrd0()** and **bw.sj()** are available in **R** software, for rule of thumb and Sheather and Jones methods, respectively.

Asymmetric kernel case

In the case of asymmetric kernel estimator, the optimal bandwidth is obtained also by the minimization of AMISE. However the bandwidth obtained depends on the unknown functions f, f' and f'' that makes the calculations more difficult. Scaillet (2004) proposed a rule of thumb method and replaced the unknown functions f, f' and f'' by choosing lognormal distribution as reference for the density f with the parameters μ and σ, i.e., $f \sim \mathcal{LN}(\mu, \sigma)$. The optimal bandwidths obtained by Scaillet (2004) using Inverse Gaussian (IG) and Reciprocal Inverse Gaussian (RIG) are presented respectively as

$$h_{IG} = \left[\frac{16\sigma^5 \exp\left\{\frac{1}{8}(7\sigma^2 - 20\mu)\right\}}{12 + 68\sigma^2 + 225\sigma^4}\right]^{2/5} n^{-2/5}$$

and

$$h_{RIG} = \left[\frac{16\sigma^5 \exp\left\{\frac{1}{8}(-17\sigma^2 + 20\mu)\right\}}{12 + 4\sigma^2 + \sigma^4}\right]^{2/5} n^{-2/5}.$$

In applied work, the unknown parameters μ and σ^2 may be estimated by the empirical mean and empirical variance computed on the algorithm of data. However, the simulation results obtained by Scaillet (2004) are not satisfactory and tends to provide bandwidths values which are very small.

The alternative popular method is the cross validation (CV) method.

2.6.2 Cross validation method (CV)

2.6.2.1 Least squared (unbiased) cross-validation

This method is a popular heuristic for selecting the smoothing parameter in kernel density estimation, introduced by Rudemo (1982) and Bowman (1984).
The basic idea in CV is to find the value of the parameter h that minimize the integrated squared error (ISE), given as

$$\begin{aligned} ISE(h) &= \int [\widehat{f}_h(t) - f(t)]^2 dt \\ &= \int \widehat{f}_h^2(t)dt - 2\int \widehat{f}_h(t)f(t)dt + \int f^2(t)dt. \end{aligned}$$

Because the last term does not depend on h, we only need to consider the first two terms. the optimal bandwidth is obtained by minimizing L given by

$$L(h) = ISE(h) - \int f^2(t)dt = \int \widehat{f}_h^2(t)dt - 2\int \widehat{f}_h(t)f(t)dt.$$

The idea is to find an estimate of $L(h)$ from the data and minimize it over h. Consider the estimator of L as

$$CV(h) = \int \left[\frac{1}{n}\sum_{i=1}^{n} K_{t,h}(T_i)\right]^2 dt - \frac{2}{n}\sum_i \widehat{f}_{h,-i}(T_i).$$

Where $\widehat{f}_{h,-i}(T_i) = \frac{1}{n-1}\sum_{j\neq i} K_{T_i,h}(T_j)$ is the density estimate (unbiased) of $\int \widehat{f}_h(t)f(t)dt$, using sample with T_i removed.
Then, the optimal value of the bandwidth h is obtained as

$$h_{CV} = \arg\min_{h>0} CV(h).$$

This method suffers from sample variation, that means; when using different samples from the same distribution, the bandwidths estimated may have large variance.

2.6.2.2 Biased cross-validation

This method was suggested by Scott and Terrell (1987). considers the asymptotic MISE in the case of symmetric kernel density estimation

$$AMISE(\widehat{f}_h) = \frac{1}{nh}\int_{\mathbb{T}} K^2(z)dz + \frac{h^4\sigma^4}{4}\int_{\mathbb{T}} f''^2(t)dt.$$

Note that $R(.) = \int_{\mathbb{T}}(.)^2$.

The main idea of this method is to replace the unknown quantity $R(f'')$ by its estimator $\widehat{R}(f'')$ as

$$\widehat{R}(f'') = R(\widehat{f}''_h) - \frac{1}{nh^5}R(K'').$$

That gives

$$BCV(h) = \frac{1}{nh}R(K) + \frac{h^4}{4}\sigma^4\left[R(\widehat{f}''_h) - \frac{1}{nh^5}R(K'')\right].$$

Then, the bandwidth selected is

$$h_{BCV} = \arg\min_h BCV(h).$$

This selector is considered as a hybrid of cross-validation and plug-in methods, since it replaces an unknown value in AMISE by a cross validation kernel estimate $\widehat{R}(f''_h)$.

2.6.3 Bayesian approach

Before presenting the bayesian approach for selection the bandwidth parameter, we first recall the concept of this approach.

Bayesian approach concepts

Consider $T_1, T_2, .., T_n$ i.i.d. r.v. with density f and observations $t = (t_1, t_2, .., t_n)$. Let $h \in \mathbb{H}$ the parameter to be estimated, with $\mathbb{H} \subset \mathbb{R}$.

This approach considers the unknown parameter h as a r.v. with a prior distribution $\pi(h)$, and combines the both of the information of the parameter h (prior information) and the information leads by the data, to provide a posterior information of the parameter h to be estimated.

The posterior distribution $\pi(h/t)$ is obtained using Bayes's theorem as

$$\pi(h/t) = \frac{\pi(t/h)\pi(h)}{\pi(t)},$$

where $\pi(t/h) = \prod_{i=1}^n f(t_i, h)$ represents the maximum likelihood function and $\pi(t) = \int_{\mathbb{H}} \prod_{i=1}^n f(t_i, h)\pi(h)dh$ is the marginal distribution.

In some cases, it is difficult to obtain an explicit form of $\pi(h/t)$, so Marcov Chain Monte Carlo MCMC are usually used to overcome this problem.

MCMC methods

MCMC methods are used to approximate the posterior distribution of a parameter of interest by random sampling. The principle is to generate a Markov chain $\{M^{(i)}\}$, ($i \in (1, .., I)$, with I is a number of iteration), using the kernel transition (law candidate) and an arbitrary initial value $M^{(0)}$. After a number of iteration I, sufficiently large, the Markov chain converges to the interest posterior density. Several algorithm to define such Markov chains exists, including Gibbs sampling, Metropolis–Hastings, ect. For more details about this method, see Zougab (2013), Ziane (2015).

The main scope in this approach, is to find an estimator $\widehat{h}$ of the parameter h that minimizes the mean cost, called also bayesian risk $\mathbb{E}[\mathcal{C}(\widehat{h} - h)]$, where $\mathcal{C}(\widehat{h} - h)$ represents the cost function, that is defined as

$$\mathbb{E}[\mathcal{C}(\widehat{h} - h)] = \int_{\mathbb{H}} \mathcal{C}(\widehat{h} - h)\pi(h/t)dh.$$

The cost function the most used is the quadratic one, that is $\mathcal{C}(\widehat{h} - h) = (\widehat{h}(t) - h)^2$. The estimator $\widehat{h}$ that minimizes the bayesian risk using the mean quadratic cost represents the posterior mean of h, given by

$$\widehat{h} = \mathbb{E}(h/t) = \int_{\mathbb{H}} h\pi(h/t)dh.$$

It exists other types of cost functions such as, the **absolute cost** $\mathcal{C}(\widehat{h} - h) = |(\widehat{h}(t) - h)|$, and the **cost 0-1** such that, for a given τ.

$$\mathcal{C}(\widehat{h} - h) = \begin{cases} 0 & \text{if } |\widehat{h}(t) - h| \leq \frac{\tau}{2} \\ 1 & \text{Otherwise.} \end{cases}$$

Now, for the selection of the optimal bandwidth parameter of the kernel estimator $\widehat{f_h}$ using the bayesian approach. It exists three essential technics: global, local and adaptive.

2.6.3.1 Global bayesian approach

This technic is proposed by Brewer (1998), Zhang et al.(2006) for multivariate kernel density estimation, and Zougab et al. (2013) by using gaussian kernel estimator in the case of continuous data and binomial kernel in discrete case. That consists to proceed as the following steps

1. Define the maximum likelihood estimator of the data $t_1, t_2, .., t_n$ knowing the parameter h, as given below
$$\widehat{\pi}(t_1, t_2, .., t_n/h) = \prod_{i=1}^{n} \widehat{f}_h(t_i).$$
By sing the leave-one-out technic to estimate $f(t_i)$ excluding the observation t_i, we get
$$\widehat{\pi}(t_1, t_2, .., t_n/h) = \prod_{i=1}^{n} \frac{1}{n-1} \sum_{j=1, j\neq i}^{n} K_{t_i,h}(t_j).$$

2. Choose the prior distribution of the bandwidth h, noted by $\pi(h)$.

3. Establish the posterior distribution estimator of h through the Bayes's theorem, as given below
$$\widehat{\pi}(h/t_1, t_2, .., t_n) = \frac{\widehat{\pi}(t_1, t_2, .., t_n/h)\pi(h)}{\widehat{\pi}(t_1, t_2, .., t_n)} \tag{2.9}$$
Where $\widehat{\pi}(t_1, t_2, .., t_n) = \int \widehat{\pi}(t_1, t_2, .., t_n/h)\pi(h)dh$. Note that, in many situation it is not easy to calculate this integral, so the explicite form of (2.9) cannot be obtained.

4. Finally, the bandwidth h is estimated by the posterior mean, mode or median by using the Marcov Chain Monte Carlo (MCMC). Fore more details see for instance Zougab (2013) and Ziane (2015).

2.6.3.2 Local bayesian approach

Here, the main idea is to estimate the bandwidth h locally at t, i.e. estimate h on each point t where the density is estimated. By using the Bayes's formula, the posterior distribution of h locally at t gives the following expression

$$\pi(h/t) = \frac{f(t)\pi(h)}{\int f(t)\pi(h)dh}.$$

As the model $f(t)$ is unknown, we use its kernel estimator $\widehat{f}_h(t) = \frac{1}{n}\sum_{i=1}^{n} K_{t,h}(T_i)$. Hence the posterior distribution of h takes the following form

$$\widehat{\pi}(h/t, T_1, T_2, .., T_n) = \frac{\widehat{f}_h(t)\pi(h)}{\int \widehat{f}_h(t)\pi(h)dh}. \tag{2.10}$$

In the case where the expressions (2.10) is not explicit, we can use the MCMC approximation methods.

2.6.3.3 Adaptive bayesian approach

Here, the objective is to estimate the bandwidth h for each observation t_i, noted by h_i. Let $\widehat{f}$ be the adaptive associated kernel estimator of f, defined as follows

$$\widehat{f}(t) = \frac{1}{n}\sum_{i=1}^{n} K_{t,h_i}(t_i),$$

where K_{t,h_i} is the associated kernel and h_i is the adaptive bandwidth parameter associated for each observation t_i.

By using the leave-one-out technic, the function $f(t_i)$ is estimated excepting the observation t_i, and can be written as

$$\widehat{f}_{-i}(t_i) = \widehat{f}(t_i/\{t_{-i}\}, h_i) = \frac{1}{n-1}\sum_{j=1, j\neq i} K_{t_i,h_i}(t_j).$$

By using the Bayes's formula, the posterior distribution for each h_i takes the following form

$$\widehat{\pi}(h_i/t_i, \{t_{-i}\}) = \frac{\widehat{f}(t_i/\{t_{-i}\}, h_i)\pi(h_i)}{\int \widehat{f}(t_i/\{t_{-i}\}, h_i)\pi(h_i)dh_i}. \tag{2.11}$$

The expression (2.11) can give the explicit results when using the conjugate priors, see Brewer (2000), Zougab (2013) and Ziane et al. (2015) for more details.

Kernel	Kernel function	$\mathbb{E}(\mathcal{Z}_{t,h})$	$\mathrm{Var}(\mathcal{Z}_{t,h})$	References
Gamma	$K_{G(\frac{t}{h}+1,h)}(u) = \frac{u^{\frac{t}{h}} \exp(\frac{-u}{h})}{h^{\frac{t}{h}+1}\Gamma(\frac{t}{h}+1)}$	$t+h$	$h(t+h)$	Chen (2000)
Inv. Gamma(IGam)	$K_{IGam(\frac{1}{ht}-1,\frac{1}{h})}(u) = \frac{1}{\Gamma(\frac{1}{ht}-1)} u^{\frac{-1}{ht}} \exp \frac{-1}{hu} \left(\frac{1}{h}\right)^{\frac{1}{ht}-1}$	$\frac{1}{\frac{1}{t}-2}$	$\frac{\left(\frac{1}{h}\right)^2}{\left(\frac{1}{ht}-2\right)^2\left(\frac{1}{ht}-3\right)}, \quad ht \leq \frac{1}{3}$	Mousa et al. (2016)
Inv. Gauss. (IG)	$K_{IG(t,\frac{1}{h})}(u) = \frac{1}{\sqrt{2\pi h u^3}} \exp\left[-\frac{1}{2ht}\left(\frac{u}{t} - 2 + \frac{t}{4}\right)\right]$	t	ht^3	Scaillet (2004)
Reciprocal IG	$K_{RIG(\frac{1}{t-h},\frac{1}{h})}(u) = \frac{1}{\sqrt{2\pi h u}} \exp\left[-\frac{t-h}{2h}\left(\frac{u}{t-h} - 2 + \frac{t-h}{4}\right)\right]$	t	$h(t+h)$	Scaillet (2004)
Lognormal	$K_{LN[\log t-4,\log(1+h)]}(u) = \frac{1}{\sqrt{8\pi \log(1+h) u}} \exp\left[-\frac{(\log u - \log t)}{8\log(1+h)}\right]$	$t(1+2h+h^2)$	$t^2[(1+h)^8 - (1+h)^4]$	Jin & Kawczak (2003)
Weibull	$K_{W(t,\frac{1}{h})}(u) = \frac{\Gamma(1+h)}{ht}\left[\frac{u\Gamma(1+h)}{t}\right]^{\frac{1}{h}-1} \exp\left\{-\left[\frac{u\Gamma(1+h)}{t}\right]^{\frac{1}{h}}\right\}$	t	$t^2\left[\Gamma(1+2h) - \Gamma^2(1+h)\right]$	Salha et al. (2014)
GBS	$K_{GBS(h^{\frac{1}{2}},t,g)}(u) = cg\left[\frac{1}{h}\left(\frac{u}{t} + \frac{t}{u} - 2\right)\right] \frac{1}{\sqrt{4h}}\left(\frac{1}{\sqrt{ut}} + \sqrt{\frac{t}{u^3}}\right)$	$\frac{[t+tu_1]h}{2}$	$ht^2(u_1 + \frac{hu_2}{2} - \frac{hu_1^2}{4})$	Marchant et al. (2013)

Table 2.3: Statistic properties of some asymmetric kernels, with $\mathbb{T} = [0, \infty)$.

Dist of Z	Dist of T	u_1	u_2	c_g	$g(y),\ \ y>0$
Normal	BS	1	3	$\frac{1}{\sqrt{2\pi}}$	$\exp(-\frac{\sqrt{y}}{2})$
Power exp.	BS-PE	$\frac{2^{\frac{1}{\nu}}\Gamma(\frac{2}{2\nu})}{\Gamma(\frac{1}{2\nu})}$	$\frac{2^{\frac{2}{\nu}}\Gamma(\frac{5}{2\nu})}{\Gamma(\frac{1}{2\nu})}$	$\frac{\nu}{[2^{\frac{1}{2\nu}}\Gamma(\frac{1}{2\nu})]}$	$\exp(-\frac{1}{2}y^{\nu}),\ \ \nu>0$
Laplace	BS-lap	$2\,!$	$4\,!$	$\frac{4}{exp(-\frac{\sqrt{y}}{2})}$	$\exp(-\|y\|)$
Logistic	BS-log	≈ 0.7957	≈ 1.5097	1	$\frac{\exp(\sqrt{y})}{[1+\exp(\sqrt{y})]^2}$
Student, $\nu>4$	BS-t	$\frac{\nu}{(\nu-2)}$	$\frac{3\nu^2}{[(\nu-2)(\nu-4)]}$	$\frac{1}{2}$	$\frac{\Gamma(\frac{\nu+1}{2})}{[\sqrt{\nu/\pi}\Gamma(\frac{\nu}{2})](1+\frac{\sqrt{y}}{\nu})^{-(\nu+1)/2}}$

Table 2.4: Values of u_1, u_2, c_g and $g(y)$ for the indicated distribution of Z.

Distribution of Z	Kernel, $(y > 0)$, $(t > 0)$
Normal	$K_{BS}(y) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2h}\left(\frac{y}{t} + \frac{t}{y} - 2\right)\right) \frac{1}{2\sqrt{h}}\left(\frac{1}{\sqrt{ty}} + \sqrt{\frac{t}{y^3}}\right)$
PE	$K_{BS-PE}(y) = \frac{\nu}{2^{\frac{1}{2\nu}}\Gamma(\frac{1}{2\nu})} \exp\left(-\frac{1}{2h^{\nu}}\left(\frac{y}{t} + \frac{t}{y} - 2\right)^{\nu}\right) \frac{1}{2\sqrt{h}}\left(\frac{1}{\sqrt{ty}} + \sqrt{\frac{t}{y^3}}\right), \nu > 0$
Student	$K_{BS-t}(y) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{\pi})}\left[1 + \frac{1}{\nu h}\left(\frac{y}{t} + \frac{t}{y} - 2\right)\right]^{-\left(\frac{\nu+1}{2}\right)} \frac{1}{2\sqrt{h}}\left(\frac{1}{\sqrt{ty}} + \sqrt{\frac{t}{y^3}}\right), \nu > 0$
Logistic	$K_{BS-log}(y) = \frac{\exp\left(\frac{1}{\sqrt{h}}\left[\sqrt{\frac{y}{t}} - \sqrt{\frac{t}{y}}\right]\right)}{\left[1+\exp\left(\frac{1}{\sqrt{h}}\left[\sqrt{\frac{y}{t}} - \sqrt{\frac{t}{y}}\right]\right)\right]^2} t\left(\frac{1}{\sqrt{ty}} + \sqrt{\frac{t}{y^3}}\right)$
Laplace	$K_{BS-lap}(y) = \frac{1}{4t\sqrt{h}} \exp\left(\frac{-1}{\sqrt{h}}\left\lvert\sqrt{\frac{y}{t}} - \sqrt{\frac{t}{y}}\right\rvert\right)\left[\sqrt{\frac{t}{y}} + \sqrt{\frac{t^3}{y^3}}\right]$

Table 2.5: GBS kernel for the indicated distribution.

Chapter 3

Hazard rate function estimation using kernel method

3.1 Introduction

In this chapter, we present an overview on hazard rate (HR) function estimation using kernel method, then we propose an estimator of HR in the context of of positively skewed data using the class of GBS kernels. This class is considered because of its several interesting properties and flexibility. Some asymptotic properties, such as bias, variance and mean integrated squared error (MISE) are established for the proposed estimator. In addition, we demonstrate that, the GBS-HR estimator is strongly consistent and asymptotically normal. The choice of bandwidth is investigated by rule of thumb and unbiased cross validation approaches. Finally, the performances of the HR estimator based on GBS kernels are illustrated by a simulation study and real applications.

3.2 An overview on kernel estimation of HR function

Recently, the hazard rate function HR function estimation has received considerable attention in the literature and many applications in several fields, such as, medical, biomedical, finance, ect. This, in the parametric case, see for instance Azevedo (2012) and Athayde et al. (2019), or nonparametric case, to cite a few, Bouezmarni et al. (2008) and Bouezmarni

(2011) by using gamma kernel in the context of censored data, Salha (2012) by using Inverse Gaussian (IG), Salha et al.(2014) with Weibull and Erlang kernels, Altun and Comert (2016) used Weibull-Exponential models to represent the typical L-shaped hazard rates of electronic products, Brazzale et al (2018) introducing a new method for estimating a change point for hazard function and Moriyama and Maesono (2018) proposed a new kernel estimator of the hazard ratio. In this section, we summarize some of these results of nonparametric case using kernel method for either censored and complete data.

3.2.1 Case of censoring data

Let $T_1, ..., T_n$ are r.v. representing observed survival times and $C_1, ..., C_n$ are r.v. representing censoring times, be two nonnegative random sequences with distribution functions F and G, respectively. We assume that the censoring times C_i are i.i.d. and independent of the the survival times T_i. Considering right censoring, that is instead of observing T_i, we observe the pair $(X_i, \delta_{(i)})$, where $X_i = \min(T_i, C_i)$, $\delta_{(i)} = I(Ti \leq Ci)$ and $I(.)$ is the indicator function. We denote by f the density function of F and by $\lambda(\cdot) = \frac{f(\cdot)}{(1-F(\cdot))}$ the corresponding hazard function. The hazard rate function estimator in the case of censoring data is defined in Bouezmarni (2008) as

$$\widehat{\lambda}(x) = \sum_{i=1}^{n} \frac{\delta_{(i)}}{n-i+1} K_{(x,h)}(X_{(i)}), \quad x > 0, \tag{3.1}$$

with $X_{(1)} \leq X_{(2)} \leq ... \leq X_{(n)} \leq$ are the order statistics of $X_1, X_2, ..., X_n$ and $\delta_{(i)}$ is the concomitant of $X_{(i)}$, where h is the bandwidth parameter and K is the associated asymmetric kernel.

Definition 3.2.1. *Let $(T_i, i \geq 1)$ be a sequence of random variables. Given a positive integer n, set*

$$\alpha(n) = \sup_k |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|, \quad A \in \mathcal{F}_k^1(T) \ \ and \ \ B \in \mathcal{F}_{k+n}(T),$$

with $\mathcal{F}_k^i(T)$ be the σ-field of events generated by $T_j, i \leq j \leq k$.
The sequence (T_i) is α-mixing if the mixing coefficient $\alpha(n) \to 0$ as $n \to 0$,

Bouezmarni (2008) supposed the survival times T_i are α-mixing (strong dependence), and established some results of convergence of the HR function estimator defined in (3.1) using gamma kernel, see Proposition 3.2.1 and 3.2.2, under the following conditions

B1. The survival times $(T_j; j \geq 1)$ are stationary $\alpha-$mixing sequence of random variables.

B2. The censoring time $(C_j; j \geq 1)$ are i.i.d. random variables and independent of $(T_j; j \geq 1)$.

B3. $\alpha(n) = O(n^{-\beta})$, for some $\beta > 3$.

Proposition 3.2.1. *(Asymptotic normality)(Bouezmarni, 2008)*
Let λ be twice continuously differentiable. Under the conditions B1-B3 and assume that $h = O(n^{\frac{2}{5}})$. For all x such that $f(x) > 0$, $\exists\tau$ such that, $x \leq \tau$. As $n \to \infty$, we have

$$n^{1/2}h^{1/4}\left(\frac{\widehat{\lambda}(x) - \mathbb{E}[\widehat{\lambda}(x)]}{\sqrt{\mathrm{V^*(x)}}}\right) \xrightarrow{d} N(0,1), \quad n \to \infty.$$

with $\xrightarrow{d}$ denotes the convergence in distribution.

$$\mathrm{V^*(x)} = \begin{cases} \dfrac{1}{2\sqrt{\pi}}\dfrac{x^{-1/2}\lambda(x)}{1 - H(x)} & \text{if } \frac{x}{h} \to \infty \\ \\ \dfrac{\Gamma(2k+1)h^{-1/2}\lambda(x)}{2^{1+2k}\Gamma^2(k+1)[1 - H(x)]} & \text{if } \frac{x}{h} \to k, \end{cases} \tag{3.2}$$

where H is a distribution function of the r.v. X_i and k is a nonnegative constant.

Proposition 3.2.2. *(Convergence almost sure)(Bouezmarni, 2008)*
Let f be a continuous density. Assume that the conditions B1-B3 are satisfied and $h = O(n^{2/5})$. Then, for all $x \leq \tau$ and as $n \to \infty$ we have

$$\widehat{\lambda}(x) \xrightarrow{a.s.} \lambda(x),$$

where $\xrightarrow{a.s.}$ denotes the almost sure convergence.

Still in the case of censoring data, Bouezmarni (2011) has studied the HR function estimator (3.1) using gamma kernel with independent r.v. T_i, and established the mean integrated squared error MISE given in the following theorem.

Theorem 3.2.1. *(Bouezmarni, 2011)*
Assume that f is twice continuously differentiable. If $\frac{log^2 n}{nh^{3/2}} \to 0$ as $n \to \infty$, then the mean integrated squared error of $\widehat{\lambda}$ is

$$\text{MISE}(\widehat{\lambda}) = h^2 \int B^2(x)dx + n^{-1}h^{-1/2} \int V(x)dx + o(h^2) + o(n^{-1}h^{-1/2}),$$

where B and V are given by

$$B(x) = \frac{xf''(x)}{2(1-F(x))} \quad \text{and} \quad V(x) = \frac{x^{-1/2}f(x)}{2\sqrt{\pi}(1-G(x))(1-F(x))^2}.$$

3.2.2 Case of complete data

HR function estimation using IG, Erlang and Weibull kernels

Consider $T_1, T_2, ..., T_n$ be a random sample from a distribution with an unknown probability density function f defined on $[0, +\infty)$. Salha (2012) and Salha et al. (2014^a,2014^b) studied the following HR function estimator

$$\widehat{\lambda}(t) = \frac{\frac{1}{n}\sum_{i=1}^{n} K_{t,h}(T_i)}{\frac{1}{n}\sum_{i=1}^{n}\int_0^t K_{x,h}(T_i)dx}, \quad t \geq 0,$$

by using: Inverse Gaussian (IG) kernel, Erlang kernel and Weibull kernel, respectively. Some results of asymptotic convergence are established in the theorems below.

Theorem 3.2.2. *Let $\widehat{\lambda}$ be the HR function estimator using IG kernel and if (i) f is twice continuously differentiable, (ii) $\int_0^\infty [t^3 f''(t)]^2 dt < \infty$, $h + \frac{1}{nh} \to 0$ and (iii) $nh^{5/2} \to 0$ as $n \to \infty$, the following holds*

$$\sqrt{nh^{1/2}}\left[\widehat{\lambda}(t) - \lambda(t)\right] \xrightarrow{d} N\left(0, \frac{1}{\sqrt{2\pi}}t^{-3/2}\frac{\lambda(t)}{1-F(t)}\right), \quad \forall t > 0.$$

Theorem 3.2.3. *Let $\widehat{\lambda}$ be the HR function estimator using Erlang kernel and if (i) the density f has a continuous second derivative, (ii) $0 < \int t^4 f''^2(t) < \infty$ and $\int \frac{f(t)}{t}dt < \infty$ and (iii) the bandwidth h satisfying, $h + \frac{1}{nh^5} \to 0$ as $n \to \infty$. The following holds*

$$2^{1/h}\sqrt{\frac{nh^5}{h+2}}\left[\widehat{\lambda}(t) - \lambda(t)\right] \xrightarrow{d} N\left(0, \frac{\lambda(t)}{8t[1-F(t)]}\right), \quad \forall t > 0.$$

Theorem 3.2.4. *Let $\widehat{\lambda}$ be HR function estimator using Weibull kernel, if (i) the density f has a continuous second derivative, (ii) $\int t^4 f''^2(t)dt < \infty$ and $\int \frac{f(t)}{t}dt < \infty$ and (iii) $\frac{\int t^4 f''^2(t)dt}{\int \frac{f(t)}{t}dt} > \frac{\exp(-\gamma)\ln(8)}{4n\gamma}$, where $\gamma = 0.5772156649$ is Euler's constant. The following holds*

$$2^{3h/2}\sqrt{\frac{nh}{[\Gamma(1+h)]^{1/h}}}\left[\widehat{\lambda}(t) - \lambda(t)\right] \xrightarrow{d} N\left(0, \frac{\lambda(t)}{4t[1-F(t)]}\right), \forall t > 0.$$

HR function estimation using new kernel

Moriyama and Maesono (2018) proposed a new kernel estimator of hazard rate function, that is based on a modification of Čwik and Mielniczuk method (Čwik and Mielniczuk, 1989). First, we describe the principal of this method. Let $X_1, X_2, ..., X_n$ be independently and identically distributed (i.i.d.) random variables with a distribution function $F(\cdot)$, and $Y_1, Y_2, ..., Y_n$ be i.i.d. random variables with a distribution function $G(\cdot)$, $f(\cdot)$ and $g(\cdot)$ are the density functions of X and Y variables, and we assume that $g(x_0) \neq 0, (x_0 \in R)$. A naive estimator of the density ratio $f(x_0)/g(x_0)$ at the point x_0, ($x_0 \in \mathbb{R}$)is given by $\widehat{f}(x_0)/\widehat{g}(x_0)$ where

$$\widehat{f}(x_0) = \frac{1}{h}\int_{-\infty}^{+\infty} K\left(\frac{x_0 - w}{h}\right) dF_n(w),$$

and

$$\widehat{g}(x_0) = \frac{1}{h}\int_{-\infty}^{+\infty} K\left(\frac{x_0 - z}{h}\right) dG_n(z),$$

with $K(\cdot)$ is a kernel function, h is a bandwidth that satisfies $h \to 0$ and $nh \to \infty$, as $n \to \infty$), and $F_n(\cdot)$ and $G_n(\cdot)$ are the empirical distribution functions of $X_1, ..., X_n$ and $Y_1, ..., Y_n$, respectively. We call $\widehat{f}(x_0)/\widehat{g}(x_0)$ an 'indirect' estimator. Čwik and Mielniczuk (1989) proposed a direct estimator, as

$$\widehat{\frac{f(x_0)}{g(x_0)}} = \frac{1}{h}\int_{-\infty}^{\infty} K\left(\frac{G_n(x_0) - G_n(w)}{h}\right) dF_n(w).$$

For more details about this estimator, see Čwik and Mielniczuk (1989), Chen et al. (2009) and Igarashi (2020).

Moriyama and Maesono (2018) have extend the idea of C´wik and Mielniczuk (1989), by develop a new 'direct' estimator of the hazard ratio function which is defined as

$$\widehat{\lambda}(x_0) = \int_{-\infty}^{+\infty} K\left(\frac{M_n(x_0) - M_n(w)}{h}\right) dF_n(w), \tag{3.3}$$

where

$$M_n(w) = w - \int_{-\infty}^{w} F_n(u)du.$$

Asymptotic properties are also investigated for this estimator, given in theorems 3.2.5 and 3.2.6.

Theorem 3.2.5. *Let us assume that (i) $f(.)$ is three-times differentiable at x_0 and $f^{(3)}(x_0)$ is bounded, (ii) K is symmetric and the support is given by a closed interval, (iii) $K^{(3)}$ is bounded, and (iv) $A_{1,4}$ and $A_{2,0}$ are bounded. Then, the MSE of $\widehat{\lambda}(x_0)$ is given by*

$$\begin{aligned}\mathbb{E}\left[\left(\widehat{\lambda}(x_0) - \lambda(x_0)\right)^2\right] \quad = \quad & \frac{h^4}{4}A^2_{(1,2)}\left\{\frac{[(1-F(x_0))(1-F(x_0)f'' + 4ff') + 3f^3]^2}{[1-F(x_0)]^{10}}\right\}(x_0) + \\ & \frac{A_{(2,0)}}{nh}\lambda(x_0) + O(h^6 + \frac{1}{nh^{1/2}}), \quad x_0 \in \mathbb{R},\end{aligned}$$

note that

$$A_{i,j} = \int_{-\infty}^{\infty} u^j K^i(u)du$$

Theorem 3.2.6. *When $h = cn^{-\varepsilon}$ $(0 < c, \frac{1}{5} \leq \varepsilon < \frac{1}{2})$, the following asymptotic normality of $\widehat{\lambda}(x_0)$ holds:*

$$\sqrt{nh}\left[\widehat{\lambda}(x_0) - \lambda(x_0)\right] \xrightarrow{d} N(B, V_1),$$

where $B = \lim_{n\to\infty} \sqrt{nh^5}B_1$

$$B_1 = \frac{A_{1,2}}{2}\left[\frac{(1-F(x_0))(1-F(x_0)f'' + 4ff') + 3f^3}{[1-F(x_0)]^5}\right](x_0),$$

and

$$V_1 = A_{2,0}\lambda(x_0).$$

In addition, Moriyama and Maesono (2018) have compared the proposed direct estimator of HR function (3.3) with the naive estimator defined by $\lambda_n(x_0) = \frac{\widehat{f}(x_0)}{1-F_n(x_0)}$, where F_n is the empirical distribution function. The authors concluded that the direct estimator of HR function performs asymptotically better than the naive estimator, especially in exponential or gamma cases, which play a central role in survival analysis and the asymptotic variance of the new estimator is usually smaller than that of the naive one. Although, the bias of the direct estimator is large in some cases, and the asymptotic variance is always small when both bandwidth parameters are the same (see, Moriyama and Maesono, 2018).

3.3 HR function estimation using GBS kernel

In this section, we develop the HR function estimation based on GBS kernels and we discuss some properties of this estimator. The bandwidth selection for HR estimator is also investigated using the popular rule of thumb and unbiased cross validation approaches.

3.3.1 Construction of the estimator

As mentioned previously, the HR function of survival time T (nonnegative random variable) has the following form

$$\lambda(t) = \lim_{dt \to 0} \frac{\Pr(t < T \leq t + dt \,|_{T>t})}{dt}, \quad t > 0,$$

which can be written as

$$\lambda(t) = \frac{f(t)}{1 - F(t)}, \quad t > 0,$$

where f is unknown pdf of the r.v. T and F its cumulative distribution function (cdf). Let $\widehat{f}_{GBS}$ be the GBS kernel estimator of the unknown pdf f, and let

$$\widehat{F}_{GBS}(t) = \int_0^t \widehat{f}_{\mathrm{GBS}}(x)dx = \int_0^t \frac{1}{n}\sum_{i=1}^{n} K_{\mathrm{GBS}(h^{1/2},x)}(T_i)dx$$

be the GBS kernel estimator of the cdf F. Then, based on a (complete) random sample $T_1, T_2, \ldots, T_n$ distributed as T, a natural GBS-HR estimator is simply given by

$$\widehat{\lambda}_{\mathrm{GBS}}(t) = \frac{\frac{1}{n}\sum_{i=1}^{n} \frac{c_g}{2\sqrt{h}} \left(\frac{1}{\sqrt{tT_i}} + \sqrt{\frac{t}{T_i^3}}\right) g\left(\frac{1}{h}\left(\frac{T_i}{t} + \frac{t}{T_i} - 2\right)\right)}{1 - \frac{1}{n}\sum_{i=1}^{n} \int_0^t \frac{c_g}{2\sqrt{h}} \left(\frac{1}{\sqrt{xT_i}} + \sqrt{\frac{x}{T_i^3}}\right) g\left(\frac{1}{h}\left(\frac{T_i}{x} + \frac{x}{T_i} - 2\right)\right) dx}, \quad t > 0, \tag{3.4}$$

where the bandwidth h controls the smoothness of the estimator $\widehat{\lambda}_{GBS}$ as for the case of the pdf estimation by kernel method. This important issue of bandwidth choice will be investigated in the Section 3.3.3. Note that for each generator g and constant c_g given in the Table 2.4, we obtain the following specific HR estimator according to each kernel

- BS kernel

$$\widehat{\lambda}_{\mathrm{BS}}(t) = \frac{\sum_{i=1}^{n} \left(\frac{1}{\sqrt{tT_i}} + \sqrt{\frac{t}{T_i^3}}\right) \exp\left[\frac{-1}{2h}\left(\frac{T_i}{t} + \frac{t}{T_i} - 2\right)\right]}{2^{\frac{3}{2}} n\sqrt{\pi h} - \sum_{i=1}^{n} \int_0^t \left(\frac{1}{\sqrt{xT_i}} + \sqrt{\frac{x}{T_i^3}}\right) \exp\left[\frac{-1}{2h}\left(\frac{T_i}{x} + \frac{x}{T_i - 2}\right)\right] dx}, \quad t > 0.$$

- BS-PE kernel

$$\widehat{\lambda}_{\mathrm{BS-PE}}(t) = \frac{\sum_{i=1}^{n}\left(\frac{1}{\sqrt{tT_i}} + \sqrt{\frac{t}{T_i^3}}\right)\exp\left[\frac{-1}{2h^v}\left(\frac{T_i}{t} + \frac{t}{T_i} - 2\right)\right]}{\frac{2^{\frac{1}{2v}+1}\Gamma(\frac{1}{2v})n\sqrt{h}}{v} - \sum_{i=1}^{n}\int_0^t\left(\frac{1}{\sqrt{xT_i}} + \sqrt{\frac{x}{T_i^3}}\right)\exp\left[\frac{-1}{2h^v}\left(\frac{T_i}{x} + \frac{x}{T_i} - 2\right)^v\right]dx}, \quad t > 0.$$

- BS-t kernel

$$\widehat{\lambda}_{\mathrm{BS-t}}(t) = \frac{\sum_{i=1}^{n}\left(\frac{1}{\sqrt{tT_i}} + \sqrt{\frac{t}{T_i^3}}\right)\left[1 + \frac{1}{2vh}\left(\frac{T_i}{t} + \frac{t}{T_i} - 2\right)\right]^{-(\frac{v+1}{2})}}{\frac{2n\sqrt{hv\pi}\Gamma(\frac{v}{2})}{\Gamma(\frac{v+1}{2})} - \sum_{i=1}^{n}\int_0^t\left(\frac{1}{\sqrt{xT_i}} + \sqrt{\frac{x}{T_i^3}}\right)\left[1 + \frac{1}{vh}\left(\frac{T_i}{x} + \frac{x}{T_i} - 2\right)\right]^{-(\frac{v+1}{2})}dx}, \quad t > 0.$$

- BS-lap kernel

$$\widehat{\lambda}_{\mathrm{BS-lap}}(t) = \frac{\sum_{i=1}^{n}\left(\frac{1}{\sqrt{tT_i}} + \sqrt{tT_i^3}\right)\exp\left[\left|\frac{-1}{h}\left(\frac{T_i}{t} + \frac{t}{T_i} - 2\right)\right|\right]}{\frac{4n}{\sqrt{h}} - \sum_{i=1}^{n}\int_0^t\left(\frac{1}{\sqrt{xT_i}} + \sqrt{\frac{x}{T_i^3}}\right)\exp\left[-\left|\frac{1}{h}\left(\frac{T_i}{t} + \frac{t}{T_i} - 2\right)\right|\right]dx}, \quad t > 0.$$

3.3.2 Convergence properties

In this section, we establish the bias, variance, mean integrated squared error (MISE), of the proposed GBS-HR kernel estimator using Proposition 3.3.1 and under the conditions C1-C3 (given in page 39) and the following assumptions

A1. $\int_0^\infty \left(\frac{tf'(t)}{1-F(t)}\right)^2 dt < \infty$;

A2. $\int_0^\infty \left(\frac{t^2f''(t)}{1-F(t)}\right)^2 dt < \infty$;

A3. $\int_0^\infty \frac{t^{-1}\lambda(t)}{1-F(t)}dt < \infty$.

Proposition 3.3.1. *Under the conditions C1, C3 and* $\mathbb{E}\left[\int_0^t K_{\mathrm{GBS}(h^{\frac{1}{2}},t,g)}(T)dx\right] < \infty$, *for all* n, *the following holds*

$$\widehat{F}_{GBS}(t) \xrightarrow{a.s.} F(t), \quad as \quad n \to \infty \tag{3.5}$$

where $\xrightarrow{a.s.}$ *denotes the almost sure convergence.*

Proof.

First, we have

$$\begin{aligned}
\mathbb{E}(\widehat{F}_{\text{GBS}}(t)) &= \mathbb{E}\int_0^t K_{\text{GBS}(h^{\frac{1}{2}},x,g)}(T)dx = \int_0^\infty \int_0^t K_{\text{GBS}(h^{\frac{1}{2}},x,g)}(y)f(y)dxdy \\
&= \int_0^t \int_0^\infty K_{\text{GBS}(h^{\frac{1}{2}},x,g)}(y)f(y)dydx \\
&= \int_0^t \mathbb{E}[f(\mathcal{K}_{x,h})]dx, \qquad (3.6)
\end{aligned}$$

where $\mathcal{K}_{x,h} \sim \text{GBS}(h^{\frac{1}{2}}, x, g)$. The mean and the variance of $\mathcal{K}_{x,h}$ are respectively; see Marchant et al. (2013)

$$\mathbb{E}(\mathcal{K}_{x,h}) = x + \frac{xhu_1(g)}{2}, \quad Var(\mathcal{K}_{x,h}) = x^2hu_1(g) + \frac{x^2h^2u_2(g)}{2} - \frac{x^2h^2u_1^2(g)}{4}.$$

By the Taylor expansion around x and using the condition C1, we obtain

$$\mathbb{E}[f(\mathcal{K}_{x,h})] = f(x) + h\left[\frac{u_1(g)}{2}(xf'(x) + x^2f''(x))\right] + o(h).$$

By replacing in (3.6), then we have

$$\begin{aligned}
\mathbb{E}(\widehat{F}_{\text{GBS}}(t)) &= \int_0^t \left\{f(x) + h\left[\frac{u_1(g)}{2}(xf'(x) + x^2f''(x))\right] + o(h)\right\} dx \\
&= \int_0^t f(x)dx + \frac{hu_1(g)}{2}\int_0^t (xf'(x) + x^2f''(x))dx + o(h). \\
&= F(t) + h\left[\frac{u_1(g)(t^2f'(t) - tf(t) + F(t))}{2}\right] + o(h) \\
&= F(t) + O(h). \qquad (3.7)
\end{aligned}$$

Second, note that $\int_0^t K_{\text{GBS}(h^{\frac{1}{2}},x)}(T_i)dx$ are i.i.d. and $\mathbb{E}\left[\int_0^t K_{\text{GBS}(h^{\frac{1}{2}},x)}(T)dx\right] < \infty, \forall n$. Hence, by the strong law of large numbers, we obtain

$$\widehat{F}_{\text{GBS}}(t) - \mathbb{E}\left[\int_0^t K_{\text{GBS}}(T_i)dx\right] \xrightarrow{a.s.} 0 \quad as \quad n \to \infty. \qquad (3.8)$$

Then, by using (3.7), (3.8) and the following classical decomposition

$$\widehat{F}_{\text{GBS}}(t) - F(t) = \left[\mathbb{E}(\widehat{F}_{\text{GBS}}(t)) - F(t)\right] + \left[\widehat{F}_{\text{GBS}}(t) - \mathbb{E}(\widehat{F}_{\text{GBS}}(t))\right],$$

we complete the proof of the proposition.

The first theorem presents the asymptotic bias and variance of the GBS-HR estimator given by the equation (3.4) and the second one gives the mean integrated squared error

(MISE) (global property) of the estimator $\widehat{\lambda}_{GBS}$ and the optimal bandwidth which minimizes the MISE criterion.

Theorem 3.3.1. *(Bias and variance of $\widehat{\lambda}_{\mathrm{GBS}}$)*

Let $\widehat{\lambda}_{\mathrm{GBS}}$ be the estimator of the hazard rate function λ with GBS kernels. Under the conditions C1, C2 given previously, the bias and variance of this estimator are given by

$$\mathrm{Bias}\left[\widehat{\lambda}_{\mathrm{GBS}}(t)\right] = \frac{m^*}{1-F(t)} + o(h), \tag{3.9}$$

and

$$\mathrm{Var}\left[\widehat{\lambda}_{\mathrm{GBS}}(t)\right] = \frac{\sigma^{*2}}{[1-F(t)]^2} + o(n^{-1}h^{-\frac{1}{2}}), \tag{3.10}$$

where $m^* = \frac{hu_1}{2}\left[tf'(t) + t^2 f''(t)\right]$ *and* $\sigma^* = \frac{c_g^2}{(c_{g^2} n h^{1/2})} t^{-1} f(t)/.$

Proof

From Proposition 3.3.1, we can write

$$\widehat{\lambda}_{\mathrm{GBS}}(t) = \frac{\widehat{f}_{\mathrm{GBS}}(t)}{1-\widehat{F}_{\mathrm{GBS}}(t)} = \frac{\widehat{f}_{GBS}(t)}{1-F(t)} \quad a.s.$$

Then for n enough large, the mean of $\widehat{\lambda}_{\mathrm{GBS}}(t)$ is simply given by

$$\mathbb{E}\left[\widehat{\lambda}_{\mathrm{GBS}}(t)\right] = \frac{\mathbb{E}\left[\widehat{f}_{\mathrm{GBS}}(t)\right]}{1-F(t)}. \tag{3.11}$$

Hence, the bias is expressed as

$$\mathrm{Bias}\left[\widehat{\lambda}_{\mathrm{GBS}}(t)\right] = \mathbb{E}\left[\widehat{\lambda}_{\mathrm{GBS}}(t)\right] - \lambda(t) = \frac{\mathbb{E}\left[\widehat{f}_{\mathrm{GBS}}(t)\right]}{1-F(t)} - \frac{f(t)}{1-F(t)} = \frac{\mathrm{Bias}\left[\widehat{f}_{\mathrm{GBS}}(t)\right]}{1-F(t)}.$$

Similarly, the variance is given by

$$\mathrm{Var}\left[\widehat{\lambda}_{\mathrm{GBS}}(t)\right] = \mathrm{Var}\left[\frac{\widehat{f}_{\mathrm{GBS}}(t)}{1-F(t)}\right] = \frac{\mathrm{Var}\left[\widehat{f}_{\mathrm{GBS}}(t)\right]}{[1-F(t)]^2}.$$

Now, by replacing the expressions of the bias and variance of $\widehat{f}_{\mathrm{GBS}}(t)$ given in the formulas (2.7) and (2.8) respectively, we obtain the desired result given in Theorem 3.3.1.

Theorem 3.3.2. *(*MISE *of $\widehat{\lambda}_{\mathrm{GBS}}$)*

Under the conditions C3 and A1-A3, we obtain the MISE *of $\widehat{\lambda}_{\mathrm{GBS}}$ as follows*

$$\begin{aligned}\mathrm{MISE}(\widehat{\lambda}_{\mathrm{GBS}}) &= \frac{u_1^2(g)h^2}{4}\int_0^\infty \left[\frac{(tf'(t)+t^2f''(t))}{1-F(t)}\right]^2 dt + \frac{c_g^2}{c_{g^2} n h^{\frac{1}{2}}}\int_0^\infty \frac{t^{-1}\lambda(t)}{[1-F(t)]}dt \\ &\quad + o\left(h^2 + \frac{1}{nh^{\frac{1}{2}}}\right),\end{aligned} \tag{3.12}$$

and the optimal bandwidth h is given by

$$h_{opt} = \left[\frac{c_g^2 \int_0^\infty \frac{t^{-1} f(t)}{(1-F(t))^2} dt}{c_{g^2} u_1^2(g) \int_0^\infty \left(\frac{tf'(t)+t^2 f''(t)}{1-F(t)} \right)^2 dt} \right]^{\frac{2}{5}} n^{-\frac{2}{5}}. \tag{3.13}$$

Proof.

The MISE (3.12) is obtained by substituting the formulas of the bias and variance of the estimator $\widehat{\lambda}_{\text{GBS}}(t)$ in $\int_0^\infty \text{Bias}^2 \left[\widehat{\lambda}_{GBS}(t) \right] dt + \int_0^\infty \text{Var}(\widehat{\lambda}_{GBS}(t))dt$, and by minimizing the MISE (3.12) on h, we obtain the optimal bandwidth (3.13). Note that the bandwidth (3.13) depends on the unknown pdf f and on the cdf F, then it can not be directly exploited in practice. The rule of thumb (RT) using BS distribution as a reference model and the Unbiased Cross Validation (UCV) approaches, will be developed in the Section 3.3.3.

Now, the two following theorems establish the strong consistency and the convergence in distribution of $\widehat{\lambda}_{\text{GBS}}$. The first consistency result concerns the almost sure convergence, and the second theorem deals with the asymptotic normality.

Theorem 3.3.3. *(Strong convergence of $\widehat{\lambda}_{\text{GBS}}$)*
Let $\widehat{\lambda}_{\text{GBS}}$ be the estimator of λ defined in (3.4). Then for fixed $t > 0$, we have

$$\widehat{\lambda}_{\text{GBS}}(t) \xrightarrow{a.s.} \lambda(t), \quad as \quad n \to \infty.$$

Proof

Recall that $\widehat{f}_{\text{GBS}}(t) = \frac{1}{n} \sum_{i=1}^n K_{\text{GBS}}(T_i)$ and $\mathbb{E}\left[\widehat{f}_{\text{GBS}}(t)\right] = \mathbb{E}\left[K_{\text{GBS}}(T)\right]$, such that $\lim_{n\to\infty} \mathbb{E}\left[K_{\text{GBS}}(T)\right] = f(t)$. Note that $K_{\text{GBS}}(T_i)$ are i.i.d. and $\mathbb{E}\left[K_{\text{GBS}}(T)\right] < \infty$. Then, by the strong law of large numbers, we have

$$\widehat{f}_{\text{GBS}}(t) = \frac{1}{n} \sum_{i=1}^n K_{\text{GBS}}(T_i) \xrightarrow{a.s.} f(t), \quad n \to \infty.$$

Now, from Proposition 3.3.1, we get

$$\widehat{\lambda}_{GBS}(t) \xrightarrow{a.s.} \lambda(t).$$

The proof of Theorem 3.3.3 is complete.

Theorem 3.3.4. *(Asymptotic normality of $\widehat{\lambda}_{\text{GBS}}$)*
For fixed $t > 0$, the estimator $\widehat{\lambda}_{\text{GBS}}$ converges in distribution to the normal distribution as

follows

$$[1-F(t)]\sigma^{*-1}n^{\frac{1}{2}}\left[\widehat{\lambda}_{\mathrm{GBS}}(t)-\lambda(t)-\frac{m^*}{1-F(t)}\right]\xrightarrow{\mathcal{D}}\mathcal{N}(0,1),$$

where $\xrightarrow{\mathcal{D}}$ *denotes the convergence in distribution.*

Proof

Following Bouezmarni and Rombouts (2010) and using that $h=o(n^{-1})$, we can write

$$\sigma^{*-1}n^{\frac{1}{2}}\left[\widehat{f}_{\mathrm{GBS}}(t)-f(t)-m^*\right]=\sum_{i}^{n}Y_i+O(n^{\frac{-1}{4}}),$$

where

$$Y_i=\sigma^{*-1}\left\{n^{\frac{-1}{2}}\left[K_{\mathrm{GBS}}(T_i)-\mathbb{E}(K_{\mathrm{GBS}}(T))\right]\right\}.$$

The central limit theorem asserts that as $n\to\infty$, the distribution of $V_n=\sum_{i=1}^{n}Y_i$ tends to the normal distribution with zero mean and unit variance, i.e. $V_n\overset{\mathcal{D}}{\to}\mathcal{N}(0,1)$, by the fact that $\mathbb{E}(V_n)=0$ and $\mathrm{Var}(V_n)=1+o(1)$. That leads to deduce that

$$\sigma^{*-1}n^{\frac{1}{2}}\left[\widehat{f}_{\mathrm{GBS}}(t)-f(t)-m^*\right]\overset{\mathcal{D}}{\to}\mathcal{N}(0,1),\quad \text{as}\quad n\to\infty.$$

Now, from Proposition 3.3.1 we can write

$$\widehat{\lambda}_{\mathrm{GBS}}(t)-\lambda(t)=\frac{\widehat{f}_{\mathrm{GBS}}(t)-f(t)}{1-F(t)}\quad a.s.$$

Hence, we write explicitly

$$\widehat{\lambda}_{\mathrm{GBS}}(t)-\lambda(t)-\frac{m^*}{1-F(t)}=\frac{\widehat{f}_{\mathrm{GBS}}(t)-f(t)-m^*}{1-F(t)}\quad a.s.$$

Consequently,

$$[1-F(t)]\,\sigma^{*-1}n^{\frac{1}{2}}\left[\widehat{\lambda}_{\mathrm{GBS}}(t)-\lambda(t)-\frac{m^*}{1-F(t)}\right]=\sigma^{*-1}n^{\frac{1}{2}}\left[\widehat{f}_{\mathrm{GBS}}(t)-f(t)-m^*\right]\quad a.s.$$

Finally, we conclude that $[1-F(t)]\,\sigma^{*-1}n^{\frac{1}{2}}\left[\widehat{\lambda}_{GBS}(t)-\lambda(t)-\frac{m^*}{1-F(t)}\right]$ is asymptotically normally distributed, which gives the desired result.

3.3.3 Bandwidth selection

The performance of the GBS kernel HR estimator given by (3.4) depends on the bandwidth h, which controls the smoothness of this estimator. Then, we investigate the bandwidth choice on the GBS kernel estimator for HR. First, recall that the optimal bandwidth (3.13) depends on the unknown quantities f, f', f'' and F. In order to overcome this problem, we propose to use both of RT with BS reference model and UCV methods, then compare the results in the simulation study given in Section 3.4.

3.3.3.1 RT method

Here we suggest to use the rule of thumb method that replaces the unknown density f in (3.13) by a known reference BS parametric model with parameters a and b ($T \sim \text{BS}(a,b)$); this approach can be called BS-referenced bandwidth (see, e.g., Silverman 1986; Jones and Henderson 2007; Hirukawa and Sakudo 2014) in the context of density estimation. Note that the parameters a and b are replaced by the corresponding estimators $\widehat{a}$ and $\widehat{b}$, which can be obtained in explicit forms, using the modified moment estimation (MME) as

$$\widehat{a} = \left[2\left\{\left(\frac{\bar{T}}{\bar{S}}\right)^{1/2} - 1\right\}\right]^{1/2}, \quad \widehat{b} = \left(\bar{T}\bar{S}\right)^{1/2}, \tag{3.14}$$

with

$$\bar{T} = \frac{1}{n}\sum_{i=1}^{n} T_i \quad \text{and} \quad \bar{S} = \left[\frac{1}{n}\sum_{i=1}^{n} T_i^{-1}\right]^{-1}.$$

Therefore, the BS-referenced bandwidth is given by

$$h_{\text{RT}} = \left[\frac{c_g^2 \int_0^\infty \frac{t^{-1} f_{\text{BS}(\widehat{a},\widehat{b})}(t)}{\left(1-\Phi\left[\frac{1}{\widehat{a}}\left\{\left(\frac{t}{\widehat{b}}\right)^{\frac{1}{2}} - \left(\frac{\widehat{b}}{t}\right)^{\frac{1}{2}}\right\}\right]\right)^2} dt}{c_{g^2} u_1^2 \int_0^\infty \left(\frac{t f'_{\text{BS}(\widehat{a},\widehat{b})}(t) + t^2 f''_{\text{BS}(\widehat{a},\widehat{b})}(t)}{1-\Phi\left[\frac{1}{\widehat{a}}\left\{\left(\frac{t}{\widehat{b}}\right)^{\frac{1}{2}} - \left(\frac{\widehat{b}}{t}\right)^{\frac{1}{2}}\right\}\right]}\right)^2 dt}\right]^{\frac{2}{5}} n^{-\frac{2}{5}}, \tag{3.15}$$

where $\Phi\left[\frac{1}{\widehat{a}}\left\{\left(\frac{t}{\widehat{b}}\right)^{\frac{1}{2}} - \left(\frac{\widehat{b}}{t}\right)^{\frac{1}{2}}\right\}\right]$ represents the cdf of BS distribution, with $\Phi(\cdot)$ is the standard normal cdf,

$$f_{BS(a,b)}(t) = \frac{1}{2ab\sqrt{2\pi}}\left(\left(\frac{b}{t}\right)^{\frac{1}{2}} + \left(\frac{b}{t}\right)^{\frac{3}{2}}\right)\exp\left(-\frac{1}{2a^2}\left[\frac{t}{b} + \frac{b}{t} - 2\right]\right), \quad t > 0,$$

$$f'_{BS(a,b)}(t) = \frac{\left(\frac{3b\sqrt{\frac{b}{t}}}{2t^2} - \frac{b}{2t^2\sqrt{\frac{b}{t}}}\right)\exp\left(-\frac{\frac{b}{t}+\frac{t}{b}-2}{2a^2}\right)}{2\sqrt{2\pi}ab} - \frac{\left(\frac{1}{b} - \frac{b}{t^2}\right)\left(\left(\frac{b}{t}\right)^{3/2} + \sqrt{\frac{b}{t}}\right)\exp\left(-\frac{\frac{b}{t}+\frac{t}{b}-2}{2a^2}\right)}{4\sqrt{2\pi}a^3 b}, \quad t > 0,$$

and

$$
\begin{aligned}
f''_{BS(a,b)}(t) &= \frac{\left(-\frac{3b^2}{4t^4\sqrt{\frac{b}{t}}} - \frac{b^2}{4t^4\left(\frac{b}{t}\right)^{3/2}} - \frac{3b\sqrt{\frac{b}{t}}}{t^3} + \frac{b}{t^3\sqrt{\frac{b}{t}}}\right)\exp\left(-\frac{\frac{b}{t}+\frac{t}{b}-2}{2a^2}\right)}{2\sqrt{2\pi}ab} \\
&+\frac{\left(\left(\frac{b}{t}\right)^{3/2}+\sqrt{\frac{b}{t}}\right)\left(\frac{1}{b}-\frac{b}{t^2}\right)^2\exp\left(-\frac{\frac{b}{t}+\frac{t}{b}-2}{2a^2}\right)}{8\sqrt{2\pi}a^5b} - \frac{\left(\left(\frac{b}{t}\right)^{3/2}+\sqrt{\frac{b}{t}}\right)\exp\left(-\frac{\frac{b}{t}+\frac{t}{b}-2}{2a^2}\right)}{2\sqrt{2\pi}a^3t^3} \\
&-\frac{\left(\frac{3b\sqrt{\frac{b}{t}}}{2t^2}-\frac{b}{2t^2\sqrt{\frac{b}{t}}}\right)\left(\frac{1}{b}-\frac{b}{t^2}\right)\exp\left(-\frac{\frac{b}{t}+\frac{t}{b}-2}{2a^2}\right)}{4\sqrt{2\pi}a^3b}, \; t>0.
\end{aligned}
$$

Note that we can use another reference model, such as gamma, lognormal, etc.

3.3.3.2 UCV method

The UCV method is based on the optimization of the integrated squared error (ISE), that is given in our case by

$$ISE(\widehat{\lambda}_{GBS}) = \int_0^\infty \left[\widehat{\lambda}_{GBS}(t)\right]^2 dt - 2\int_0^\infty \widehat{\lambda}_{GBS}(t)\lambda(t)dt + \int_0^\infty [\lambda(t)]^2\, dt.$$

The last term does not depend on bandwidth h, so we need to minimize the score function CV given by

$$
\begin{aligned}
CV(h) &= \int_0^\infty \left[\widehat{\lambda}_{GBS}(t)\right]^2 dt - 2\int_0^\infty \widehat{\lambda}_{GBS}(t)\lambda(t)dt \\
&= \int_0^\infty \left[\frac{\widehat{f}_{GBS}(t)}{1-\widehat{F}_{GBS}(t)}\right]^2 dt - 2\mathbb{E}\left\{\frac{\widehat{f}_{GBS}(t)}{[1-F(t)][1-\widehat{F}_{GBS}(t)]}.\right\}
\end{aligned}
$$

We replace F by its estimator $\widehat{F}_{GBS}$, then we get the new expression of $CV(h)$, given by

$$
\begin{aligned}
UCV(h) &= \int_0^\infty \left[\frac{\sum_{i=1}^n K_{GBS(t,h)}(T_i)}{n - \int_0^t \sum_{i=1}^n K_{GBS(x,h)}(T_i)dx}\right]^2 dt - \\
&\quad \frac{2(n-1)}{n}\sum_{i=1}^n \frac{\sum_{j=1,j\neq i}^n K_{GBS(T_i,h)}(T_j)}{\left[(n-1) - \int_0^{T_i}\sum_{j=1,j\neq i}^n K_{GBS(X_i,h)}(T_j)dX_i\right]^2}.
\end{aligned}
$$

The UCV optimal bandwidth is defined as

$$h_{UCV} = \arg\min_{h>0} UCV(h).$$

3.4 Simulation study

This section investigates the GBS (BS, BS-PE, BS-t and BS-Lap) kernel HR estimators developed in Section 3.3, and compare their performances with the Reciprocal Inverse Gaussian (RIG) and Gamma kernels HR estimators through simulation study. We note that the optimal bandwidth of the HR estimator using RIG and gamma kernels is calculated using RT and UCV methods. The comparison of these two methods is also investigated.

We simulate data from four nonnegative life distributions. We consider the lognormal, BS, gamma and BS-Student (GBS-t) distributions. The corresponding pdfs are listed in Table 3.1. For each target density, 100 replications of sample size n =50, 200, 500 and 1000 are generated. We compare the performance of the estimators using the ISE criterion given by

$$ISE[\widehat{\lambda}(t)] = \int_0^\infty \left[\lambda(t) - \widehat{\lambda}(t)\right]^2 dt.$$

	Distribution	pdf expression, $t > 0$	Parameters
D1	lognormal	$\frac{1}{t\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{\sigma^2}(\ln(t) - \mu)^2\right)$	$(\mu, \sigma) = (2, 3)$
D2	gamma	$\frac{1}{\beta^\alpha \Gamma(\alpha)} t^{\alpha-1} \exp(-\frac{t}{\beta})$	$(\alpha, \beta) = (3, 0.5)$
D3	BS	$\frac{1}{2\alpha\beta\sqrt{2\pi}} \left(\left(\frac{\beta}{t}\right)^{\frac{1}{2}} + \left(\frac{\beta}{t}\right)^{\frac{3}{2}}\right) \exp\left(-\frac{1}{2\alpha^2}\left[\frac{t}{\beta} + \frac{\beta}{t} - 2\right]\right)$	$(\alpha, \beta) = (2, 3)$
D4	BS-Student	$\frac{\Gamma(\frac{\nu+1}{2})}{2\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})\alpha\beta} \left(\left(\frac{\beta}{t}\right)^{\frac{1}{2}} + \left(\frac{\beta}{t}\right)^{\frac{3}{2}}\right) \left(1 + \frac{\frac{t}{\beta} + \frac{\beta}{t} - 2}{\nu\alpha^2}\right)^{-\frac{\nu+1}{2}}$	$(\alpha, \beta, t_\nu) = (1, 2, t_5)$

Table 3.1: Distributions used in simulation study.

Table 3.2 presents the average ISE based on 100 replications for the HR estimators of the models D1, D2, D3 and D4. For bandwidth choice, which is an important practical issue in nonparametric HR kernel estimation, we used RT and UCV procedures, developed in Section 3.3.3, for the purpose of selecting the one that gives best results.

Size	Kernels \ Models	BS(2,3)		lognormal(2,3)		gamma(3,0.5)		GBS(1, 2; $t_{v=5}$)	
		h_{RT}	h_{UCV}	h_{RT}	h_{UCV}	h_{RT}	h_{UCV}	h_{RT}	h_{UCV}
n=50	BS	**0.02608**	**0.05044**	**0.09258**	0.20487	**0.03692**	0.21300	**0.03061**	0.11378
	BS-PE($\nu = 2$)	0.05359	0.05628	0.09841	0.26722	0.04060	0.12588	0.07373	0.12567
	BS-t($\nu = 5$)	0.02618	0.05419	0.07130	**0.06313**	0.03337	0.26009	0.04082	0.07847
	BS-lap	0.03541	0.26491	0.14983	0.12800	0.04767	0.25904	0.04417	**0.07411**
	RIG	0.05507	0.11706	0.13853	0.11592	0.03761	**0.09739**	0.07605	0.12397
	gamma	0.05940	0.06396	0.09339	0.06872	0.03917	0.11679	0.10101	0.14568
n=200	BS	**0.00807**	**0.03860**	**0.02682**	0.06756	0.01956	0.19084	0.01564	0.04011
	BS-PE($\nu = 2$)	0.01700	0.09110	0.04774	**0.05680**	0.02740	0.11424	0.02337	**0.02136**
	BS-t($\nu = 5$)	0.00981	0,15111	0.04835	0.12515	0.02825	0.26264	**0.00990**	0.04406
	BS-lap	0.01228	0.27653	0.05037	0.07218	0.04140	0.25695	0.01266	0.04054
	RIG	0.01463	0.11497	0.05290	0.12381	**0.01026**	**0.04350**	0.02348	0.08176
	gamma	0.02040	0.06334	0.03656	0.06668	0.01779	0.11092	0.02316	0.14428
n=500	BS	**0.00407**	0.01056	0.02501	0.04410	0.01819	0.10905	0.00828	0.03865
	BS-PE($\nu = 2$)	0.00740	**0.00933**	0.06187	**0.02362**	0.02035	0.10932	0.01137	**0.03804**
	BS-t($\nu = 5$)	0.00543	0.01086	0.04201	0.03960	0.02447	0.26854	**0.00444**	0.04167
	BS-lap	0.00583	0.19160	**0.02032**	0.04062	0.03731	0.26217	0.00865	0.03974
	RIG	0.00999	0.09817	0.02525	0.11634	**0.00590**	**0.04961**	0.01332	0.16962
	gamma	0.00978	0.06189	0.02101	0.06602	0.01289	0.10905	0.01163	0.14297
n=1000	BS	**0.00217**	0.00582	0.01200	0.02464	0.01552	0.10242	**0.00277**	0.02537
	BS-PE($\nu = 2$)	0.00424	**0.00444**	0.04130	**0.01337**	0.01874	0.08560	0.00287	**0.02244**
	BS-t($\nu = 5$)	0.00436	0.00694	0.03491	0.03676	0.02301	0.26688	0.00339	0.03787
	BS-lap	0.00237	0.12540	**0.00884**	0.01340	0.03602	0.26209	0.00502	0.03349
	RIG	0.00571	0.09860	0.01535	0.11524	**0.00380**	**0.04175**	0.00613	0.07871
	gamma	0.00526	0.06022	0.01171	0.05954	0.00971	0.10900	0.00702	0.14573

Table 3.2: Some expected values of ISE for HR estimators, based on 100 replications for the considered models in simulation, using the bandwidths, h_{RT} and h_{UCV}.

In terms of average ISE, the obtained results based on GBS, RIG and gamma kernels reveal that:

- In general, the average ISE values, decreases as the sample size n increases, and the proposed class of GBS kernels outperforms the RIG and gamma kernels, whatever the bandwidth selection method, the sample size n and the distribution considered in the simulation study, except for gamma model.

- In the case of RT method, the BS kernel performs better than the other kernels for BS model, and also for small and moderate sample sizes n for lognormal distribution. However, the BS-lap kernel works better than the other for large sample size n in the case of lognormal distribution. In the case of BS-Student model, the BS-t and BS kernels have presented the best results. The performances of BS-t and BS HR estimators are mixed depending on the sample size n.

- In the case of UCV method, the BS kernel is better for small and moderate sample sizes n for BS model. However the BS-PE kernel gives best results for moderate and large sample sizes n for BS, lognormal and BS-Student distributions. The BS-t and BS-lap kernels perform well in the case of lognormal and BS-Student models, respectively, in the case of small sample sizes.

- The RIG kernel seems to be the suitable one in the case of gamma distribution for UCV method, in particular for moderate and large sample sizes in the case of RT method.

- In addition, the results show that the RT bandwidth selection method is more appropriate than UCV method, for the models considered in simulation study excepting the gamma model.

The comparison is also given in Figures 3.1 and 3.2. These figures indicate the estimates of the HR function for BS and lognormal models for the sample size $n = 200$. Globally, we can see graphically that the smoothing quality of the HR estimators, given in the Figure 3.1, where the parameter h is selected with RT method, is better compared to those in Figure 3.2, and it is very satisfactory for the GBS kernel compared to RIG and gamma kernels. We can also note that the best fit is obtained by using the BS and BS-lap kernels in comparison with the fit provided by BS-t and BS-PE kernels.

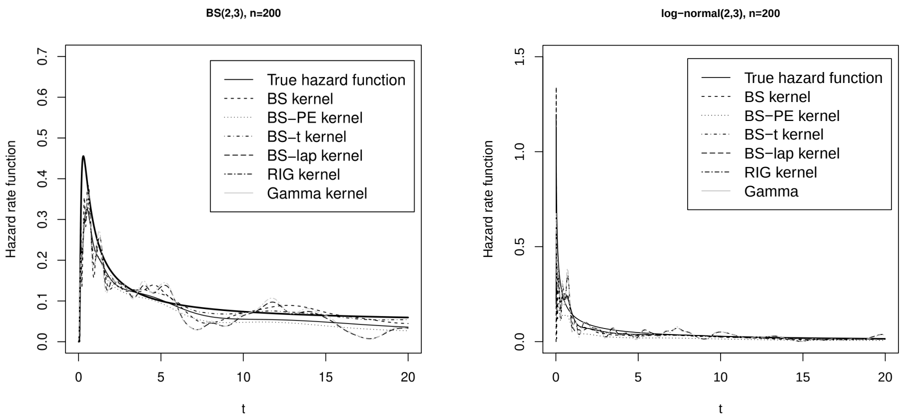

Figure 3.1: HR estimators for BS and lognormal models with $n = 200$, using the bandwidth h_{RT}.

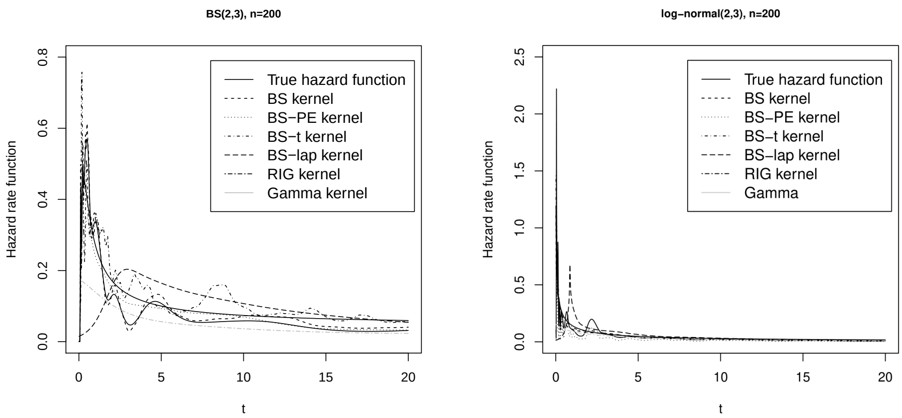

Figure 3.2: HR estimators for BS and lognormal models with $n = 200$, using the bandwidth h_{UCV}.

Computer programs used in this simulation study are given in appendices.

3.5 Exemple with real data

In this section, three real lifetime data sets are analyzed by our proposed approach. These three data sets are already discussed using the parametric BS and GBS distributions by Kundu et al. (2008) and Athayde et al. (2019), respectively; see also Paula et al. (2012).

(S1) represents survival times (in days) of $n = 72$ pigs injected with the same dose of tubercle bacilli, corresponding to 4.0×106 bacillary units per 0.5 ml, that is, to a regimen number of 237 the base-10 logarithm of bacillary units in 0.5 ml of challenge solution. These data are discussed by Kundu et al. (2008) using the life BS distribution, and analyzed more recently by Athayde et al. (2019) with the GBS life distribution.

(S2) denotes the total phosphorus 240 (mg/l) in Melides lagoon, Portugal, measured from 05-April-2004 to 26-Jan-2013 with sample size $n = 104$. Athayde et al. (2019) have investigated the HR of GBS distributions. Note that in our study, we have multiplied these data by 10, see Table 3.3.

(S3) concerns the claim amounts corresponding to $n = 542$ injuries paid by an insurance Australian. These data have been analyzed earlier by using the BS-Student regression model, and more recently, Athayde et al. (2019) have investigated the HR of GBS distributions in parametric estimation. Note that in our investigation, we have divided these data by 1000, see Table 3.3.

Data	Mean	Median	SD	CV	CS	CK	Min	Max	n
S1	99.82	70	81.11	81.26%	1.79	5.61	12	376	72
S2	1.39	0.99	1.49	107.14%	3.52	16.67	0.1	9.91	240
S3	8.99	6.76	8.79	97.71%	5.59	54.31	0.109	116.58	542

Table 3.3: Descriptive statistics for the indicated data set.

Note that

CV: coefficient of variation.

CS: coefficient of skewness.

CK: coefficient of kurtosis.

The Table 3.3 gives a descriptive statistics for lifetime data sets *S1*, *S2* and *S3*. We can see that all these data sets are positively skewed and present a high kurtosis level, in particular for *S2* and *S3*. Then, we apply the GBS, RIG and gamma kernels to estimate the density and the HR function for these considered data. Table 3.4 provides the bandwidth selectors given by the RT method for the density estimator $\widehat{f}_{GBS}$, and both of RT and UCV methods for the HR function estimator $\widehat{\lambda}_{GBS}$, respectively, according to the real data sets *S1*, *S2* and *S3*. Note that the RT method is based on BS reference model, see Section 3.3.3 for HR estimators. Figures 3.3, 3.4 and 3.5 show the estimates of the pdf and HR function for *S1*, *S2* and *S3* data sets based on GBS (BS, BS-PE, BS-t and BS-lap), RIG and gamma kernels combined with RT and UCV bandwidth selectors. In general, we can observe that in term of smoothing quality, the GBS kernel perform better than RIG and gamma kernels for the data *S1* in the case of density estimation, and the data *S1*, *S2* and *S3* in the case of HR estimation using RT bandwidth selection method, see Figures 3.3 and 3.4, respectively. The first best performance is obtained with BS kernel, and the second best result can be attributed to BS-t and BS-lap, in particular for HR function estimation. The BS-PE and gamma kernels tend to under or over smooth the HR function of the considered data sets, especially in the case of UCV bandwidth selector, see Figure 3.5.

Data	Kernel	$\widehat{f}_{GBS}$	$\widehat{\lambda}_{GBS}$	
			h_{RT}	h_{UCV}
	BS	0.02800	0.05635	0.00621
S1	BS-PE($\nu = 2$)	0.01019	0.05954	0.48943
	BS-t($\nu = 5$)	0.01752	0.06873	0.01626
	BS-lap	0.02670	0.04237	0.07517
	RIG	0.03524	0.63086	1.99782
	Gamma	0.02368	0.46079	0.55600
	BS	0.02454	0.03992	1.26476
S2	BS-PE($\nu = 2$)	0.05756	0.04277	0.67289
	BS-t($\nu = 5$)	0.03564	0.87868	1.22220
	BS-lap	0.02354	0.03048	0.81522
	RIG	0.06524	0.02578	0.48830
	Gamma	0.08524	0.02545	1.99708
	BS	0.14089	0.03667	0.82671
S3	BS-PE($\nu = 2$)	0.32045	0.03650	1.81204
	BS-t($\nu = 5$)	0.08463	0.03728	0.99120
	BS-lap	0.26548	0.02314	0.61790
	RIG	0.12563	0.08665	1.74301
	Gamma	0.09648	0.07141	1.83027

Table 3.4: Bandwidth of HR estimators computed with RT and UCV methods, and those of pdf computed with RT method.

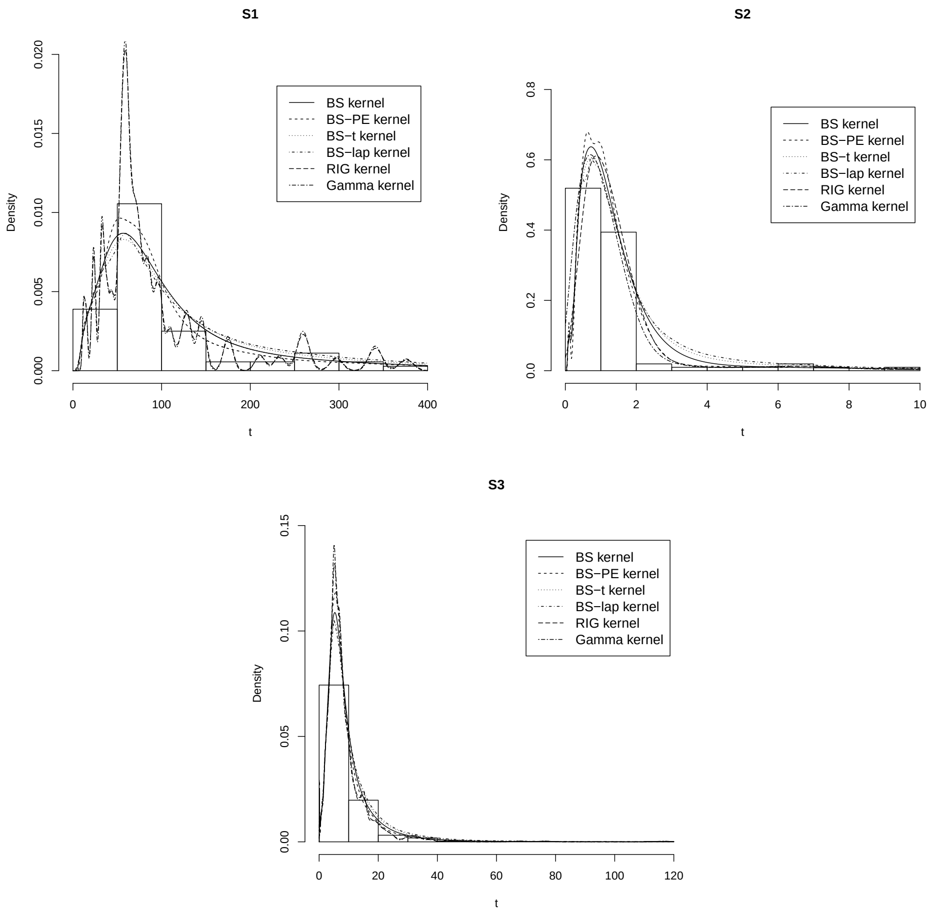

Figure 3.3: Pdf estimation for *S1*, *S2* and *S3* data.

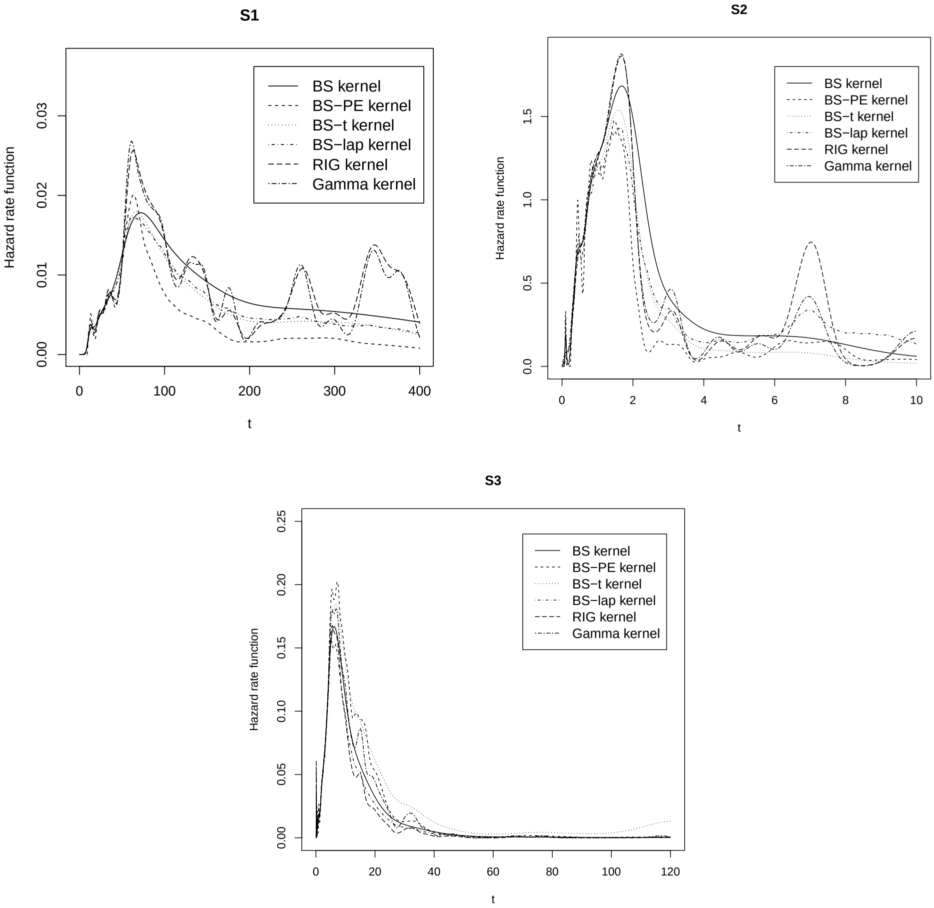

Figure 3.4: HR function estimation for *S1*, *S2* and *S3* data, using RT bandwidth selector.

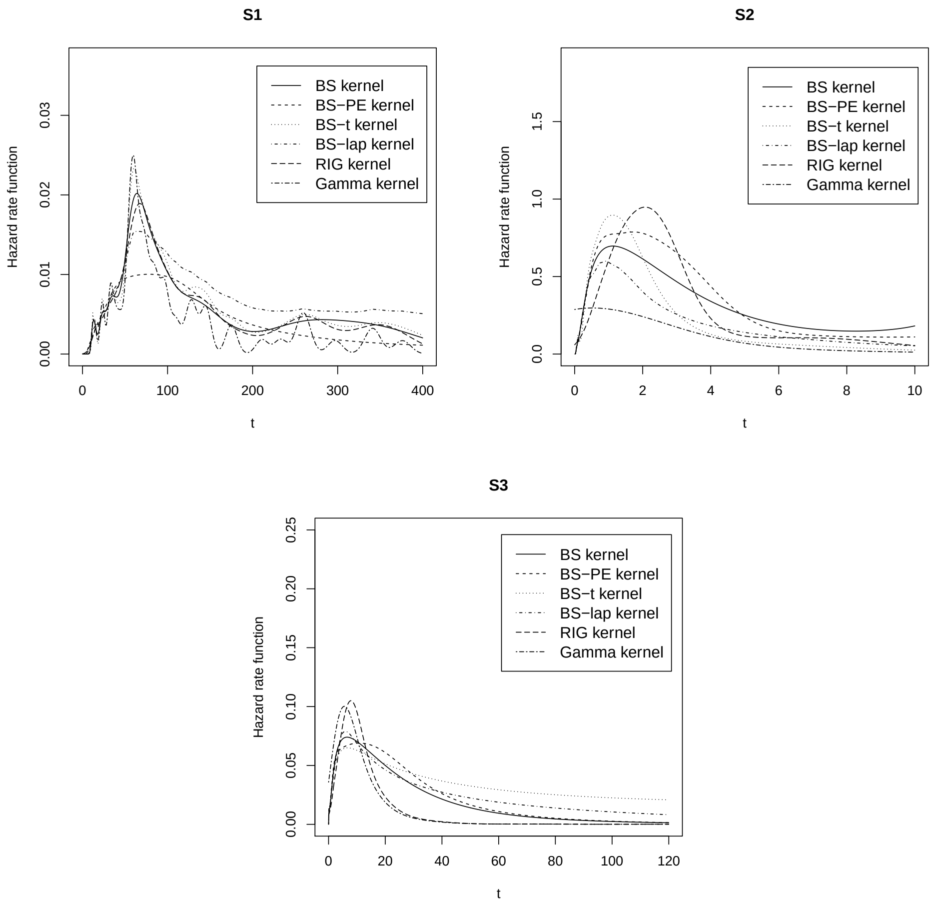

Figure 3.5: HR function estimation for *S1*, *S2* and *S3* data, using UCV bandwidth selector.

Chapter 4

Reliability and reversed hazard rate function estimation using GBS kernel

4.1 Introduction

According to Marchant at al. (2013) and Chekkal et al. (2021), the class of GBS kernels gives a good estimates for density pdf and hazard rate HR function, respectively. Since these functions are directly related to the reliability and reversed hazard rate (RHR) functions, it will be interesting to test if the class of GBS kernels performs in the case of reliability and reversed hazard rate functions. That is why we propose to study these two functions using GBS kernels, and establish the asymptotic properties for each estimator. Finally, simulation study is investigated to test their performance.

4.2 GBS-Reliability estimation

The natural estimator of reliability function of nonnegative r.v. T, is given as

$$\widehat{R}(t) = 1 - \int_0^t \widehat{f}(x)dx, \quad t > 0, \tag{4.1}$$

with $\widehat{f}$ is the pdf estimator of f.

Let $T_1, T_2, .., T_n$ be a independent identically distributed (i.i.d.) random sample distributed as a nonnegative r.v. T. By using the class of GBS kernel, the estimator GBS-

Reliability can be rewritten as

$$\widehat{R}_{GBS}(t) = 1 - \int_0^t \frac{1}{n}\sum_{i=1}^{n} K_{GBS(x,h)}(T_i)dx, \quad t > 0, \tag{4.2}$$

where t is the target, h is the bandwidth parameter and K_{GBS} represents the class of GBS kernels, presented previously in Chapter 2, Section 2.4.

By replacing the expression of the kernel K_{GBS}, we obtain the following GBS-Reliability estimator

$$\widehat{R}_{GBS}(t) = 1 - \frac{c_g}{2n\sqrt{h}}\sum_{i=1}^{n}\int_0^t \left(\frac{1}{\sqrt{xT_i}} + \sqrt{\frac{x}{T_i^3}}\right) g\left(\frac{1}{h}\left(\frac{T_i}{x} + \frac{x}{T_i} - 2\right)\right) dx, \quad t > 0. \tag{4.3}$$

The asymptotic properties and the methods for selecting the bandwidth h of this estimator, are illustrated in the next.

4.2.1 Asymptotic properties

The propositions above are established under the conditions C1-C3 given in the page 39 and the following hypothesis

F1. $\int \frac{t^{-1}R(t)^2}{f(t)}dt < \infty$;

F2. $\int \left[\frac{tf'(t)R(t)}{f(t)}\right]^2 dt < \infty$;

F3. $\int \left[\frac{t^2 f''(t)R(t)}{f(t)}\right]^2 dt < \infty$.

Proposition 4.2.1. *Let $\widehat{R}_{GBS}$ be the GBS kernel of the reliability function. Under the conditions C1 and C2, the following holds*

$$\text{Bias}[\widehat{R}_{GBS}(t)] = m^* \frac{R(t)}{f(t)} + o(h), \quad t > 0.$$

$$\text{Var}[\widehat{R}_{GBS}(t)] = \sigma^{*2}\frac{R^2(t)}{f(t)} + o\left(\frac{1}{nh^{1/2}}\right), \quad t > 0.$$

Where $m^* = \dfrac{hu_1}{2}\left[tf'(t) + t^2 f''(t)\right]$ *and* $\sigma^* = \dfrac{c_g^2}{(c_{g^2} nh^{1/2})} t^{-1} f(t)$.

Proof The bias and the variance of $\widehat{R}_{GBS}$ are obtained by using Theorem 3.3.3 given in Section 3.3.2, Chapter 3, (displaying the convergence almost sure of $\widehat{\lambda}_{GBS}$, $\widehat{\lambda}_{GBS}(t) \xrightarrow{a.s.} \lambda(t)$ as $n \to \infty$), and the fact that $\widehat{R}_{GBS}(t) = \frac{\widehat{f}_{GBS}(t)}{\widehat{\lambda}_{GBS}(t)}$. Then we replace the expressions of $\text{Bias}(\widehat{f}_{GBS})$ and $\text{Var}(\widehat{f}_{GBS})$ given in Chapter 2, Section 2.4.

Proposition 4.2.2. *The mean integrated squared error* (MISE) *of the estimator GBS-R is given under the conditions C1-C3 and the hypothesis F1-F3 by*

$$\text{MISE}[\widehat{R}_{GBS}] = \frac{h^2u_1^2}{4}\int_0^{+\infty}[tf'(t)+t^2f''(t)]^2\frac{R^2(t)}{f^2(t)}dt+\frac{c^2}{c_{g^2}nh^{1/2}}\int_0^{+\infty}t^{-1}\frac{R^2(t)}{f(t)}dt+o\left(h^2+\frac{1}{n\sqrt{h}}\right). \tag{4.4}$$

The $\text{MISE}[\widehat{R}_{GBS}]$ is obtained by replacing the $\text{Bias}[\widehat{R}_{GBS}(t)]$ and $\text{Var}[\widehat{R}_{GBS}(t)]$ in $\text{MISE}[\widehat{R}_{GBS}(t)] = \int_0^{+\infty}\text{Bias}^2[\widehat{R}_{GBS}(t)] + \int_0^{+\infty}\text{Var}[\widehat{R}_{GBS}(t)]dt$, with the expressions of m^* and σ^{*2}.

Using the hypothesis F1-F3, $h \to 0$ and $\frac{1}{n\sqrt{h}} \to \infty$, as $n \to \infty$, the $\text{MISE}[\widehat{R}_{GBS}]$ tend to 0.

4.2.2 Bandwidth selection

For the selection of a bandwidth parameter h, we propose to use the same methods used in the case of GBS-HR estimator, that are; rule of thumb (RT) and unbiased cross validation (UCV).

4.2.2.1 RT method

By minimizing the MISE criterion in h, given in (4.4) we obtain the following optimal bandwidth parameter

$$h^* = \left[\frac{c^2\int t^{-1}\frac{R^2(t)}{f(t)}dt}{u_1^2(g)c_{g^2}\int[tf'(t)+t^2f''(t)]^2\frac{R^2(t)}{f^2(t)}dt}\right]^{2/5} n^{-2/5}.$$

The main obstacle here is that the bandwidth above depends on the unknown quantities f, f', f'' and R. That makes the calculation more difficult especially in the practical case, see for instance, Hirukawa and Sakudo (2014) in the case of density function. We propose to use the BS parametric distribution as a reference model with parameters a and b, $T \sim BS(a, b)$ (as in the case of GBS-HR estimator), and replace the unknown functions by f_{BS}, f'_{BS}, f''_{BS} and R_{BS}, respectively. The parameters a and b are estimated using the modified moment method (MME) and their estimators $\hat{a}$ and $\hat{b}$ are given in the formula (3.14), Section 3.3.3, Chapter 3.

Therefore the BS-referenced bandwidth of GBS-R estimator is given by

$$h^R_{RT} = \left[\frac{c_g^2 \int_0^{+\infty} t^{-1} \frac{\left(1-\Phi\left[\frac{1}{\widehat{a}}\left\{\left(\frac{t}{\widehat{b}}\right)^{\frac{1}{2}} - \left(\frac{\widehat{b}}{t}\right)^{\frac{1}{2}}\right\}\right]\right)^2}{f_{BS}(t)} dt}{u_1^2 c_{g^2} \int_0^{+\infty} \left[t f'_{BS}(t) + t^2 f''_{BS}(t)\right]^2 \frac{\left(1-\Phi\left[\frac{1}{\widehat{a}}\left\{\left(\frac{t}{\widehat{b}}\right)^{\frac{1}{2}} - \left(\frac{\widehat{b}}{t}\right)^{\frac{1}{2}}\right\}\right]\right)^2}{f^2_{BS}(t)} dt} \right]^{2/5} n^{-2/5}, \qquad (4.5)$$

where $\Phi\left[\frac{1}{\widehat{a}}\left\{\left(\frac{t}{\widehat{b}}\right)^{\frac{1}{2}} - \left(\frac{\widehat{b}}{t}\right)^{\frac{1}{2}}\right\}\right]$ represents the cdf of BS distribution, with $\Phi(\cdot)$ is the standard normal cdf and the expressions of f_{BS}, f'_{BS}, f''_{BS} are given previously in Chapter 3, Section 3.3.3.

4.2.2.2 UCV method

This method consists to minimize the integrated squared error (ISE) of the GBS-R estimator, given by

$$ISE(\widehat{R}_{GBS}) = \int_0^\infty \left[\widehat{R}_{GBS}(t)\right]^2 dt - 2\int_0^\infty \widehat{R}_{GBS}(t)R(t)dt + \int_0^\infty [R(t)]^2\, dt.$$

The last term does not depend on bandwidth h, so we need to minimize the score function CV_R given by

$$\begin{aligned} CV_R(h) &= \int_0^\infty \left[\widehat{R}_{GBS}(t)\right]^2 dt - 2\int_0^\infty \widehat{R}_{GBS}(t)R(t)dt \\ &= \int_0^\infty (1-\widehat{F}_{GBS})^2 dt - 2\mathbb{E}\left[\frac{\widehat{f}_{GBS}(t)}{\lambda(t)\widehat{\lambda}_{GBS}(t)}\right]. \end{aligned}$$

We replace λ by its estimator $\widehat{\lambda}_{GBS}$, then we get the new expression of $CV_R(h)$, given by

$$\begin{aligned} UCV_R(h) &= \int_0^\infty \left[1 - \int_0^t \frac{1}{n}\sum_{i=1}^n K_{GBS(x,h)}(T_i)dx\right]^2 - \\ & \frac{2(n-1)}{n}\sum_{i=1}^n \frac{\left[1 - \frac{1}{n-1}\int_0^{T_i}\sum_{j\neq i} K_{GBS(X_i,h)}(T_j)dX_i\right]^2}{\sum_{j\neq i} K_{GBS(T_i,h)}(T_j)}. \end{aligned}$$

The UCV_R optimal bandwidth is

$$h^R_{UCV} = \arg\min_{h>0} UCV_R(h).$$

A simulation study is conducted in the Section 4.4.1, in order to test the performance of the reliability estimator $\widehat{R}_{GBS}$ and select the suitable bandwidth parameter for the samples considered.

4.3 GBS-RHR estimator

In this section, we present the reversed hazard rate RHR function estimator based on the class of GBS kernels. Recall the expression of the reversed hazard rate function of the lifetime r.v. T

$$\rho(t) = \frac{f(t)}{F(t)}, \quad t > 0,$$

with pdf f and cdf F.

Let $T_1, T_2, ..., T_n$ a set of r.v. distributed as T. The kernel estimator GBS-RHR of RHR function using GBS kernels is given as

$$\begin{aligned}\widehat{\rho}_{\text{GBS}}(t) &= \frac{\widehat{f}_{\text{GBS}}(t)}{\widehat{F}_{\text{GBS}}(t)} \\ &= \frac{\sum_{i=1}^{n}\left(\frac{1}{\sqrt{tT_i}} + \sqrt{\frac{t}{T_i^3}}\right) g\left(\frac{1}{h}\left(\frac{T_i}{t} + \frac{t}{T_i} - 2\right)\right)}{\sum_{i=1}^{n}\int_0^t \left(\frac{1}{\sqrt{xT_i}} + \sqrt{\frac{x}{T_i^3}}\right) g\left(\frac{1}{h}\left(\frac{T_i}{x} + \frac{x}{T_i} - 2\right)\right) dx}, \quad t > 0, \end{aligned} \tag{4.6}$$

with $\widehat{f}_{\text{GBS}}(t)$ and $\widehat{F}_{\text{GBS}}(t)$ are GBS kernel estimator of pdf f and cdf F, given respectively by

$$\widehat{f}_{\text{GBS}}(t) = \frac{c_g}{2n\sqrt{h}} \sum_{i=1}^{n} \left(\frac{1}{\sqrt{tT_i}} + \sqrt{\frac{t}{T_i^3}}\right) g\left(\frac{1}{h}\left(\frac{T_i}{t} + \frac{t}{T_i} - 2\right)\right), \quad t > 0,$$

$$\widehat{F}_{\text{GBS}}(t) = \frac{c_g}{2n\sqrt{h}} \sum_{i=1}^{n} \int_0^t \left(\frac{1}{\sqrt{xT_i}} + \sqrt{\frac{x}{T_i^3}}\right) g\left(\frac{1}{h}\left(\frac{T_i}{x} + \frac{x}{T_i} - 2\right)\right) dx, \quad t > 0,$$

and h is the parameter that controls the smoothness of the estimator $\widehat{\rho}_{\text{GBS}}(t)$. Two methods will be investigated to select the optimal one, in the Section 4.3.2.
The asymptotic properties of the estimator (4.6) are given in the next section.

4.3.1 Asymptotic properties

The asymptotic properties of the estimator GBS-RHR defined in (4.6) are established under the conditions C1-C3 and the following assumptions

G1. $\int_0^\infty \left(\frac{tf'(t)}{F(t)}\right)^2 dt < \infty$;

G2. $\int_0^\infty \left(\frac{t^2 f''(t)}{F(t)}\right)^2 dt < \infty$;

G3. $\int_0^\infty \frac{t^{-1} f(t)}{F^2(t)} dt < \infty$.

Proposition 4.3.1. *Under the assumptions C1-C3, the bias and the variance of the estimator GBS-RHR are given by*

$$\text{Bias}\left[\widehat{\rho}_{\text{GBS}}(t)\right] = \frac{m^*}{F(t)} + o(h), \quad t > 0.$$

$$\text{Var}\left[\widehat{\rho}_{\text{GBS}}(t)\right] = \frac{\sigma^{*2}}{F^2(t)} + o\left(\frac{1}{n\sqrt{h}}\right), \quad t > 0,$$

where the expressions of m^* and σ^* are given in Proposition 4.2.1.

Proof.The bias and the variance of $\widehat{\rho}_{GBS}$ are obtained by using Proposition 3.3.1 given in Chapter 3, Section 3.3.2, displaying the convergence almost sure of the estimator $\widehat{F}_{GBS}$, $\widehat{F}_{GBS}(t) \xrightarrow{a.s.} F(t),$ as $n \to \infty$. Then we replace the expressions of Bias$(\widehat{f}_{GBS})$ and Var$(\widehat{f}_{GBS})$ given in Chapter 2, Section 2.4, page 36.

Proposition 4.3.2. *The mean integrated squared error* MISE *is given under G1-G3*

$$\begin{aligned} \text{MISE}(\widehat{\rho}_{\text{GBS}}) &= \frac{u_1^2(g)h^2}{4}\int_0^\infty \left[\frac{(tf'(t)+t^2f''(t))}{F(t)}\right]^2 dt + \frac{c_g^2}{c_{g^2}nh^{\frac{1}{2}}}\int_0^\infty \frac{t^{-1}f(t)}{[F(t)]^2}dt \\ &\quad + o\left(h^2 + \frac{1}{nh^{\frac{1}{2}}}\right), \end{aligned} \tag{4.7}$$

The MISE (4.7) is obtained by substituting the formulas of the bias and variance of the estimator $\widehat{\rho}_{\text{GBS}}(t)$ in $\int_0^\infty \text{Bias}^2[\widehat{\rho}_{GBS}(t)]dt + \int_0^\infty \text{Var}[\widehat{\rho}_{GBS}(t)]dt$ with the expressions of m^* and σ^{*2}.

Using the hypothesis G1-G3, $h \to 0$ and $\frac{1}{n\sqrt{h}} \to \infty$, as $n \to \infty$, the MISE$[\widehat{R}_{GBS}]$ tend to 0.

4.3.2 Bandwidth selection

For the selection of a bandwidth parameter h, we propose to use the same methods used in the case of GBS-HR estimator; RT and UCV methods.

4.3.2.1 RT method

This method consists to minimize the MISE$(\widehat{\rho}_{GBS})$ given in (4.7), to obtain the following expression of h

$$h^{**} = \left[\frac{c_g^2\int_0^\infty \frac{t^{-1}f(t)}{F(t)^2}dt}{c_{g^2}u_1^2(g)\int_0^\infty\left(\frac{tf'(t)+t^2f''(t)}{F(t)}\right)^2 dt}\right]^{\frac{2}{5}} n^{-\frac{2}{5}}.$$

We see clearly that the bandwidth above depends on the unknown quantities f, f', f'', and F. That makes the calculation more difficult especially in the practical case, see for instance, Hirukawa and Sakudo (2014) in the case of density function. This consists to substitute the unknown functions by those of BS parametric model, i.e., f_{BS}, f'_{BS}, f'_{BS}, and F_{BS}, with parameters a and b ($T \sim BS(a,b)$) that are also estimated by modified moment method (MME), noted by $\widehat{a}$ and $\widehat{b}$ respectively, see their expressions in (3.14), Chapter 3, Section 3.3.3. Finally we obtain the following optimal bandwidth as

$$h^{\rho}_{\mathrm{RT}} = \left[\frac{c_g^2 \int_0^{\infty} \frac{t^{-1} f_{\mathrm{BS}(\widehat{a},\widehat{b})}(t)}{\left(\Phi\left[\frac{1}{\widehat{a}}\left\{\left(\frac{t}{\widehat{b}}\right)^{\frac{1}{2}} - \left(\frac{\widehat{b}}{t}\right)^{\frac{1}{2}}\right\}\right]\right)^2} dt}{c_{g^2} u_1^2(g) \int_0^{\infty} \left(\frac{t f'_{\mathrm{BS}(\widehat{a},\widehat{b})}(t) + t^2 f''_{\mathrm{BS}(\widehat{a},\widehat{b})}(t)}{\Phi\left[\frac{1}{\widehat{a}}\left\{\left(\frac{t}{\widehat{b}}\right)^{\frac{1}{2}} - \left(\frac{\widehat{b}}{t}\right)^{\frac{1}{2}}\right\}\right]} \right)^2 dt} \right]^{\frac{2}{5}} n^{-\frac{2}{5}}, \tag{4.8}$$

where $\Phi\left[\frac{1}{\widehat{a}}\left\{\left(\frac{t}{\widehat{b}}\right)^{\frac{1}{2}} - \left(\frac{\widehat{b}}{t}\right)^{\frac{1}{2}}\right\}\right]$ represents the cdf of BS distribution, with $\Phi(\cdot)$ is the standard normal cdf. See the expressions of f_{BS}, f'_{BS}, f''_{BS}, F_{BS} in the Chapter 3, Section 3.3.3.

4.3.2.2 UCV method

This method consists in optimization of the integrated squared error (ISE), that is given by

$$ISE(\widehat{\rho}_{GBS}) = \int_0^{+\infty} \left[\widehat{\rho}_{GBS}(t)\right]^2 dt - 2\int_0^{+\infty} \widehat{\rho}_{GBS}(t)\rho(t)dt + \int_0^{+\infty} \left[\rho(t)\right]^2 dt.$$

The last term does not depend on bandwidth h, so we need to minimize the score function CV_ρ given by

$$\begin{aligned} CV_\rho(h) &= \int_0^{+\infty} \left[\widehat{\rho}_{GBS}(t)\right]^2 dt - 2\int_0^{+\infty} \widehat{\rho}_{GBS}(t)\rho(t)dt \\ &= \int_0^{\infty} \left[\frac{\widehat{f}_{GBS}(t)}{\widehat{F}_{GBS}(t)}\right]^2 dt - 2\mathbb{E}\left[\frac{\widehat{f}_{GBS}(t)}{F(t)\widehat{F}_{GBS}(t)}\right]. \end{aligned}$$

We replace F by its estimator $\widehat{F}_{GBS}$, then we get the new expression of $CV_\rho(h)$, given by

$$UCV_\rho(h) = \int_0^{\infty} \left[\frac{\sum_{i=1}^n K_{GBS(t,h)}(T_i)}{\int_0^t \sum_{i=1}^n K_{GBS(x,h)}(T_i)dx}\right]^2 dt - \frac{2(n-1)}{n} \sum_{i=1}^n \frac{\sum_{j=1,j\neq i}^n K_{GBS(T_i,h)}(T_j)}{\left[\int_0^{T_i} \sum_{j=1,j\neq i}^n K_{GBS(X_i,h)}(T_j)dX_i\right]^2}.$$

The UCV optimal bandwidth is defined as

$$h^{\rho}_{UCV} = \arg\min_{h>0} UCV_\rho(h).$$

A simulation study is conducted in the Section 4.4.2 to test the performance of the RHR function estimator $\widehat{\rho}_{GBS}$, and select the appropriate bandwidth for the corresponding generated samples.

4.4 Simulation study

In this section, we test the performance of GBS kernels comparing to the RIG gamma kernels in the case of reliability and RHR estimations, and compare the two methods of bandwidth selection; RT and UCV.

We proceed as in the case of GBS-HR estimator by using the kernels: BS, B-PE, BS-lap, BS-t, gamma, RIG and generating sample sizes $n = 50$, $n = 200$, $n = 500$ and $n = 1000$ from the nonnegative distributions: BS (2,3), lognormal (2,3), gamma (3,1/2) and BS-Student (1,2; $t_{v=5}$), using 100 replications. See their corresponding densities in the Chapter 3, Section 3.4, Table 3.1.

4.4.1 GBS-Reliability estimator

We calculate the ISE criterion in both case of RT and UCV bandwidth selection methods and the results of the average ISE are illustrated in the Table 4.1.

According to the results given in Table 4.1, we can see globally that, the average ISE values decreases as the sample size n increases, whatever the bandwidth selection method. GBS and RIG kernels perform well and the results obtained in the case of RT method are better than those obtained for UCV method, except the case of gamma distribution. In fact

- In the case of BS, lognormal and BS-Student distributions, the values of ISE obtained for RT method are good comparing to those obtained through UCV method. Indeed, for RT method and in the case of BS distribution, the values of ISE are nearly closed for all the kernels, however, the best results are obtained for BS, BS-PE and BS-t depending on the sample size n. Similarly to lognormal and BS-Student distributions, the values of ISE are almost closed but the smallest ones are presented in RIG kernel for all the sizes n, except the small sample size n for BS-Student distribution, where the BS-PE is the most efficient.

Size	Models / Kernels	BS(2,3)		lognormal(2,3)		gamma(3,0.5)		GBS$(1,2;t_{v=5})$	
		h_{RT}	h_{UCV}	h_{RT}	h_{UCV}	h_{RT}	h_{UCV}	h_{RT}	h_{UCV}
	BS	**0.04070**	0.17779	0.08596	0.08188	2.70045	0.45230	0.08857	0.17573
	BS-PE($\nu = 2$)	0.04354	**0.09446**	0.06382	**0.06527**	2.45184	0.26535	**0.07415**	**0.16583**
	BS-t($\nu = 5$)	0.05397	0.21528	0.06686	0.17824	2.33854	0.51446	0.09384	0.17407
n=50	BS-lap	0.05336	0.22030	0.06878	0.24423	2.38445	0.52074	0.10135	0.16709
	RIG	0.04866	0.26938	**0.05370**	0.13927	**0.03618**	0.25433	0.07682	0.33264
	gamma	0.04150	0.31413	0.06891	0.27139	0.03833	**0.05243**	0.07694	0.22923
	BS	0.01417	0.01848	0.02379	0.01924	2.71708	0.43191	0.06451	0.11651
	BS-PE($\nu = 2$)	**0.01170**	**0.01424**	0.01699	**0.01736**	2.58081	0.24639	0.05745	0.11902
	BS-t($\nu = 5$)	0.01169	0.02116	0.01795	0.01914	2.45229	0.50392	0.06455	0.12106
n=200	BS-lap	0.01360	0.02957	0.02746	0.01912	2.44789	0.49206	0.07924	**0.11638**
	RIG	0.01497	0.24960	**0.01666**	0.12814	**0.00803**	0.23625	**0.05298**	0.35137
	gamma	0.01459	0.29682	0.02453	0.27011	0.01030	**0.04857**	0.05319	0.22979
	BS	0.00580	0.00574	0.01238	0.01249	2.73422	0.43616	0.06414	0.09786
	BS-PE($\nu = 2$)	0.00810	**0.00567**	0.01141	0.01292	2.53538	0.25031	0.05760	0.11353
	BS-t($\nu = 5$)	**0.00565**	0.00578	0.01339	**0.01221**	2.45207	0.50496	0.05856	**0.09653**
n=500	BS-lap	0.00665	0.01056	0.01345	0.12347	2.41191	0.49603	0.05813	0.10808
	RIG	0.00688	0.24870	**0.01009**	0.11465	**0.00368**	0.23614	**0.05076**	0.34616
	gamma	0.00665	0.30200	0.01653	0.27481	0.00421	**0.05061**	0.05086	0.23525
	BS	0.00537	0.00586	0.00875	0.00690	2.7098	0.43361	0.05569	0.09219
	BS-PE($\nu = 2$)	**0.00459**	**0.00584**	0.00619	0.00781	2.65970	0.24877	0.05567	0.10524
	BS-t($\nu = 5$)	0.00483	0.00585	0.00884	**0.00689**	2.41113	0.50355	**0.05321**	**0.08333**
n=1000	BS-lap	0.00518	0.00585	0.00746	0.00955	2.37221	0.49503	0.05988	0.09389
	RIG	0.00544	0.24810	**0.00549**	0.12616	**0.00224**	0.23469	0.05390	0.35589
	gamma	0.00530	0.30260	0.01034	0.26393	0.00270	**0.05186**	0.05403	0.22835

Table 4.1: Some expected values of ISE for reliability estimators, based on 100 replications for the considered models in simulation, using the bandwidths, h_{RT}^{R} and h_{UCV}^{R}.

- In regards to gamma distribution, the UCV method works well than RT method. In that case the most efficient kernel is gamma.

The comparison is also given in Figures 4.1 and 4.2. These figures indicate, respectively the estimates of the reliability function for BS and lognormal models for the sample size $n = 200$ for h^R_{RT} and h^R_{UCV}. Globally, we can see graphically that the smoothing quality of the reliability estimators in the case of RT method (Figure 4.1) is satisfactory compared to the case where the bandwidth h is obtained with UCV method (Figure 4.2), and we can note that the shape of all the estimators are nearly closed to the true reliability function, for BS and lognormal distributions in the case of RT method.

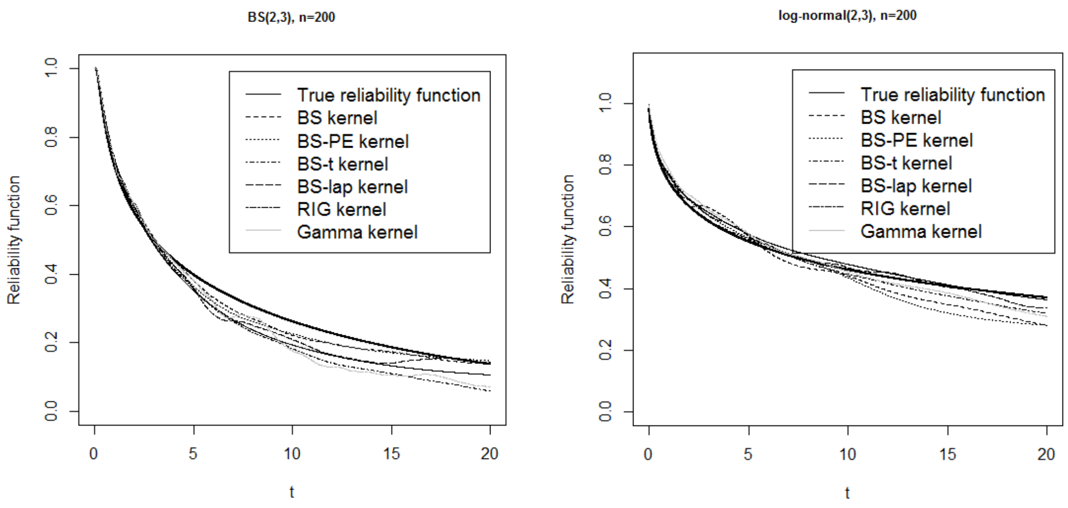

Figure 4.1: Reliability estimators for BS and lognormal models with $n = 200$, using the bandwidth h^R_{RT}.

The computer programs used in this simulation study are given in appendices.

4.4.2 GBS-RHR estimator

According to the results given in Table 4.2, we can see globally that, the average ISE values decreases as the sample size n increases, and in the case of the class of GBS kernels the results obtained for RT method are better than those obtained for UCV method. In fact

- In the case of BS distribution, the RT method gives better results than UCV method, and that is for all the kernels. Thus, in the case RT method, all the kernels perform well, however, the gamma kernel presents the smallest values of ISE for all the sizes n.

- In the case of lognormal distribution, the smallest values of ISE are obtained mostly in the case of RT method. In that case, BS kernel is the best one for small sizes n and BS-lap for moderate and large sample sizes n.

- In the case of BS-Student distribution the RT method is also better than UCV for all the kernels, except the BS kernel witch gives the best results in the case of UCV approach. The BS-t kernel is the most efficient in the case of RT method.

- In the case of gamma distribution, The RT method performs well for GBS kernels.

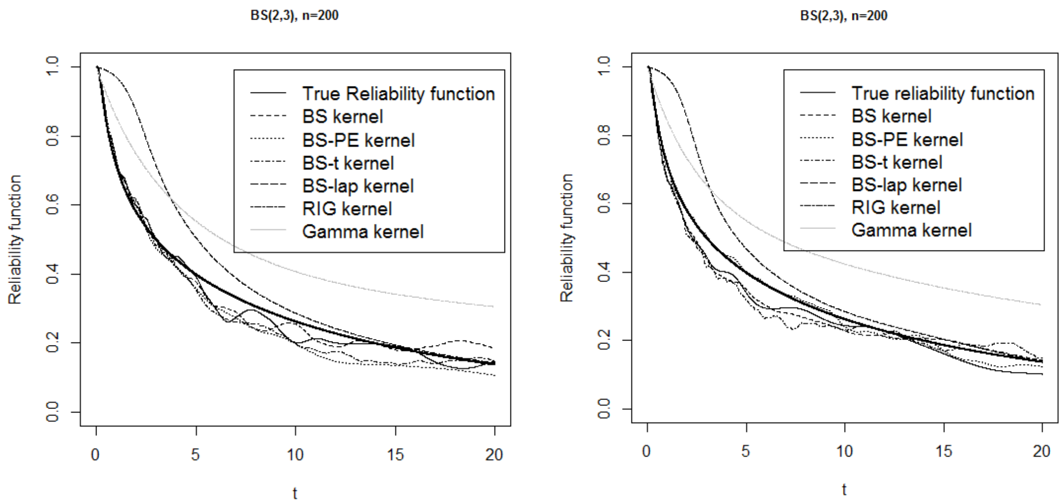

Figure 4.2: Reliability estimators for BS and lognormal models with $n = 200$, using the bandwidth h^R_{UCV}.

Hence, BS kernel is the best one for small and large sizes n and BS-t for moderate sample sizes n.

The comparison is also given in Figures 4.3 and 4.4. These figures indicate the estimates of the RHR function for BS and lognormal models for the sample size $n = 200$. Globally, we can see graphically that the smoothing quality of the RHR estimators in the case of RT method (Figure 4.1) is satisfactory compared to the case of UCV method (Figure 4.2). In that case, we remark that the shape of all the estimators are nearly closed to the true RHR function in the case of lognormal distribution, however the BS-PE kernel does not show a good fit in the case of BS distribution.

The computer programs used in this simulation study are given in appendices.

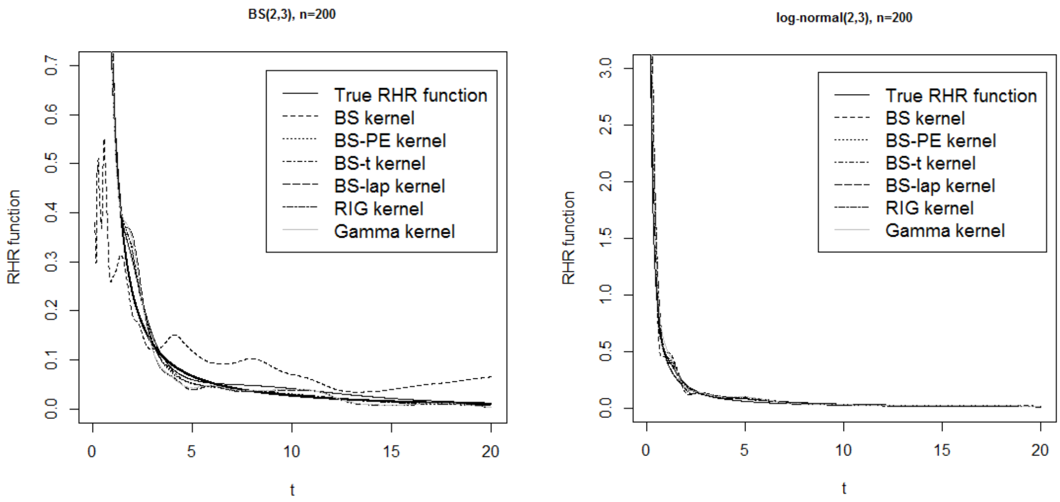

Figure 4.3: RHR function estimators for BS and log-normal models with $n = 200$, using the bandwidth h^{ρ}_{RT}.

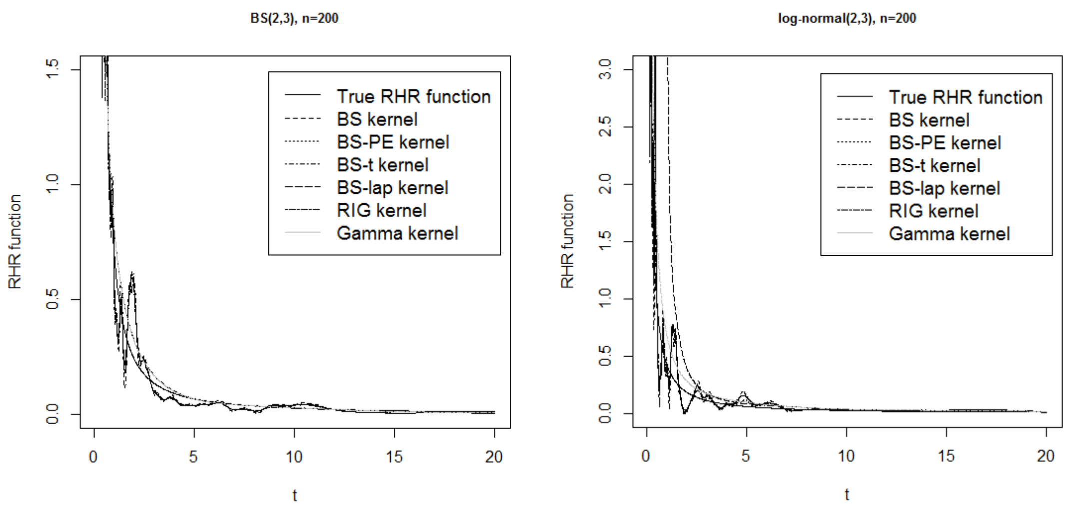

Figure 4.4: RHR function estimators for BS and log-normal models with $n = 200$, using the bandwidth h^{ρ}_{UCV}.

Size	Kernels \ Models	BS(2,3)		lognormal(2,3)		gamma(3,0.5)		GBS$(1, 2; t_{v=5})$	
		h_{RT}	h_{UCV}	h_{RT}	h_{UCV}	h_{RT}	h_{UCV}	h_{RT}	h_{UCV}
n=50	BS	0.02070	0.17251	**0.01671**	0.09621	**0.91893**	0.28370	0.34567	**0.07317**
	BS-PE($\nu = 2$)	0.04828	0.11557	0.10153	0.18185	2.23470	0.31090	0.34936	0.51407
	BS-t($\nu = 5$)	0.02401	0.12196	0.15655	0.08449	1.00151	0.35747	**0.02898**	0.33493
	BS-lap	0.02033	0.05954	0.02025	0.11642	2.03049	0.41204	0.04203	0.52510
	RIG	0.03469	1.85236	0.04166	1.41419	2.87615	**0.27142**	0.23496	0.41791
	gamma	**0.01572**	**0.03920**	0.02814	**0.04165**	3.1658	0.84553	0.14702	0.37401
n=200	BS	0.00793	0.03698	0.0946	**0.02515**	0.53728	**0.20392**	0.08654	**0.03767**
	BS-PE($\nu = 2$)	0.01488	0.034135	0.03635	0.07228	0.87553	0.28026	0.13980	0.27829
	BS-t($\nu = 5$)	0.00861	0.05142	0.08881	0.02597	**0.51246**	0.33595	**0.01038**	0.24630
	BS-lap	0.00741	0.04874	**0.00500**	0.02599	2.01892	0.41144	0.01081	0.20421
	RIG	0.01252	1.12219	0.2967	1.54882	0.77096	0.27010	0.01848	0.45590
	gamma	**0.00519**	**0.03365**	0.04053	0.04220	1.25365	0.48279	0.10377	0.23919
n=500	BS	0.00529	**0.014601**	0.08393	0.01140	0.26970	0.22698	0.03861	**0.02178**
	BS-PE($\nu = 2$)	0.00895	0.02343	0.02692	0.04576	0.50043	**0.16404**	0.08049	0.14724
	BS-t($\nu = 5$)	0.00567	0.03305	0.08026	**0.01016**	**0.24889**	0.41624	**0.00522**	0.09442
	BS-lap	0.00315	0.04954	**0.00201**	0.01337	1.99272	0.41559	0.00586	0.09379
	RIG	0.00617	1.86213	0.01515	1.51131	0.47761	0.33932	0.00894	0.4599
	gamma	**0.00252**	0.03356	0.03095	0.02887	0.55134	0.43483	0.07322	0.10695
n=1000	BS	0.00311	**0.01204**	0.00565	0.00750	**0.16547**	0.30002	0.02139	**0.01423**
	BS-PE($\nu = 2$)	0.00499	0.01807	0.02292	0.04913	0.29233	**0.25919**	0.00525	0.058807
	BS-t($\nu = 5$)	0.00324	0.03119	0.08185	0.00673	0.16894	0.42911	**0.00371**	0.06119
	BS-lap	0.003954	0.05954	**0.00145**	**0.00672**	1.80250	0.44969	0.00571	0.05487
	RIG	0.00462	1.85439	0.01242	1.08808	0.23610	0.33796	0.0056	0.39145
	gamma	**0.00136**	0.03771	0.02598	0.03869	0.31158	0.30410	0.05878	0.03962

Table 4.2: Some expected values of ISE for RHR function estimators, based on 100 replications for the considered models in simulation, using the bandwidths, h^{ρ}_{RT} and h^{ρ}_{UCV}.

Conclusion and perspectives

Our thesis deals with the nonparametric hazard rate function (HR) estimation in the case of complete data, with kernel method and by using the family of Generalized Birbaum-Sauders (GBS) kernels. The choice of this class of kernels is motivated by its good properties and flexibility.

Firstly, we have introduced the basic concepts of reliability theory, in particular the general properties of HR, reliability and reversed hazard rate RHR functions. Then, we have presented the kernel method with its two parameters; the kernel K and the bandwidth h. The kernel is chosen according to the support of the function to be estimated. In fact, when the unknown function has an unbounded support, the suitable kernel is the symmetric one, however, when the unknown function has a bounded support in $\mathbb{R}_+$, we use an asymmetric kernel, whose the method of construction is also displayed. In addition, details about the class of GBS kernel and different methods for selection bandwidth are given too. Afterwards, we gave an overview of some results of kernel estimation of HR function in both cases of complete and censored data, and introduced our proposed HR kernel estimator using the class of GBS kernel. Under some conditions, the convergence properties such as bias, variance and mean integrated squared error are established. In addition, we have proved that, the GBS-HR estimator is strongly consistent and asymptotically normal. The choice of bandwidth is investigated by rule of thumb (RT) and unbiased cross validation (UCV) approaches. The performance of the proposed estimator compared to RIG and gamma HR kernels estimators, and the comparison of the two bandwidth selection methods are illustrated by a simulation study and real applications. In the sense of integrated squared error (ISE), the GBS-HR estimator dominates the other HR estimators based in RIG and gamma kernels, and the bandwidth parameter obtained using the RT method outperforms the one obtained with UCV.

We have also conducted a study on the reliability and RHR functions with the class of GBS kernel. Asymptotic properties are investigated for each estimator and the bandwidth parameter is calculated using RT and UCV methods. Finally, simulation study is investigated to test the performance of GBS-Reliability and GBS-RHR estimators compared to RIG and gamma kernel estimators of reliability and RHR functions, respectively. In the sense of ISE criterion, we have noted globally that, the GBS-Reliability estimator performs well in the case of RT method compared to UCV method for BS, lognormal, BS-Student and gamma distributions, and it is also the case for GBS-RHR estimator excepting the gamma distribution.

From the three simulation studies that we have conducted for GBS-HR, GBS-Reliability and GBS-RHR, we consider that the class of GBS kernel can be a good candidate for estimating the HR, reliability and RHR for positively skewed data, by using the RT bandwidth selection method.

As a perspective, we can cite

- Bayesian approach in Generalized Birnbaum-Saunders kernel estimator of the hazard rate function.
- Application of bias correction technics for GBS hazard rate function estimator.
- Comparison study between the GBS hazard rate function estimator and that based on the $\check{C}$wik and Mielniczuk method.
- New kernel estimator of hazard rate function without using the ratio of the density function and the reliability function.
- Hazard rate function estimation in discrete case.

Appendix

4.5 Computer programs

The computer programs that we have used in our simulation study are implemented by using the R software. We illustrated above the computer program used for GBS-HR, GBS-Reliability and GBS-RHR estimators, that represent hazard rate, reliability and reversed hazard rate estimators using the class of GBS kernels. That is, in the case of rule of thumb (RT) and Unbiased Cross Validation (UCV) bandwidth selection methods. In all three cases, we take an example of BS distribution with BS kernel.

4.5.1 GBS-HR estimator

Case of RT method

```
n=50
X=rgbs(n, 2 , 3 , nu = 1.0, kernel = "normal")
###BS kernel###
noyau=function(u,x,h){
NBS=1/(2*sqrt(2*h*pi))*((1/sqrt(x*u))+sqrt(x/(u^3)))*exp((-1/(2*h))*((u/x)-2+(x/u)))}
###GBS density estimator###
ESTD=function(X,x,h){
ESTG=mean(noyau(X,x,h))}
###GBS Reliability estimator###
y=runif(1000,0,1)
E=0
```

```
ESTF=function(X,x,y,h){
for (j in 1:length(y)){
E[j]=ESTD(X,x*y[j],h)
E}
E2=1-x*mean(E)}
###################################### RT method for GBS-HR #################################
###Estimate the parameters a and b###
mlegbs(X, kernel = "normal")
a=mlegbs(X)$alphaEstimate
b=mlegbs(X)$betaEstimate
###pdf and cdf of BS distribution###
f=function(x){
ff=dgbs(x, a , b , nu = 1.0, kernel = "normal",log = FALSE)
}
Frep=function(x){FF=pgbs(x, a , b, nu = 1.0, kernel = "normal",lower.tail = TRUE, log.p = FALSE)
}
fprim=function(t){
fp=1/(2*a*b*sqrt(2*pi))*(3*b*sqrt(b/t)/(2*t^2)-b/(2*t^2*sqrt(b/t)))*exp(-1/(2*a^2)*(b/t+t/b-2))
-1/(4*b*a^3*sqrt(2*pi))*(1/b-b/(t^2))*(sqrt(b/t)+(b/t)^(3/2))*exp(-1/(2*a^2)*(b/t+t/b-2))
}
f2prim=function(t){
f2p=1/(2*a*b*sqrt(2*pi))*(-3*b^2/(4*t^4*sqrt(b/t))-b^2/(4*t^4*(b/t)^(3/2))-(3*b*sqrt(b/t)/t^3)
+b/(t^3*sqrt(b/t)))*exp(-1/(2*a^2)*(b/t+t/b-2))+1/(8*sqrt(2*pi)*b*a^5)*(sqrt(b/t)+(b/t)^(3/2))
*(1/b-b/t^2)^2*exp(-1/(2*a^2)*(b/t+t/b-2))-1/(2*sqrt(2*pi)*a^3*t^3)*((b/t)^(3/2)+sqrt(b/t))
*exp(-1/(2*a^2)*(b/t+t/b-2))-1/(4*sqrt(2*pi)*b*a^3)*(-3*b*sqrt(b/t)/(2*t^2)-b/(2*t^2*sqrt(b/t)))
*(1/b-b/t^2)*exp(-1/(2*a^2)*(b/t+t/b-2))-1/(4*sqrt(2*pi)*a^3*b)*(3*b*sqrt(b/t)/(2*t^2)
-b/(2*t^2*sqrt(b/t)))*(1/b-b/t^2)*exp(-1/(2*a^2)*(b/t+t/b-2))
}
u1g=1
c=1/sqrt(2*pi)
cg2=1/sqrt(pi)
###Numerator###
u1=runif(500,0,1)
num=0
for(i in 1:length(u1)){
num[i]=((c^2)*(10*u1[i])^(-1)*f(10*u1[i]))/(1-Frep(10*u1[i]))^2
```

```
num }
Mnum=10*mean(num)
###Denominator###
u2=runif(500,0,1)
deno=0
for(i in 1:length(u2)){
deno[i]=cg2*(u1g^2)*((10*u2[i]*fprim(10*u2[i])+((10*u2[i])^2)*f2prim(10*u2[i]))
/(1-Frep(10*u2[i])))^2
deno
Mdeno=10*mean(deno)
}
###Optimal h###
hplug=(Mnum/Mdeno)^(2/5)*n^(-2/5)
###GBS-HR estiamtor###
tauxdef=function(X,x,y,h){
ESTTD=(ESTD(X,x,hplug))/(ESTF(X,x,y,hplug))
ESTTD}
```

Case of UCV method

```
n=50
X=rgbs(n, 2 , 3 , nu = 1.0, kernel = "normal")
###BS kernel###
noyau=function(u,x,h){
NBS=1/(2*sqrt(2*h*pi))*((1/sqrt(x*u))+sqrt(x/(u^3)))*exp((-1/(2*h))*((u/x)-2+(x/u)))}
###GBS density estimator###
ESTD=function(X,x,h){
ESTG=mean(noyau(X,x,h))}
###GBS reliability estimator###
y=runif(1000,0,1)
E=0
ESTF=function(X,x,y,h){
for (j in 1:length(y)){
E[j]=ESTD(X,x*y[j],h)
E}
E2=1-x*mean(E)}
####################################UCV method for GBS-HR estimator ############################
```

```
###terme1###
ter=0
x=runif(100,0,1)
terme1=function(h){
for (j in 1:length(x)){
ter[j]=(ESTD(X,20*x[j],h)/ESTF(X,20*x[j],y,h))^2
ter}
tt=20*mean(ter)
}
#### terme2###
MM1=0
MM2=0
terme2=function(h){
M1=matrix(0,nrow=n, ncol=n)
M2=matrix(0,nrow=n, ncol=n)
for (i in 1:n){
M1[i, ]=(n^2)*ESTD(X,X[i],h)
M2[i, ]=1-ESTF(X,X[i],y,h)
diag(M1)=0
diag(M2)=0
MM1[i]=sum(M1[i,])
MM2[i]=(((n-1)*n^2)-sum(M2[i,]))^2 }
MM3=sum(MM1/MM2)
}
CV=function(h){
U=terme1(h)-((2*(n-1))/n)*terme2(h)
}
hcvopt=optimize(CV, c(0, 2), tol = 0.005, maximum = FALSE )
hucv=hcvopt$minimum
hucv
###GBS-HR estimator###
tauxdef=function(X,x,y,h){
ESTTD=(ESTD(X,x,hucv))/(ESTF(X,x,y,hucv))
ESTTD}
```

4.5.2 GBS-Reliability estimator

Case of RT method

```
n=50
X=rgbs(n, 2 , 3 , nu = 1.0, kernel = "normal")
### BS kernel ###
noyau=function(u,x,h){
NBS=1/(2*sqrt(2*h*pi))*((1/sqrt(x*u))+sqrt(x/(u^3)))*exp((-1/(2*h))*((u/x)-2+(x/u)))}
###GBS density estimator###
ESTD=function(X,x,h){
ESTG=mean(noyau(X,x,h))}
########################################## RT method #########################################
###Estimate the parameters a and b###
mlegbs(X, kernel = "normal")
a=mlegbs(X)$alphaEstimate
b=mlegbs(X)$betaEstimate
### pdf and cdf of BS distribution ###
f=function(x){
ff=dgbs(x, a , b , nu = 1.0, kernel = "normal",log = FALSE)
}
Frep=function(x){
FF=pgbs(x, a , b, nu = 1.0, kernel = "normal",lower.tail = TRUE, log.p = FALSE)
}
###Derivatives###
fprim=function(t){
fp=1/(2*a*b*sqrt(2*pi))*(3*b*sqrt(b/t)/(2*t^2)-b/(2*t^2*sqrt(b/t)))*exp(-1/(2*a^2)*(b/t+t/b-2))
-1/(4*b*a^3*sqrt(2*pi))*(1/b-b/(t^2))*(sqrt(b/t)+(b/t)^(3/2))*exp(-1/(2*a^2)*(b/t+t/b-2))
}
f2prim=function(t){
f2p=1/(2*a*b*sqrt(2*pi))*(-3*b^2/(4*t^4*sqrt(b/t))-b^2/(4*t^4*(b/t)^(3/2))-(3*b*sqrt(b/t)/t^3)
+b/(t^3*sqrt(b/t)))*exp(-1/(2*a^2)*(b/t+t/b-2))+1/(8*sqrt(2*pi)*b*a^5)*(sqrt(b/t)+(b/t)^(3/2))
*(1/b-b/t^2)^2*exp(-1/(2*a^2)*(b/t+t/b-2))-1/(2*sqrt(2*pi)*a^3*t^3)*((b/t)^(3/2)+sqrt(b/t))
*exp(-1/(2*a^2)*(b/t+t/b-2))-1/(4*sqrt(2*pi)*b*a^3)*(-3*b*sqrt(b/t)/(2*t^2)-b/(2*t^2*sqrt(b/t)))
*(1/b-b/t^2)*exp(-1/(2*a^2)*(b/t+t/b-2))-1/(4*sqrt(2*pi)*a^3*b)*(3*b*sqrt(b/t)/(2*t^2)
-b/(2*t^2*sqrt(b/t)))*(1/b-b/t^2)*exp(-1/(2*a^2)*(b/t+t/b-2))
}
```

```
u1g=1
c=1/sqrt(2*pi)
cg2=1/sqrt(pi)
###Numerator###
u1=runif(500,0.01,1)
num=0
for(i in 1:length(u1)){
num[i]=((c^2)*(10*u1[i])^(-1)*(1-Frep(10*u1[i]))^2)/f(10*u1[i])
num }
Mnum=10*mean(num)
###Denominator###
deno=0
for(i in 1:length(u1)){
deno[i]=cg2*(u1g^2)*((1-Frep(10*u1[i]))*(10*u1[i]*fprim(10*u1[i])+((10*u1[i])^2)
*f2prim(10*u1[i]))/f(10*u1[i]))^2
deno
Mdeno=10*mean(deno)
}
###Optimal h###
hplug=(Mnum/Mdeno)^(2/5)*n^(-2/5)
###GBS-Reliability estimator###
y=runif(1000,0,1)
E=0
ESTF=function(X,x,y,h){
for (j in 1:length(y)){
E[j]=ESTD(X,x*y[j],hplug)
E}
E2=1-x*mean(E)}
```

Case of UCV method

```
n=50
X=rgbs(n, 2 , 3 , nu = 1.0, kernel = "normal")
###BS kernel###
noyau=function(u,x,h){
NBS=1/(2*sqrt(2*h*pi))*((1/sqrt(x*u))+sqrt(x/(u^3)))*exp((-1/(2*h))*((u/x)-2+(x/u)))}
### GBS density estimator ###
```

```
ESTD=function(X,x,h){
ESTG=mean(noyau(X,x,h))}
###UCV method for GBS-Reliability estimator###
####terme1#####
ter=0
x=runif(100,0,1)
terme1=function(h){
for (j in 1:length(x)){
ter[j]=ESTF(X,20*x[j],y,h)^2
ter}
tt=20*mean(ter)
}
#### terme2###
MM1=0
MM2=0
terme2=function(h){
M1=matrix(0,nrow=n, ncol=n)
M2=matrix(0,nrow=n, ncol=n)
for (i in 1:n){
M1[i, ]= (n-1)-n*ESTF(X,X[i],y,h)
M2[i, ]=n*ESTD(X,X[i],h)
diag(M1)=0
diag(M2)=0
MM1[i]=((1/(n-1))sum(M1[i,]))^2
MM2[i]=(1/(n-1))*sum(M2[i,]) }
MM3=sum(MM1/MM2)
}
CV=function(h){
U=terme1(h)-(2/n)*terme2(h)
}
hcvopt=optimize(CV, c(0, 2), tol = 0.005, maximum = FALSE )
hucv=hcvopt$minimum
hucv
###GBS-Reliability estimator###
y=runif(1000,0,1)
E=0
```

```
ESTF=function(X,x,y,h){
for (j in 1:length(y)){
E[j]=ESTD(X,x*y[j],hucv)
E}
E2=1-x*mean(E)}
```

4.5.3 GBS-RHR estimator

Case of RT method

```
n=50
X=rgbs(n, 2 , 3 , nu = 1.0, kernel = "normal")
### BS kernel ###
noyau=function(u,x,h){
NBS=1/(2*sqrt(2*h*pi))*((1/sqrt(x*u))+sqrt(x/(u^3)))*exp((-1/(2*h))*((u/x)-2+(x/u)))}
###GBS density estimator###
ESTD=function(X,x,h){
ESTG=mean(noyau(X,x,h))}
###GBS kernel estimator of cdf###
y=runif(1000,0,1)
E=0
ESTF=function(X,x,y,h){
for (j in 1:length(y)){
E[j]=ESTD(X,x*y[j],h)
E}
E2=x*mean(E)}
########################################### RT method #######################################
###Estimate the parameters a and b###
mlegbs(X, kernel = "normal")
a=mlegbs(X)$alphaEstimate
b=mlegbs(X)$betaEstimate
###pdf and cdf of BS distribution###
f=function(x){
ff=dgbs(x, a , b , nu = 1.0, kernel = "normal",
     log = FALSE)
```

```
}
Frep=function(x){
FF=pgbs(x, a , b, nu = 1.0, kernel = "normal",
     lower.tail = TRUE, log.p = FALSE)
}
###Derivatives###
fprim=function(t){
fp=1/(2*a*b*sqrt(2*pi))*(3*b*sqrt(b/t)/(2*t^2)-b/(2*t^2*sqrt(b/t)))*exp(-1/(2*a^2)*(b/t+t/b-2))-
1/(4*b*a^3*sqrt(2*pi))*(1/b-b/(t^2))*(sqrt(b/t)+(b/t)^(3/2))*exp(-1/(2*a^2)*(b/t+t/b-2))
}
f2prim=function(t){
f2p=1/(2*a*b*sqrt(2*pi))*(-3*b^2/(4*t^4*sqrt(b/t))-b^2/(4*t^4*(b/t)^(3/2))-(3*b*sqrt(b/t)/t^3)
+b/(t^3*sqrt(b/t)))*exp(-1/(2*a^2)*(b/t+t/b-2))+1/(8*sqrt(2*pi)*b*a^5)*(sqrt(b/t)+(b/t)^(3/2))
*(1/b-b/t^2)^2*exp(-1/(2*a^2)*(b/t+t/b-2))-1/(2*sqrt(2*pi)*a^3*t^3)*((b/t)^(3/2)+sqrt(b/t))
*exp(-1/(2*a^2)*(b/t+t/b-2))-1/(4*sqrt(2*pi)*b*a^3)*(-3*b*sqrt(b/t)/(2*t^2)-b/(2*t^2*sqrt(b/t)))
*(1/b-b/t^2)*exp(-1/(2*a^2)*(b/t+t/b-2))-1/(4*sqrt(2*pi)*a^3*b)*(3*b*sqrt(b/t)/(2*t^2)
-b/(2*t^2*sqrt(b/t)))*(1/b-b/t^2)*exp(-1/(2*a^2)*(b/t+t/b-2))
}
u1g=1
c=1/sqrt(2*pi)
cg2=1/sqrt(pi)
###numérateur###
u1=runif(500,0,1)
num=0
for(i in 1:length(u1)){
num[i]=((c^2)*(10*u1[i])^(-1)*f(10*u1[i]))/(Frep(10*u1[i]))^2
num }
Mnum=10*mean(num)
###Dénominateur###
u2=runif(500,0,1)
deno=0
for(i in 1:length(u2)){
deno[i]=cg2*(u1g^2)*((10*u2[i]*fprim(10*u2[i])+((10*u2[i])^2)*f2prim(10*u2[i]))/(Frep(10*u2[i])))^2
deno
Mdeno=10*mean(deno)
}
```

```
###Optimal h###
hplug=(Mnum/Mdeno)^(2/5)*n^(-2/5)
###GBS-RHR estimator###
tauxdef=function(X,x,y,h){
ESTTD=(ESTD(X,x,hplug))/(ESTF(X,x,y,hplug))
ESTTD}
```

Case of UCV method

```
n=50
X=rgbs(n, 2 , 3 , nu = 1.0, kernel = "normal")
###BS kernel###
noyau=function(u,x,h){
NBS=1/(2*sqrt(2*h*pi))*((1/sqrt(x*u))+sqrt(x/(u^3)))*exp((-1/(2*h))*((u/x)-2+(x/u)))}
###GBS density estimator###
ESTD=function(X,x,h){
ESTG=mean(noyau(X,x,h))}
###GBS kernel estimator of cdf###
y=runif(500,0,1)
E=0
ESTF=function(X,x,y,h){
for (j in 1:length(y)){
E[j]=ESTD(X,x*y[j],h)
E}
E2=x*mean(E)}
####################################UCV method for GBS-HR estimator ##############################
####terme1#####
ter=0
x=runif(100,0,1)
terme1=function(h){
for (j in 1:length(x)){
ter[j]=(ESTD(X,20*x[j],h)/ESTF(X,20*x[j],y,h))^2
ter}
tt=20*mean(ter)
}
#### terme2###
MM1=0
```

```
MM2=0
terme2=function(h){
M1=matrix(0,nrow=n, ncol=n)
M2=matrix(0,nrow=n, ncol=n)
for (i in 1:n){
M1[i, ]=n*ESTD(X,X[i],h)
M2[i, ]=n*ESTF(X,X[i],y,h)
diag(M1)=0
diag(M2)=0
MM1[i]=(1/(n-1))*sum(M1[i,])
MM2[i]=((1/(n-1))*sum(M2[i,]))^2 }
MM3=sum(MM1/MM2)
}
CV=function(h){
U=terme1(h)-((2)/n)*terme2(h)
}
hcvopt=optimize(CV, c(0, 3), tol = 0.005, maximum = FALSE )
hucv=hcvopt$minimum
hucv
###GBS-RHR estimator###
tauxdef=function(X,x,y,h){
ESTTD=(ESTD(X,x,hucv))/(ESTF(X,x,y,hucv))
ESTTD}
```

4.6 Real data

Data S1

12, 15, 22, 24, 24, 32, 32, 33, 34, 38, 38, 43, 44, 48, 52, 53, 54, 54, 55, 56, 57, 58, 58, 59, 60, 60, 60, 60, 61, 62, 63, 65, 65, 67, 68, 70, 70, 72, 73, 75, 76, 76, 81, 83, 84, 85, 87, 91, 95, 96, 98, 99, 109, 110, 121, 127, 129, 131, 143, 146, 146, 175, 175, 211, 233, 258, 258, 263, 297, 341, 341, 376.

Data S2

0.010, 0.059, 0.089, 0.121, 0.142, 0.081, 0.127, 0.194, 0.071, 0.047, 0.071, 0.097, 0.166,

0.134, 0.553, 0.709, 0.183, 0.047, 0.111, 0.072, 0.074, 0.073, 0.070, 0.190, 0.128, 0.170, 0.185, 0.122, 0.194, 0.445, 0.169, 0.134, 0.081, 0.991, 0.300, 0.139, 0.098, 0.092, 0.040, 0.164, 0.030, 0.240, 0.070, 0.085, 0.150, 0.160, 0.110, 0.130, 0.170, 0.150, 0.092, 0.670, 0.029, 0.077, 0.073, 0.120, 0.071, 0.096, 0.092, 0.092, 0.110, 0.130, 0.064, 0.041, 0.070, 0.050, 0.050, 0.039, 0.047, 0.075, 0.110, 0.100, 0.170, 0.110, 0.130, 0.140, 0.059, 0.150, 0.099, 0.081, 0.096, 0.091, 0.150, 0.120, 0.160, 0.091, 0.130, 0.310, 0.041, 0.031, 0.042, 0.045, 0.048, 0.039, 0.050, 0.075, 0.110, 0.160, 0.690, 0.140, 0.140, 0.100, 0.062, 0.093.

Data S3

109.00, 253.26, 529.40, 624.38, 878.37, 1000.00, 1005.42, 1018.42, 1325.00, 1350.00, 1500.00, 1570.82, 1727.03, 1750.00, 1800.00, 1800.00, 1838.00, 1994.00, 2000.00, 2000.00, 2000.00, 2000.00, 2000.00, 2118.15, 2242.32, 2265.15, 2269.09, 2382.61, 2397.92, 2444.78, 2484.25, 2500.00, 2500.00, 2500.00, 2510.12, 2521.00, 2551.25, 2598.10, 2652.85, 2672.25, 2780.55, 2915.07, 2920.00, 2932.40, 3000.00, 3000.00, 3000.00, 3000.00, 3000.00, 3000.00, 3000.00, 3001.00, 3015.00, 3100.00, 3148.00, 3211.00, 3250.00, 3251.70, 3323.00, 3382.64, 3455.00, 3480.30, 3480.30, 3493.50, 3500.00, 3500.00, 3500.00, 3503.00, 3515.80, 3556.32, 3579.36, 3581.50, 3591.00, 3658.50, 3666.00, 3666.00, 3718.50, 3750.00, 3750.00, 3800.00, 3878.00, 3892.00, 3897.00, 3900.00, 3976.00, 3994.88, 4000.00, 4000.00, 4000.00, 4000.00, 4000.00, 4000.00, 4000.00, 4000.00, 4020.00, 4020.18, 4075.85, 4146.00, 4151.85, 4248.30, 4250.00, 4250.00, 4257.45, 4269.00, 4285.00, 4300.24, 4346.00, 4358.00, 4380.77, 4410.06, 4472.85, 4476.00, 4476.00, 4500.00, 4500.00, 4500.00, 4500.00, 4530.00, 4550.00, 4581.10, 4590.00, 4650.00, 4672.10, 4676.35, 4691.16, 4691.50, 4699.00, 4700.00, 4717.00, 4729.45, 4746.00, 4751.41, 4774.50, 4908.00, 4920.01, 4923.36, 4938.00, 4989.50, 4991.75, 4999.05, 5000.00, 5000.00, 5000.00, 5000.00, 5000.00, 5000.00, 5000.00, 5000.00, 5000.00, 5000.00, 5000.00, 5000.00, 5000.00, 5000.00, 5000.00, 5000.00, 5000.00, 5000.00, 5000.11, 5002.21, 5021.00, 5022.45, 5025.00, 5025.00, 5026.00, 5039.00, 5045.35, 5064.45, 5091.83, 5098.00, 5098.52, 5099.60, 5119.00, 5136.11, 5136.19, 5144.15, 5171.80, 5208.92, 5275.00, 5275.00, 5303.45, 5309.10, 5372.30, 5379.15, 5381.51, 5388.00, 5393.95, 5428.00, 5444.15, 5470.22, 5497.00, 5500.00, 5500.00, 5500.00, 5500.00, 5500.00, 5500.00, 5506.00, 5522.05, 5527.50, 5527.50, 5541.00, 5543.00, 5544.90, 5575.74, 5598.00, 5600.00, 5600.00, 5606.15, 5625.00, 5628.75, 5650.00, 5664.50, 5792.21, 5793.17, 5882.90, 5929.54, 5955.00, 5955.00, 5963.20, 5982.53, 6000.00, 6000.00, 6000.00, 6000.00, 6000.00, 6000.00, 6000.00, 6021.00, 6033.17, 6038.00, 6078.50, 6113.00, 6149.45, 6154.70, 6164.47, 6179.49, 6202.50, 6248.35, 6295.30, 6300.00, 6316.50, 6326.50, 6333.00, 6342.15, 6400.00, 6424.00, 6458.05, 6460.00, 6500.00, 6500.00, 6507.65, 6524.56, 6548.33, 6550.00, 6553.00, 6602.50, 6607.15, 6631.40, 6647.15, 6670.50, 6673.30, 6673.50, 6695.65, 6700.00, 6704.35, 6713.71, 6719.50, 6724.50, 6750.00, 6750.00, 6780.50, 6791.82, 6795.50, 6798.67, 6800.00, 6800.00, 6810.00, 6815.15, 6874.25, 6875.15, 6887.40, 6901.60, 6920.35, 6933.49, 6958.50, 6958.97, 7000.00, 7000.00, 7000.00, 7005.05, 7022.65, 7038.93, 7055.51, 7063.00, 7073.50, 7074.55, 7097.00, 7099.50, 7101.10, 7141.02, 7213.50, 7248.36, 7293.35, 7303.60, 7338.10, 7384.95, 7448.50, 7460.00, 7490.50, 7494.50, 7496.88, 7500.00, 7500.00, 7500.00, 7500.00, 7500.00, 7527.30, 7572.17, 7642.15, 7642.26, 7666.65, 7670.05, 7673.00, 7675.20, 7693.56, 7704.79, 7762.35, 7765.00, 7781.36, 7791.32, 7802.55, 7827.01, 7850.95, 7890.39, 7915.30, 8000.00, 8000.00, 8001.67, 8043.15, 8070.30, 8104.00, 8117.84, 8135.00, 8168.90, 8171.12, 8172.50, 8212.83, 8225.66, 8229.40, 8296.95, 8301.51, 8331.20, 8375.10, 8437.35, 8500.00, 8500.00, 8500.00, 8500.00, 8564.00, 8637.40, 8639.80, 8706.32, 8742.00, 8782.55, 8828.56, 8831.05, 8855.00, 8908.40, 8999.75, 9000.00, 9009.05, 9061.50, 9084.00, 9138.90, 9152.10, 9163.74, 9227.21, 9248.35, 9314.08, 9319.67, 9326.95, 9354.90, 9423.05, 9437.85, 9456.65, 9467.75, 9481.94, 9520.14, 9646.50, 9664.62, 9665.85, 9742.50, 9805.10, 9839.50, 9900.00, 9950.00, 9990.00, 10000.00, 10000.00, 10000.00, 10000.00, 10000.00, 10061.21, 10081.15, 10152.40, 10156.65, 10186.92, 10213.00, 10213.55, 10315.40, 10329.00, 10342.86, 10344.90, 10356.32, 10395.38, 10457.00, 10687.42, 10746.16, 10854.90, 10979.39, 11000.00, 11000.00, 11056.34, 11133.30, 11245.32, 11250.00, 11363.00, 11377.15, 11500.00, 11500.00, 11512.49, 11850.50, 11860.92, 11874.50, 11878.95, 11949.74, 12000.00, 12000.00, 12000.00, 12000.00, 12140.00, 12310.25, 12358.00, 12417.00, 12555.35, 12783.42, 12804.25, 12853.15, 13000.00, 13000.00, 13184.65, 13321.14, 13386.93, 13500.00, 13592.41, 13688.55, 13715.24, 13946.00, 14000.00, 14000.00, 14138.89, 14336.51, 14546.60, 14709.60, 14725.81, 14736.74, 14909.15, 14950.00, 14958.35, 14979.07, 15000.00, 15000.00, 15000.00, 15060.15, 15066.15, 15104.00, 15128.32, 15163.10, 15174.50, 15268.64, 15460.00, 15466.15, 15591.00, 15654.21, 16000.00, 16038.25, 16149.55, 16164.15, 16368.50, 16999.90, 17000.00, 17215.00, 17708.09, 17734.58, 17901.50, 17904.38, 17977.10, 18286.00, 18358.00, 18707.16, 18849.90, 18872.70, 19000.00, 19500.00,19839.78, 19912.57, 20000.00, 20000.00, 20000.00, 20245.58, 21200.00, 21241.09, 21450.00, 21502.25, 22175.00, 22400.00, 22575.50, 22800.00, 23535.88, 24135.00, 24435.00, 24495.15, 25000.00, 26275.83, 29000.00, 29279.00, 30579.89, 30732.68, 31207.99, 32500.00, 32691.00, 33000.00, 33796.00, 34465.40, 37413.25, 37806.20, 43600.00, 47446.50, 67750.00, 76255.76, 116586.72.

Bibliography

[1] Aitchison, J. and Aitken, C. G. G. 1976. Nonparametric density estimation for multivariate bounded data. *Multivariate binary discrimination by the kernel method* 63:413–420.

[2] Altun, M. and Comert, S. V. 2016. A change-point based reliability prediction model using field return data. *Reliability Engineering and System Safety* 156:175–184.

[3] Andersen, P. K., Borgan, O. and Gill, R. D. 1993. Statistical methods based on counting processes. *New York: Springer.*

[4] Athayde, E., Azevedo, A., Barros, M. and Leiva, V. 2019. Failure rate of Birnbaum-Saunders distributions: Shape, change-point, estimation and robustness. *Brazilian Journal of Probabilty and Statistics* 33:301-328.

[5] Block, H., Savits, T. H. and Singh, H. 1998. The reversed hazard rate function. *Probability in Engineering and Informational Sciences* 12:69–90.

[6] Bouezmarni, T. and Rolin, J. M. 2003. Consistency of the beta kernel density function estimator. *Canadian Journal of Statistics* 31:89–98.

[7] Bouezmarni, T. and Rombouts, J. V. K. 2008. Density and hazard rate estimation for censored and α-mixing data using gamma kernels. *Journal of Nonparametric Statistics* 20:627-643.

[8] Bouezmarni, T. and Rombouts, J. V. K. 2010. Nonparametric density estimation for multivariate bounded data. *Journal of Statistical Planning and Inference* 140:139-152.

[9] Bouezmarni, T. and Scaillet, O. 2005. Consistency of asymmetric kernel density estimators and smoothed histograms with application to income data. *Econometric Theory* 21:390–412.

[10] Bouezmarni, T., El Ghouch, A. and Mesfioui, M. 2011. Gamma kernel estimators for density and hazard rate of right-censored data. *Journal of Probability and Statistics*. doi: 10.1155/2011/937574.

[11] Bowman, A. W. 1984. An Alternative Method of Cross-validation for the Smoothing of Density Estimates. *Biometrika* 71:353–360.

[12] Brazzale, A. R., Küchenhoff, H., Krügel, S., Schiergens, T. S., Trentzsch, H. and Hartl, W. 2018. Nonparametric change point estimation for survival distributions with a partially constant hazard rate. *Lifetime data analysis* 25:301-321.

[13] Brewer, M. J. 1998. A modelling approach for bandwidth selection in kernel density estimation. *In : Proceedings of COMPSTAT Physica Verlag, Heidelberg* 203–208.

[14] Brewer, M. J. 2000. A bayesian model for local smoothing in kernel density estimation. *Statistics and Computing* 10:299–309.

[15] Brunel, E., Comte, F. and Genon-Catalot, V. 2016. Nonparametric density and survival function estimation in the multiplicative censoring model. *Test* 25: 570-590.

[16] Chandra, N. K. and Roy, D. 2001. Some results on reversed hazard rate. *Probability in the Engineering and Informational Sciences* 15:95-102.

[17] Chandra, N. K. and Roy, D. 2005. Classification of distributions based on reversed hazard rate. *Calcutta Statistical Association Bulletin.* 56:231-250.

[18] Chekkal, S., Lagha, K. and Zougab, N. 2021. Generalized Birnbaum–Saunders kernel for hazard rate function estimation. *Communications in Statistics-Simulation and Computation.* doi: 10.1080/03610918.2021.1887228.

[19] Chen, S. X. 1999. Beta kernel estimators for density functions. *Computational Statistics and Data Analysis* 31:131-145.

[20] Chen, S. X. 2000. Gamma kernel estimators for density functions. *Annals of the Institute of Statistical Mathematics* 52:471-480.

[21] Chen, S. M., Hsu, Y. S. and Liaw, J. T. 2009. On kernel estimators of density ratio. *Statistics* 43:463-479.

[22] Cheng, M.Y. 1997. Boundary—aware estimators of integrated density derivative. *Journal of the Royal Statistical Society: Series B* 59:191–203.

[23] Cline, D. B. H. and Hart, J. D. 1991. Kernel estimation of densities of discontinuous derivatives. *Statistics* 22:69–84.

[24] Cowling, A. and Hall, P. 1996. On pseudodata methods for removing boundary effects in kernel density estimation. *Journal of the Royal Statistical Society: Series B* 58:551–563.

[25] de Montricher, G. F., Tapia, R. A. and Thompson, J. R. 1975. Nonparametric maximum likelihood estimation of probability densities by penalty function methods. *The Annals of Statistics* 3:1329-1348.

[26] Desai, D., Mariappan, V. and Sakhardande, M. 2011. Nature of reversed hazard rate: An investigation. *International Journal of Performability Engineering* 7:165-171.

[27] Epanechnicov, E. A. 1969. Nonparametric estimation of a multidimensional probability density. *Theory. Probab. Appl*, 14:153–158.

[28] Fang, K. T., Kotz, S. and Ng, K. W. 1990. Symmetric Multivariate and Related Distributions. *Chapman and Hall*. London.

[29] Ferraty, F. and Vieu, P. 2003. Curves discrimination : a nonparametric functional approach. special issue in honour of stan azen : a birthday celebration. *Computational Statistics and Data Analysis* 44:161–173.

[30] Finkelstein, M. S. 2002 . On the reversed hazard rate. *Reliability Engineering and System Safety* 78:71-75.

[31] Finkelstein, M. S. 2008. Failure rate modelling for reliability and risk. *Springer Science and Business Media.*

[32] Jin, X. and J. Kawczak. 2003. Birnbaum-Saunders and lognormal kernel estimators for modelling durations in high frequency financial data. *Annals of Economics and Finance* 4:103-124.

[33] Jones, M. 1993. Simple boundary correction for kernel density estimation. *Statistics and Computing* 3:135–146.

[34] Jones, M. and Henderson, D. A. 2007. Kernel-type density estimation on the unit interval. *Biometrika* 94:977-984.

[35] Jones, M. C., Marron, J. S. and Sheather, S. J. 1996. A brief survey of bandwidth selection for density estimation. *Journal of the American Statistical Association* 91:401–407.

[36] Hall, P. and Park, B. U. 2002. New methods for bias correction at endpoints and boundaries. *The annals of Statistics* 30:1460–1479.

[37] Härdle, W. 1990. Applied Nonparametric Regression. *Cambridge University Press, UK*

[38] Hirukawa, M. 2018. Asymmetric Kernel Smoothing: Theory and Applications in Economics and Finance. *Springer.*

[39] Hirukawa, M. and Sakudo, M. 2014. Nonnegative bias reduction methods for density estimation using asymmetric kernels. *Computational Statistics and Data Analysis* 75:112-123.

[40] Igarashi, G. 2016. Weighted log-normal kernel density estimation. *Communications in Statistics-Theory and Methods* 45:6670-6687.

[41] Igarashi, G. 2020. Nonparametric direct density ratio estimation using beta kernel. *Statistics* 54:257-280.

[42] Gimiz, M. L., Kulasekera, K. B., Limnios, N. and Lindqvist, B. H. 2011. Applied nonparametric statistics in reliability. *Springer Science and Business Media.*

[43] Gu, C. and Qiu, C. 1993. Smoothing spline density estimation: Theory. *The Annals of Statistics* 217-234.

[44] Gupta, R. D. and Nanda, A. K. 2001. Some results on reversed hazard rate ordering. *Communications in Statistics-Theory and Methods* 30:2447-2457.

[45] Kakizawa, Y. 2018. Nonparametric density estimation for nonnegative data, using symmetrical-based inverse and reciprocal inverse Gaussian kernels through dual transformation. *Journal of Statistical Planning and Inference* 193:117-135.

[46] Kaplan, E. L. and Meier, P. 1958. Nonparametric estimation from incomplete observations. *Journal of the American statistical association* 53:457-481.

[47] Karunamuni, R.J. and Alberts, T. 2006. A locally adaptive transformation method of boundary correction in kernel density estimation. *Journal of Statistical Planning and Inference* 136:2936–2960.

[48] Keilson, J. and Sumita, U. 1982. Uniform stochastic ordering and related inequalities. *Canadian Journal of Statistics* 10:181–98.

[49] Kiessé, S. 2008. Approche non-paramétrique par noyaux associés discrets des données de d´enombrement. *Doctoral thesis* University of Pau, France.

[50] Kissell, R. and Poserina, J. 2017. Optimal sports math, statistics, and fantasy. *(*Academic Press).

[51] Kokonendji, C. C. and Libengué, F. G. B. 2018. Asymptotic Results For Continuous Associated Kernel Estimators of Density Functions. *African Diaspora Journal of Mathematics. New Series* 21:87-97.

[52] Kokonendji, C. C. and Somé, S. M. 2018. On multivariate associated kernels to estimate general density functions. *Journal of the Korean Statistical Society* 47:112-126.

[53] Kundu, D., Kannan, N. and Balakrishnan, N. 2008. On the hazard function of Birnbaum Saunders distribution and associated inference. *Computational Statistics and Data Analysis* 52:2692-2702.

[54] Lai, C. D. and Xie, M. 2006. Stochastic ageing and dependence for reliability. *Springer Science and Business Media.*

[55] Lejeune, M. and Sarda, P. 1992. Smooth estimators of distribution and density functions. *Computational Statistics and Data Analysis* 14:457–471.

[56] Libengué, F. G. B. 2013. Méthodes non paramétriques des noyaux associés mixtes et applications. *PhD thesis* University of Franche-Comté (France) and University of Ouagadougou (Burkina-Faso).

[57] Libengué, F. G. B. and Kokonendji, C. C. 2017. The mode-dispersion approach for constructing continuous associated kernels. *Afrika Statistika* 12:1417-1446.

[58] Marchant, C., Bertin, K., Leiva, V. and Saulo, H. 2013. Generalized Birnbaum-Saunders kernel density estimators and an analysis of financial data. *Computational Statistics and Data Analysis* 63:1-15.

[59] Marron, J. and Ruppert, D. 1994. Transformation to reduce boundary bias in kernel density estimation. *Journal of the Royal Statistical Society: Series B* 56:653–671.

[60] Marshall, A. W. and Olkin, I. 2007. Life distributions. Vol. 13 of *Springer*. New York.

[61] Moriyama, T. and Maesono, Y. 2018. New kernel estimators of the hazard ratio and their asymptotic properties. *Annals of the Institute of Statistical Mathematics* 72:187-211.

[62] Nadaraya, E. A. 1964. On estimating regression. *Theory of Probability and its Applications* 9:141–142.

[63] Nelson, W. 2000. Theory and applications of hazard plotting for censored failure data. *Technometrics* 42:12-25.

[64] O'Connor, A. N. 2011. Probability distributions used in reliability engineering. RiAC. ISBN: 978-0-9966468-1-9

[65] Parzen, E. 1962. On Estimation of a Probability Density Function. *The Annals of Mathematical Statistics* 33:1065-1076.

[66] Paula, G. A., Leiva, V., Barros, M. and Liu, S. 2012. Robust statistical modeling using the Birnbaum-Saunders-t distribution applied to insurance. *Applied Stochastic Models in Business and Industry* 28:16-34.

[67] Rinne, H. 2008. The Weibull Distribution: A Handbook 1st ed. *Chapman and Hall/CRC.*

[68] Rosenblatt, M. 1956. Remarks on some nonparametric estimates of a density function. The Annals of Mathematical Statistics 832-837.

[69] Rudemo, M. 1982. Empirical choice of histograms and kernel density estimators. *ScandinavianJournal of Statistics* 9:65–78.

[70] Salha, R. B. 2012. Hazard rate function estimation using inverse Gaussian kernel. *The Islamic university of Gaza journal of Natural and Engineering studies* 20:73-84.

[71] Salha, R. B. 2013. Estimating the Density and Hazard Rate Functions Using the Reciprocal Inverse Gaussian Kernel. *Proceedings of the 15th Applied Stochastic Models and Data Analysis (ASMDA2013)*, 25-28. Mataro (Barcelona). .

[72] Salha, R. B., El Shekh Ahmed, H. I. and Alhoubi, I. M. 2014. Hazard Rate Function Estimation Using Weibull Kernel. *Open Journal of Statistics* 4:650-661.

[73] Salha, R. B., El Shekh Ahmed, H. I. and Alhoubi, I. M. 2014. Hazard rate function estimation using Erlang kernel. *Pure Mathematical Sciences* 3:141-152.

[74] Scaillet, O. 2004. Density estimation using inverse and reciprocal inverse Gaussian kernels. *Journal of Nonparametric Statistics* 16:217-226.

[75] Scott, D. W. and Terrel, G. R. 1987. Biased and unbiased cross-validation in density estimation. *Journal of the American Statistical Association* 82:1131–1146.

[76] Silverman, B. W. 1986. Density Estimation for Statistics and Data Analysis. *Monographs on Statistics and Applied Probability, Chapman and Hall.* New York.

[77] Srivastava, A. K. 2020. Estimation of Parameters and Reliability Function of Log Gompertz Model: Bayesian Approach under Gamma Prior. *International Journal of Advanced Research in Science, Communication and Technology* 10:183-195.

[78] Tiago de Oliviera, J. 1963. Estatistica de densidades : resultados assintoticos. *Revista de la Facultad de Ciencias, Universidad de Lisboa* A9:111-206.

[79] Veres-Ferrer, E. J. and Pavía, J. M. 2014. On the relationship between the reversed hazard rate and elasticity. *Statistical Papers* 55:275-284.

[80] Vilca, F., Romeiro, R. G. and Balakrishnan, N. 2016. A bivariate Birnbaum–Saunders regression model. *Computational Statistics and Data Analysis* 97:169-183.

[81] Xie, M., Goh, T. N. and Tang, Y. 2004. On changing points of mean residual life and failure rate function for some generalized Weibull distributions. *Reliability Engineering and System Safety* 84:293–299.

[82] Zhang, S. and Karunamuni, R. J. 2000. On nonparametric density estimation at the boundary. *Journal of Nonparametric Statistics* 12:197–221.

[83] Zhang, S. and Karunamuni, R. J. 1998. On kernel density estimation near endpoints. *Journal of Statistical Planning and Inference* 70:301–316.

[84] Zhang, X., King, M. L. and Hyndman, R. J. 2006. A bayesian approach to bandwidth selection for multivariate kernel density estimation. *Computational Statistics and Data Analysis* 50:3009–3031.

[85] Ziane, Y. 2015. Sur l'estimation non paparémtrique de l'indice de variabilité et la distribution des densités heavy tailed. *Doctoral thesis*, University of Bejaia, Algeria.

[86] Ziane, Y., Adjabi, S. and Zougab, N. 2015. Adaptive bayesian bandwidth selection in asymmetric kernel density estimation for nonnegative heavy-tailed data. *Journal of Applied Statistics* 42:1645–1658.

[87] Zougab, N. 2013. Approche bayésienne dans l'estimation non paramétrique de la densité de probabilité et la courbe de régression de la moyenne. *Thesis, Univ. A. Mira of Bejaïa.*

[88] Zougab, N., Adjabi, S., and Kokonendji, C. C. 2013. A bayesian approach to bandwidth selection in univariate associate kernel estimation. *Journal of Statistical Theory and Practice* 7:8–23.

Abstract

The main objective of this thesis is to propose the nonparametric kernel method for the hazard rate (HR) function estimation in the context of positively skewed data. The class of generalized Birnbaum-Saunders (GBS) kernels is considered because of its several interesting properties and flexibility. Some asymptotic properties, such as bias, variance and mean integrated squared error (MISE) are established for the proposed estimator. In addition, we demonstrate that, the GBS-HR estimator is strongly consistent and asymptotically normal. The choice of bandwidth is also investigated by rule of thumb and unbiased cross validation approaches. Finally, performances of the HR estimator based on GBS kernels and comparison of the two bandwidth selection methods are illustrated by a simulation study and real applications.

Résumé

L'objectif principal de cette thèse est de proposer la méthode non paramétrique de noyaux pour l'estimation de la fonction de hazard (HR) dans le contexte de données positives et asymétriques. La classe des noyaux de Birnbaum-Saunders généralisés (GBS) est considérée en raison de ses nombreuses propriétés intéressantes et de sa flexibilité. Certaines propriétés asymptotiques, telles que le biais, la variance et l'erreur quadratique moyenne intégrée (MISE) de l'estimateur proposé sont établies. En outre, nous démontrons la consistance forte et la normalité asymptotique de l'estimateur GBS-HR .Le choix du paramètre de lissage est également étudié par la méthode de réinjection et de validation croisée non biaisée. Enfin, la performance de l'estimateur HR basé sur les noyaux GBS et la comparaison des deux méthodes de sélection de paramètre de lissage sont illustrées par une étude de simulation et des applications sur des données réelles.

ملخص

الهدف الرئيسي من هذه الأطروحة هو اقتراح طريقة النواة اللامعلمية لتقدير وظيفة معدل الخطر (HR) في سياق البيانات المنحرفة بشكل إيجابي. يُنظر إلى فئة نواة بيرنبومسوندرز المعممة (GBS) نظرًا لخصائصها ومرونتها العديدة المثيرة للاهتمام. تم إنشاء بعض الخصائص المقاربة، مثل التحيز والتباين ومتوسط الخطأ التربيعي المتكامل (MISE) للمقدر المقترح. بالإضافة إلى ذلك، نوضح أن مقدر GBS-HR متسق بشدة وطبيعي بشكل مقارب. يتم أيضًا التحقق من اختيار النطاق الترددي من خلال قواعد التجربة ونهج التحقق المتبادل غير المتحيز. أخيرًا، يتم توضيح أداء مقدر،استنادًا إلى نواة GBS والمقارنة بين طريقتين لاختيار النطاق الترددي من خلال دراسة محاكاة وتطبيقات حقيقية.