

République Algérienne Démocratique et Populaire  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

**UNIVERSITÉ MOHAMED KHIDER, BISKRA**

FACULTÉ des SCIENCES EXACTES et des SCIENCES de la NATURE et de la VIE

**DÉPARTEMENT DE MATHÉMATIQUES**



Mémoire présenté en vue de l'obtention du Diplôme :

**MASTER en Mathématiques**

Option : **Statistique**

Par

**SADAOUI Sihem Salima**

Titre :

**Choix bayésien du paramètre de lissage dans  
l'estimation à noyaux discrets.**

Membres du Comité d'Examen :

Dr. SAYAH ABDALLAH	UMKB	Président
Dr. CHERFAOUI Mouloud	UMKB	Encadreur
Dr. OUANOUGHY YASMINA	UMKB	Examinatrice

Juin 2020

## DÉDICACE

*Je dédie ce modeste travail :*

*A mes chers parents qui ont toujours été présent pour moi et qui m'ont donné la vie.*

*Pour de millions des raisons, ils m'ont donné jour après jour autant d'amour et de confiance, ils ont veillé à m'encourager tout au long de ma vie, à me donner de l'aide et à me protéger.*

*Que le grand Dieu me les protèges.*

*A mes chers frères et soeurs, ma source de joie et de bonheur,*

*A mes oncles, mes tantes, mes cousins et cousines,*

*A tous mes amis et camarades,*

*A tous ceux qui me sont chers,*

*A tous ceux qui m'aiment,*

*Et à tous ceux que j'aime.*

Sihem

## REMERCIEMENTS

*Merci ALLAH (mon Dieu) de m'avoir donné : la capacité d'écrire et de réfléchir, la force d'y croire, le courage et la patience durant mes longues années d'étude et le bonheur d'en progresser, et de réussir et d'arriver jusqu'au bout de mes rêves.*

*Je tiens à remercier vivement ma très chère mère pour sa patience et ses sacrifices, mon père, et toute ma famille.*

*Je tiens à remercier chaleureusement mon encadreur, Monsieur **CHERFAOUI Mouloud**, pour la qualité de son encadrement, du soutien qu'il m'a accordé durant ma formation, et de sa contribution à l'élaboration de ce mémoire.*

*Je voudrais remercier également les membres de jury de m'avoir fait l'honneur en acceptant de juger ce travail.*

*Je tiens à remercier, tous les enseignants qui ont contribué dans ma formation du primaire jusqu'à l'université.*

*Merci également à la doctorante :**BOUREDJI Hind** pour ses encouragements, ses conseils et sa précieuse aide.*

*Enfin, je tiens à exprimer ma profonde gratitude à mes amis et toutes les personnes qui m'ont apportée une aide pour la réalisation de ce travail.*

# Table des matières

Remerciements	ii
Table des matières	ii
Table des figures	v
Liste des tables	vi
Introduction	1
<b>1 Estimation à noyau discret d'une fonction de masse</b>	<b>4</b>
Introduction . . . . .	4
1.1 Notion de noyau discret . . . . .	4
1.2 Choix du noyau . . . . .	7
1.2.1 Noyau binomial . . . . .	7
1.2.2 Noyau binomial négatif . . . . .	8
1.2.3 Noyau de Poissonien . . . . .	9
1.2.4 Noyau Triangulaire . . . . .	9
1.3 Choix du paramètre de lissage . . . . .	10
Conclusion . . . . .	12
<b>2 Choix local et global du paramètre de lissage par l'approche bayésienne</b>	<b>13</b>
Introduction . . . . .	13

2.1	Principe Bayésien . . . . .	13
2.1.1	Quelques définitions . . . . .	14
2.1.2	Choix de lois a priori . . . . .	14
2.1.3	Estimateur de Bayes . . . . .	17
2.1.4	Méthodes de Monté Carlo par Chaîne de Markov (MCMC) . . . . .	18
2.2	Approche bayésienne pour la sélection du paramètre de lissage . . . . .	20
2.2.1	Approche bayésienne globale . . . . .	20
2.2.2	Approche bayésienne locale . . . . .	25
	Conclusion . . . . .	26
<b>3</b>	<b>Etude comparative : Application Numérique</b>	<b>27</b>
	Introduction . . . . .	27
3.1	Plan de simulation . . . . .	27
3.1.1	Paramètre de simulation . . . . .	28
3.2	Résultats et discussion . . . . .	30
	Conclusion . . . . .	35
	<b>Conclusion générale</b>	<b>37</b>
	<b>Bibliographie</b>	<b>39</b>
	<b>Annexe B : Abréviations et Notations</b>	<b>41</b>

# Table des figures

3.1	Diagrammes des distributions cibles . . . . .	29
3.2	Variation de $h^*$ et $\overline{ISE}$ en fonction de la taille de l'échantillon $n$ , cas : $K_T$ et $f_1$ . . . . .	34
3.3	Variation de $h^*$ et $\overline{ISE}$ en fonction de la taille de l'échantillon $n$ , cas : $K_T$ et $f_2$ . . . . .	35
3.4	Variation de $h^*$ et $\overline{ISE}$ en fonction de la taille de l'échantillon $n$ , cas : $K_T$ et $f_3$ . . . . .	35

# Liste des tableaux

2.1	Quelques densités de lois a priori conjuguée . . . . .	16
3.1	Résultats obtenus pour le cas $f_1$ . . . . .	32
3.2	Résultats obtenus pour le cas $f_2$ . . . . .	33
3.3	Résultats obtenus pour le cas $f_3$ . . . . .	33
3.4	Résultats obtenus pour le cas $f_4$ . . . . .	34

# Introduction Générale

Il existe plusieurs méthodes d'estimation non-paramétrique de la fonction de densité. Nous citons la méthode d'estimation par histogramme, la méthode des  $k$  plus proches voisins, la méthode des fonctions orthogonales et la méthode d'estimation par noyau. Mais ça reste que l'estimateur le plus populaire et qui a connu beaucoup de succès est bien que l'estimateur à noyau. Ce dernier peut s'expliquer par au moins trois raisons : premièrement, l'expression théorique de l'estimateur est très simple puisqu'il s'écrit comme la somme de  $n$  variables aléatoires indépendantes et identiquement distribuées (*i.i.d*), en utilisant la fonction noyau  $K$  et le paramètre de lissage  $h$ . Deuxièmement, elle possède de bonnes propriétés asymptotiques. Enfin, l'estimateur à noyau est flexible, car il laisse à l'utilisateur une grande latitude dans le choix du noyau  $K$  et du paramètre  $h$ . L'estimateur à noyau a été proposé initialement par Rosenblatt [12] et Parzen [13] pour estimer une fonction de densité  $f$  à support non borné et généralisée par Kokonendji et al. [10] et [9] pour des densités à support quelconque (non borné, semi-borné, borné, discret).

L'estimation de la fonction de probabilité par la méthode du noyau est principalement conditionnée par un choix préalable du noyau  $K$  et du paramètre de lissage  $h$ . Quoique le choix du noyau engendre un petit problème mais le vrai problème est dans le choix du paramètre de lissage. En effet, ce paramètre est un facteur important, il représente en quelque sorte une fenêtre qui permet de déterminer le degré de lissage de la courbe de l'estimateur où de petites ou grandes valeurs de de cette fenêtre peuvent conduire à une estimation sous-lissée ou sur-lissée.

Pour un choix approprié du paramètre de lissage, dans la littérature deux catégories de méthodes ont été proposées. La première catégorie repose sur la minimisation de l'erreur quadratique moyenne intégrée *MISE*. Cette classe de méthodes est intéressante en théorie, mais sa difficulté majeure réside dans les applications. La seconde catégorie est de type validation croisée, elle est intéressante en pratique car elle se base uniquement sur les observations. Cependant, l'inconvénient principal de ces deux catégories est qu'elles ont tendance à fournir des estimateurs sous ou sur-lisses lorsque les échantillons sont de petite ou de moyenne taille, ou lorsque on veut estimer des densités complexes.

L'objectif de ce travail est de présenter une approche alternative pour le choix du paramètre de lissage dite "approche bayésienne" introduite pour remédier aux insuffisances des méthodes classiques. Contrairement aux méthodes classiques cette approche est caractérisée par la considération du paramètre de lissage  $h$  comme une variable aléatoire positive, en lui associant une loi a priori qui sert à compenser le manque d'information. Plus précisément, nous proposons l'approche bayésienne globale pour le choix du paramètre de lissage  $h$ , dans l'estimation à noyau de la fonction de densité à support discret et pour surmonter le problème de calcul de la loi a posteriori qui est de forme complexe nous faisons appel aux méthodes de Monte Carlo par Chaînes de Markov (*MCMC*). Pour répondre à notre objectif, en plus de la présente introduction, nous avons réparti le présent travail en trois chapitres, une Conclusion générale et une liste de références bibliographiques.

- Dans le premier chapitre nous présentons quelques détails sur la méthode d'estimation à noyau d'une fonction de densité discrètes (définitions, propriétés statistique,...) et nous présentons également quelques procédures classiques proposées dans la littérature pour la sélection du paramètre de lissage dans ce genre de situations.
- Le deuxième chapitre concerne l'approche bayésienne pour l'estimation du paramètre de lissage dans l'estimation à noyau de la fonction de masse. Nous rappelons dans un premier lieu quelques notions liées particulièrement à l'inférence bayésienne. Par la

suite, nous présentons en détail l'adaptation de l'inférence bayésienne au problème du choix du paramètre de lissage dans deux cas, à savoir : le choix global et le choix local.

- Avant de conclure et de présenter quelques perspectives complémentaire au présent travail, dans le troisième chapitre nous présentons les résultats de l'étude de simulation réalisée dans le cadre de la comparaison des performances de la méthode de validation croisée (*UCV*) et de l'approche bayésienne globale.

# Chapitre 1

## Estimation à noyau discret d'une fonction de masse

### Introduction

Dans ce chapitre, nous exposons la notion du noyau associé  $K_{x,h}$  de cible  $x$ , et du paramètre de lissage  $h$ . Cette notion a été introduite par Kokonendji et al. [10] et [9] dans le cas discret. En particulier, notre intérêt sera orienté vers la question du choix du noyau et du paramètre de lissage dans le cas discret. Les propriétés de quelques estimateurs conçus à l'aide des noyaux discrets usuels seront également présentées.

### 1.1 Notion de noyau discret

Les deux définitions suivantes présentent la notion du *noyau discret* et celle du *l'estimateur à noyau discret*, respectivement.

**Définition 1** Soit  $\mathfrak{N}_x$  le support d'une fonction de masse de probabilité  $f$  à estimer. Étant donné  $x \in \mathfrak{N}$  et  $h > 0$ , on appelle "**noyau discret**"  $K_{x,h}(\cdot)$ , toute fonction de masse de probabilité liée à la variable aléatoire discrète  $K_{x,h}$  de support  $\mathfrak{N}_x$  contenant au moins  $x$  et

indépendant de  $h$ , vérifiant les quatre conditions :

1.  $\aleph \subseteq \cup \aleph_x$  .
2.  $E(\mathcal{K}_{x,h}) \sim x$  lorsque  $h \rightarrow 0$ .
3.  $Var(\mathcal{K}_{x,h}) < +\infty$ .
4.  $Var(\mathcal{K}_{x,h}) \rightarrow 0$  lorsque  $h \rightarrow 0$ .

**Définition 2** Soit  $X_1, X_2, \dots, X_n$  un  $n$ -échantillon indépendant et identiquement distribué (i.i.d) issu d'une variable aléatoire  $X$  de la fonction de masse de probabilité inconnue  $f$  sur  $\aleph$ . L'**estimateur à noyau discret** de  $f$  est défini par :

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

où  $h$  est le paramètre de lissage (fenêtre) et  $K_{x,h}$  est le noyau discret dépendant de  $x$  et  $h$  de support  $\aleph_{x,h} = \aleph_x$  (ne dépend pas de  $h$ ).

**Proposition 1** Soit  $X_1, X_2, \dots, X_n$  un  $n$ -échantillon i.i.d issu d'une variable aléatoire  $X$  de la fonction de masse de probabilité inconnue  $f$  sur  $\aleph$ . Si  $\hat{f}$  est l'estimateur à noyau discret de  $f$ , alors, pour  $x \in \aleph$  et  $h > 0$  on a :

$$E(\hat{f}(x)) = E(f(\mathcal{K}_{x,h})), \quad (1.2)$$

avec  $\mathcal{K}_{x,h}$  est la variable aléatoire de loi  $K_{x,h}$  définie sur  $\aleph_x$ . De plus, on a  $\hat{f}(x) \in [0 ; 1]$  pour  $x \in \aleph$  et

$$E(\hat{f}(x)) = \sum_{y \in \aleph \cap \aleph_x} f(y)K_{x,h}(y) \rightarrow f(x) \text{ quand } h \rightarrow 0 \text{ lorsque } n \rightarrow +\infty \quad (1.3)$$

La proposition suivante résume les expressions du *MSE* et du *MISE* de l'estimateur à noyau discret d'une fonction de masse ainsi que les conditions de convergence de ces deux quantités.

**Proposition 2** Soit  $X_1, X_2, \dots, X_n$  un  $n$ -échantillon i.i.d issu d'une variable aléatoire  $X$  de la fonction de masse inconnue  $f$  sur  $\mathbb{N}$ . Si  $\hat{f}$  l'estimateur à noyau associé discret de  $f$ , alors, pour  $x \in \mathbb{N}$  et  $h > 0$  on a :

– **L'erreur quadratique moyenne (MSE) :**

$$\begin{aligned} MSE(\hat{f}(x)) &= E \left[ \hat{f}(x) - f(x) \right]^2 = Var(\hat{f}_h(x)) + Biases^2(\hat{f}(x)), \\ &= \frac{1}{n} f(x) \left[ (\Pr(\mathcal{K}_{x,h} = x))^2 - f(x) \right] \\ &+ \left[ f(E(\mathcal{K}_{x,h})) - f(x) + \frac{1}{2} Var(\mathcal{K}_{x,h}) f^{(2)}(x) \right]^2 + o\left(\frac{1}{nh} + h^2\right). \end{aligned} \quad (1.4)$$

– **L'erreur quadratique moyenne intégrée (MISE) :**

$$\begin{aligned} MISE(\hat{f}) &= \sum_{x \in \mathbb{N}} MSE(f(x), \hat{f}(x)) = \sum_{x \in \mathbb{N}} Var(\hat{f}(x)) + \sum_{x \in \mathbb{N}} Biases^2(\hat{f}_h(x)), \\ &= MISE(n, h, K, f). \end{aligned} \quad (1.5)$$

où quantité  $f^{(2)}$  représentent la différence finie d'ordre 2, par :

$$f^{(2)}(x) = \begin{cases} \{f(x+2) - 2f(x) + f(x-2)\} / 4, & \text{si } x \in \mathbb{N} / \{0, 1\}; \\ \{f(3) - 3f(1) + 2f(0)\} / 4, & \text{si } x = 1; \\ \{f(2) - 2f(1) + f(0)\} / 2, & \text{si } x = 0. \end{cases} \quad (1.6)$$

De plus,  $MSE \rightarrow 0$  et  $MISE \rightarrow 0$  quand  $nh \rightarrow +\infty$  et  $h \rightarrow 0$ .

Rappelons que la différence finie d'ordre 1 est donnée par :

$$f^{(1)}(x) = \begin{cases} \{f(x+1) - f(x-1)\} / 2, & \text{si } x \in \mathbb{N} / \{0\}; \\ f(1) - f(0), & \text{si } x = 0. \end{cases} \quad (1.7)$$

## 1.2 Choix du noyau

Dans cette partie, nous allons présenter les noyaux les plus usités dans le cadre d'estimation d'une fonction de masse ainsi que les propriétés des estimateurs fournis par chaque noyau.

### 1.2.1 Noyau binomial

Soit  $K_{B(x+1,(x+h)/x+1)}$  le noyau de loi Binomiale définie sur le support  $\mathbb{N}_{x,h} = \{0, 1, \dots, x+1\}$  est donné par :

$$K_{B(x+1,(x+h)/x+1)}(y) = \frac{(x+1)!}{y!(x+1-y)!} \left(\frac{x+h}{x+1}\right)^y \left(\frac{1-h}{1+x}\right)^{x+1-y} \mathbf{1}_{\{y \leq x+1\}}.$$

avec  $x \in \mathbb{N}$ ,  $h \in ]0 ; 1]$  et  $\mathbf{1}$  est la fonction indicatrice.

**Propriété 1** Soit  $\hat{f}(x)$  l'estimateur d'une fonction de masse. Si l'estimateur  $\hat{f}(x)$  est construit à l'aide d'un noyau **Binomial**, alors l'expression de cet estimateur, celles de son biais ponctuelle et de sa variance ponctuelle et celle de son MISE sont données respectivement par :

$$\begin{aligned} \hat{f}(x) &= \frac{1}{n} \sum_{i=1}^n \frac{(x+1)}{X_i! (x+1-X_i)!} \left(\frac{x+h}{x+1}\right)^{X_i} \left(\frac{1-h}{1+x}\right)^{x+1-X_i}, \\ \text{Biais}(\hat{f}(x)) &= f(x) \{K_{B(x,h)}(x) - 1\} + \sum_{y \in \mathbb{N}_x \setminus \{x\}} f(y) K_{B(x,h)}(y), \\ \text{Var}(\hat{f}(x)) &= \frac{1}{n} \left( f(x) K_{B(x,h)}^2(x) + \sum_{y \in \mathbb{N}_x \setminus \{x\}} f(y) K_{B(x,h)}^2(y) \right. \\ &\quad \left. - \left\{ f(x) + \sum_{y \in \mathbb{N}_x} [f(y) - f(x)] K_{B(x,h)}(y) \right\}^2 \right), \\ \text{MISE}(\hat{f}(x)) &= \sum_{x \in \mathbb{N}} \left\{ h f^{(1)}(x) + \frac{1}{2} (x+h) \left(\frac{1-h}{x+1}\right) f^{(2)}(x) + o(h) \right\}^2 \\ &\quad + \frac{1-h}{n} \sum_{x \in \mathbb{N}} \left(\frac{x+h}{x+1}\right)^x f(x). \end{aligned}$$

où  $f^{(1)}(x)$  et  $f^{(2)}(x)$  sont données respectivement par (1.7) et (1.6).

## 1.2.2 Noyau binomial négatif

Soit  $K_{BN(x+1,(x+1)/(2x+1+h))}$  le noyau associé de loi Binomiale négative définie sur  $\aleph_{x,h} = \mathbb{N}$  tel que :

$$K_{BN(x+1,(x+1)/(2x+1+h))}(y) = \frac{(x+y)!}{y!x!} \left( \frac{x+h}{2x+1+h} \right)^y \left( \frac{x+1}{2x+1+h} \right)^{x+1},$$

avec  $x, y \in \mathbb{N}$  et  $h > 0$ .

**Propriété 2** Soit  $\hat{f}(x)$  l'estimateur d'une fonction de masse. Si l'estimateur  $\hat{f}(x)$  est construit à l'aide d'un noyau **binomial négatif**, alors l'expression de cet estimateur, celles de son biais ponctuelle, de sa variance ponctuelle et celle de son MISE sont données respectivement par :

$$\begin{aligned} \hat{f}(x) &= \frac{1}{n} \sum_{i=1}^n \frac{(x+X_i)!}{X_i!x!} \left( \frac{x+h}{2x+1+h} \right)^{X_i} \left( \frac{x+1}{2x+1+h} \right)^{x+1}, \\ \text{Biais}(\hat{f}(x)) &= f(x) \{K_{BN(x,h)}(x) - 1\} + \sum_{y \in \mathbb{N} - \{x\}} f(y) K_{BN(x,h)}(y), \\ \text{Var}(\hat{f}(x)) &= \frac{1}{n} \left( f(x) K_{BN(x,h)}^2(x) + \sum_{y \in \mathbb{N} - \{x\}} f(y) K_{BN(x,h)}^2(y) \right. \\ &\quad \left. - \left\{ f(x) + \sum_{y \in \mathbb{N}} [f(y) - f(x)] K_{BN(x,h)}(y) \right\}^2 \right), \\ \text{MISE}(\hat{f}) &= \frac{1}{n} \sum_{x \in \mathbb{N}} \frac{2}{x!} \left( \frac{x+h}{2x+1+h} \right)^x \left( \frac{x+1}{2x+1+h} \right)^{x+1} f(x) \\ &\quad + \sum_{x \in \mathbb{N}} \left( h f^{(1)}(x) + \frac{1}{2} (x+h) \left( \frac{2x+1+h}{x+1} \right) f^{(2)}(x) \right)^2 + o(h). \end{aligned}$$

où  $f^{(1)}(x)$  et  $f^{(2)}(x)$  sont définies respectivement par (1.7) et (1.6).

### 1.2.3 Noyau de Poissonien

Soit  $K_{P(x+h)}$  le noyau de Poisson définie sur  $\aleph_{x,h} = \mathbb{N}$  tel que :

$$K_{P(x+h)}(y) = e^{-(x+h)} \frac{(x+h)^y}{y!},$$

avec  $(x, y) \in \mathbb{N}^2$  et  $h > 0$  est le paramètre de lissage.

**Propriété 3** Soit  $\hat{f}_h(x)$  l'estimateur d'une fonction de masse. Si l'estimateur  $\hat{f}_h(x)$  est construit à l'aide d'un noyau **Poissonien**, alors l'expression de cet estimateur, celles de son biais ponctuelle, de sa variance ponctuelle et de son MISE sont données respectivement par :

$$\begin{aligned} \hat{f}_h(x) &= \frac{1}{n} \sum_{i=1}^n \left( e^{-(x+h)} \frac{(x+h)^{X_i}}{X_i!} \right), \\ \text{Biais}(\hat{f}(x)) &= f(x) \{P_{x,h}(x) - 1\} + \sum_{y \in \mathbb{N} - \{x\}} f(y) P_{x,h}(y), \\ \text{Var}(\hat{f}_h(x)) &= \frac{1}{n} \left( f(x) P_{x,h}^2(x) + \sum_{y \in \mathbb{N} / \{x\}} f(y) P_{x,h}^2(y) - \left\{ f(x) + \sum_{y \in \mathbb{N}} [f(y) - f(x)] P_{x,h}(y) \right\}^2 \right), \\ \text{MISE}(\hat{f}) &= \frac{1}{n} \sum_{x \in \aleph_x} f(x) \frac{(x+h)^x}{x!} e^{-(x+h)} + \sum_{x \in \aleph_x} \left( h f^{(1)}(x) + \frac{1}{2} (x+h) f^{(2)}(x) + o(h) \right)^2. \end{aligned}$$

où les expressions de  $f^{(1)}(x)$  et  $f^{(2)}(x)$  sont données dans (1.7) et (1.6), respectivement.

### 1.2.4 Noyau Triangulaire

Soit  $K_{T(a,h,x)}$  le noyau triangulaire défini sur  $\aleph_{a,x} = \{x, x \pm 1, \dots, x \pm a\}$ , tel que :

$$K_{T(a,h,x)}(y) = \frac{(a+1)^h - |y-x|^h}{a} \mathbf{1}_{\{|y-x| < a\}},$$

$$(2a+1)(a+1)^h - 2 \sum_{j=0}^a j^h$$

avec  $x \in \mathbb{N}$ ,  $h > 0$  et  $a \in \mathbb{N}$ .

**Propriété 4** Soit  $\hat{f}_h(x)$  l'estimateur d'une fonction de masse. Si l'estimateur  $\hat{f}_h(x)$  est

construit à l'aide d'un noyau **Triangulaire**, alors pour tout  $x \in \mathbb{N}$ ,  $a \in \mathbb{N}$ ,  $h > 0$  l'expression de cet estimateur, celles de son biais ponctuelle, de sa variance ponctuelle et de son MISE sont données respectivement par :

$$\begin{aligned}
 \hat{f}_h(x) &= \frac{1}{n} \sum_{i=1}^n \left( \frac{(a+1)^h - |X_i - x|^h}{(2a+1)(a+1)^h - 2 \sum_{j=0}^a j^h} \mathbf{1}_{\{|y-x| < a\}} \right), \\
 \text{Biais}(\hat{f}(x)) &= \left[ \frac{a(a+1)^{h+1}(2a+1)}{6P_0(a,h)} \right] - 2 \sum_{i=0}^a i^{h+2} f^{(2)}(x) + o(h), \\
 \text{Var}(\hat{f}_h(x)) &= \frac{(a+1)^h}{nP_0(a,h)} f(x), \\
 \text{MISE}(\hat{f}) &= \sum_{x \in \mathbb{N}} \left\{ \frac{1}{2} \frac{1}{P_0(a,h)} \left[ \frac{a(a+1)^{h+1}(2a+1)}{3} \right] - 2 \sum_{i=0}^a i^{h+2} f^{(2)}(x) + o(h) \right\}^2, \\
 &\quad + \frac{(a+1)^h}{nP_0(a,h)} \sum_{x \in \mathbb{N}} f(x), \tag{1.8}
 \end{aligned}$$

avec  $P_0(a, h)$  est la constante de normalisation donnée par :

$$P_0(a, h) = (2a+1)(a+1)^h - 2 \sum_{j=0}^a j^h,$$

et  $f^{(2)}(x)$  la différence finie d'ordre 2 donnée dans (1.6).

### 1.3 Choix du paramètre de lissage

Dans cette partie nous présentons quelques méthodes classiques pour le choix du paramètre de lissage dans le cadre de l'estimation d'une fonction de masse.

1. **Minimisation du MISE** : Soit  $X = (X_1, \dots, X_n)$  un  $n$  échantillons fixé *i.i.d* de distribution inconnue  $f$  alors l'erreur quadratique intégrée (*ISE*) est donné par :

$$ISE = \sum_{x \in \mathbb{N}} (\hat{f}(x) - f(x))^2 = ISE(n, h, K, f), \tag{1.9}$$

ainsi (1.9) conduit à choisir une fenêtre adéquate :

$$h_{ise}^* = \arg \min_h ISE(X, h, K, f), \quad (1.10)$$

pour laquelle la mesure est sur un seul échantillon et la fenêtre optimale  $h_{opt}^*$  peut être obtenue, dans le cas de plusieurs échantillons, à travers

$$h_{opt}^* = \arg \min_{h>0} E(ISE(X, h, K, f)) = \arg \min_{h>0} MISE(X, h, K, f). \quad (1.11)$$

2. **Validation croisée** : Nous suggérons ici deux techniques qui se basent sur la méthode de validation croisée.

(a) **Validation croisée par les moindres carrés** : La fenêtre optimale, dans ce cas, est donnée par :

$$\begin{aligned} h_{ucv}^* &= \arg \min_{h>0} CV(h) = \arg \min_{h>0} \left[ \sum_{x \in \mathbb{N}} \left\{ \hat{f}(x) \right\}^2 - \frac{2}{n} \sum_{i=1}^n \hat{f}_i(X_i) \right] \\ &= \arg \min_{h>0} \left[ \sum_{x \in \mathbb{N}} \left\{ \frac{1}{n} \sum_{i=1}^n K_{x,h}(X_i) \right\}^2 - \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{j=1, j \neq i}^n K_{X_i,h}(X_j) \right], \end{aligned} \quad (1.12)$$

avec

$$\hat{f}_i(X_i) = \frac{1}{n-1} \sum_{j=1, j \neq i}^n K_{X_i,h}(X_j)$$

(b) **Validation croisée par le maximum de vraisemblance** : Ce critère consiste à choisir  $h$  qui maximise

$$LCV(h) = \frac{1}{n} \sum_{i=1}^n \log \left( \hat{f}_i(X_i) \right). \quad (1.13)$$

Ainsi, on détermine une fenêtre optimale  $h_{LCV}$  par :  $h_{LCV} = \arg \max_{h>0} LCV(h)$ .

3. **Excès des zéros**

Nous pouvons choisir une fenêtre adaptés  $h_0 = h_0(X, K)$  tel que  $h$  satisfait :

$$\sum_{i=1}^n \Pr(\mathcal{K}_{X_i, h_0} = 0) = n_0, \quad (1.14)$$

avec  $n_0 = \text{card}\{X_i = 0\}$  désigne le nombre des zéros dans l'échantillon  $X_1, \dots, X_n$  à condition que  $n_0 > 0$ . Quelques exemples de  $h_0$  :

- Si le noyau utilisé est Poissonnien alors  $h_0 = \log \left( \frac{1}{n_0} \sum_{i=1}^n e^{-X_i} \right)$ .
- Si le noyau utilisé est Binomial alors le  $h_0$  est la solution de  $n_0 = \sum_{i=1}^n \left( \frac{1-h}{X_i+1} \right)^{X_i+1}$ .
- Si le noyau utilisé est Binomial négatif alors le  $h_0$  est la solution de  $n_0 = \sum_{i=1}^n \left( \frac{X_i+1}{2X_i+1+h} \right)^{X_i+1}$ .

**Remarque 1** *La solution d'excès des zéros n'existe pas dans le noyau triangulaire car l'équation :*

$$\sum_{i=1}^n \Pr(\mathcal{K}_{X_i, h_0} = 0) = n_0,$$

*n'admet pas de solution (pour plus de détails voir [10, 9]).*

## Conclusion

A partir de ce chapitre on constate que pour le choix du paramètre de lissage on dispose de deux catégories de méthodes. La première catégorie repose sur la minimisation de l'erreur quadratique moyenne intégrée *MISE*. Cette classe de méthodes est intéressante en théorie, mais sa difficulté majeure réside dans les applications. La seconde catégorie est de type validation croisée, elle est intéressante en pratique car elle se guide seulement par les observations. Cependant, l'inconvénient principal de ces deux catégories est qu'elles ont tendance à fournir des estimateurs sous ou sur-lissés lorsque les échantillons sont de petite ou de moyenne taille, ou encore lorsque on veut estimer des fonctions complexes.

Une récente alternative pour le choix du paramètre de lissage  $h$  est l'approche bayésienne. Les étapes de la mise en oeuvre de cette dernière technique fait l'objet du prochain chapitre.

# Chapitre 2

## Choix local et global du paramètre de lissage par l'approche bayésienne

### Introduction

Dans ce chapitre, nous allons illustrer la démarche à suivre pour la quantification du paramètre de lissage  $h$  par l'approche bayésienne. Pour cela, nous présentons d'abord quelques notions sur l'inférence bayésienne tel : le choix de lois a priori, l'estimateur bayésien d'un paramètre,... Par la suite, les méthodes de Monte Carlo par chaînes de Markov (*MCMC*) qui visent à surpassé les problèmes des calculs dans l'inférence bayésienne seront brièvement présenter. Enfin, avons de conclure nous exposons les étapes du choix du paramètre de lissage par l'inférence bayésienne globale et locale.

### 2.1 Principe Bayésien

Supposons que  $x = (x_1, x_2, \dots, x_n)$  est le vecteur d'observation et  $\theta = (\theta_1, \theta_2, \dots, \theta_d) \in \Theta$  est le vecteur des paramètres à estimer.

L'idée principale de l'analyse bayésienne repose sur la **loi a posteriori** des paramètres en considérons  $\theta$  comme variable aléatoire. L'espace des paramètres  $\Theta$  est muni d'une loi

de probabilité  $\pi$  et nous noterons  $\theta \sim \pi$ . La loi  $\pi$  est appelée **loi a priori** de  $\theta$  choisie en fonction des connaissances disponibles (**information a priori**) sur  $\theta$  avant la prise en compte des observations.

### 2.1.1 Quelques définitions

- La loi conditionnelle de  $x$  sachant  $\theta$  sa densité est notée  $f(x/\theta)$  est appelé **Loi d'observations**.
- On entend par **information a priori** sur le paramètre  $\theta$  toute information disponible sur  $\theta$  en dehors de celle apportée par les observations.
- La loi  $\theta$  conditionnelle à  $x$ ,  $\pi(\theta/x)$  appelée **loi a posteriori** est donnée par :

$$\pi(\theta/x) = \frac{h(\theta, x)}{m(x)},$$

avec  $h(\theta, x) = f(x/\theta)\pi(\theta)$  et  $m(x) = \int_{\Theta} f(x/\theta)\pi(\theta) d\theta$ .

- **Le modèle statistique paramétrique bayésienne** consiste en la donnée d'une loi a priori  $\pi$  et de la loi des observations.

### 2.1.2 Choix de lois a priori

Le choix de la loi a priori est une étape fondamentale. Le point le plus critiquable et le plus critiqué de l'analyse bayésienne est bien que le choix de la loi a priori. En pratique cette information est rarement a priori soit suffisamment précise pour déterminer exactement la loi a priori.

#### Priories non informatives

En utilise des priori non informatives, lorsque aucune information a priori n'est disponible. Dans ce cas, on laisse les données conduire l'inférence qui peut être justifié par les résultats

théoriques qui montrent que dans les deux paradigmes bayésien et fréquentiste les statistiques (estimateur) retenus converge vers la vraie valeur lorsque la taille de l'échantillon est assez grande.

Dans le cadre de ce mémoire nous présentons les considérations, les plus usitées dans la pratique, sur les priori non informatives.

### A priori conjuguée

Souvent, il n'est pas facile à calculer les densités  $\pi(\theta)$  et  $f(x/\theta)$  pour une distribution a priori quelconque. A cet effet, il convient alors de choisir une distribution a priori qui permette facilement d'exploiter la distribution a posteriori dès le recueil d'une nouvelle information  $x$  sur le paramètre à estimer  $\theta$ . Cependant, quand l'information a priori sur le modèle est trop vague ou peu fiable, une construction subjective complète de la distribution a priori est évidemment impossible. Mais d'autres considérations plus techniques peuvent être prisent en compte. En particulier, Raiffa and Schlaifer [7] ont proposé l'utilisation des densités a priori dites conjuguées afin de faciliter l'analyse Bayésienne.

**Définition 3** *Soit  $\mathcal{F}$  une famille de distributions de densité  $f(x/\theta)$ , indexée par  $\theta$ . Une famille  $\mathfrak{S}$  de distributions a priori de densité  $\pi(\theta)$  est dite conjuguée par rapport à  $\mathcal{F}$ , si la distribution a posteriori de densité  $\pi(\theta/x)$  reste dans la même famille  $\mathfrak{S}$  pour tout  $\pi \in \mathfrak{S}$  et tout  $f \in \mathcal{F}$ . Autrement dit, la distribution a posteriori garde la même forme que la distribution a priori.*

Dans ce cas, il n'y a généralement pas besoin de calculer explicitement la loi marginale  $m(x)$  parce que les noyaux de fonctions  $f(x/\theta)\pi(\theta)$ ,  $\pi(\theta/x)$  sont les mêmes à un facteur constant près :

$$f(x/\theta)\pi(\theta) \propto \pi(\theta/x).$$

Lorsque la famille de distributions conjuguées  $\mathfrak{S}$  est paramétrée, le passage de la distribution a priori à la distribution a posteriori se réduit à un changement de leurs paramètres.

Dans ce cas, la distribution a posteriori est toujours calculable. Une telle considération est fondée sur le principe suivant : l'information apportée par des observations  $x$  sur  $\theta$  est limitée et dont la modification par  $x$  ne doit pas conduire à une remise en cause de la forme de  $\pi(\theta)$ , mais seulement de ses paramètres. Enfin, la Table suivante énumère les priori conjuguées de certaines lois classiques.

$f(x/\theta)$	$\pi(\theta)$	$\pi(\theta/x)$
$\mathcal{N}ormale(\mu, 1/\theta)$	$\mathit{Gamma}(\alpha, \beta)$	$\mathit{Gamma}\left(\alpha + \frac{1}{2}, \beta + \frac{(\mu-x)^2}{2}\right)$
$\mathcal{P}oisson(\theta)$	$\mathit{Gamma}(\alpha, \beta)$	$\mathit{Gamma}(\alpha + x, \beta + 1)$
$\mathit{B}inomiale(n, \theta)$	$\mathit{Beta}(\alpha, \beta)$	$\mathit{Beta}(\alpha + x, \beta + n - x)$
$\mathit{B}Negative(m, \theta)$	$\mathit{Beta}(\alpha, \beta)$	$\mathit{Beta}(\alpha + m, \beta + x)$

TABLE 2.1: Quelques densités de lois a priori conjuguée

**Densité a priori uniforme :** Historiquement, Laplace fut le premier à utiliser des techniques non informatives. Il munit les paramètres d'une loi a priori qui prend en compte son ignorance en donnant la même vraisemblance à chaque valeur du paramètre, donc en utilisant une loi uniforme. Ainsi, la densité a priori d'un paramètre est défini par :

$$\pi(\theta) = c,$$

où  $c$  est une constante.

**A priori impropre :**

On choisit la loi priori impropre, lorsque l'espace du paramètre est infini qui vérifie  $\int_{\Theta} \pi(\theta) d\theta = \infty$ . n'existe pas de lois uniforme sur cet espace. Ce choix de type de loi n'a donc plus d'intérêt que calculatoire et s'interprète difficilement. La construction de lois non informatives peut conduire à des lois a priori de ce type.

### 2.1.3 Estimateur de Bayes

Soit  $\hat{\theta}$  un estimateur de  $\theta$ . On définit une fonction de coût non négative  $C(\hat{\theta} - \theta)$  tel que  $\hat{\theta} - \theta$  est l'erreur d'estimation pour les observations  $x = (x_1, x_2, \dots, x_n)$  données. Le but est de déterminer l'estimateur  $\hat{\theta}$  qui minimise le coût moyen  $\mathbf{E} \left( C \left( \hat{\theta} - \theta \right) \right)$  appelée le risque bayésien défini par :

$$\mathbf{E} \left( C \left( \hat{\theta} - \theta \right) \right) = \int C \left( \hat{\theta} - \theta \right) \pi \left( \theta/x \right) d\theta.$$

Considérons le cas où  $C \left( \hat{\theta} - \theta \right)$  est le coût quadratique (la fonction coût la plus utilisée), c'est-à-dire que  $C \left( \hat{\theta} - \theta \right) = \left( \hat{\theta}(x) - \theta \right)^2$ . L'estimateur qui minimise le risque bayésien en utilisant la fonction de coût quadratique est la **moyenne a posteriori** donnée simplement par :

$$\hat{\theta} = \mathbf{E}(\theta/x) = \int_{\Theta} \theta \pi(\theta/x) d\theta = \frac{\int_{\Theta} \theta f(x/\theta) \pi(\theta) d\theta}{\int_{\Theta} f(x/\theta) \pi(\theta) d\theta}. \quad (2.1)$$

Pour la construction de l'estimateur de Bayes on peut également considérer autres fonctions de coût telles :

- **Le coût absolu :**

$$C \left( \hat{\theta}(x) - \theta \right) = \left| \hat{\theta}(x) - \theta \right|.$$

- **Le coût uniforme :**

$$C \left( \hat{\theta}(x) - \theta \right) = \begin{cases} 1, & \text{si } \left| \hat{\theta}(x) - \theta \right| \leq \frac{\delta}{2}; \\ 0, & \text{sinon,} \end{cases}$$

avec  $\delta$  est la longueur maximale (la précision) du domaine de l'erreur d'estimation.

Dans le reste document, nous ne utilisons que l'estimateur de Bayes décrit dans (2.1).

## Propriétés de l'estimateur de Bayes

Notons que parmi les principale caractéristiques de l'estimateur de Bayes on a :

- L'estimateur de Bayes est admissible.
- L'estimateur de Bayes est biaisé.

Sous certaines hypothèses de régularité, le plus souvent satisfaites en pratique, on a également les deux propriétés suivantes :

- L'estimateur de Bayes est convergent en probabilité (quand la taille de l'échantillon  $n \rightarrow \infty$ ).
- La loi a posteriori peut être asymptotiquement (c'est-à-dire pour de grandes valeurs de  $n$ ) approximée par une loi normale  $\mathbf{N}(\mathbf{E}(\theta/x), \mathbf{Var}(\theta/x))$ .

Cette dernière propriété est particulièrement utile pour construire des intervalles de confiance a posteriori.

### 2.1.4 Méthodes de Monté Carlo par Chaîne de Markov (MCMC)

Le calcul de la densité a posteriori analytiquement est très difficile car l'inférence bayésienne se heurte à des problèmes d'intégration. Pour résoudre ce problème, on peut faire recours aux méthodes d'intégration numériques ou encore les méthodes de Monté Carlo par Chaîne de Markov (MCMC). Dans la suite, nous allons rappeler quelques notions intéressantes pour les méthodes MCMC mais il convient de rappeler d'abord le principe des méthodes de Monte Carlo et la définition d'une chaîne de Markov.

**Méthodes de Monte Carlo :** Techniques d'estimation s'appuyant sur la simulation d'un grand nombre de variables aléatoires.

**Chaîne de Markov :** Une chaîne de Markov sur un espace  $D$  est un processus aléatoire où l'état future de la chaîne étant donnée l'état présent est indépendant de tous les états passés. Soit  $x_1, x_2, \dots, x_n \in D$  les observations associées à la suite de variables aléatoires  $X_1, X_2, \dots, X_n$ , on a :

$$\mathbf{P}(X_{n+1} = x_{n+1}/X_n = x_n, \dots, X_0 = x_0) = \mathbf{P}(X_{n+1} = x_{n+1}/X_n = x_n).$$

L'idée principale des méthodes MCMC est de construire une chaîne de Markov ergodique dont la distribution stationnaire est la loi a posteriori. En simulant une réalisation suffisamment grande de cette chaîne, on pourra supposer qu'à partir d'un certain rang  $N_0$  les échantillons simulés sont représentatifs de la loi a posteriori dite dans ce cadre la densité cible.

En supposant que pour  $N > N_0$ , on échantillonne exactement selon la loi cible. L'inférence sur les paramètres est basée alors sur cet échantillon. Afin de réduire le biais de l'estimateur bayésien qui est dû à l'effet des valeurs initiales, le nombre d'itérations  $N_0$  est ignoré. Le nombre d'itérations  $N_0$  est appelé phase d'échauffement ("burn-in" en Anglais). Une des difficultés rencontrée dans l'implémentation des méthodes MCMC est le diagnostic de convergence de la chaîne construite. Plusieurs méthodes ont été proposées dans la littérature pour vérifier la convergence des méthodes MCMC. Nous citons par exemple, le critère de inter-intra chaîne proposé par Gelman and Rubin [4].

### **Algorithme de Métropolis-Hastings**

Nous présentons dans ce passage l'algorithme de Metropolis-Hasting (M-H) proposé par Metropolis et al. [6] en 1953 et généralisé par Hastings [5] en 1970, qu'on peut considérer comme l'algorithme de base d'une grande partie des méthodes MCMC. L'objectif de l'algorithme de M-H est de simuler un échantillon selon une densité cible connue à une constante multiplicative près, à partir d'une loi de proposition (instrumental)  $q(\theta/\tilde{\theta})$  facilement simulable. Les étapes de l'algorithme M-H sont données comme suit :

#### ***Algorithme M-H***

**Étape (1)** Initialiser  $\theta$  à  $\theta^{(0)}$  et  $m = 1$

**Étape (2)** Si  $m > N$  Allez à l'**Étape (4)** sinon

(a) Générer  $\tilde{\theta} \sim q(\tilde{\theta}/\theta^{(m-1)})$ .

(b) Calculer la probabilité d'acceptation  $\rho = \min \left\{ 1, \frac{\pi(\hat{\theta}/x) q(\theta^{(m-1)}/\tilde{\theta})}{\pi(\theta^{(m-1)}/x) q(\tilde{\theta}/\theta^{(m-1)})} \right\}$ .

(c) mettre à jour  $\theta$  :

$$\theta^{(m+1)} = \begin{cases} \tilde{\theta}, & \text{si } u < \rho \text{ et } u \sim U[0, 1] \text{ (} U[0, 1] \text{ désigne la loi uniforme dans } [0, 1] \text{)}; \\ \theta^{(m)}, & \text{sinon.} \end{cases}$$

**Étape (3)** Poser  $m = m + 1$  et aller à l'**Étape (2)**.

**Étape (4)** Calculer l'estimateur de Bayes  $\hat{\theta} = \frac{1}{N-N_0} \sum_{m=N_0+1}^N \theta^{(m)}$ .

## 2.2 Approche bayésienne pour la sélection du paramètre de lissage

Le choix du paramètre de lissage  $h$  est d'une importance capitale dans l'estimation de la fonction densité par la méthode en générale et à noyau discret en particulier. Dans cette section nous présentons deux approches différentes pour le choix du paramètre de lissage à savoir : l'approche bayésienne globale et l'approche bayésienne locale.

### 2.2.1 Approche bayésienne globale

Nous allons présenter les étapes de l'approche bayésienne globale pour la sélection du paramètre de lissage. L'estimation bayésienne du paramètre  $h$  conditionné aux données se fait par la densité a posteriori  $\pi(h/\text{données})$ . Concrètement, nous considérons une séquence  $X_1, X_2, \dots, X_n$  de variables aléatoire naturel, *i.i.d* par une densité de probabilité inconnu  $f$  et de réalisations  $x = (x_1, x_2, \dots, x_n)$ . L'estimation bayésienne globale du paramètre de lissage dans l'estimation de la densité consiste à :

1. Donner la forme de l'estimateur de la vraisemblance  $f(x_1, x_2, \dots, x_n/h)$ , la densité des données sachant le paramètre  $h$  ;

2. Choisir la loi a priori sur le paramètre de lissage  $h$  ;
3. Calculer la densité a posteriori  $\pi(h/x_1, x_2, \dots, x_n)$  du paramètre de lissage  $h$  et les données  $x_1, x_2, \dots, x_n$  en utilisant le théorème de Bayes ;
4. Estimer le paramètre de lissage  $h$  par la moyenne a posteriori (en utilisant les méthodes de MCMC).

Les détails de l'adaptation de ces quatre étapes au cas de l'estimation à noyau discret d'une densité de probabilité son comme suit :

### Construction de l'estimateur de la vraisemblance

La vraisemblance des données  $x_1, x_2, \dots, x_n$  notée  $L(x_1, x_2, \dots, x_n)$  et s'écrit comme suit :

$$L(x_1, x_2, \dots, x_n) = f(x_1, x_2, \dots, x_n) = \prod_{i=1}^n f(x_i).$$

Vue que  $f$  est inconnue alors l'estimateur de la vraisemblance sera obtenu en utilisant un estimateur de  $f$ . Naturellement,  $f$  sera remplacer par son estimateur à noyau le fait que nous nous intéressons à cette technique dès le départ. Ainsi pour estimer  $f$  au point  $x_i$  nous on la remplace par son estimateur à noyau toute en utilisant la méthode de validation croisée et on aura :

$$f(x_1, x_2, \dots, x_n/h) = \prod_{i=1}^n \hat{f}_i(x_i).$$

Rappelons que la technique de validation croisée consiste a estimer  $f(x_i)$  a partir de l'ensemble des points sauf le point  $x_i$ . Alors, l'estimateur à noyau via la technique de validation croisée de  $f(x_i)$  est donnée par :

$$\hat{f}_i(x_i) = \frac{1}{n-1} \sum_{i=1, j \neq i} K_{x_i, h}(x_j).$$

La vraisemblance est alors approximée par :

$$f(x_1, x_2, \dots, x_n/h) = \frac{1}{(n-1)^n} \prod_{i=1}^n \sum_{j \neq i} K_{x_i, h}(x_j). \quad (2.2)$$

Ce dernier estimateur sera utilisé pour le calcul de la densité a posteriori.

### Choix de la loi a priori

L'inférence bayésienne est caractérisée par le choix d'une loi a priori. Ce choix n'est pas unique. L'utilisation de la loi a priori conjuguée n'est pas bénéfique dans ce modèle. Dans ce cas, le choix doit être fait d'une manière que les méthodes MCMC fonctionnent correctement. Si utilise, par exemple, la loi Gamma proposée par Brewer [1] comme loi a priori pour  $h$  alors la densité  $\pi(h)$  est donnée, à une constante près, par

$$\pi(h/\alpha, \beta) \propto h^{\alpha-1} \exp\left(-\frac{h}{\beta}\right),$$

où  $\alpha$  et  $\beta$  sont des hyper-paramètres.

Dans cette partie, nous ne discutons pas le choix des hyper-paramètres précisément car on propose que le choix se fasse empiriquement de sorte que les méthodes de Monte Carlo fonctionnent correctement.

### Calcul de la loi a posteriori :

En utilisant le théorème de Bayes, la loi a posteriori de  $h$  prend la forme suivante :

$$\pi(h/x_1, x_2, \dots, x_n) = \frac{f(x_1, x_2, \dots, x_n/h)\pi(h)}{\pi(x_1, x_2, \dots, x_n)} = \frac{\pi(h) \prod_{i=1}^n \hat{f}_{h,i}(x_i)}{\pi(x_1, x_2, \dots, x_n)},$$

où  $\pi(x_1, x_2, \dots, x_n) = \int f(x_1, x_2, \dots, x_n/h)\pi(h) dh$

Alors la densité a posteriori du paramètre  $h$  sachant les données  $x = (x_1, x_2, \dots, x_n)$  est donnée à une constante d'intégration près, par :

$$\begin{aligned}\pi(h/\alpha, \beta, x_1, x_2, \dots, x_n) &= \frac{f(x_1, x_2, \dots, x_n/h)\pi(h/\alpha, \beta)}{\pi(x_1, x_2, \dots, x_n)} \approx \frac{\pi(h/\alpha, \beta) \prod_{i=1}^n \hat{f}_i(x_i)}{\pi(x_1, x_2, \dots, x_n)}, \\ \pi(h/x_1, x_2, \dots, x_n) &\approx \frac{h^{\alpha-1} \exp\left(-\frac{h}{\beta}\right)}{(n-1)^n \pi(x_1, x_2, \dots, x_n)} \prod_{i=1}^n \sum_{i=1, j \neq i} K_{x_i, h}(x_j),\end{aligned}\quad (2.3)$$

où  $\pi(x_1, x_2, \dots, x_n) = \int f(x_1, x_2, \dots, x_n/h)\pi(h/\alpha, \beta) dh$  est la loi marginale de l'échantillon. Vu la complexité de la loi a posteriori (2.3), il est impossible d'obtenir directement l'estimateur bayésien de  $h$ . Pour cela, pour estimer le paramètre de lissage  $h$  on peut utiliser par exemple les méthodes MCMC.

Ainsi, nous proposons de raisonner proportionnellement afin de nous débarrasser des calculs inutiles. La densité a posteriori du paramètre  $h$  sachant les données  $x_1, \dots, x_n$  est donnée, à une constante d'intégration près, par

$$\pi(h/x_1, x_2, \dots, x_n) \propto \frac{h^{\alpha-1} \exp\left(-\frac{h}{\beta}\right)}{(n-1)^n} \prod_{i=1}^n \sum_{i=1, j \neq i} K_{x_i, h}(x_j). \quad (2.4)$$

Nous présenterons, ci-dessous l'adaptation numérique de l'algorithme de M-H, présentés dans la section 2.1.4, pour la sélection du paramètre de lissage  $h$ .

### Estimation du paramètre de lissage $h$

Nous utilisons l'algorithme M-H à marche aléatoire pour estimer le paramètre de lissage par la moyenne a posteriori. Cet algorithme est basé sur l'utilisation d'une loi génératrice de candidats de la forme  $Q(\cdot/h^{(k)})$ . Le candidat  $\tilde{h}$  est généré à partir d'une variable aléatoire  $\varepsilon$  positive de densité  $q(h/h^{(k)}, \gamma_{(k)})$ , tels que  $h^{(k)}$  est la candidate généré à la  $k^{\text{ème}}$  étape et  $\gamma_{(k)}$  est un paramètre de réglage choisi de telle sorte à obtenir un taux d'acceptation optimal  $\tau$ . La variable aléatoire  $\varepsilon$  est généralement choisie comme gaussienne de moyenne  $h^{(k)}$  et de variance  $\gamma_{(k)}^2$  (pour plus de détails voir Zougab [11]).

Ici, nous proposons d'utiliser la loi gaussienne tronquée sur  $[0, +\infty[$ , vu la positivité du paramètre de lissage  $h$ . C'est-à-dire, la densité de l'instrumentale est donnée par :

$$q(h/h^{(k)}, \gamma_{(k)}) = \frac{1}{\gamma_{(k)} \sqrt{2\pi} \phi\left(\frac{h^{(k)}}{\gamma_{(k)}}\right)} \exp\left(-\frac{(h - h^{(k)})^2}{2\gamma_{(k)}^2}\right), \quad (2.5)$$

où  $\phi(\cdot)$  est la fonction de répartition de la loi normale standard.

Garthwaite et al. [2] donnent la formule de mise-à-jour, du paramètre  $\gamma_{(k)}$ , suivante :

$$\gamma_{(k+1)} = \begin{cases} \gamma_{(k)} + \frac{\gamma_{(k)}}{k\tau}, & \text{si } \tilde{h} \text{ est accepté;} \\ \gamma_{(k)} - \frac{\gamma_{(k)}}{k(1-\tau)}, & \text{si } \tilde{h} \text{ est rejeté.} \end{cases} \quad (2.6)$$

où  $\tau$  est le taux d'acceptation optimal, il est égal à 0.44 dans l'univarié voir Gelman et al. [3], Roberts and Rosenthal [8] et Garthwaite et al. [2].

Les étapes de l'algorithme M-H à marche aléatoire associé au présent problème est donné comme suite :

**Étape (1)** Initialiser  $h^{(0)}, \gamma_{(0)}$

**Étape (2)** Si  $k > N$  allez à **Étape (4)** sinon

(a) Générer  $\tilde{h} \sim q(h/h^{(k)}, \gamma_{(k)})$

(b) Calculer la probabilité d'acceptation  $\rho = \min\left\{1, \frac{\pi(\tilde{h}/x_1, x_2, \dots, x_n)}{\pi(h^{(k)}/x_1, x_2, \dots, x_n)} * \frac{q(h^{(k)}/\tilde{h}, \gamma_{(k)})}{q(\tilde{h}/h^{(k)}, \gamma_{(k)})}\right\}$ .

(c) Calculer l'état de la chaîne :

$$h^{(k+1)} = \begin{cases} \tilde{h}, & \text{si } u < \rho, u \sim U[0, 1]; \\ h^{(k)}, & \text{sinon.} \end{cases}$$

(d) Mettre à jour le paramètre de réglage :

$$\gamma_{(k+1)} = \begin{cases} \gamma_{(k)} + \frac{\gamma_{(k)}}{k\tau}, & \text{si } \tilde{h} \text{ est accepté;} \\ \gamma_{(k)} - \frac{\gamma_{(k)}}{k(1-\tau)}, & \text{si } \tilde{h} \text{ est rejeté.} \end{cases}$$

**Étape (3)** Poser  $k = k + 1$  et aller à **Étape (2)**.

**Étape (4)** Calculer l'estimateur de Bayes :  $\hat{h} = \frac{1}{N-N_0} \sum_{k=N_0+1}^N h^{(k)}$ .

## 2.2.2 Approche bayésienne locale

Le principe du choix local du paramètre de lissage est de choisir ce paramètre de telle sorte à minimiser un certain critère d'erreur ou de maximiser une certaine information en chaque point  $x$  où on désire estimer la densité. Ainsi pour l'approche bayésienne locale, le paramètre de lissage en points  $x$  est considéré comme une variable aléatoire de loi a priori  $\pi(h)$  qu'on considère aussi comme paramètre du modèle  $f_h(x)$ . Par la suite, l'idée de l'approche consiste à utiliser  $f_h(x)$  pour construire un estimateur de  $h$  en chaque point  $x$  bien fixé.

Considérons  $X_1, X_2, \dots, X_n$  un  $n$ -échantillon *i.i.d* issu d'une variable aléatoire  $X$  de fonction de densité inconnue  $f$ , et  $x_1, x_2, \dots, x_n$  des réalisations de l'échantillon. Par la formule de Bayes, la loi a posteriori de  $h$  au point  $x$  prend la forme suivante :

$$\pi(h/x) = \frac{f_h(x) \pi(h)}{\int f_h(x) \pi(h) dh}.$$

Comme le modèle  $f_h(x)$  est inconnu, on utilise son estimateur  $\hat{f}(x)$ . La loi a posteriori de  $h$  devient alors :

$$\hat{\pi}(h/x, X_1, X_2, \dots, x_n) = \frac{\hat{f}(x) \pi(h)}{\int \hat{f}(x) \pi(h) dh}.$$

L'estimateur de Bayes sous le coût quadratique de  $h$  au point  $x$  est donné par

$$\hat{h}_n = \int h \hat{\pi}(h/x, X_1, X_2, \dots, x_n) dh.$$

## Conclusion

Dans ce chapitre, après avoir présenté les notions de base et le principe de l'inférence bayésienne, nous avons mis en évidence la démarche à suivre pour sa mise en œuvre dans l'estimation du paramètre de lissage dans le cadre d'estimation à noyau d'une densité de probabilité.

A ce stade, il reste à vérifier et à analyser l'apport de cette technique sur les performances de l'estimateur par rapport aux techniques classiques (minimisation du biais et validations croisées). Pour cela, une étude numérique basée sur des échantillons artificiels sera présentée dans le chapitre 3.

# Chapitre 3

## Etude comparative : Application

### Numérique

#### Introduction

Nous présentons dans ce chapitre le travail de simulation effectué sous **R** pour étudier, d'une part, l'influence du noyau dans l'estimation de la densité de probabilité discrète, d'autre part, les performances de la méthode MCMC pour l'approche bayésienne et la méthode de validation croisée non biaisée pour l'approche classique dans le choix du paramètre de lissage.

Afin d'illustrer l'influence du noyau et les performances des méthodes de sélection du paramètre de lissage, nous utilisons quatre densités tests présentant différentes formes à savoir : uni-modale semi-bornée, uni-modale bornée, bimodale semi-bornée et bimodale bornée.

#### 3.1 Plan de simulation

Dans le reste de ce chapitre, on adoptera les notations suivantes :

- $\hat{h}_{UCV}$  : le  $h$  optimal sélectionné par la technique de validation croisée non biaisé (*UCV*).

- $\hat{h}_{Bayes}$  : est le  $h$  optimale sélectionné par l'inférence bayésienne.
- $\overline{ISE}_{UCV}$  : est le  $ISE$  moyenne de l'estimateur fournis par  $h_{UCV}$ .
- $\overline{ISE}_{Bayes}$  : le  $ISE$  moyenne de l'estimateur fournis par  $h_{Bayes}$ .
- *Bayésienne* est la sélection de  $h$  par l'algorithme présenté dans le chapitre 2.

Afin de répondre à notre objectif nous avons implementer un programme sous **R** dont les principales étapes sont :

1. Générer un échantillon de taille  $n$  d'une loi cible.
2. Estimer  $\hat{h}_{UCV}$  par la technique  $UCV$  et calculer  $\hat{f}_{UCV}$ .
3. Estimer  $\hat{h}_{Bayes}$  par l'inférence bayésienne globale et calculer  $\hat{f}_{Bayes}$ .
4. Comparer les performances des estimateurs  $\hat{f}_{UCV}$  et  $\hat{f}_B$ .

L'étape (4) en réalité est composée des sous-étapes suivantes :

- (a) Fixer la loi a priori de  $h$ .
- (b) Calculer la loi a posteriori de  $h$ .
- (c) Estimer  $\hat{h}_{Bayes}$  par l'algorithme de Metropolis-Hasting présenté dans la section 2.2.1 à la page 22.

### 3.1.1 Paramètre de simulation

Pour l'application numérique nous avons considéré les quatre distributions suivantes :

- Une loi Poisson de paramètre  $\lambda = 1.5$  :

$$f_1(x) = \frac{\lambda^x}{x!} e^{-\lambda}, \quad x \in \mathbb{N}. \quad (3.1)$$

- Une loi binomiale de paramètres  $(m, p) = (10, 0.5)$  :

$$f_2(x) = C_m^x p^x (1-p)^{m-x}, \quad x \in \{0, 1, \dots, 10\}. \quad (3.2)$$

- Un Mélange de deux lois de Poisson de paramètres  $(\alpha, \lambda_1, \lambda_2) = (0.3, 5, 15)$  :

$$f_3(x) = \alpha \frac{\lambda_1^x}{x!} e^{-\lambda_1} + (1 - \alpha) \frac{\lambda_2^x}{x!} e^{-\lambda_2}, \quad x \in \mathbb{N}. \quad (3.3)$$

- Un Mélange de deux lois Binomiale de paramètres  $(\alpha, m_1, p_1, m_2, p_2) = (0.5, 10, 0.2, 10, 0.8)$  :

$$f_4(x) = \alpha C_{m_1}^x p_1^x (1-p_1)^{m_1-x} + (1-\alpha) C_{m_2}^x p_2^x (1-p_2)^{m_2-x}, \quad x \in \{0, 1, \dots, 10\}. \quad (3.4)$$

Les diagrammes de ces densités tests sont données dans la figure suivante.

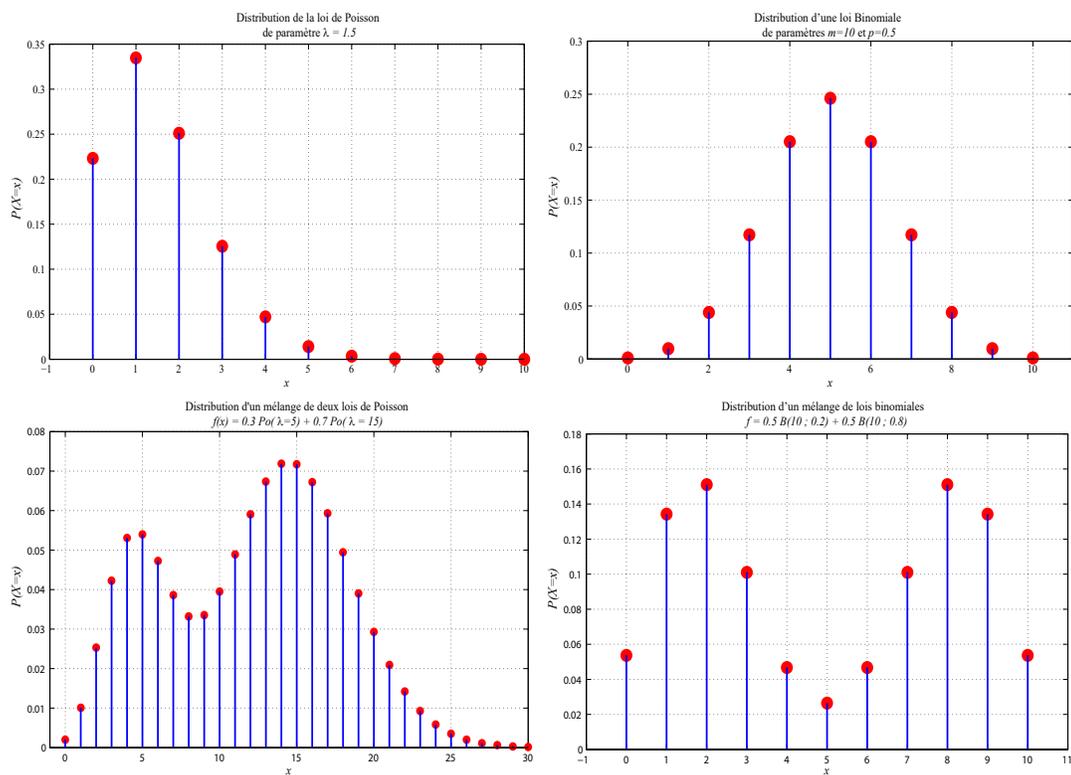


FIGURE 3.1 – Diagrammes des distributions cibles

On constate à partir de la figure 3.1 que les lois considérées peut être classer en deux catégories selon leurs modes (uni-modale et bi-modales) et en deux catégories également selon leurs support (semi-borné ou borné). Tandis que, le reste des paramètres de notre application sont fixés comme suit :

**Le noyau  $K$**  : Le noyau  $K \in \{ \text{Poisson; Binomiale Négatif, Triangulaire} \}$ .

**La loi a priori  $\pi(h)$**  : Pour la loi a priori nous avons opter pour une loi Gamma de paramètres  $(\alpha, \beta) = (2, 5)$

**La loi instrumentale  $q(h/h^{(k)}, \gamma(k))$**  : est une loi normale tronquée dont les paramètres initiales sont fixés respectivement à  $\gamma(0) = 1$  et  $h^{(0)} = 0.5$  (voir la formule (2.5)),

La taille de la chaîne  $N$  égale à 10000 et la période d'échauffement  $N_0$  est 1000.

**La taille  $n$**  : Nous avons considéré quatre différents tailles des échantillons

$$n \in \{25, 50, 100, 150, 200, 250, 500, 1000\}.$$

De plus, on a considéré 50 échantillons pour chaque taille  $n$ , c'est-à-dire le nombre de replications  $mc = 50$ .

## 3.2 Résultats et discussion

Les résultats obtenus par l'exécution de notre programme pour les différents paramètres cités dans la section précédente sont rangés dans les Tables 3.1–3.4. La variation du paramètre de lissage et du  $ISE$  moyenne lui associé en fonction de la taille de l'échantillon  $n$  lorsque l'estimateur est conçu à l'aide du noyau triangulaire est présenté dans les Figures 3.2–3.4.

Les résultats obtenus montrent que :

- Cas de  $f_1$

- ▷ Le paramètre de lissage sélectionné par la méthode de validation croisée non biaisée ne vérifie pas la propriété de convergence, c'est-à-dire le paramètre de lissage n'est pas une fonction décroissante en fonction de la taille de l'échantillon et ne tend pas vers zéro lorsque  $n$  tend vers l'infinie et ceci quel que soit le noyau utilisé. Contrairement au  $h_{UCV}$ , le paramètre de lissage  $h_{bayes}$  vérifie la propriété en question.

- ▷ Le *ISE* associée aux paramètres de lissage  $h_{UCV}$ ,  $h_{bayes}$  ne converge que dans le cas d'utilisation du noyau triangulaire  $K_T$ .
- ▷ Lors de l'utilisation du noyau de Binomial négatif,  $K_{BN}$ , les estimations conçues par les deux paramètres de lissage considéré sont pratiquement de même qualité (au sens du *ISE*) avec une légère préférence pour ceux conçue via  $h_{UCV}$ .
- ▷ Lors de l'utilisation du noyau de Poisson,  $K_{Po}$ , en générale, les meilleurs estimations au sens du *ISE* sont conçues à l'aide  $h_{Bayes}$ .
- ▷ Pour le cas du noyau triangulaire, ne dépend pas seulement de la procédure de sélection utilisée mais aussi de la taille de l'échantillon. En effet, pour de petites tailles ( $n \leq 50$ ) c'est le paramètre de lissage  $h_{Bayes}$  qui nous fournit de bonnes estimations tandis que pour des échantillon de tailles supérieur les deux paramètre nous fournis des estimateurs de même qualité avec un léger l'avantage du  $h_{UCV}$ .

D'une manière globale, on a d'une part les meilleures estimations sont obtenues dans le cas d'utilisation du noyau Triangulaire combiné avec le paramètre de lissage  $h_{Bayes}$  lorsque la taille de l'échantillon est petite et avec  $h_{UCV}$  lorsque la taille de l'échantillon est moyenne ou grande. D'autre part, les estimations les moins performantes, au sens du *ISE*, est celles conçues à l'aide du noyau Binomial négatif est cela indépendamment de la procédure de sélection du paramètre de lissage utilisée.

- Cas de  $f_2$

- ▷ Le comportement des deux paramètres de lissage est le même que dans le cas de la loi de Poisson  $f_1$ .
- ▷ D'une manière générale, lorsque la taille de l'échantillon est petite le paramètre de lissage  $h_{Bayes}$  nous fournit de meilleurs estimateurs au sens du *ISE* que le  $h_{UCV}$ . Mais, ça reste qu'en fur et à mesure que la taille de l'échantillon augmente les deux paramètres de lissage ont tendance à nous fournir des estimateurs ayant pratiquement la même qualité.

▷ Du même que dans le cas du  $f_1$ , on a d'une part les meilleures estimations sont obtenues dans le cas d'utilisation du noyau Triangulaire combiné avec le paramètre de lissage  $h_{Bayes}$ . D'autre part, les estimations les moins performantes, au sens du  $ISE$ , est celles conçues à l'aide du noyau Binomial négatif.

• Cas de  $f_3$  et  $f_4$

▷ Les quantités  $h_{Bayes}$ ,  $h_{UCV}$ ,  $\overline{ISE}_{Bayes}$  et  $\overline{ISE}_{UCV}$  convergent faiblement vers zéro en fonction de la taille de l'échantillon, lorsque nous considérons le noyau de Poisson ou le noyau Binomial négatif. Par contre, dans le cas d'utilisation du noyau Triangulaire la convergence est nette.

▷ En générale,  $h_{UCV}$  à des grandes valeurs, ceci nous laisse prévoir que les estimateurs conçus via sous paramètre seront sur-lissé c'est-à-dire ces estimateurs ont tendance à être biaisés.

▷ Les meilleures estimations sont obtenues dans le cas d'utilisation du noyau Triangulaire combiné avec le paramètre de lissage  $h_{Bayes}$ . Tandis que, les estimations les moins performantes, au sens du  $ISE$ , est celles conçues à l'aide du noyau Binomial négatif.

Noyau	$(h^*, \overline{ISE})$	Taille d'échantillon $n$							
		25	50	100	150	200	250	500	1000
$K_{Po}$	$\hat{h}_{UCV}$	0.54757	0.55148	0.50177	0.56760	0.50587	0.53964	0.21534	0.53624
	$\overline{ISE}_{UCV}$	0.06721	0.06713	0.06632	0.06655	0.06576	0.06599	0.02652	0.06617
	$\hat{h}_{Bayes}$	0.27615	0.21469	0.13580	0.10865	0.08667	0.08978	0.21534	0.53624
	$\overline{ISE}_{Bayes}$	0.06542	0.06553	0.06463	0.06499	0.06388	0.06396	0.02652	0.06617
$K_{BN}$	$\hat{h}_{UCV}$	0.46294	0.47200	0.42233	0.47618	0.42074	0.44907	0.17833	0.44240
	$\overline{ISE}_{UCV}$	0.07651	0.07729	0.07620	0.07695	0.07549	0.07574	0.03053	0.07598
	$\hat{h}_{Bayes}$	0.23311	0.20469	0.14577	0.13931	0.10334	0.09439	0.17833	0.44240
	$\overline{ISE}_{Bayes}$	0.07693	0.07776	0.07646	0.07751	0.07579	0.07610	0.03053	0.07598
$K_T$	$\hat{h}_{UCV}$	0.03730	0.01420	0.00317	0.00291	0.00223	0.00217	0.00087	0.00217
	$\overline{ISE}_{UCV}$	0.07190	0.06049	0.05570	0.05387	0.05236	0.05204	0.02032	0.05035
	$\hat{h}_{Bayes}$	0.12562	0.05096	0.02351	0.01673	0.00894	0.00895	0.00087	0.00217
	$\overline{ISE}_{Bayes}$	0.06882	0.06014	0.05611	0.05391	0.05243	0.05197	0.02032	0.05035

TABLE 3.1: Résultats obtenus pour le cas  $f_1$ .

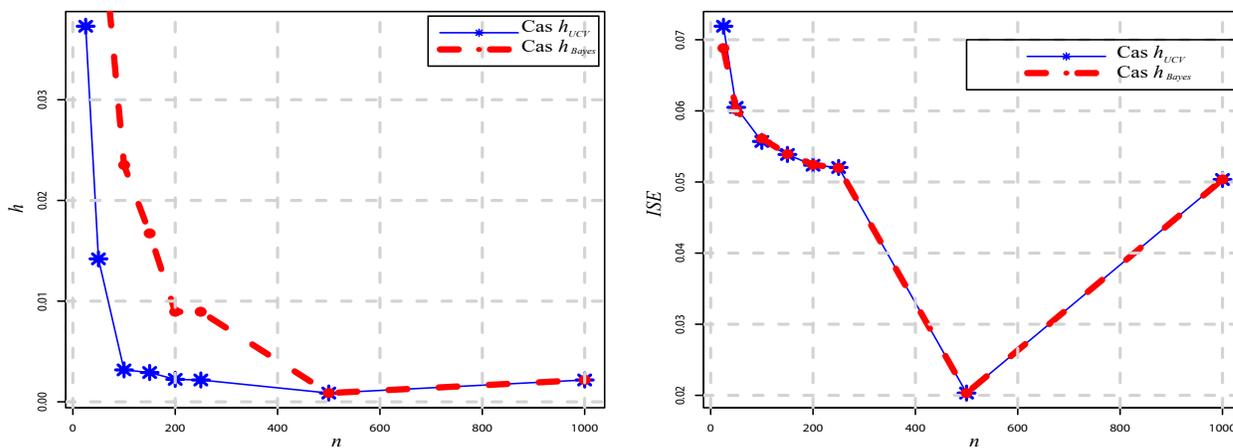
Noyau	$(h^*, \overline{ISE})$	Taille d'échantillon $n$							
		25	50	100	150	200	250	500	1000
$K_{Po}$	$\hat{h}_{UCV}$	0.11052	0.10175	0.09437	0.09327	0.09344	0.09497	0.09573	0.09428
	$\overline{ISE}_{UCV}$	0.02779	0.02778	0.02725	0.02711	0.02747	0.02723	0.02692	0.02692
	$\hat{h}_{Bayes}$	0.35197	0.38236	0.30138	0.29940	0.29304	0.30811	0.09573	0.09428
	$\overline{ISE}_{Bayes}$	0.02903	0.02958	0.02776	0.02777	0.02795	0.02790	0.02692	0.02692
$K_{BN}$	$\hat{h}_{UCV}$	0.13620	0.12873	0.11280	0.11286	0.11102	0.11479	0.11544	0.11206
	$\overline{ISE}_{UCV}$	0.04443	0.04418	0.04431	0.04426	0.04452	0.04427	0.04397	0.04410
	$\hat{h}_{Bayes}$	0.30728	0.36941	0.34503	0.30556	0.30369	0.31793	0.11544	0.11206
	$\overline{ISE}_{Bayes}$	0.04453	0.04486	0.04469	0.04446	0.04477	0.04462	0.04397	0.04410
$K_T$	$\hat{h}_{UCV}$	0.32257	0.15788	0.06257	0.03720	0.02915	0.02272	0.01158	0.00538
	$\overline{ISE}_{UCV}$	0.01830	0.01254	0.00697	0.00452	0.00401	0.00325	0.00154	0.00088
	$\hat{h}_{Bayes}$	0.30810	0.14562	0.05926	0.03890	0.02923	0.01958	0.01158	0.00538
	$\overline{ISE}_{Bayes}$	0.01580	0.01213	0.00632	0.00468	0.00412	0.00319	0.00154	0.00088

 TABLE 3.2: Résultats obtenus pour le cas  $f_2$ .

Noyau	$(h^*, \overline{ISE})$	Taille d'échantillon $n$							
		25	50	100	150	200	250	500	1000
$K_{Po}$	$\hat{h}_{UCV}$	0.54796	0.33294	0.23718	0.25448	0.22951	0.24188	0.21550	0.21175
	$\overline{ISE}_{UCV}$	0.00431	0.00385	0.00288	0.00298	0.00284	0.00294	0.00271	0.00271
	$\hat{h}_{Bayes}$	0.42737	0.41275	0.41702	0.38636	0.38184	0.24188	0.21550	0.21175
	$\overline{ISE}_{Bayes}$	0.00417	0.00389	0.00291	0.00304	0.00288	0.00294	0.00271	0.00271
$K_{BN}$	$\hat{h}_{UCV}$	0.81179	0.62045	0.50889	0.55013	0.51780	0.53622	0.49539	0.48992
	$\overline{ISE}_{UCV}$	0.00587	0.00564	0.00493	0.00496	0.00491	0.00501	0.00486	0.00486
	$\hat{h}_{Bayes}$	0.41101	0.44781	0.38794	0.41008	0.40948	0.53622	0.49539	0.48992
	$\overline{ISE}_{Bayes}$	0.00573	0.00563	0.00493	0.00493	0.00491	0.00501	0.00486	0.00486
$K_T$	$\hat{h}_{UCV}$	1.44350	1.01095	0.72732	0.53076	0.46105	0.34210	0.22056	0.12371
	$\overline{ISE}_{UCV}$	0.00523	0.00430	0.00201	0.00187	0.00161	0.00144	0.00086	0.00058
	$\hat{h}_{Bayes}$	0.62188	0.44120	0.32770	0.21996	0.19603	0.34210	0.22056	0.12371
	$\overline{ISE}_{Bayes}$	0.00552	0.00408	0.00229	0.00231	0.00199	0.00144	0.00086	0.00058

 TABLE 3.3: Résultats obtenus pour le cas  $f_3$ .

Noyau	$(h^*, \overline{ISE})$	Taille d'échantillon $n$							
		25	50	100	150	200	250	500	1000
$K_{Po}$	$\hat{h}_{UCV}$	0.71941	0.67680	0.61723	0.59153	0.63521	0.64160	0.63601	0.62595
	$\overline{ISE}_{UCV}$	0.02633	0.02529	0.02425	0.02381	0.02387	0.02382	0.02370	0.02373
	$\hat{h}_{Bayes}$	0.37557	0.32179	0.27056	0.24344	0.19740	0.22216	0.63601	0.62595
	$\overline{ISE}_{Bayes}$	0.02431	0.02412	0.02305	0.02269	0.02231	0.02245	0.02370	0.02373
$K_{BN}$	$\hat{h}_{UCV}$	0.72867	0.68165	0.66837	0.66438	0.65151	0.67052	0.67166	0.65335
	$\overline{ISE}_{UCV}$	0.02907	0.02814	0.02776	0.02771	0.02742	0.02753	0.02748	0.02742
	$\hat{h}_{Bayes}$	0.38526	0.35902	0.30820	0.24416	0.19332	0.22448	0.67166	0.65335
	$\overline{ISE}_{Bayes}$	0.02819	0.02763	0.02715	0.02710	0.02659	0.02677	0.02748	0.02742
$K_T$	$\hat{h}_{UCV}$	0.37548	0.14172	0.05319	0.04113	0.02917	0.02552	0.01431	0.00821
	$\overline{ISE}_{UCV}$	0.01981	0.01485	0.01005	0.00725	0.00616	0.00544	0.00439	0.00372
	$\hat{h}_{Bayes}$	0.19424	0.10555	0.03261	0.01939	0.01484	0.01135	0.01431	0.00821
	$\overline{ISE}_{Bayes}$	0.01870	0.01483	0.01023	0.00716	0.00619	0.00537	0.00439	0.00372

 TABLE 3.4: Résultats obtenus pour le cas  $f_4$ .

 FIGURE 3.2 – Variation de  $h^*$  et  $\overline{ISE}$  en fonction de la taille de l'échantillon  $n$ , cas :  $K_T$  et  $f_1$ .

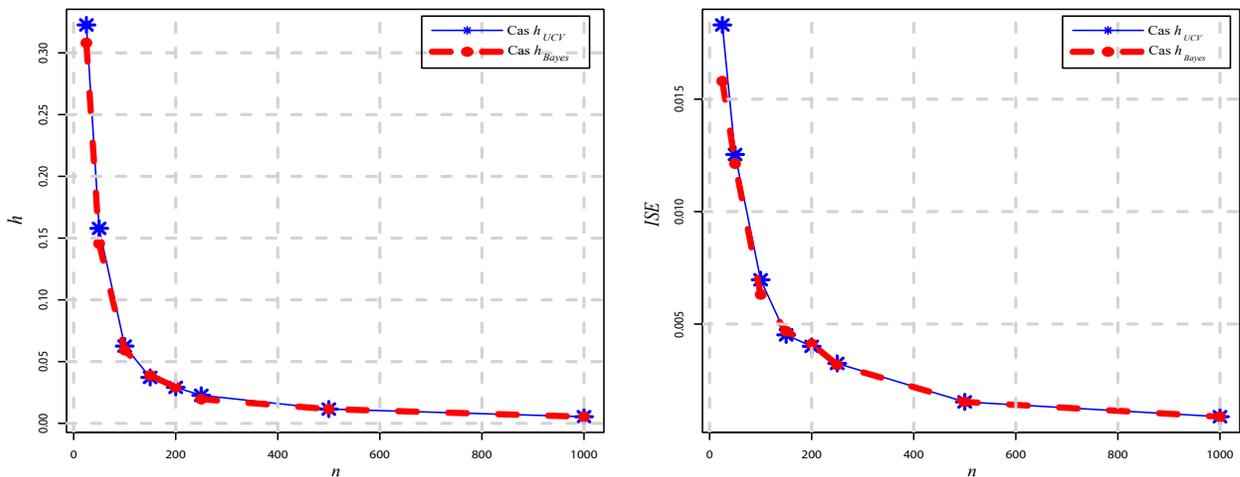


FIGURE 3.3 – Variation de  $h^*$  et  $\overline{ISE}$  en fonction de la taille de l'échantillon  $n$ , cas :  $K_T$  et  $f_2$ .

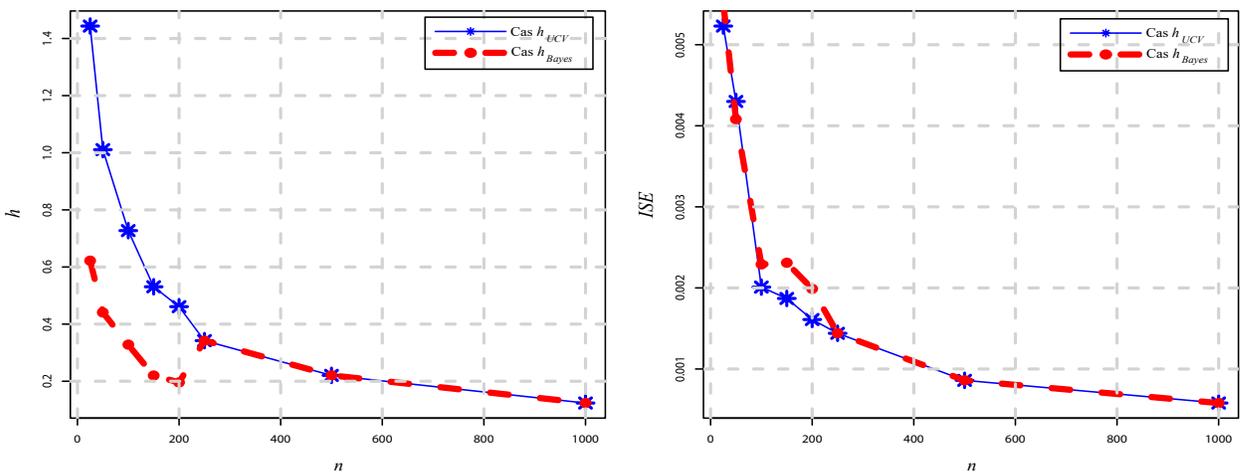


FIGURE 3.4 – Variation de  $h^*$  et  $\overline{ISE}$  en fonction de la taille de l'échantillon  $n$ , cas :  $K_T$  et  $f_3$ .

## Conclusion

Les résultats obtenus en comparant l'approche bayésienne globale et la technique de validation croisée ( $UCV$ ) montrent que :

D'une part, Le choix du paramètre de lissage global, dans l'estimation à noyau discret, par l'inférence bayésienne nous fournis, en générale, de meilleurs estimateurs au

sens du  $ISE$  que les paramètres de lissage sélectionnés par la technique de validation croisée  $UCV$  pour des échantillons de petites tailles ou moyenne taille en particulier lors de l'utilisation du noyau Triangulaire.

D'autre part, lors de l'estimation à noyau discret d'une fonction de masse de probabilité, si on retient le  $ISE$  en tant que critère de performance de l'estimateur, l'utilisation du noyau binomiale négatif est déconseillé et ceci quel que soit notre choix de la méthode de sélection du paramètre de lissage.

# Conclusion générale

L'approche bayésienne est une alternative introduite dans la littérature pour remédier aux problèmes des techniques classiques de sélection du paramètre de lissage dans l'estimation à noyau d'une manière générale et d'une densité de probabilité en particulier.

L'objectif du présent travail est de se familiariser avec les procédures de sélection du paramètre de lissage  $h$  par l'inférence bayésienne dans le cadre d'estimation à noyau d'une densité de probabilité discrète. Afin de répondre à notre objectif, nous avons procédé comme suite : Dans la première partie, nous avons présenté les principales notions liées à l'estimation à noyau d'une densité discrète où nous avons mis l'accent particulièrement sur ses propriétés, le choix du noyau et les procédures classique suggérées dans la littérature pour l'estimation du paramètre de lissage  $h$  dans une telle situation.

En deuxième lieu, l'idée est d'illustré l'adaptation du principe bayésien pour l'estimation du paramètre de lissage, dans le cas d'estimation à noyau discret d'une densité de probabilité, lorsque ce dernier est vu comme étant une variable aléatoire et en lui associant une distribution de probabilité pour décrire cette variabilité.

La troisième partie, afin de mesurer l'efficacité de l'approche bayésienne globale pour le choix du paramètre de lissage dans l'estimation à noyau d'une densité discrète, nous avons réalisé une étude numérique basée sur des échantillons de différent taille  $n = [25, 50, 100, 150, 200, 500, 1000]$  et qui sont simulés à partir de différentes loi (Poisson, binomiale, mélange de deux lois de Poisson, mélange de deux lois binomiales). Les résultats

obtenus dans cette étude montrent qu'en générale, les estimateurs associés aux paramètres de lissage sélectionnés par l'inférence bayésienne sont légèrement meilleures, au sens du ISE, que ceux obtenus par les paramètres de lissage sélectionnés la méthode classique *UCV*.

Le travail réalisé dans ce mémoire ouvre de nombreuses perspectives intéressantes, entre autres on peut citer :

- Réaliser des simulations extensives tout en considérons d'autres type de distributions (multimodale, multi-dimensionnel, à queue lourde,...) et d'autres paramètres (autres fonctions a priori, autres fonctions instrumentales, autres valeurs initiales,...).
- Revoir le même travail dans le cadre d'autres fonctions de coût.
- Considérer d'autres approches bayésiennes, à savoir : l'approche bayésienne locale, l'approche bayésienne adaptative.
- Se familiariser avec les techniques de contrôle de convergence de l'algorithme de MCMC.
- Considérer d'autres méthodes d'échantillonnage que le MCMC telles : Échantillonnage préférentiel, Population Monte Carlo.
- etc

# Bibliographie

- [1] M.J. Brewer, (1998). A modelling approach for bandwidth selection in kernel density estimation. In : Proceedings of COMPSTAT Physica Verlag, Heidelberg, pages 203–208.
- [2] P. H. Garthwaite, Y. Fan, and S. A. Sisson. Adaptive optimal scaling of metropolisshastings algorithms using the robbins-monro process. Working paper, (2010).
- [3] A. Gelman, G. O. Roberts, and W. R. Gilks, (1996). Efficient metropolis jumping rules. *Bayesian Statistics*, 5, 599-608.
- [4] A. Gelman and D. B. Rubin, (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7, 457–511.
- [5] W. K. Hastings, (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1), 97–109.
- [6] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, (1953). Equations of state calculations by fast computing machines. *Journal of Chemical physics*, (21) 1087-1091.
- [7] H. Raiffa and R. Schlaifer. *Applied statistical decision theory*. Boston, Massachusetts : Harvard Business School, 1961.
- [8] G. O. Roberts and J. S. Rosenthal, (2009). Examples of adaptive MCMC. *Journal of Computational and Graphical Statistics*, 18(2), 349–367.

- [9] T. S. Kiessé, (2008). Approche non-paramétrique par noyaux associés discrets des données de dénombrement (Doctoral dissertation, Université de Pau et des Pays de l'Adour).
- [10] Kokonendji, C. C., Senga Kiessé, T., & Zocchi, S. S. (2007). Discrete triangular distributions and non-parametric estimation for probability mass function. *Journal of Nonparametric Statistics*, 19(6-8), 241-254.
- [11] Zougab, N. (2013). Approche bayésienne dans l'estimation non paramétrique de la densité de probabilité et la courbe de régression de la moyenne Thèse de doctorat, Université de Bejaïa.
- [12] M. Rosenblatt, (1956). Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics*, 27(3), 832–837.
- [13] Parzen, E. (1962). On estimation of a probability density function and mode. *The annals of mathematical statistics*, 33(3), 1065-1076.

# Annexe A : Abréviations et Notations

Les différentes abréviations et notations utilisées tout au long de ce mémoire sont expliquées ci-dessous :

$\propto$	: Proportionnel à.
$v.a$	: Variable aléatoire.
$iid$	: Indépendant et identiquement distribués.
$\pi(\theta), \pi(\theta/x)$	: A priori, A posteriori.
$f(x/\theta)$	: Loi d'observations.
$MSE$	: L'erreur quadratique moyenne.
$MISE$	: L'erreur quadratique moyenne intégrée.
$ISE$	: L'erreur quadratique intégrée.
$CV$	: La validation croisée.
$UCV$	: La validation croisée non biaisée.
$LCV$	: La validation croisée par le maximum de vraisemblance.
$MCMC$	: Méthode de Mont carlo par chaîne de Markov.

## Résumé

Ce travail, porte sur la sélection du paramètre de lissage lors de l'estimation d'une densité de probabilité discrète. Plus précisément, nous sommes intéressés au choix global du paramètre de lissage par l'inférence bayésienne. L'étude numérique réalisée sur des échantillons issus de différentes distributions et ayant de différentes tailles, montre que la sélection du paramètre de lissage par l'approche bayésienne globale, peut améliorer le *ISE* des estimateurs en les comparant avec ceux conçus par des paramètres de lissage sélectionnés par la méthode classique *UCV*.

**Mots clés :** Paramètre de lissage, Noyaux discrets, Méthodes validation croisées, Approche bayésienne, Les Méthodes MCMC.

## Abstract

This work deals with the selection of the smoothing parameter when estimating a discrete probability density. More precisely, we are interested in the construction of a procedure for the selection of this parameter by global Bayesian inference. The numerical application realized, on simulated data, enabled us to highlight the impact of the choice of the parameters of the kernel estimator, of a discrete density, on its performances. In addition, the result show that the global Bayesian inference give us better estimators, in sense of the *ISE*, compared with those conceived through the smoothing parameter selected by the *UCV* procedure.

**Keywords :** Smoothing parameter ; Discrete-kernels ; Cross-validation methods ; Bayesian approach ; MCMC methods.