People's Democratic Republic of Algeria

Ministry of Higher Education and Scientific Research

Mohamed Khider University of Biskra

Faculty of Exact Sciences

Department of Mathematics

Thesis Submitted in Partial Execution of the Requirements of the Degree of

Master in" **Statistics & probability** "

Option : Statistics

Submitted and Defended By

ABBI Rim lalouia

Title :

Comparison of Different Methods for Selecting the Smoothing Parameter in Asymmetric Kernel Estimation

Examination Committee Members :

Pr.	SAYAH Abdallah	U. Biskra	President
Dr.	BERKANE Hassiba	U. Biskra	Supervisor
Dr.	TOUR Madiha	U. Biskra	Examiner

03/06/2025

Dedication

To my parents,

To all my family,

To all those who are dear to me.

Acknowledgment

First, I would like to thank Allah, the Almighty, for granting me the patience, courage, and determination to complete this work.

I extend my gratitude and appreciation to my supervisor, Dr. Berkane Hassiba, for her valuable advice and guidance.

I would also like to express my deep thanks to all the members of the examination committee for accepting to evaluate this work. I truly appreciate their feedback and constructive criticism.

I express my gratitude to my family, who supported, encouraged, and stood by me throughout my years of study.

I would also like to extend my sincere thanks to all my professors in the Department of Mathematics for their continuous support.

Thank you all!

Notations and symbols

$K(.)$	:	kernel.
$\pi(h)\,,\ \pi(h\|x)$	:	*A prior distribution .A posteriori.*
$MCMC$	:	Markov chain Monte Carlo methods.
MSE	:	The mean square error.
ISE	:	The integrated squared error.
$MISE$	:	The integrated -mean square error.
LCV	:	likelihood cross-validation.
UCV	:	Unbiased cross-validation.
$\pi(x_1,...,x_n)$	:	Normalization constant.
h_{lcv}	:	Bandwidth using likelihood cross-validation.
h_{ucv}	:	Bandwidth using Unbiased Cross-Validation (UCV).
$\xrightarrow{L^2}$	:	convergence in the L^2 norm.
$\xrightarrow{p.s}$	:	convergence in the almost sure convergence.
$\xrightarrow{P}$	:	convergence in probability.

Contents

List of Figures

List of Tables

Introduction

Due to their simplicity, practitioners have widely used classical (symmetric) kernels, introduced by *Rosenblatt (*1956*)* and *Parzen (*1962*)*, for estimating probability density functions with unbounded support. However, the literature has shown that classical kernel estimators exhibit boundary bias when applied to data from distributions with bounded support. To address this issue, researchers have proposed alternative estimators, often based on modifications of classical kernels. Notable contributions include the reflection method of *Schuster (*1985*)*, the boundary kernels of *Müller (*1991*,* 1993*)*, and the empirical transformation of *Marron and Ruppert (*1994*)*.

*Chen (*1999*)* proposed replacing classical kernels with beta kernels for kernel density estimation when data are supported on $[0, 1]$. Inspired by beta kernels, *Chen (*2000*)* extended this approach to density estimation for semi-bounded support, introducing gamma kernels. The literature presents a variety of asymmetric kernels linked to beta, gamma, inverse Gaussian, and lognormal probability densities. Specifically, beta and gamma kernels were proposed by *Chen (*1999*,* 2000*)*, while Birnbaum-Saunders (BS) and lognormal kernels were introduced by *Jin and Kawczak (*2003*)*. Additionally, *Scaillet (*2004*)* proposed inverse Gaussian and reciprocal inverse Gaussian kernels.

This memoir focuses on kernel-based estimation techniques, particularly for densities with positive support, addressing key challenges such as boundary bias and smoothing parameter selection. Traditional symmetric kernels often prove inadequate for bounded or semi-bounded data, motivating the use of asymmetric kernels, including gamma, inverse Gaussian, and lognormal kernels. This work examines the construction of these associated kernels, their properties, and the convergence of the corresponding estimators. It also explores Bayesian approaches for smoothing parameter selection, leveraging Monte Carlo methods such as Markov Chain Monte Carlo ($MCMC$). These approaches are compared with classical methods, including cross-validation, and their performance is evaluated on both simulated and real-world datasets.

The main contributions of this memoir include:

Theoretical insights into kernel construction.

Bias reduction and normalization techniques.

Empirical evaluations of various kernels and smoothing methods.

This memoir consists of a general introduction, three chapters, and a conclusion:

Chapter 1 : introduces asymmetric continuous associated kernels, their construction, and the associated kernel density estimator, along with its convergence properties.

Chapter 2 : addresses smoothing parameter selection, presenting both frequentist and Bayesian approaches.

Chapter 3 : presents simulation results using well-known target densities with varying characteristics, as well as real-world datasets commonly referenced in the literature.

All numerical results and visualizations were generated using R software.

Chapter 1

Asymmetric kernel

1.1 Associated kernel estimator

The following definitions present the concepts of the associated kernel and the associated kernel estimator for the density function f Unknown on the support $\aleph$.

Definition 1.1.1 *Let $x \in \aleph$ and $h > 0$. The associated kernel is called $K_{x,h}$. Any probability density associated with a continuous or discrete random variable $\mathcal{K}_{x,h}$ of support $\aleph_{x,h}$. Verifying the following four conditions :*

$$\aleph_{x,h} \cap \aleph \neq \phi, \tag{1}$$

$$\underset{x}{\cup} \aleph_{x,h} \supseteq \aleph, \tag{2}$$

$$\lim_{h\to 0} E\left(\mathcal{K}_{x,h}\right) = x, \tag{3}$$

$$\lim_{h\to 0} Var\left(\mathcal{K}_{x,h}\right) = 0. \tag{4}$$

1.2 Univariate asymmetric continuous associated kernel

The asymmetric continuous kernel estimator is suitable for estimating densities with compact and bounded support. Let $X_1, X_2, ..., X_n$ .A sample of $(i.i.d)$ random variables from an unknown continuous probability density function f with support $\aleph = [a; b]$, where $(a \in \mathbb{R}$ and $b \in \mathbb{R})$In general, the continuous kernel estimator is of the form:

$$\hat{f}_n(x)=\frac{1}{n}\sum_{i=1}^{n}K_{x,h}(X_i) = \hat{f}_{n,h,K}(x), \tag{1.1}$$

where $x \in \aleph$ set, $h \in \mathbb{R}; (h > 0)$: is the smoothing parameter, $K_{x,h}$ is associated with a verified asymmetric continuous kernel :

$$K(u) \geq 0 \quad , \quad K_{x,h}(\cdot) = \frac{1}{h}K(\frac{x-\cdot}{h}) \quad \textit{and} \quad \int_{\mathbb{R}} K(u)du = 1.$$

Remark 1.2.1 *A symmetric kernel also satisfies the definition of the associated asymmetric kernel.*

1.2.1 Examples of Associated Asymmetric Kernels

Gamma Kernel:

The associated kernel Gamma was introduced by *Chen* (*2000*) for the estimation of densities with support $S_{x,h} = [0, +\infty)$. He used the Gamma distribution to construct asymmetric continuous associated kernels. The Gamma kernel is defined

:

$$K_{GA_{(1+x/h,h)}}(y) = \frac{y^{x/h}}{\Gamma(1+x/h)\, h^{1+x/h}} \exp\left(-\frac{y}{h}\right), \tag{1.2}$$

where $\Gamma(.)$is the classical gamma function . It is the probability density function of the gamma distridution with scale parameter $1 + x/h$ and shape parameter h ; see *Chen* (*2000*)and also *Libenguè* (*2013*). The Gamma kernel estimator is given by:

$$\hat{f}_n(x) = \frac{1}{n}\sum_{i=1}^{n} K_{GA_{(1+x/h,h)}}(X_i). \tag{1.3}$$

Modified Gamma Kernel:

The second class was proposed by the same author to improve the performance (reduce the bias) of the estimator. The form of the modified Gamma kernel is given by:

$$K_{GAM(\rho_h(x),h)}(y) = \frac{y^{x/h}}{\Gamma(\rho_h(x))\, h^{\rho_h(x)}} \exp\left(-\frac{y}{h}\right), \tag{1.4}$$

where $\Gamma(\alpha) = \int_0^\infty e^{(-t)} t^{\alpha-1} dt$, $\alpha > 0$ is the gamma function and h is the parameter of smoothing satisfying the conditions $h \to 0$ and $nh \to \infty$ when $n \to \infty$ where:

$$\rho_h(x) = \begin{cases} \frac{x}{h} & si \quad x \geq 2h \\ \\ \frac{1}{4}\left(\frac{x}{h}\right)^2 + 1 & si \quad x \in [0, 2h[\end{cases},$$

The modified Gamma kernel estimator is written as:

$$\hat{f}_n(x) = \frac{1}{n}\sum_{i=1}^{n} K_{GAM(\rho_h(x),h)}(X_i) = \frac{1}{n}\sum_{i=1}^{n} \frac{X_i^{x/h} \exp(-X_i/h)}{\Gamma(\rho_h(x))\, h^{\rho_h(x)}}. \tag{1.5}$$

Beta Kernel:

The extended beta kernel is defined on $S_{x,h,a,b} = [a, b]$ with $a \langle b \langle \infty$, and $h > 0$:

$$K_{BE_{x,h,a,b}}(y) = \frac{(y-a)^{(x-a)/\{(b-a)h\}}(b-y)^{(b-x)/\{(b-a)h\}}}{(b-a)^{1+h^{-1}} B(1+(x-a)/(b-a)h, 1+(b-x)/(b-a)h)}, \tag{1.6}$$

where $B(r, s) = \int_0^1 t^{r-1}(1-t)^{s-1} dt$ is the usual beta function with $r > 0, s > 0$. For $a = 0$ and $b = 1$, the extended beta kernel corresponds to the beta kernel whith is the probability density function of the beta distribution with shape parameters $1 + x/h$; see *Libenguè(2013)*.

The Beta kernel estimator $\hat{f}_n(x)$ is denoted:

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^{n} K_{BE_{x,h,a,b}}(X_i). \tag{1.7}$$

Gaussian-Inverse and Gaussian-Inverse-Reciprocal Kernels:

Scaillet (2004) introduced the associated Gaussian-inverse and Gaussian-reciprocal-inverse kernels defined on $]0, \infty[$. The Gaussian-inverse and Gaussian-reciprocal-inverse kernels are given respectively as follows:

$$K_{IG(x,1/h)}(y) = \frac{1}{\sqrt{2\pi h y^3}} \exp\left(-\frac{1}{2hx}\left(\frac{y}{x} - 2 + \frac{x}{y}\right)\right), \tag{1.8}$$

$$K_{RIG(1/(x-h);1/h)}(y) = \frac{1}{\sqrt{2\pi h y}} \exp\left(-\frac{x-h}{2h}\left(\frac{y}{x-h} - 2 + \frac{x-h}{y}\right)\right), \tag{1.9}$$

where $h > 0$ and $x \in \mathbb{R}_{+}$. see *Igarashi and Kakizawa (2015)* and also *Libenguè (2013)*.

The estimators of the unknown probability density f using the inverse-Gaussian

kernel	$\aleph_{x,h}$	$K_{x,h}(u)$	Expectation	Variance
$Gamma(a,b)$	$\mathbb{R}_+$	$\frac{1}{\Gamma(a)b^a}u^{a-1}e^{-\frac{u}{a}}$	ab	ab^2
$B\hat{e}ta(a,b)$	$[0;1]$	$\frac{1}{\beta(a,b)}u^{a-1}(1-u)^{b-1}$	$\frac{a}{(a+b)}$	$\frac{a}{\left\{(a+b)^2(a+b+1)\right\}}$
$IG(a,b)$	$\mathbb{R}_+$	$\frac{\sqrt{b}}{\sqrt{2\pi u^3}}\exp\left\{-\frac{b}{2a}\left(\frac{u}{a}-2+\frac{a}{u}\right)\right\}$	a	$\frac{a^3}{b}$
$RIG(a,b)$	$\mathbb{R}_+$	$\frac{\sqrt{b}}{\sqrt{2\pi u}}\exp\left\{-\frac{b}{2a}\left(au-2+\frac{1}{au}\right)\right\}$	$\frac{1}{a}+\frac{1}{b}$	$\frac{1}{ab}+\frac{2}{b^2}$

Table 1.1: Some asymmetric continuous kernels.

and inverse-Gaussian kernels are given, respectively by:

$$\hat{f}_n(x) = \frac{1}{n}\sum_{i=1}^{n} K_{IG(x,1/h)}(X_i) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\sqrt{2\pi h X_i^3}}\exp\left(-\frac{1}{2hx}\left(\frac{X_i}{x}-2+\frac{x}{X_i}\right)\right). \tag{1.10}$$

$$\hat{f}_n(x) = \frac{1}{n}\sum_{i=1}^{n} K_{RIG(1/(x-h);1/h)}(X_i) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\sqrt{2\pi h X_i}}\exp\left(-\frac{x-h}{2h}\left(\frac{X_i}{x-h}-2+\frac{x-h}{X_i}\right)\right). \tag{1.11}$$

Log-Normal Kernels:

The lognormal kernel is defined on $S_{x,h} = [0,\infty)$, and $h > 0$:

$$K_{LN(\ln(x),4\ln(1+h))}(y) = \frac{1}{\sqrt{8\pi\ln(1+h)y}}\exp\left[-\frac{(\ln(y)-\ln(x))^2}{8\ln(1+h)}\right], \tag{1.12}$$

Using this kernel, the unknown density f is estimated by:

$$\hat{f}_n(x) = \frac{1}{n}\sum_{i=1}^{n} K_{LN(\ln(x),4\ln(1+h))}(X_i). \tag{1.13}$$

It is the probability density function of the classical lognormal distribution with mean $log(x) + h^2$ and standard deviation h; see *Igarashi and Kakizawa* (2015) and also *Libenguè* (2013) .In this section, we provide the different fundamental properties of the associated kernel estimator.

where: $IG(a,b)$, $RIG(a,b)$ are the inverse Gaussian laws and the reciprocal inverse

Gaussian laws, respectively, and $\Gamma(\alpha) = \int_0^{+\infty} t^{\alpha-1}e^{-t}dt$ and $\beta(a,b) = \int_0^1 (1-t)^{b-1}t^{a-1}dt$, with $a, b \in \mathbb{R}_+^*$.

1.2.2 The properties of the asymmetric kernel estimator

Point Bias For a fixed x, the bias of the associated asymmetric continuous kernel estimator is generally calculated.

Point Variance For a fixed x, we generalize the expression of the variance of $\hat{f}_n$.

Mean Integrated Squared Error (MISE) We assume throughout that f has a continuous second derivative on the support $\aleph$ and that the following terms are finite: $\int_{\aleph} [f'(u)]^2 du$, $\int_{\aleph} [xf"(u)]^2 du$ and $\int_{\aleph} [x^3 f"(u)]^2 du$.

Example 1.2.1 *The Gamma kernel estimator is denoted:*

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^{n} K_{GA_{(1+x/h,h)}}(X_i)$$

The asymptotic bias of $\hat{f}_n(x)$ is expressed by the following formula:

$$Biais\left[\hat{f}_n(x)\right] = hf'(x) + \frac{1}{2}h^2 f"(x) + \circ(h),$$

The asymptotic variance of $\hat{f}_n(x)$ is given by:

$$Var\left[\hat{f}_n(x)\right] \approx \begin{cases} \frac{1}{2\sqrt{\pi}} n^{-1}h^{-1/2}x^{-1/2}f(x) + \circ\left(n^{-1}h^{-1/2}\right) & si \quad \frac{x}{h} \to \infty \\ \\ \frac{\Gamma(k+1)}{2^{2k+1}\Gamma^2(k+1)} h^{-1}n^{-1}x^{-1/2}f(x) + \circ\left(n^{-1}h^{-1}\right) & si \quad \frac{x}{h} \to K \end{cases},$$

The $MISE$ is measured as:

$$MISE\left[\hat{f}_n(x)\right] = h^2 \int_0^{+\infty} \left\{ xf'(x) + \frac{xf"(x)}{2} \right\}^2 dx + \frac{n^{-1}h^{-1/2}}{2\sqrt{\pi}} \int_0^{+\infty} x^{-1/2} f(x)dx + \circ(\frac{1}{n\sqrt{h}} + h^2)$$

1.2.3 Convergence of associated kernel estimators

A large body of literature has been devoted particularly to the issues of boundary estimation by asymmetric kernel estimators. for example *Bouezmarni and Scaillet (*2005*)* analyze the behavior of the asymmetric kernel estimator at the point $x = 0$ in the case of densities with a pole at $x = 0$.

Theorem 1.2.1 (Bouezmarni and Scaillet, 2005) *Let f be a probability density function on $[0, +\infty[$ that is not bounded at $x = 0$, and let $\hat{f}_h$ be its associated asymmetric kernel estimator. If :*

$$\lim_{n\to\infty} h = 0 \text{ and } \lim_{n\to\infty} nh^{2a} = +\infty \qquad (a > 0),$$

Then:

$$\hat{f}_h(0) \xrightarrow{P} +\infty, \qquad as\ n \to \infty,$$

and if for every $\delta > 0$,

$$\delta \int_0^{\delta} K_{\theta(0,h)}(u)du \to 1 \qquad as \quad h \to 0$$

where $\xrightarrow{P}$ denotes convergence in probability, it is noted that the additional condition in the previous theorem is verified for the Gamma kernel estimator. Indeed, for any $\delta > 0$ and $x = 0$, we have:

$$\int_0^\delta K_{Gam(1,h)}(u)du = 1 - exp\left(-\frac{\delta}{h}\right) \to 1 \qquad as\ h \to 0$$

Consequently, the Gamma kernel estimator assigns significant weight to the boundary points when the bandwidth is small. This is due to the particular property of the Gamma kernel at $x = 0$.

Thus, *Bouezmarni and Scaillet* (2005) advise against using other asymmetric kernels that do not satisfy the above property for estimating densities with a pole at $x = 0$ (for example, the inverse Gaussian and reciprocal inverse Gaussian kernels). Similarly, *Malec and Schienle* (2014) have recommended the use of the Gamma kernel for the estimation of such densities.

The following theorem addresses the pointwise conditions for both weak and strong consistency of the associated kernel estimator.

Theorem 1.2.2 *(*Kokonendji and al.*, 2012) Let $f \in C^2(\mathbb{T})$ be a probability density function and $\hat{f}_h$ its associated kernel estimator for a given kernel type K . For any fixed $x \in \mathbb{T}$ and for $h = h(n)$, assume there exists a positive real number $r = r(K, x)$ such that $nh^r \to +\infty$. Then,*

$$\hat{f}_h(x) \overset{\mathbb{L}^2\ and\ P.s}{\rightarrow} f(x), \qquad as\ n \to \infty.$$

where $\overset{\mathbb{L}^2 \text{ and } P.s}{\rightarrow}$ denote convergence in the $\mathbb{L}^2$ norm and almost sure convergence, respectively.

1.3 Multivariate asymmetric kernels

This section presents the method of multidimensional density estimation using continuous asymmetric kernels. The concept of a multivariate associated kernel $K_{\mathrm{x},H}$of target vector x and the smoothing window matrix H was introduced by *Kokonendji and Somé (*2015*)* In the continuous case, the observed variables $X_1, ..., X_n$ are independent and identically distributed random vectors (iid),Taking real values with the same density f. The multivariate asymmetric kernel estimator of f can be defined by:

$$\hat{f}_n(\mathrm{x}) = \frac{1}{n}\sum_{i=1}^{n} K_{\mathrm{x},H}(X_i) \qquad \forall x \in \aleph_d \subset \mathbb{R}_d \tag{1.14}$$

where $K_{\mathrm{x},H}(.) = \frac{1}{(\det H)} K(H^{\frac{1}{2}}(\mathrm{x} - .))$.

Definition 1.3.1 *Let* $\mathrm{x} \in \aleph_d \subseteq \mathbb{R}_d$ *and* H *a smoothing matrix, with* $\aleph_d$ *is the support of the function* f *to be estimated.* $K_{\mathrm{x},H}(.)$ *of support* $\aleph_{\mathrm{x},H} \subseteq \mathbb{R}_d$ *is called a multivariate associated kernel if the following conditions are satisfied:*

$$\mathrm{x} \in \aleph_{\mathrm{x},H}, \qquad E(\mathcal{K}_{\mathrm{x},h}) = \mathrm{x} + a(\mathrm{x}, H), \qquad Cov(\mathcal{K}_{\mathrm{x},h}) = B(\mathrm{x}, H). \tag{1.15}$$

where $a(\mathrm{x}, H) \underset{H \to 0_d}{\longrightarrow} 0_d$, $B(\mathrm{x}, H) \underset{H \to 0_d}{\longrightarrow} 0_d$ *(* 0_d *is a null square matrix of order* d*)* $K_{\mathrm{x},h}$ *is a vector of discrete random variables with a distribution* $K_{\mathrm{x},H}$.

Chapter 2

Smoothing Parameter Selection

2.1 Bandwidth Selection for Univariate asymmetric Kernel

At this stage, we present methods for selecting window sizes to approximate the optimal window value h defined by:

$$h_{opt} = \arg\min_{h>0} MISE\left(\hat{f}_n(x)\right). \quad (2.1)$$

2.1.1 Minimization of the integrated mean squared error

This method involves minimizing the integrated mean squared error ($MISE$) or asymptotic mean integrated squared error ($AMISE$). We recall that the $MISE$ is given by:

$$MISE\left(\hat{f}_n(x)\right) = \sum_{x\in\mathbb{N}} Biais^2\left[\hat{f}_n(x)\right] + \sum_{x\in\mathbb{N}} Var\left[\hat{f}_n(x)\right] \quad (2.2)$$

The variance can be approximated as follows:

$$Var\left(\hat{f}_n(x)\right) = \frac{1}{n}Var\left(K_{x,h}\left(X\right)\right) = \frac{1}{n}\left[f(u)\sum_{x\in\mathbb{N}}\mathcal{K}_{x,h}^2 - f^2\left(x\right)\right] + \circ\left(\frac{h}{n}\right), \quad (2.3)$$

Under the condition $\lim_{h\to 0}\sum_{u\in\aleph_{x,h}} u\ K\left(u\right) = x$, We have the following approximation:

$$\tilde{V}ar\left(\hat{f}_n(x)\right) = \frac{1}{n}f(x)P(\kappa_{x,h} = x). \quad (2.4)$$

where $\mathcal{K}_{x,h}$ is the discrete random variable with density $K_{x,h}$. The bias of $\hat{f}_n$ is approximated using the second-order discrete *Taylor* expansion.

$$Biais\left(\hat{f}_n(x)\right) = E\left[\hat{f}_n(x)\right] - f(x) = f\left[E(\mathcal{K}_{x,h})\right] - f(x) + \frac{1}{2}Var\left(\mathcal{K}_{x,h}\right)f\text{"}\ (x) + \circ\left(h\right) \quad (2.5)$$

Finally, the $MISE$ can be approximated by:

$$AMISE(h) = \frac{1}{n}\sum_{x\in\mathbb{N}} f\left(x\right)P(\mathcal{K}_{x,h} = x) \quad (2.6)$$

$$+\sum_{x\in\mathbb{N}}\left[f\left[E(\mathcal{K}_{x,h})\right] - f(x) + \frac{1}{2}Var\left(\mathcal{K}_{x,h}\right)f\text{"}\ (x) + \circ\left(h\right)\right]^2 \quad (2.7)$$

The smoothing parameter h_{AMISE} In this case, it can be obtained as follows:

$$h_{AMISE} = \arg\min_h AMISE\left(h\right). \quad (2.8)$$

The smoothing parameter h_{AMISE} is not directly usable in practice, because $AMISE(h)$ it depends on the unknown discrete density f.

2.1.2 Classical Methods for Bandwidth Selection

Plug-In Method

The popular plug-in (or re-injection) selector in kernel density estimation is due to *Sheather and Jones (*1991*)*. This method adopts the asymptotic $MISE$ $(AMISE)$ as criterion, defined by:

$$AMISE = \frac{h^4}{4}\delta_K^4 \int f"^2 (x)\, dx + \frac{\int K^2 (y)\, dy}{nh}, \tag{2.9}$$

The optimal bandwidth can be obtained by minimizing $AMISE$ as:

$$h^* = \left[\frac{R(K)}{\delta_K^4 R(f")}\right]^{1/5} n^{-1/5}, \tag{2.10}$$

where: $R(K) = \int K^2(y)dy$ and $R(f") = \int f"^2(x)dx$.

The expression of the optimal bandwidth h^* depends on unknown density f through $R(f")$. This ideal bandwidth is not directly calculable.*Sheather and Jones(*1991*)* choose to estimate $R(f") = \int f"^2(x)dx$ by:

$$\hat{R}_a (f") = \frac{1}{n^2 a^5} \sum_{i=1}^{n} \sum_{j=1}^{n} L^{(4)} \left(\frac{X_i - X_j}{a}\right), \tag{2.11}$$

Where: $L^{(4)}$ is the fourth derivative of the kernel function L, a is the pilot bandwidth parameter,The estimator $\hat{R}_a (f")$ is obtained as follows:

$$R(f") = \int f"^2(x)dx = \int f^{(4)} (x)\, f (x)\, dx$$

$$E\left[f^{(4)}(x)\right],$$

the quantity $R(f")$ can be estimated by:

$$\hat{R}_a(f") = \frac{1}{n}\sum_{i=1}^{n} f_a^{(4)}(X_i),$$

The optimal bandwidth is given by:

$$h_{sj} = \left(\frac{R(K)}{\delta_a^4 \hat{R}_a(f")}\right)^{1/5} n^{-1/5}. \tag{2.12}$$

These authors select a to minimize the asymptotic MSE, which is denoted by $AMSE(a)$.

In the case of using asymmetric associated kernels, the ideal smoothing parameter in the sense of $AMISE$ generally depends on three unknown quantities through f, $f^{'}$ and $f^{''}$. This makes it more difficult to choose the smoothing parameter by *plug-in methods*. The only method that has been proposed in this case is the *reference rule method* by analogy with the case of symmetric kernels. *Scaillet [2004]* suggests estimating the unknown quantities by replacing f by a log-normal reference model with parameters μ_f and σ_f^2.

Example 2.1.1 *For example, the smoothing parameters obtained by* Scaillet [2004] *for the reference rule using the inverse-reciprocal-Gaussian kernel is given by:*

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

In practice, the two parameters μ_f and σ_f^2 are estimated using the observations

$X_1, ..., X_n$ by the empirical mean and variance. This method tends to provide very small valuesfor the smoothing parameter, which leads to the phenomenon of under-smoothing.

Excess of zeros

In this section, the choice of the window is based on count data with $\aleph$ which is none other than the excess of zeros in the sample $X = (X_1, X_2, ..., X_n)$. One can choose a suitable window $h_0 = h_0(X, K)$ of h satisfying:

$$\sum_{i=1}^{n} P(\kappa_{X_i,h_0} = 0) = n_0 \tag{2.13}$$

Where n_0 denotes the number of zeros in X ($n_0 = Card(X_i = 0)$). The equation (2.12) is obtained from the expression:

$$E\left(\hat{f}_n(x)\right) = \sum_{u \in \aleph_{x,h}} f(u) P(\mathcal{K}_{x,h} = u). \tag{2.14}$$

In which we take $u = 0$ and $f(0) = 1$ n order to match the number of theoretical zeros with the number of empirical zeros n_0. The window h_0 adjusts the number of theoretical zeros to the number of observed zeros.

Cross-Validation Methods

Unbiased Cross-Validation *Rudemo (*1982*)* and *Bowman (*1984*)* introduced the idea of this method. By using the expanded formula of the integrated squared error $ISE(h)$, the smoothing parameter h is chosen to minimize this error. The cross-validation criterion $UCV(h)$ is applicable to both symmetrical and asymmetrical kernels.

$$ISE(h) = \int (\hat{f}_h(x) - f(x))^2 dx = \int \hat{f}_h^2(x)dx - 2\int \hat{f}_h(x)f(x)dx + \int f^2(x)dx. \tag{2.15}$$

We note that $\int f^2(x)dx$ does not depend on h, so we can choose h in such a way that it minimizes the *Unbiased cross-validation* criterion defined by:

$$UCV(h) = ISE(h) - \int f^2(x)dx = \int \hat{f}_h^2(x)dx - 2\int \hat{f}_h(x)f(x)dx. \tag{2.16}$$

We must therefore find an estimator for $\int \hat{f}_h(x)f(x)dx$. Let us note that:

$$\int \hat{f}_h(x)f(x)dx = E\left[\hat{f}_h(x)\right] \tag{2.17}$$

The empirical estimator of $\int \hat{f}_h(x)f(x)dx$ is then: $\frac{1}{n}\sum_{i=1}^{n} f_{h,i}(x_i)$.

The criterion to be optimized is then:

$$UCV(h) = \int \hat{f}_h(x)^2 dx - \frac{2}{n}\sum_{i=1}^{n} f_{h,i}(x_i),$$

where $f_{h,i}(x_i) = \frac{1}{(n-1)h}\sum\limits_{j\neq i,j=1} K(\frac{X_i - X_j}{h})$ is the density estimator constructed from the set of points except for the point x_i.

Using the explicit formulas of $\hat{f}_h(x)$ and $f_{h,i}(x_i)$, the criterion $UCV(h)$ is given by:

$$UCV(h) = \frac{R(K)}{nh} + \sum_{i=1}^{n}\sum_{j\neq i,j=1}\left(\frac{1}{(nh)^2}\int K\left(\frac{x-X_i}{h}\right)K\left(\frac{x-X_j}{h}\right)dx - \frac{2K\left(\frac{X_i-X_j}{h}\right)}{n(n-1)h}\right) \tag{2.18}$$

where $R(K) = \int K^2(x)dx$.

The optimal bandwidth h_{UCV} is thus obtained as follows:

$$\hat{h}_{ucv} = \arg\min_{h} \left(UCV(h) \right). \tag{2.19}$$

Likelihood cross-validation The simplest of the cross-validation methods is that proposed by *Habbema and al. [1974]*, called maximum likelihood cross-validation. It consists in maximizing with respect to h a likelihood estimator given by:

$$LCV(h) = \prod_{i=1}^{n} \hat{f}_{h,i}(x_i) = \frac{1}{(n-1)^n} \prod_{i=1}^{n} \sum_{j=1, i \neq j}^{n} K_{x_i,h}(x_j), \tag{2.20}$$

The optimal smoothing parameter provided by this method is given by:

$$\hat{h}_{lcv} = \arg\max_{h} \left(LCV(h) \right). \tag{2.21}$$

2.1.3 Bayesian Approach to Bandwidth Selection

Global Bayesian approach

This approach was proposed by *Kuroita and al. (*2010*)* to estimate the *pdf*. By the natural logarithmic kernel and then by *Zoghab and al. (*2013*b)* to estimate pdf. By separate connected kernels. $\hat{f}_n$ estimated with linked kernels f. The idea is to treat the smoothing parameter h as a random variable. We then choose the priori distribution $\pi(.)$ for h.

Bayesian inference concerning a parameter h conditional on data is made via the posterior density $\Pi(h|data)$. In particular, we consider a sequence $X_1, ..., X_n$ of $i.i.d.$ real random variables with density of probability or probability mass function (pmf) f and $x_1, ..., x_n$ an independent random sample drawn from f .

The likelihood function is written as:

$$L(x_1, ..., x_n; h) = \pi(x_1, x_2..., x_n|h) = \prod_{i=1}^{n} \hat{f}_n(x_i) = \frac{1}{(n-1)^n} \prod_{i=1}^{n} \sum_{j=1, i \neq j}^{n} K_{x_i,h}(x_j), \tag{2.22}$$

Thus, the likelihood cross-validation is given by the Bayes theorem, the posterior of h takes the form:

$$LCV(x_1, ..., x_n; h) = \pi(x_1, ..., x_n|h) = \prod_{i=1}^{n} \hat{f}_{h,i}(x_i) \tag{2.23}$$

$$\pi(h \mid x_1, x_2..., x_n) = \frac{\pi(x_1, ..., x_n|h)\pi(h)}{\pi(x_1, x_2, ..., x_n)} = \frac{\pi(h) \prod_{i=1}^{n} \hat{f}_{h,i}(x_i)}{\pi(x_1, x_{2,}..., x_n)} \tag{2.24}$$

where $\pi(x_1, x_2..., x_n) = \int \pi(x_1, x_{2,}..., x_n|h)\pi(h)dh$. We can also write:

$$\pi(h \mid x_1, x_2..., x_n) \; \alpha \; \pi(x_1, ..., x_n|h)\pi(h) = \pi(h) \prod_{i=1}^{n} \hat{f}_{h,i}(x_i), \tag{2.25}$$

Therefore, the posterior is given by:

$$\pi(h \mid x_1, x_{2,}..., x_n) \; \alpha \; \pi(h) \prod_{i=1}^{n} \frac{1}{(n-1)} \sum_{j=1, i \neq j}^{n} K_{x_i,h}(x_j) \tag{2.26}$$

The Bayes estimator is the posterior mean given by:

$$\hat{h}_{GBay} = \int h\pi(h \mid x_1, x_2..., x_n)dh. \tag{2.27}$$

Local Bayesian approach

We derive the Bayesian bandwidth at each point x where the density is being estimated (*Bayesian local approach*) in the context of kernel density estimation with positive support using the asymmetric kernel. Our goal is to estimate the bandwidth h at each point x by using the Bayesian approach. Now, consider h as a scale parameter for $f_h(x)$.

$$\hat{f}_h(x) = \frac{1}{n}\sum\nolimits_{i=1}^{n} \frac{1}{h}K\left(\frac{x - X_i}{h}\right) = \frac{1}{n}\sum\nolimits_{i=1}^{n} K_{x,h}(X_i) \tag{2.28}$$

Our approach consists in using $f_h(x)$ and constructing a *Bayesian local* estimator for h. Let $\pi(h)$ be a prior distribution, then by the Bayes rule, the posterior of h at the point of estimation x takes the form:

$$\pi(h \mid x) = \frac{f_h(x)\pi(h)}{\int f_h(x)\pi(h)dh}, \tag{2.29}$$

Since $f_h(x)$ is unknown, we use a suitable estimator $\hat{f}_h(x)$ we can estimate a bayesian model which depend on the prior and likelihood function by:

$$\hat{\pi}(h \mid x, X_1, X_2..., X_n) = \frac{\hat{f}_h(x)\pi(h)}{\int \hat{f}_h(x)\pi(h)dh} \tag{2.30}$$

The Bayes estimates of the smoothing parameter h can be computed under squared error loss. The Bayes estimator is the posterior mean given by:

$$\hat{h}(x) = \int h\hat{\pi}(h \mid x, X_1, X_2, ..., X_n)dh. \tag{2.31}$$

2.2 Multivariate bandwidth selectors

2.2.1 Excess of Zeros

The choice of the window matrix for this method is based on a particularity of count data ($\aleph_d$), We will generalize this technique from the one-dimensional case to the multidimensional case. Given a multivariate associated kernel $K_{\mathrm{x},H}$, we can choose H_0 such that:

$$\sum_{i=1}^{n} P\left(\mathcal{K}_{X_1,H_0}^{[1]} = 0, ..., \mathcal{K}_{X_d,H_0}^{[d]} = 0\right) = n_0.$$

2.2.2 Cross-validation methods

In the multivariate case of the UCV criterion of least squares validation the generalization of the univariate form devised by *Rudemo (*1982*) and Bowman (*1984*)* (see also *Duong and Hazelton [*2005*]*),

$$UCV(H) = \int_{\mathbb{R}^d} \left(\hat{f}_n(\mathrm{x})\right)^2 d\mathrm{x} - \frac{2}{n}\sum_{i=1}^{n} \hat{f}_{n,-i}(X_i) \tag{2.32}$$

where $\hat{f}_{n,-i}(X_i;H) = (n-1)^{-1}\sum_{j\neq i}^{n} K_H(\mathrm{x}-X_j)$ is being computed as $\hat{f}_n(X_i)$ excluding the observation X_i . The bandwidth matrix obtained as follows:

$$\hat{H}_{UCV} = \arg\min_{H\in\mathcal{M}} UCV(H). \tag{2.33}$$

where $\mathcal{M}$ is the set of all positive definite full bandwidth matrices. The optimal window matrix, denoted H_{LSCV} ,is given by:

$$H_{LSCV} = \arg\min_{\mathcal{M}} LSCV(H). \tag{2.34}$$

where $\mathcal{M}$ is the space of symmetric positive-definite smoothing matrices and

$$LSCV(H) = \int_{\mathbb{S}_d} \hat{f}_n(\mathrm{x})dx - \frac{2}{n}\sum\nolimits_{i=1}^{n} \hat{f}_{n,-i}(X_i)d\mathrm{x}. \tag{2.35}$$

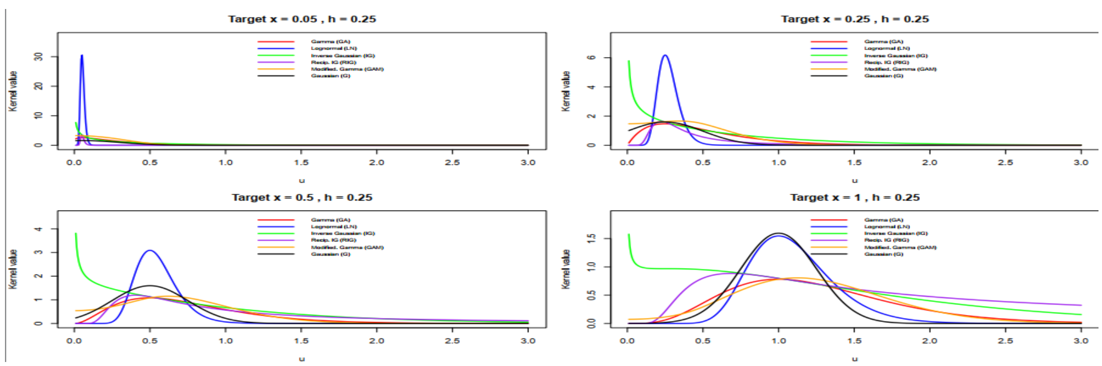

Figure 2.1: Asymmetric associated kernels for different smoothing parameters and for a fixed target $x = 1.5$

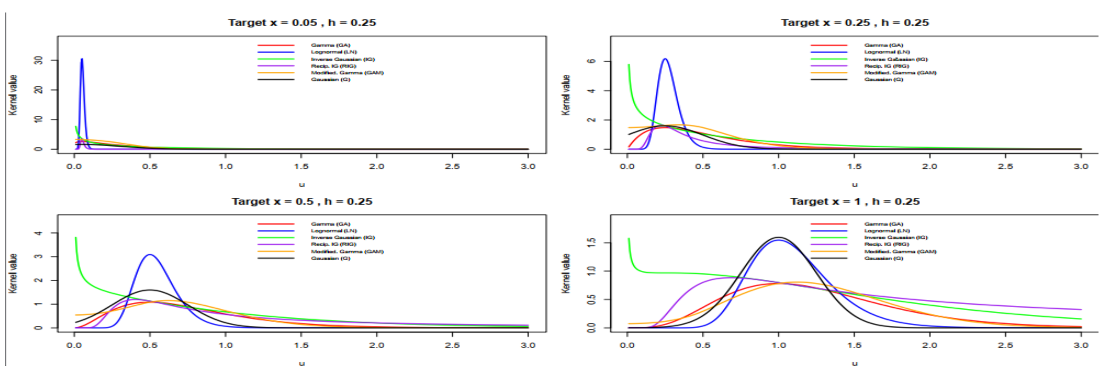

Figure 2.2: Asymmetric associated kernels for different targets x and for a fixed smoothing parameter $h = 0.25$

Chapter 3

Simulation and numerical results

This chapter presents a simulation study comparing several methods for selecting the smoothing parameter h, with a focus on asymmetric continuous kernel functions. All estimators are developed in a univariate setting and evaluated using the Integrated Squared Error and Mean Integrated Squared Error.

The simulation emphasizes the performance of Bayesian inference in selecting the smoothing parameter $\hat{h}_{bays}$, especially for density estimation with positive support on $]0, +\infty[$. The results are compared to classical approaches, including unbiased cross-validation (UCV) and the theoretical optimal bandwidth h^*. This comparison highlights the strengths and weaknesses of each method.

3.1 Simulation Study

In this section, we illustrate the performance of several asymmetric continuous kernel estimators: Gamma, Modified Gamma, Reciprocal Inverse Gaussian, and Lognormal kernels. We simulate samples of sizes $n \in \{25, 100, 200, 400\}$. For each kernel, the optimal bandwidth is selected using UCV and Bayesian $(MCMC)$

methods.

3.1.1 Methods Used

Among the methods and criteria presented in previous chapters, we use:

1. Kernel Types: Gamma (GA), Modified Gamma (GAM), Inverse Reciprocal Gaussian (RIG), and Lognormal (LN).

2. Smoothing Parameter Selection:

   - Classical: Using UCV.
   - Bayesian: Using $MCMC$.

3. Error Criteria: We chose to evaluate the estimation error using the ISE and $MISE$.

Let us denote:

n : Sample size.

h^* : Theoretical optimal smoothing parameter

$\hat{h}_{ucv}$: Bandwidth from unbiased cross-validation.

$\hat{h}_{bays}$: Bandwidth from Bayesian inference.

$MISE^*$: Theoretical mean integrated squared error.

ISE^*: Relative deviation from optimal ISE.

$\widehat{ISE}_{UCV}$: The empirical mean ISE with the smoothing parameter $\hat{h}_{ucv}$.

$\widehat{ISE}_{Bays}$: The empirical mean ISE with the smoothing parameter $\hat{h}_{bays}$.

For the numerical application we considered the following distributions: Density set to $]0, +\infty[$

1. Gamma model with parameter $(2, 2)$:

$$f_1(x) = \frac{x}{4} \exp\left(-\frac{x}{2}\right), \qquad x > 0.$$

2. Log-normal with parameters $\mu = 2$ and $\sigma = 1$:

$$f_2(x) = \frac{1}{x\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\log(x) - 2\right)^2\right), \qquad x > 0.$$

3. Weibull model with shape parameter $\alpha = 2$ and scale parameter $\beta = 2$:

$$f_3(x) = \frac{x}{2} \exp\left(-\frac{x^2}{4}\right) \quad , \qquad x \geq 0.$$

3.1.2 Algorithm

- Simulate a sample of size n from the target density,
- Select an asymmetric kernel (GA, GAM, RIG, or LN).,
- Calculate the optimal smoothing parameter ($\hat{h}_{ucv}$, $\hat{h}_{bays}$) using one of the selection algorithms (UCV, $MCMC$),
- Construct the kernel estimator using the observations and the calculated bandwidth.
- Compute the ISE for the estimated density,
- Plot the theoretical and estimated densities for comparison.

3.2 Discussion of results

Gamma Model

n	K	h^*	$\hat{h}_{ucv}$	$\hat{h}_{bays}$	$MISE^*$	ISE^*	$\widehat{ISE}_{UCV}$	$\widehat{ISE}_{Bays}$
25	GA	0.15227	0.75383	0.24293	0.02264	0.00894	0.00153	0.00608
	GAM	0.55189	0.75382	0.22766	0.01570	0.00227	0.00153	0.00644
	RIG	0.55189	0.22361	0.53715	0.01190	0.02220	0.06437	0.02244
	LN	0.11959	0.45769	0.19806	0.01760	0.05144	0.01397	0.02064
100	GA	0.08746	0.35214	0.35980	0.00747	0.00319	0.00128	0.00129
	GAM	0.31698	0.35212	0.33980	0.00518	0.00127	0.00128	0.00127
	RIG	0.31698	0.33808	0.65980	0.00392	0.00917	0.00820	0.00322
	LN	0.06869	0.39264	0.35980	0.00581	0.02130	0.00300	0.00244
200	GA	0.06628	0.28743	0.22482	0.00429	0.00190	0.00152	0.00132
	GAM	0.24022	0.28744	0.21722	0.00297	0.00136	0.00152	0.00130
	RIG	0.24022	0.52477	0.14884	0.00225	0.00403	0.00373	0.00582
	LN	0.05206	0.29443	0.05404	0.00334	0.00725	0.00089	0.00692
400	GA	0.05023	0.11081	0.23934	0.00246	0.00128	0.00070	0.00093
	GAM	0.18206	0.11080	0.21934	0.00171	0.00074	0.00070	0.00085
	RIG	0.18205	0.40206	0.53934	0.00129	0.00361	0.00218	0.00267
	LN	0.03945	0.22575	0.23934	0.03813	0.00812	0.00064	0.00065

Table 3.1: Simulation results for the selection of the smoothing parameter by UCV and Bays, case of Gamma model with parameter (2,2)

Reading table *(3.1)* and estimated and theoretical curves presented in figures *(3.1), (3.2), (3.3) and (3.4)* show that:

For $n = 25$:

* GAM kernel: gives the lowest $ISE^* = 0.00227$, and the Bayesian estimate $\widehat{ISE}_{Bays} = 0.00644$ remains very competitive.

* GA kernel provides the best $\widehat{ISE}_{UCV} = 0.00153$.

* RIG and LN performs poorly for small samples (especially LN in ISE^*).

For $n = 100$:

* GAM : gives the lowest theoretical error $ISE^* = 0.00127$, and the Bayesian estimate $\widehat{ISE}_{Bays} = 0.00127$ matches it, showing strong consistency and accuracy.

* GA :provides the best $\widehat{ISE}_{UCV} = 0.00128$, indicating competitive performance in data-driven bandwidth selection.

* RIG and LN performs poorly for small samples (especially LN in ISE^*).

Observations:

- $\hat{h}_{ucv}$tends to be larger than h^*, especially for the GA and GAM kernels.

- All selection methods yield $\hat{h}_{bays}$ values close to h^* for the GA and GAM.

- $MISE^*$ is small but ISE values vary: particularly large for LN.

- Relative errors $\widehat{ISE}_{UCV}$ and $\widehat{ISE}_{Bays}$ are higher for the RIG and LN kernels (indicating unstable estimation for small n).

For $n \in \{200, 400\}$:

* LN : consistently outperforms all others when the UCV and $Bayesian$ methods are used, with $\widehat{ISE}_{UCV}$ as low as 0.00089 for $n = 200$ and $\widehat{ISE}_{UCV} = 0.00064$, $\widehat{ISE}_{Bays} = 0.00065$ for $n = 400$, indicating superior performance and stability with large sample sizes

  Although the $MISE$ for the LN kernel remains relatively high for large sample sizes (0.03813..), the low ISE indicates good sample stability.

* GAM exhibits balanced performance, but it is not the best option for large samples.

■ *How does performance change with increasing sample size?*

When n increases to 400, the results shift:

* LN kernel outperforms all others, especially in terms of ISE using UCV and $Bayes$. Although the $MISE$ for the LN kernel remains relatively high (0.03813), the low ISE indicates good stability for this sample.

* GAM exhibits balanced performance, but it is not the best option for large samples.

* RIG remains unstable, often yielding the worst results.

* GA exhibits balanced performance, but it is not the best option for large samples *(Figure 3.1)* because shows a tendency to over smooth for small n, especially with $\hat{h}_{ucv}$, due to the large bandwidth (the estimated curve is flatter than the true density.when $n = 25$, but it aligns better as n increases).

Associated Figures

Gamma Kernel Estimators of a Gamma Density (*Figure 3.1)*:

* Shows performance of the GA kernel across all sample sizes ($n = 25 - 400$):

* Demonstrates close matching to true density for large samples.

* For small samples, the estimated curve tends to oversmooth when using UCV bandwidths.

Modified Gamma Kernel Estimators of a Gamma Density (*Figure 3.2*):

* Shows the performance of the GAM kernel.

* Displays high variability, especially at small sample sizes ($n = 25$).

* Shows potential over-smoothing/under-fitting due to unstable bandwidths selection .

Log-Normal Kernel Estimators of a Gamma Density (*Figure 3.3*):

* Demonstrates the superiority of the LN kernel for large sample sizes ($n \geq 200$).

* Showing excellent stability and accuracy at $n = 400$.

* Reveals poor performance for small samples ($n = 25$),though better than RIG.

RIG Kernel Estimators of a Gamma Density (*Figure 3.4*):

* Displays RIG kernel instability across all sample sizes.

* It Shows significant fluctuations and a weaker fits, especially with small n.

* Indicates slight improvement for larger samples but still less stable than GA/LN estimators.

Key Visual Observations:

Small samples ($n = 25$):

GAM Kernel: Shows the closest fit to the true density as shown in (*Figure 3.2).*

RIG and LN Kernels :exhibit significant deviations from the true density *(Figures 3.3-3.4).*

Large samples ($n = 400$):

LN Kernel: The curve aligns nearly perfectly with the true density (*Figure 3.3*).

GA Kernel: Shows a balanced fit, but it's less precise compared to LN, as seen in Figure (*Figure 3.1*).

Conclusions :

Performance by Method:

* Bayesian methods excellent for small to medium samples ($n \leq 100$).

* The UCV method performs best for larger samples ($n \geq 200$).

Kernel Performance:

* The LN kernel demonstrated superior estimation performance in large samples ($n \geq 200$).

* The GAM kernel shows excellent performance for small samples ($n \in \{25, 100\}$).

* The RIG kernel demonstrates instability across all sample sizes.

Bandwidth Observations:

* $\hat{h}_{ucv} > h^*$ in most cases, except for the RIG kernel when $n = 25$.

* The $MCMC$ method usually provides a bandwidth value between $h^*$and $\hat{h}_{ucv}$.

* $\hat{h}_{bays}$ tends to be closer to h^* for the GAM and GA kernels, while for the LN and RIG kernels, it shows higher variation especially with smaller sample sizes.

* $MCMC$ shows more fluctuation compared to h^* and $\hat{h}_{ucv}$, particularly with the LN and RIG kernels for small n ($n = 25$).

n	Best Kernel	Best Method	Notes
25	GAM	Bayesian	Lowest $\widehat{ISE}_{Bays} = 0.00644$.
100	GAM	Bayesian	$\widehat{ISE}_{Bays} = 0.00127$.
200	LN	UCV	$\widehat{ISE}_{UCV} = 0.00089$.
400	LN	Both	$\widehat{ISE}_{UCV} = 0.00064$ and $\widehat{ISE}_{Bays} = 0.00065$, respectively.

Table 3.2: Summary of Best Performers in case of the Gamma model.

Final Recommendations :

* For small sample sizes (≤ 100), the GAM kernel is excellent, especially with the $\hat{h}_{bay}$ to improve accuracy of the estimation .

* As the sample size increases, the LN kernel started to outperform others, particularly with adaptive methods (UCV and $Bayes$).

* The RIG kernel provides weak and unstable results.

* Selecting the suitable method for the smoothing parameter: is crucial UCV performs best when $n \geq 200$.

Log-normal Model

n	K	h^*	$\hat{h}_{ucv}$	$\hat{h}_{bays}$	$MISE^*$	ISE^*	$\widehat{ISE}_{UCV}$	$\widehat{ISE}_{Bays}$
25	GA	0.26815	0.19986	0.26846	0.01135	0.01512	0.01914	0.01511
	GAM	0.85061	0.86152	0.87467	0.00841	0.02342	0.02305	0.02262
	RIG	0.85061	0.80429	0.49406	0.00638	0.02211	0.02378	0.03934
	LN	0.04121	0.66543	0.26853	0.00523	0.10208	0.01303	0.01854
100	GA	0.15416	0.32833	0.26993	0.00374	0.00702	0.00477	0.00539
	GAM	0.48855	0.51834	0.49622	0.00277	0.01064	0.01021	0.01052
	RIG	0.48855	0.49186	0.47767	0.00210	0.00982	0.00976	0.01002
	LN	0.02367	0.36201	0.28614	0.00144	0.03697	0.00752	0.00782
200	GA	0.11672	0.42826	0.19209	0.00215	0.00195	0.00094	0.00153
	GAM	0.37025	0.99994	0.29946	0.00159	0.00357	0.00274	0.00422
	RIG	0.37025	0.95349	0.39921	0.00121	0.00346	0.00180	0.00326
	LN	0.01794	0.36989	0.20535	0.00083	0.02277	0.00202	0.00216
400	GA	0.08846	0.37166	0.20966	0.00123	0.00104	0.00070	0.00078
	GAM	0.28060	0.93130	0.04512	0.00092	0.00207	0.00183	0.01490
	RIG	0.28060	0.30843	0.34369	0.00069	0.00203	0.00068	0.00111
	LN	0.01359	0.31264	0.32236	0.00048	0.01363	0.00050	0.00050

Table 3.3: Simulation results for the selection of the smoothing parameter by UCV and Bays, case of the Log-normal model with parameter (2,1)

Reading table *(3.3)* allows us to notice that:

when $n = 25$:

- The GA kernel with $\hat{h}_{bays}$ yields the lowest $ISE \approx 0.01511$.

- The LN kernel provides the best $\widehat{ISE}_{UCV} = 0.01303$.

- Bayesian estimates $\hat{h}_{bays}$ align closely with h^* for the GA and GAM kernels, but show instability for the LN and RIG kernels indicating unreliable smoothing for small samples.

when $n = 100$:

- The GA kernel with $\hat{h}_{bays}$achieves the lowest $ISE = 0.00539$.

- Both UCV and *Bayesian* methods show improvement, with Bayesian smoothing performing better, especially for the GA and GAM kernels.

when $n \in \{200, 400\}$:

For $n = 200$:

- The GA kernel with $\hat{h}_{ucv}$ yields the lowest ISE (0.00094).

- The LN kernel begins to outperform others as n increases.

For $n = 400$:

- The LN kernel achieves the lowest ISE (0.00050) with both UCV and *Bayesian* methods, demonstrating excellent stability. Although the $MISE$ for the LN kernel is very small (0.00048), the low ISE indicating high accuracy and stability in this sample.

- The GA kernel shows balanced performance but it is not the best for large samples.

- The GAM and RIG kernels remain unstable.

Key Observations:

* In most cases, the Bayesian estimated bandwidth $\hat{h}_{bays}$ is close to the optimal h^* , particularly for the GA and GAM kernels. However, for the RIG and especially the LN kernels, greater deviations are observed, notably in small n.

* UCV bandwidth $\hat{h}_{ucv}$ is generally greater than h^* except for the GA and RIG kernels when $n = 25$, due to the higher sensitivity of UCV to fluctuations in small samples.

* The GA kernel with $\hat{h}_{bays}$ offers the best performance for $n \in \{25, 100\}$, while at $n = 200$, the GA kernel with $\hat{h}_{ucv}$ gives the best result.when $n = 400$,the LN kernel (with UCV or $Bayes$) achieves the highest accuracy.

■ *How does performance change with increasing sample size?*

When n increases to 400, the results shift:

* LN kernel outperforms all others, especially in terms of ISE using UCV and $Bayes$. Although the $MISE$ for the LN kernel is relatively very small (0.00048), the low ISE indicates good stability and accuracy for this sample.

* GAM, RIG remains unstable, often yielding the worst results.

* GA exhibits balanced performance, but it is not the best option for large samples.

* The $MCMC$ method usually provides a bandwidth value between h^* and $\hat{h}_{ucv}$, but it tends to exhibit more fluctuations,particularly with the LN and RIG kernels when $n = 25$ due to the instability of the posterior distribution in small samples and the inherent asymmetry of these kernels.

Associated Figures:

Gamma Kernel (Figure 3.5):

- Closely matches the true density for small n, aligning with its low ISE values.

- Slight over-smoothing for larger n (e.g., $n = 400$) where LN outperforms.

Modified Gamma (Figure 3.6):

- High variability, especially for $n = 25$ (spiky estimates).

- Poor fit due to unstable bandwidths (e.g., $\hat{h}_{bays} = 0.87467$ at $n = 25$).

RIG Kernel (Figure 3.7):

- Similar to GAM, erratic for small n but improves slightly for $n = 400$.

- Struggles with tail behavior (underestimates peaks).

Lognormal Kernel (Figure 3.8):

- Best fit for large n, with smooth, accurate curves.

- Near-perfect alignment with f_2 at $n = 400$ (lowest ISE).

n	Best Kernel	Best Method	Lowest ISE
25	GA	Bayesian $\left(\hat{h}_{bays}\right)$	0.01511
100	GA	Bayesian $\left(\hat{h}_{bays}\right)$	0.00539
200	GA	UCV $\left(\hat{h}_{ucv}\right)$	0.00094
400	LN	with similar performance	0.00050, excellent stability.
		UCV/Bayesian (tie)	

Table 3.4: Summary of Best Performers in case of the Log-normal model.

Final Recommendations :

- Small samples($n = 25$) :the GA kernel provides the best performance, especially with $\hat{h}_{bays}$.

- Medium samples ($n \in \{100, 200\}$) :the GA kernel remains strong, but as the size grows, the LN kernel becomes more distinguished, particularly with UCV and *Bayesian* methods.

- Large samples ($n = 400$): the LN kernel is the best, with UCV and *Bayesian* providing excellent performance and higher accuracy.

- Avoid: GAM and RIG for small samples due to instability.

- UCV method: More stable and reliable for larger sample sizes, showing greater consistency across different kernel types.

- *Bayesian* method: Effective for small to medium sample sizes, but its performance can be sensitive to the choice of kernel function.

Weibull Model

n	K	h^*	$\hat{h}_{ucv}$	$\hat{h}_{bays}$	$MISE^*$	ISE^*	$\widehat{ISE}_{UCV}$	$\widehat{ISE}_{Bays}$
25	GA	0.12307	0.13561	0.05486	0.03484	0.02149	0.01900	0.04393
	GAM	0.18704	0.11521	0.21580	0.03729	0.08686	0.17257	0.07141
	RIG	0.18704	0.11574	0.19312	0.02826	0.08830	0.17333	0.08435
	LN	0.02455	0.32685	0.34857	0.03023	0.37785	0.00423	0.00387
100	GA	0.07069	0.08702	0.06204	0.01149	0.02834	0.02485	0.03042
	GAM	0.10743	0.29081	0.22734	0.01230	0.06137	0.03177	0.03921
	RIG	0.10743	0.27281	0.19909	0.00932	0.06140	0.03466	0.04162
	LN	0.01410	0.33924	0.42038	0.00997	0.21136	0.00362	0.00674
200	GA	0.05357	0.08656	0.04777	0.00660	0.04467	0.03726	0.04616
	GAM	0.08141	0.26366	0.19700	0.00706	0.06900	0.04510	0.05322
	RIG	0.08141	0.03573	0.31666	0.00573	0.06904	0.10553	0.04963
	LN	0.01069	0.26489	0.32024	0.00334	0.14955	0.00245	0.00433
400	GA	0.04060	0.05247	0.03992	0.00379	0.06005	0.05621	0.06028
	GAM	0.06170	0.21143	0.19394	0.00406	0.08117	0.06416	0.06635
	RIG	0.06170	0.12928	0.17496	0.00308	0.08084	0.07201	0.06873
	LN	0.00810	0.15267	0.22322	0.00329	0.08022	0.00303	0.00230

Table 3.5: Simulation results for the selection of the smoothing parameter by UCV and Bays, case of the Weibull model

The simulation results for the Weibull model are summarized in *Table 3.5*. Key findings are as follows:

Sample size $n = 25$:

- GA Kernel: Achieves the lowest $ISE^* = 0.02149$. The Bayesian method yields $\widehat{ISE}_{Bays} = 0.04393$, making it highly reliable for small samples.

- LN Kernel: performs well using the UCV and *Bayesian* methods with very low values $\widehat{ISE}_{UCV} = 0.00423$ and $\widehat{ISE}_{Bays} = 0.00387$,but struggles with high $ISE^* = 0.37785$, showing sensitivity to bandwidth selection.

- GAM and RIG Kernels: shows weak performance with small n, exhibiting large fluctuations and high ISE values with (e.g., $\widehat{ISE}_{UCV} = 0.17257$ for GAM), but good with Bayesian method yields (e.g.,$\widehat{ISE}_{Bays} = 0.07141$).

Medium Sample Size $n = 100$:

- The LN kernel with UCV provides the best results, with the lowest $\widehat{ISE}_{UCV} = 0.00362$. This kernel outperforms the GA kernel in medium-sized samples.
- Bayesian method is still competitive for small samples but shows slightly higher variability in results.

Large Sample Size $n \in \{200{,}400\}$:

- LN Kernel: Outperforms all others with $\widehat{ISE}_{UCV}$ as low as 0.00245 for $n = 200$ and $\widehat{ISE}_{Bays} = 0.00230$ for $n = 400$, demonstrating high accuracy and stability.
- GA :performs reasonably well , but doesn't achieve the lowest errors compared to LN.
- GAM and RIG Kernels: Continue to show instability and higher variability.
- The Bayesian method achieves the most stable results for GA.

Key Observations:

* In most cases, the Bayesian bandwidth $\hat{h}_{bays}$ is close to the optimal h^* for the GA kernel, although high variability is observed in small samples, particularly for the LN and GAM kernels.
* For the GA kernel, $\hat{h}_{bay}$ aligns closely with h^*, while the LN and GAM kernels exhibit greater variability, especially in small samples.
* The GA kernel with $\hat{h}_{bays}$ shows very good performance at $n = 25$,

the LN kernel gives the best result when $n \in \{100, 200, 400\}$, (with UCV or $Bayes$) achieves the highest accuracy and best stability.

Associated Figures

The figures illustrate the estimated density curves (using GA, GAM, and RIG kernels) superimposed on the true Weibull density (f_3). Key insights:

Gamma Kernel (Figure 3.9):

- Closely matches the true density for small n, aligning with its low ISE values in *Table 3.6*.

- Smoothness depends on $\hat{h}_{bays}$ or $\hat{h}_{ucv}$; Bayesian estimates may show tighter fits.

Modified Gamma (Figure 3.10):

- Higher variability, especially for small n, consistent with its large ISE^* values.

- Potential over-smoothing or under-fitting due to unstable $\hat{h}_{bays}$.

RIG Kernel (Figure 3.11):

- Similar to GAM, with fluctuations and poorer fits for small n.

- Slight improvement for larger n but still less stable than GA or LN.

Conclusion:

- The Bayesian approach is highly competitive, especially for small to medium samples.

- The GA kernel deliver the best results, depending on sample size and method but he GAM and RIG kernels exhibit instability.

- The UCV method is the most stable and accurate in bandwidth estimation, particularly for large n.

Comprehensive comparison of all models:

Aspect	UCV Method	Bayesian Method
Small $n = 25$	High variability, poor approximation especially for complex kernels (LN, RIG)	provides greater stability,lowers ISE, making it more reliable in such situations.
Moderate $n \in \{100, 200\}$	Can be effective with larger n, improvement but still sensitive to kernel choice	Robust smoothing, generally closer to h^* offering stronger smoothing .
Large $n = 400$	Good smoothing, small ISE, still slightly worse than Bayesian	Excellent performance, minimal errors, stable across kernels.
Kernel effect	LN and RIG require careful tuning ,may exhibit performance fluctuations,less reliable with small n.	GA and GAM Provide high stability,reliability with all h selection methods, across all n.

Table 3.7: General Observations Across All Models.

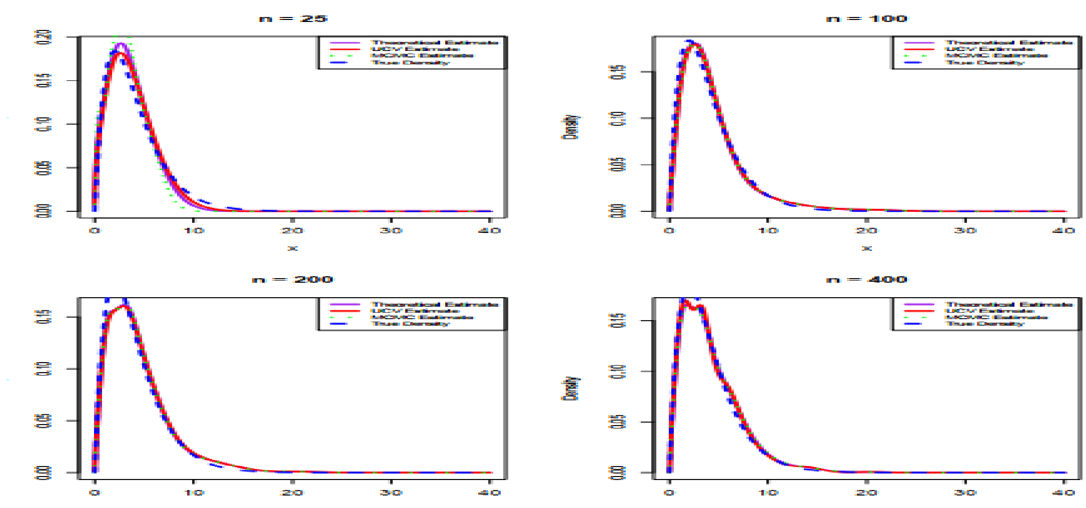

Figure 3.1: Modified Gamma Kernel Estimators of a Gamma Density (f_1).

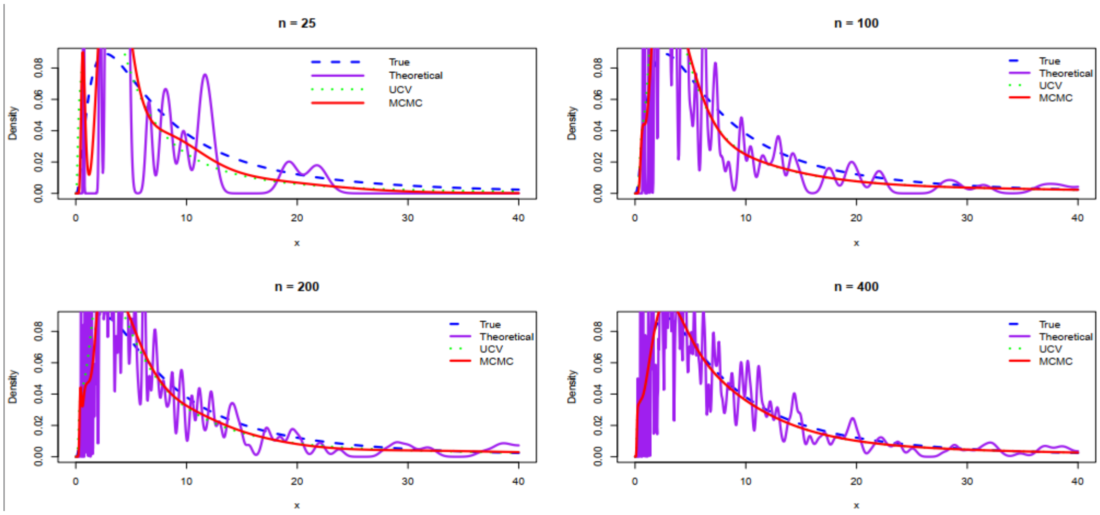

Figure 3.2: Log-Normal Kernel Estimators of a Gamma Density (f_1).

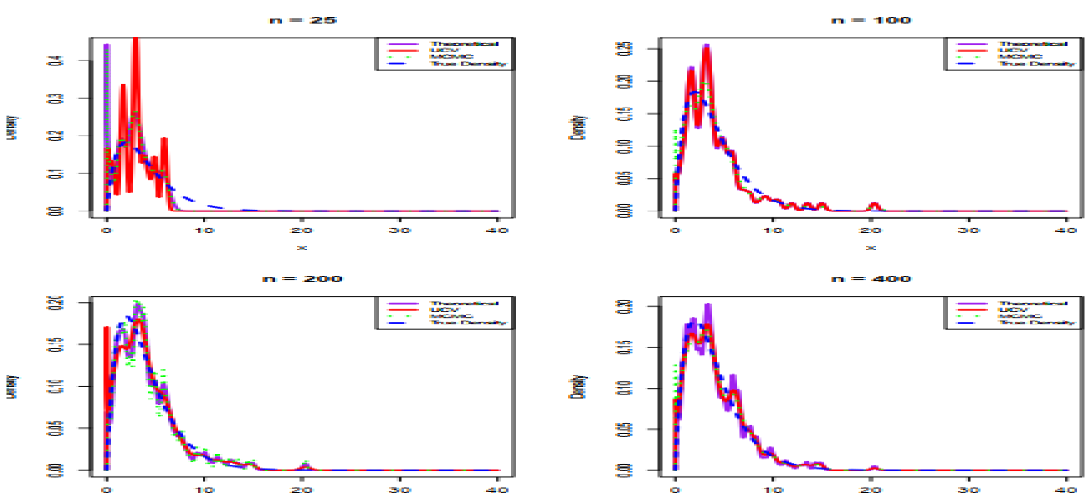

Figure 3.3: RIG Kernel Estimators of a Gamma Density (f_1).

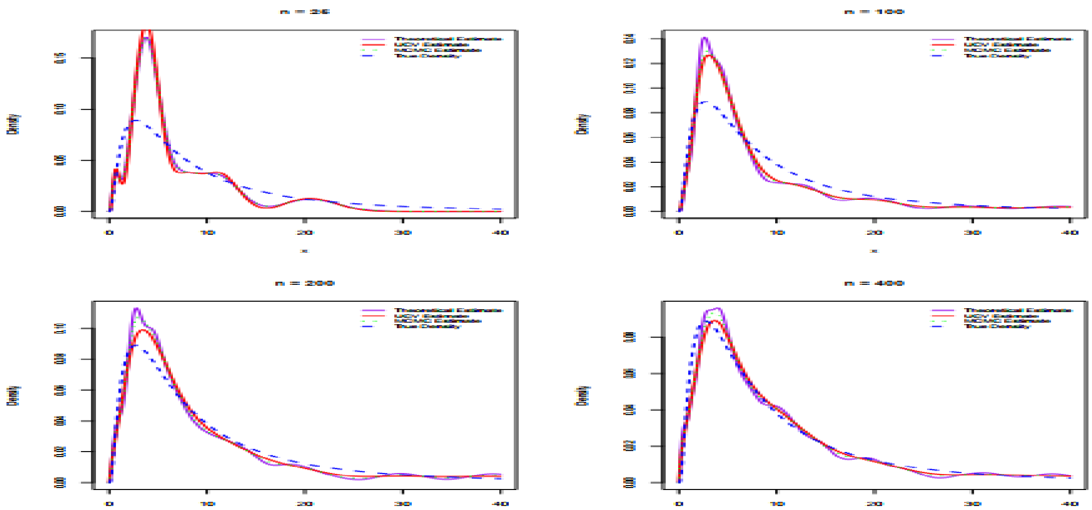

Figure 3.4: Gamma Kernel Estimators of a Log-Normal Density (f_2).

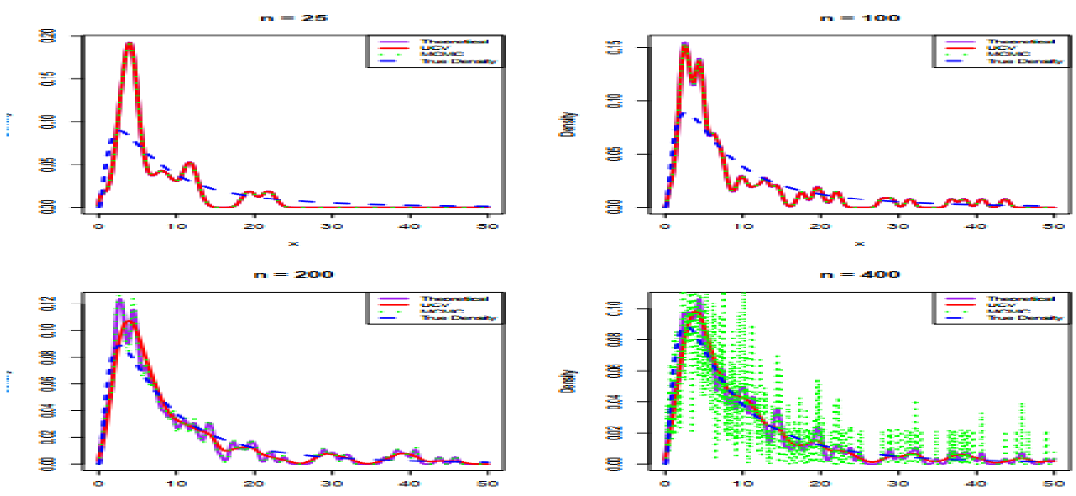

Figure 3.5: Modified Gamma Kernel Estimators of a Log-Normal Density (f_2).

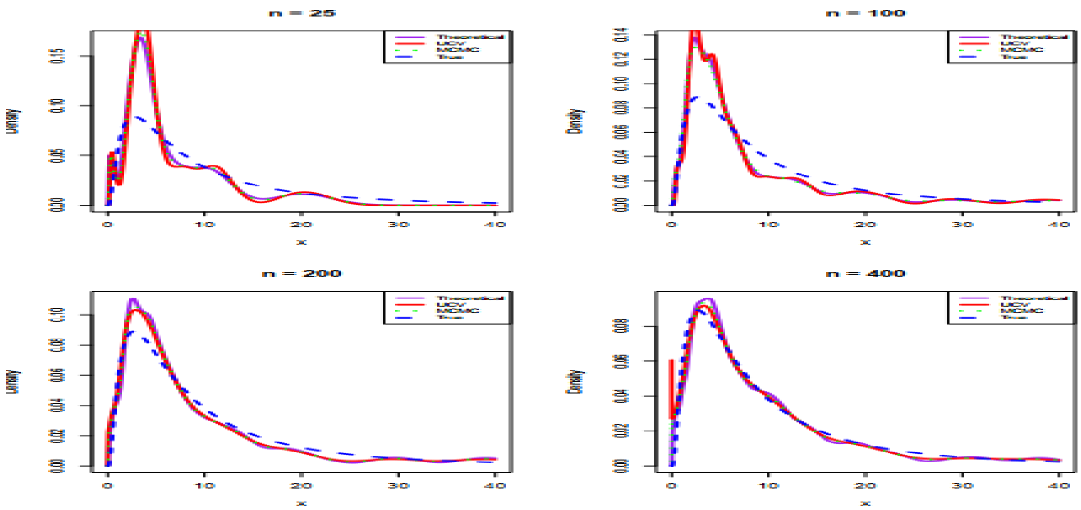

Figure 3.6: RIG Kernel Estimators of a Log-Normal Density (f_2).

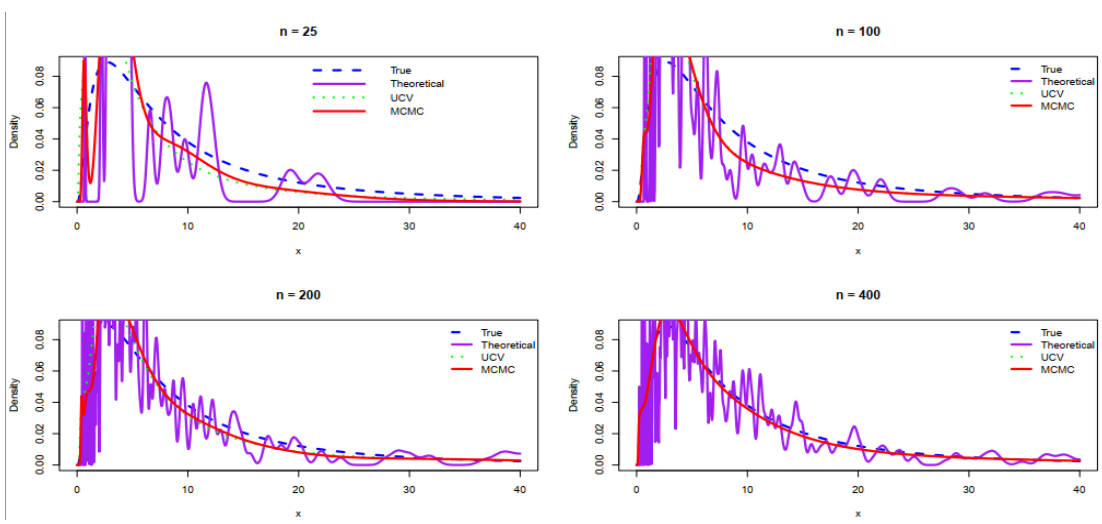

Figure 3.7: LN Kernel Estimators of a Log-Normal Density (f_2).

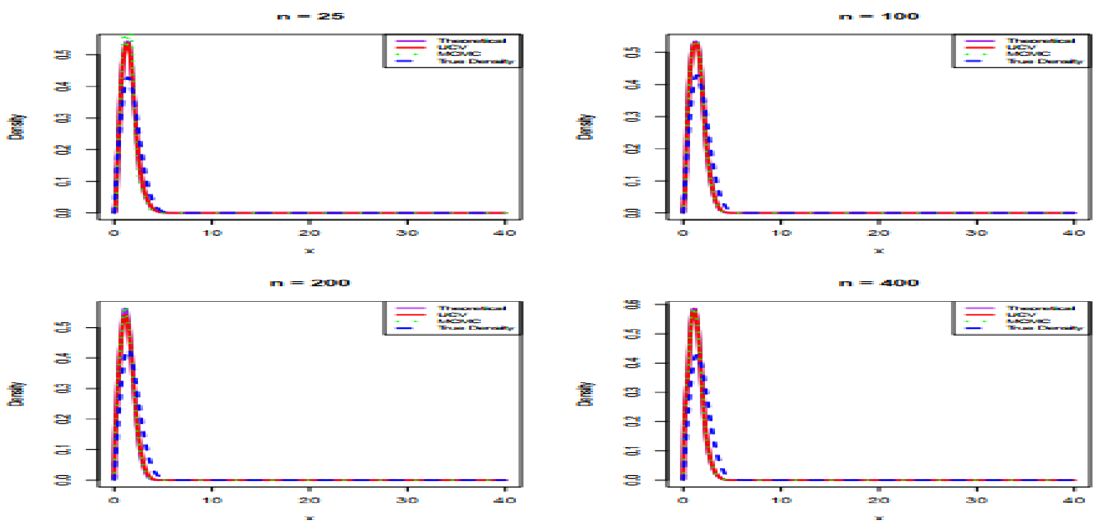

Figure 3.8: Gamma Kernel Estimators of a Weibull Density (f_3).

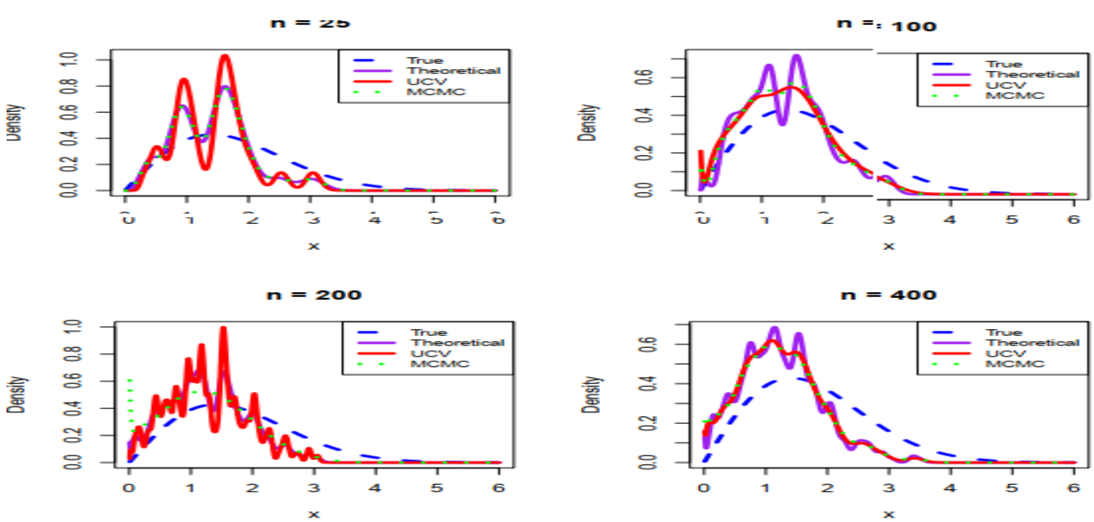

Figure 3.9: RIG Kernel Estimators of a Weibull Density (f_3).

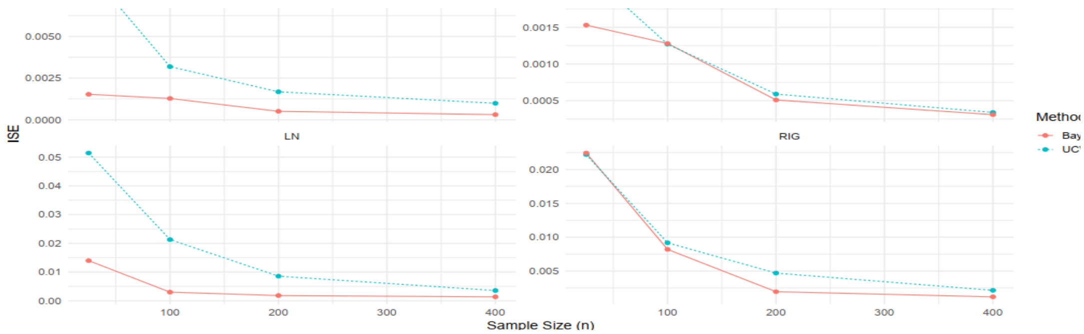

Figure 3.10: ISE vs Sample Size by Method and Kernel (Gamma Model).

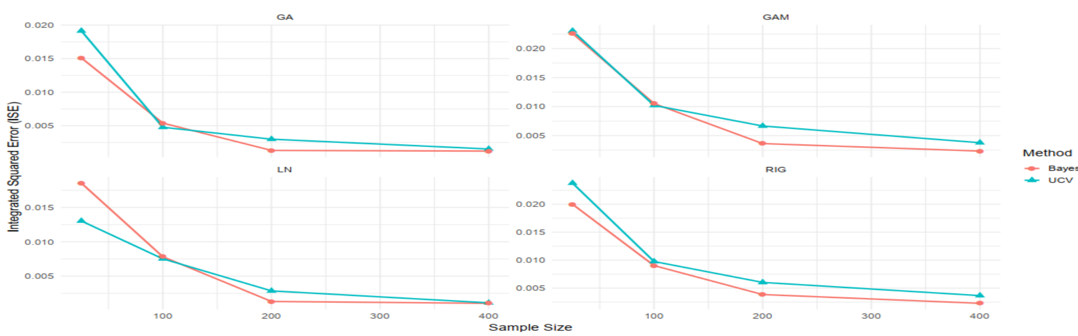

Figure 3.11: ISE vs. Sample Size by Method (Log-normal Model).

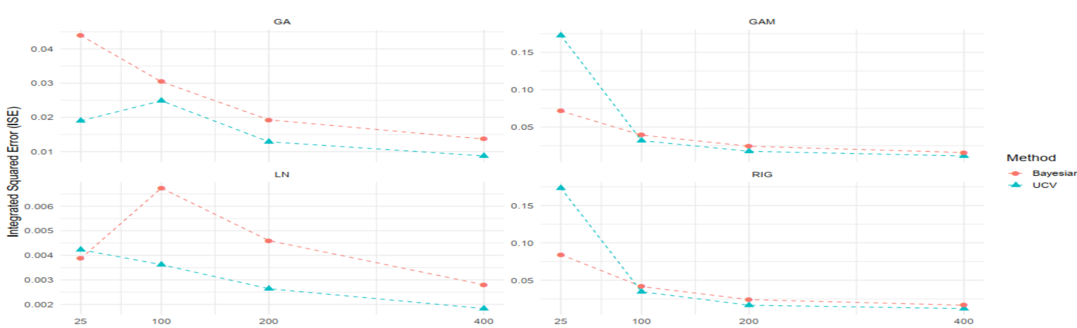

Figure 3.12: ISE vs. Sample Size by Method and Kernel (Weibull Model).

Conclusion

This study focused on comparing classical methods, such as the Plug-in and Cross-Validation approaches, with Bayesian methods for bandwidth selection. The results indicated that classical methods face practical challenges, such as the need for unavailable information or difficult-to-verify assumptions, in addition to instability when applied to small samples. In contrast, the Bayesian approach provides a flexible framework that accounts for uncertainty; however, it requires advanced computational resources, such as *MCMC* techniques, which may pose limitations in some practical scenarios.based on these findings, the study recommends a hybrid strategy: classical methods can be effectively employed in stable situations where the underlying assumptions hold, while the Bayesian approach is more suitable in complex cases or when data is limited. Furthermore, it is recommended to develop new tools and techniques to improve computational efficiency and reduce the cost associated with Bayesian methods, thereby enhancing estimation accuracy across a wider range of practical applications.

Bibliography

[1] *Abadir, K. M., & Lawford, S.* (2004). Optimal asymmetric kernels. Economics Letters, 83(1), 61 − 68.

[2] *Boreux, J. J., Parent, E., & Bernier, J.* (2010). Pratique du calcul bayésien. Springer-Verlag.

[3] *Bouezmarni, T., & Scaillet, O.* (2005). Consistency of asymmetric kernel density estimators and smoothed histograms with application to income data. Econometric Theory, 21, 390 − 412.

[4] *Bowman, A. W.* (1984). An alternative method of cross-validation for the smoothing of density estimates. Biometrika, 71(2), 353 − 360.

[5] *Chen, S. X.* (2000). Gamma kernel estimators for density functions. Annals of the Institute of Statistical Mathematics, 52, 471 − 480.

[6] *Hirukawa, M.* (2018). Asymmetric kernel smoothing: Theory and applications in economics and finance. Springer.https://doi.org/10.1007/978 − 981 − 10 − 5466 − 2.

[7] *Scaillet, O.* (2004). Density estimation using inverse and reciprocal inverse Gaussian kernels. Journal of Nonparametric Statistics, 16, 217 − 226.

[8] *Scott, D. W., & Terrell, G. R.* (1987). Biased and unbiased cross-validation in density estimation. Journal of the American Statistical Association, 82, 1131 – 1146.

ملخص

في هذه الدراسة، قمنا بتقييم أداء الطرائق الكلاسيكية والبايزية في اختيار معامل التنعيم باستخدام أنوية غير متماثلة.أظهرت النتائج أن الطرائق الكلاسيكية تعاني من ضعف في الدقة مع العينات الصغيرة، بينما تقدم الطرائق البايزية، خاصة المعتمدة على تقنيات MCMC، نتائج أفضل في العينات الصغيرة إلى المتوسطة، رغم احتياجها إلى موارد حسابية أكبر. ومن بين الأنوية المدروسة، تفوقت نواة غاما في أداء التقدير.وتشير الدراسة إلى أن الجمع بين الطريقتين يمكن أن يحقق مزايا عملية، خاصة مع تطوير خوارزميات بايزية أكثر كفاءة.

الكلمات المفتاحية: تقدير الكثافة غير المعلمي، الأنوية غير المتماثلة، اختيار معامل التنعيم، النهج البايزي، طرق مونتي كارلو، التحقق المتقاطع، طريقة الإدراج.

Résumé

Dans cette étude, nous avons évalué les performances des méthodes classiques et bayésiennes pour la sélection de la largeur de bande à l'aide de noyaux asymétriques. Les résultats ont montré que les méthodes classiques présentent une faible précision avec de petits échantillons, tandis que les méthodes bayésiennes, en particulier celles basées sur les techniques MCMC, ont donné de meilleurs résultats pour des tailles d'échantillons petites à moyennes, malgré des besoins plus importants en ressources de calcul. Parmi les noyaux étudiés, le noyau Gamma a démontré la meilleure performance en estimation. L'étude suggère que la combinaison des approches classique et bayésienne pourrait offrir des avantages pratiques, notamment avec le développement d'algorithmes bayésiens plus efficaces.

Mots-clés : Estimation de densité non paramétrique, noyaux asymétriques, sélection de bande passante, approche bayésienne,méthodes de Monte Carlo, validation croisée, méthode plug-in.

Abstract

In this study, we evaluated the performance of classical and Bayesian methods for selecting the bandwidth using asymmetric kernels.The results showed that classical methods suffer from low accuracy with small samples, while Bayesian methods, especially those based on MCMC techniques, provided better results for small to medium sample sizes, despite requiring greater computational resources.Among the kernels studied,the Gamma kernel demonstrated the best estimation performance. The study suggests that combining classical and Bayesian approaches could offer practical advantages, particularly with the development of more efficient Bayesian algorithms.

Keywords: Nonparametric density estimation, asymmetric kernels, bandwidth selection, Bayesian approach,Monte Carlo methods, cross-validation, plug-in method.