

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**

folder/New folder (7)/33/sara finel/UMKB<sub>l</sub>ogo.wmf

Mémoire présenté en vue de l'obtention du Diplôme

**MASTER en Mathématiques**

Option : **Statistique**

Par

**Ben Harkat *Raouia***

Titre :

**Approche bayésienne par monte carlo pour  
estimation à noyau**

Membres du Comité d'Examen :

Dr. <b>DJABRAN YAHIA</b>	UMKB	Président
Dr. <b>BERKANE Hassiba</b>	UMKB	Encadreur
Dr. <b>BENATIA FATEH</b>	UMKB	Examineur

22 Juin 2019

## *Dédicace*

*je dédie ce modeste travail :*

*A mes parents, mes estimes pour eux sont immenses, je vous remercie pour tout ce que vous avez fait pour moi. Que dieu vous préserve une longue vie heureuse.*

*Avec toute ma tendresse.*

*Pour vous, mes chères frères : **walid***

*Pour vous, mes chères sœurs : **Sara, Asma, Wafa, Chaima.***

*pour vous, **iyad et loay et ayoub et mouadh***

*Aux personnes que je n'oublierais jamais mes amies : **hind, Sakina, houria, Samah,**  
*pour son soutien moral et sa présence dans bons et les mauvais moments.**

*A tous les membres de ma promotion.*

*A tous mes professeurs*

**Raouia**

## *REMERCIEMENTS*

*Je tiens à remercier ALLAH pour la volonté, la santé, et la patience. Je tiens à adresser mes remerciement à mon encadreur BERKANE Hassiba de m'avoir appris le sens de la responsabilité et de travail toujours bien fait et sur tout pour les connaissances, ces conseils judicieux et son apport précieux tout au long de la réalisation de ce mémoire et pour leur disponibilité.*

*Nous adressons également nos remerciements à tous les membres de jury, nous sommes très reconnaissant à leurs remarques et commentaires qui nous aidions beaucoup pour mieux présenter ce document.*

*Grand remerciement à monsieur le professeur BENATIA Fateh pour son aide.*

*Merci également à l'étudiante de doctorat : BOUREDJI Hind*

*Mes remerciements vont aussi à tous mes professeurs, enseignants et toutes les personnes qui m'ont soutenu jusqu'au bout, et qui n'ont pas cessé de me donner des conseils très importants en signe de reconnaissance.*

# Table des matières

Dédicace	i
Remerciements	ii
Table des matières	iii
Liste des figures	v
Liste des tableaux	vi
Introduction	1
<b>1 Estimation non-paramétrique d'une densité de probabilité à noyau</b>	<b>3</b>
1.1 Estimateur à noyau . . . . .	3
1.1.1 Quelques définitions . . . . .	3
1.1.2 Exemples de noyaux usuels . . . . .	6
1.2 Expression du variance et le biais d'un estimateurs á noyau . . . . .	7
1.3 Propriétés asymptotique de l'estimateur . . . . .	8
1.4 Choix du paramètre de lissage . . . . .	10
1.4.1 Méthode de plug-in (ré-injection) . . . . .	10
1.4.2 Méthodes de Validation croisée par moindres carrés . . . . .	12
1.4.3 Validation croisée biais . . . . .	13
1.4.4 Validation croisée de la vraisemblance . . . . .	14

<b>2</b>	<b>Le choix de paramètre de lissage par l'approche bayésienne</b>	<b>16</b>
2.1	Les principes de la démarche bayésienne . . . . .	16
2.2	Choix de loi a priori . . . . .	18
2.2.1	Loi a priori conjuguée . . . . .	19
2.2.2	Priories non informatives . . . . .	20
2.2.3	A priori Uniforme . . . . .	20
2.2.4	A priori impropre . . . . .	20
2.3	Estimateur bayésien . . . . .	20
2.3.1	Estimation intrinsèque . . . . .	21
2.3.2	Propriétés de l'estimateur de Bayes . . . . .	23
2.4	Les méthodes de Monte Carlo . . . . .	23
2.4.1	Méthodes de Monte Carlo par chaînes de Markov (MCMC) . . . . .	23
2.5	Étapes de l'approches bayésienne pour la sélection du paramètre de lissage	26
<b>3</b>	<b>Simulation et résultats numériques</b>	<b>31</b>
3.1	Application Numérique . . . . .	31
3.1.1	Introduction . . . . .	31
3.1.2	Etude de Simulation . . . . .	31
	<b>Conclusion</b>	<b>34</b>
	<b>Bibliographie</b>	<b>35</b>
	<b>Annexe A : Abréviations et Notations</b>	<b>37</b>

# Table des figures

1.1	La Forme des noyaux usuels. . . . .	6
1.2	Les courbes des différents noyaux usuels. . . . .	7

# Liste des tableaux

1.1	Quelques noyaux classiques. . . . .	6
2.1	Lois a priori conjuguées pour quelques familles exponentielles usuelles. . . .	19
3.1	résultats obtenus pour le cas $g_1$ . . . . .	33
3.2	résultats obtenus pour le cas $g_2$ . . . . .	33
3.3	résultats obtenus pour le cas $g_3$ . . . . .	33

# Introduction

La théorie de l'estimation est une des préoccupations majeures des statisticiens. On trouve dans la littérature deux types d'approches d'estimations de la densité de probabilité : l'approche paramétrique et l'approche non-paramétrique. L'approche paramétrique a comme inconvénient principal de nécessiter une connaissance préalable du phénomène aléatoire considéré. L'approche non paramétrique estime la densité de probabilité directement à partir de l'information disponible sur l'ensemble d'observations.

L'estimation à noyau (ou encore méthode de *Parzen-Rosenblatt*) est une méthode non-paramétrique d'estimation de la densité de probabilité d'une variable aléatoire. Elle se base sur un échantillon d'une population statistique et permet d'estimer la densité en tout point du support. Cette estimateur sont des fonctions de deux paramètres  $K$ , appelé noyau, et  $h$  dit paramètre de lissage (largeur de fenêtre). Si le choix du noyau n'est pas un problème dans l'estimation de la densité, il n'en est pas de même pour le choix de paramètre de lissage qui ne dépend que de la taille  $n$  de l'échantillon. Plusieurs travaux ont montré que les estimateurs peuvent changer dramatiquement pour de petites variations du paramètre de lissage. Actuellement il n'existe pas de choix optimal pour ce paramètre, le choix optimal qui minimise l'erreur quadratique moyenne intégrée dépend de la dérivée seconde de la densité inconnue.

L'objectif de ce travail est d'utiliser l'approche *Bayésienne* pour le choix du paramètre de lissage  $h$  dans l'estimation de la fonction densité de probabilité par la méthode des noyaux. Cette alternative *Bayésienne* est dirent des méthodes classiques qui consistent souvent à

minimiser ou à maximiser directement un certain critère. Cette loi sert à compenser le manque d'information, lorsqu'il s'agit de données de petite ou moyenne taille. L'estimation *Bayésienne* de  $h$  peut-être obtenue par la moyenne a posteriori, en calculant la loi a posteriori à partir de la règle de Bayes. Parmi les trois variantes de l'approche *Bayésienne*, nous proposons l'approche *Bayésienne globale* pour le choix du paramètre de lissage  $h$ . Dans l'estimation *Bayésienne globale*, la loi a posteriori est souvent de forme complexe. Cette difficulté est, généralement, surmontée grâce aux méthodes populaires de Monte Carlo par Chaînes de Markov (*MCMC*).

Ce mémoire est composé d'une introduction, de trois chapitres et d'une conclusion. Le premier chapitre présente de l'estimation non paramétrique de la densité de probabilité. En effet, après la présentation de la définition de l'estimateur à noyau de cette fonction de densité et ses propriétés et expression (biais, variances,...), nous avons abordé le problème du choix du noyau et du paramètre de lissage par les techniques classiques. Dans le deuxième chapitre présente le cadre théorique dans lequel les outils statistique *bayésienne*, sont développés, ainsi qu'un rappel de la théorie des méthodes usuelles d'approximation exactement *MCMC*, ensuite nous avons présenté l'approche *bayésienne globale* pour l'estimation du paramètre de lissage. Le troisième chapitre regroupera les résultats de simulation des différentes méthodes de sélection du paramètre de lissage. Tous les résultats numériques et graphiques sont effectués à l'aide du logiciel *R*.

# Chapitre 1

## Estimation non-paramétrique d'une densité de probabilité à noyau

La méthode du noyau est l'une des méthodes d'estimation non paramétrique la plus utilisée. *Rosenblatt (1956)*, *suivie de Parzen (1962)*, ont proposé une classe d'estimateurs à noyau d'une densité de probabilité. Cet estimateur est une fonction de deux paramètres : le noyau  $K$  et le paramètre de lissage  $h$ .

Dans ce chapitre, nous avons présenté la notion d'une densité de probabilité par la méthode du noyau. En effet, après la présentation de l'estimateur à noyau de cette densité, nous citons quelques unes de ses propriétés et expression (le biais, la variance,...). Ensuite, nous allons aborder le problème de choix du noyau et du paramètre de lissage dans ce cas.

### 1.1 Estimateur à noyau

#### 1.1.1 Quelques définitions

Dans cette section nous allons nous intéresser à l'estimateur à noyau symétrique. Avant d'exposer ses différentes composantes nous présenterons d'abord sa construction d'origine qui se fait à travers de l'estimateur empirique d'une fonction de répartition.

**Définition 1.1.1** *On observe  $X_1, X_2, X_3, \dots, X_n$  variable aléatoires réelles indépendant et identiquement distribué (i.i.d) de fonction de répartition :*

$$F : x \rightarrow F(x) = P(X_1 \leq x)$$

*L'estimateur de la fonction de répartition  $F$  est la fonction de répartition empirique notée  $F_n$  et définie par :*

$$F_n(x) = \hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n I_{\{X_i \leq x\}} = \begin{cases} 0 & \text{si } x \leq X_1 \\ \frac{k}{n} & \text{si } X_k \leq x < X_{k+1} \\ 1 & \text{si } x \geq X_n \end{cases}$$

*C'est un estimateur non paramétrique de fonction de répartition  $F$ .*

**Définition 1.1.2** *Soit  $X_1, X_2, X_3, \dots, X_n$  un échantillon de variables aléatoires (i.i.d) de fonction de répartition  $F$  on a :*

$$f(x) = \lim_{h \rightarrow 0} \frac{F(x+h) - F(x)}{2h}$$

$$f_n(x) = \lim_{h \rightarrow 0} \frac{F_n(x+h) - F_n(x-h)}{2h} \tag{1.1}$$

L'estimation à noyau de *Rosenblatt* en 1956 notée  $\hat{f}_n(x)$  et définie par :

$$\begin{aligned} \hat{f}_n(x) &= \lim_{h \rightarrow 0} \frac{\hat{F}_n(x+h) - \hat{F}_n(x-h)}{2h} \\ &= \frac{1}{2h} \left[ \frac{1}{n} \sum_{i=1}^n I_{\{X_i \leq x+h\}} - \frac{1}{n} \sum_{i=1}^n I_{\{X_i \leq x-h\}} \right] \\ &= \frac{1}{2nh} \left[ \sum_{i=1}^n I_{\left\{\frac{X_i-x}{h} \leq 1\right\}} - \sum_{i=1}^n I_{\left\{\frac{X_i-x}{h} \leq -1\right\}} \right] \\ &= \frac{1}{nh} \left[ \sum_{i=1}^n \frac{1}{2} I_{\left\{\frac{|X_i-x|}{h} \leq 1\right\}} \right]. \end{aligned}$$

Le premier exemple d'estimateur à noyau construit à l'aide du noyau :

$$K(z) = \frac{1}{2} I_{\{|z| \leq 1\}}$$

ou plus généralement (*Parzen*, 1962) :

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

où  $K$  est le noyau de cet estimateur telle que :

- $K(z) \geq 0$
- $\int_{\mathbb{R}} K(z) dz = 1$
- $h > 0$  est une paramètre de lissage il dépend de  $n$  et il vérifie  $h(n) \rightarrow 0$  lorsque  $n \rightarrow \infty$

### 1.1.2 Exemples de noyaux usuels

Les noyaux les plus couramment utilisés en pratique sont :

Noyau	Fonction	Domain
Rectangulaire	$\frac{1}{2}$	$[-1, 1]$
Gaussien	$\frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$	$\mathbb{R}$
Triangulaire	$(1 -  u )$	$[-1, 1]$
D'épanechnikove	$\frac{3}{4}(1 - u^2)$	$[-1, 1]$

TAB. 1.1 – Quelques noyaux classiques.

Les courbes de ces noyaux sont présentées ci-dessous :

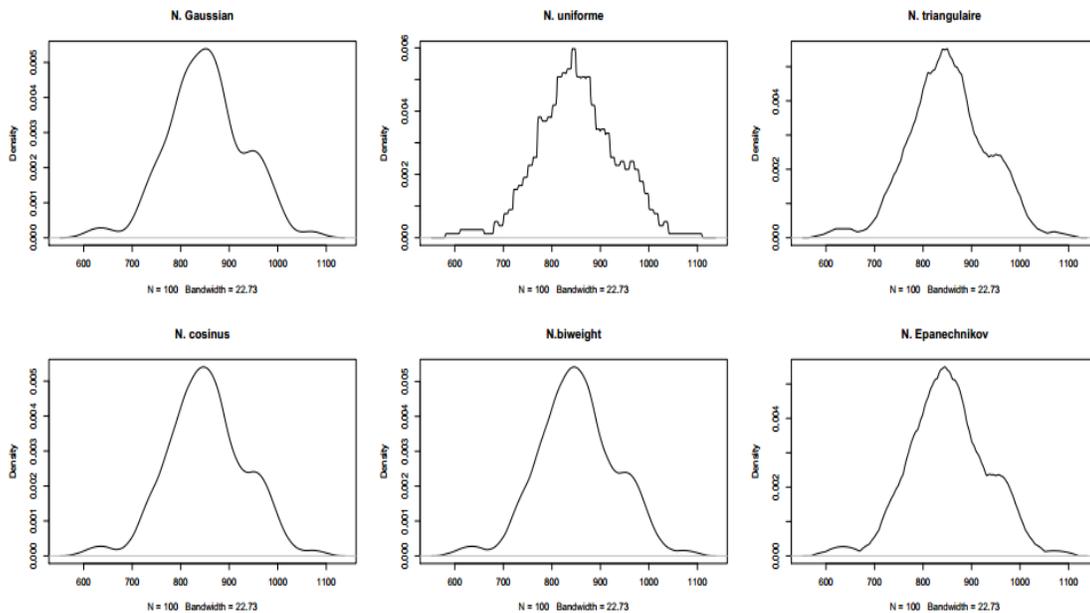


FIG. 1.1 – La Forme des noyaux usuels.

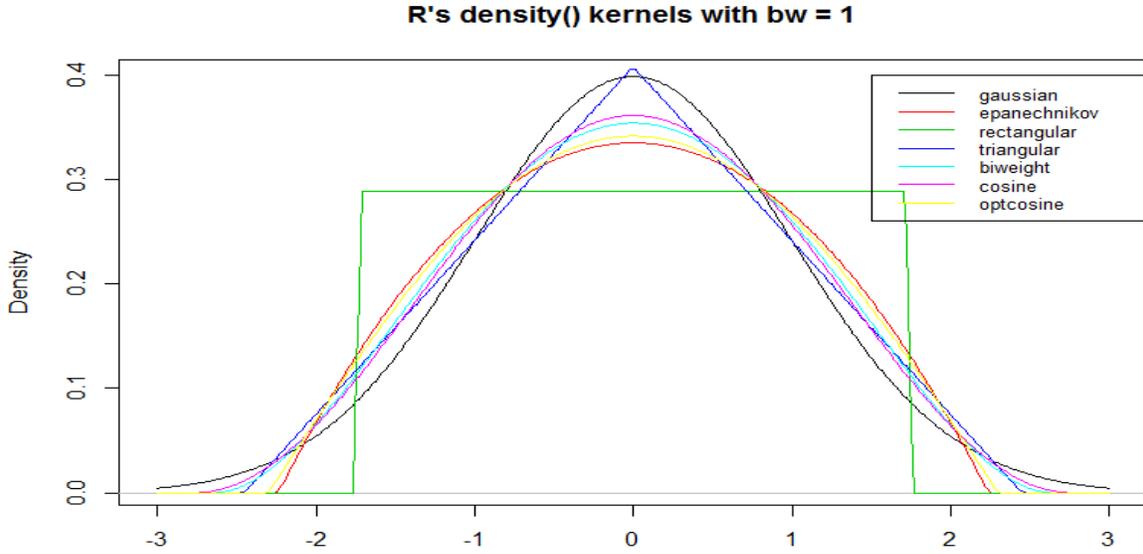


FIG. 1.2 – Les courbes des différents noyaux usuels.

## 1.2 Expression du variance et le biais d'un estimateurs à noyau

Pour que  $\hat{f}(x)$  soit une densité, on suppose que  $K(u) \geq 0$  et  $\int_{\mathbb{R}} K(u)du = 1$ .

La fonction noyau est supposée être symétrique autour de zéro, c.à.d  $\int_{\mathbb{R}} uK(u)du = 0$  et possède un moment d'ordre 2 fini, c.à.d,  $\int_{\mathbb{R}} u^2K(u)du < \infty$ .

Nous avons établi que :

$$E \left[ \hat{f}_n(x) \right] = f(x) + \frac{h^2}{2} f^{(2)}(x) \mu_2(K) + o(h^2).$$

$$\text{biais} \left( \left( \hat{f}_n(x) \right) \right) = E \left[ \hat{f}(x) \right] - f(x) = \frac{h^2}{2} f^{(2)}(x) \mu_2(K) + o(h^2).$$

$$\text{var} \left[ \hat{f}_n(x) \right] = \frac{f(x)}{nh} \int_{-\infty}^{+\infty} K^2(z) dz - \frac{f^{(1)}(x)}{n} \int_{-\infty}^{+\infty} zK^2(z)dz - \frac{1}{n} \left( f(x) + \text{biais} \left( \hat{f}(x) \right) \right)^2.$$

Les expressions asymptotiques du biais et de la variance nous permettent de trouver des expressions pour l'erreur quadratique moyenne ( $MSE$ ) et l'erreur quadratique moyenne

intégrée (*MISE*) et par définition le *MSE* et le *MISE* d'un estimateur sont donnée respectivement par :

$$\begin{aligned}
 MSE(f(x), \hat{f}_n(x)) &= E \left[ \left( \hat{f}_n(x) - f(x) \right)^2 \right] \\
 &= E \left[ \left( \hat{f}_n(x) - E \left( \hat{f}_n(x) \right) \right)^2 \right] + 2E \left[ \hat{f}_n(x) - E \left( \hat{f}_n(x) \right) \right] \\
 &= \left[ E \left( \hat{f}_n(x) \right) - f(x) \right] + \left[ E \left( \hat{f}_n(x) \right) - f(x) \right]^2 \\
 &= var \left[ \hat{f}_n(x) \right] + biais^2 \left( \left( \hat{f}_n(x) \right) \right)
 \end{aligned}$$

$$\begin{aligned}
 MISE(f, \hat{f}) &= \int_{\mathbb{R}} MSE(f(x), \hat{f}_n(x)) dx \\
 &= \int_{\mathbb{R}} var \left[ \hat{f}_n(x) \right] + \int_{\mathbb{R}} biais^2 \left( \left( \hat{f}_n(x) \right) \right)
 \end{aligned}$$

En reprenant l'expression du biais de l'estimateur  $f_n$  et la variance et en remplaçant dans l'expression *MSE* on obtient alors :

$$MSE(f(x), \hat{f}_n(x)) = \frac{h^4}{4} \delta_K^4 (f^{(2)}(x))^2 + \frac{f(x)}{nh} \int K^2(y) dy + o(h^4 + \frac{1}{nh})$$

$$MISE(f, \hat{f}) = \frac{h^4}{4} \delta_K^4 \int (f^{(2)}(x))^2 dx + \frac{1}{nh} \int K^2(x) dx + o(h^5 + \frac{1}{nh}) \quad (1.2)$$

avec  $\delta_K^4 = (\int y^2 K(y) dy)^2$ .

### 1.3 Propriétés asymptotique de l'estimateur

Nous présentons dans cette partie les propriétés statistiques de l'estimateur de densité  $f_n$ .

1. Si la fonction  $K$  est positive et continue et  $\int K(z)dz = 1$  alors  $\hat{f}_n$  est une densité de probabilité.
2. La fonction noyau  $K$  est une densité de probabilité sur  $\mathbb{R} \rightarrow \mathbb{R}_+$  est symétrique par rapport à zéro :

$$K(-x) = K(x)$$

ce qui implique

$$\int tK(t) dt = 0 \tag{1.3}$$

de plus

$$\int K^2(t) dt < \infty \tag{1.4}$$

$$\int t^2 |K(t)| dt < \infty \tag{1.5}$$

- 3 Sous 1.3 et 1.4 1.5 et supposons que  $f$  est une densité bornée dont la dérivée seconde est bornée, alors on a :

$$|biais(\hat{f}_n(x))| \leq c_1 h^2$$

Si, de plus, la condition 1.5 de l'hypothèse  $K$  est satisfaite, alors :

$$var(\hat{f}_n(x)) \leq \frac{c_2}{nh}$$

où  $c_1$  et  $c_2$  est une constantes.

**Proof.** d'après 1.2 on a :

$$\begin{aligned} |biais(\hat{f}_n(x))| &= \left| \frac{h^2}{2} f^{(2)}(x) \mu_2(K) + o(h^2) \right| \\ &\leq h^2 \sup_z f^{(2)}(z) \int_{\mathbb{R}} \frac{t^2}{2} |f^{(2)}(t)| K(t) dt \leq c_1 h^2 \end{aligned}$$

et on a :

$$\begin{aligned}
 \text{var} \left( \hat{f}_n(x) \right) &= E \left[ \left( \hat{f}_n(x) \right)^2 \right] - E^2 \left[ \hat{f}_n(x) \right] \\
 &\leq E \left[ \left( \hat{f}_n(x) \right)^2 \right] \\
 &= E \left[ \left( \frac{1}{nh} \sum_{i=1}^n K \left( \frac{X_i - x}{h} \right) \right)^2 \right] \\
 &= \frac{1}{nh^2} \int_{\mathbb{R}} K^2 \left( \frac{y - x}{h} \right) f(y) dy
 \end{aligned}$$

en prend  $\frac{y-x}{h} = t$  donc  $y = x + th$  et  $dy = hdt$  alors :

$$\begin{aligned}
 \text{var} \left( \hat{f}_n(x) \right) &\leq \frac{1}{nh} \int_{\mathbb{R}} K^2(t) f(x + th) dt \\
 &\leq \frac{1}{nh} \sup_z f(z) \int_{\mathbb{R}} K^2(t) dt \\
 &\leq \frac{c_2}{nh}
 \end{aligned}$$

■

## 1.4 Choix du paramètre de lissage

Dans cette partie, Elles comparent plusieurs méthodes pour choisir le paramètre de lissage pour plusieurs distributions différentes. Toutes ces méthodes nous donnent un paramètre de lissage qui est optimale pour la distribution à estimer. Celles-ci diffèrent au niveau du choix du critère à optimiser. Dans ce chapitre on va étudier les méthodes suivantes :

### 1.4.1 Méthode de plug-in (ré-injection)

La méthodes plug-in(ré-injection) est constitué des méthodes purement théoriques qui sont basées sur la minimisation de l'erreur quadratique moyenne intégrée ( $MSE$ ) donnée par (1.2) où  $\delta_K^4 = \left( \int y^2 K(y) dy \right)^2$ .

Mais vue la difficulté de calcul du  $MISE$ , dans la pratique on utilise l'erreur quadratique moyenne intégrée asymptotique, noté  $AMISE$ , obtenu en ignorant les termes d'ordre supérieurs du  $MISE$ .

$$AMISE = \frac{h^4}{4} \delta_K^4 \int (f^{(2)}(x))^2 dx + \frac{\int K^2(x) dx}{nh}.$$

L'expression de paramètre de lissage optimale, au sens  $dAMISE$ , se détermine par une minimisation directe de ce critère par rapport à  $h$ . En effet, il suffit de résoudre le système suivant :

$$\begin{cases} \frac{\partial AMISE}{\partial h} = 0 \\ \frac{\partial^2 AMISE}{\partial h} > 0 \end{cases}$$

La résolution de ce dernier nous conduit à :

$$h^* = \left[ \frac{\int K^2(y) dy}{\delta_K^4 \int (f^{(2)})^2 dx} \right] n^{-\frac{1}{5}}$$

On constate que le paramètre de lissage  $h^*$  dépend de la densité inconnue  $f$  à travers  $\int (f^{(2)})^2 dx$ . une façon classique de remédier à ce problème consiste à remplacer la quantité  $\int (f^{(2)})^2 dx$  par un estimateur approprié, nous citons par exemple, la méthode de règle de référence (*Rule of Thumb*) qui consiste à choisir  $f$  comme étant la distribution normale de moyenne  $\mu_f$  et de variance  $\delta_f^2$ . Les paramètres  $\mu_f$  et  $\delta_f^2$  sont estimés à l'aide des observations  $x_1, x_2, \dots, x_n$  respectivement par la moyenne empirique :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

et la variance empirique :

$$S_n^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

Lorsque le noyau  $k$  est Gaussian, alors le paramètre de lissage optimale, au sens du *AMISE* est de la forme suivante :

$$h^* = 1.06 S_n^2 n^{-\frac{1}{5}}.$$

### 1.4.2 Méthodes de Validation croisée par moindres carrés

*Rudemo* (1982) et *Bowman* (1984) ont donné l'idée de cette méthode. En utilisant la formule développée de l'erreur quadratique intégrée  $ISE(h)$ , on choisit le paramètre de lissage  $h$  qui minimise cette erreur.

$$ISE(h) = \int_{\mathbb{R}} (\hat{f}_n(x) - f(x))^2 dx = \int_{\mathbb{R}} \hat{f}_n^2(x) dx - 2 \int_{\mathbb{R}} \hat{f}_n(x) f(x) dx + \int_{\mathbb{R}} f^2(x) dx.$$

On remarque que  $\int_{\mathbb{R}} f^2(x) dx$  ne dépend pas de  $h$ , donc on peut choisir le  $h$  de façon à ce qu'il minimise le critère de la validation croisée défini par :

$$UCV(h) = ISE(h) - \int_{\mathbb{R}} f^2(x) dx = \int_{\mathbb{R}} \hat{f}_n^2(x) dx - 2 \int_{\mathbb{R}} \hat{f}_n(x) f(x) dx.$$

On doit donc trouver un estimateur pour  $\int_{\mathbb{R}} \hat{f}_n(x) f(x) dx$  remarquons que :

$$\int_{\mathbb{R}} \hat{f}_n(x) f(x) dx = E \left[ \hat{f}_n(x) \right].$$

L'estimateur empirique de  $\int_{\mathbb{R}} \hat{f}_n(x) f(x) dx$  est alors :

$$\frac{1}{n} \sum_{i=1}^n \hat{f}_{h,i}(x_i)$$

Le critère à optimiser est alors :

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

où

$$f_{h,i}(x_i) = \frac{1}{(n-1)h} \sum_{\substack{j=1 \\ j \neq i}}^n K\left(\frac{x_i - x_j}{h}\right)$$

est l'estimateur de la densité construit à partir de l'ensemble de points sauf le point  $x_i$ .

En utilisant les formules explicite de  $\hat{f}_n(x)$  et  $f_{h,i}(x_i)$ , le critère  $UCV(h)$  est donné par :

$$UCV(h) = \frac{R(K)}{nh} + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \left[ \frac{1}{(nh)^2} \int K\left(\frac{x - x_i}{h}\right) K\left(\frac{x - x_j}{h}\right) dx - \frac{2}{n(n-1)h} K\left(\frac{x_i - x_j}{h}\right) \right]$$

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

### 1.4.3 Validation croisée biais

La notion de validation croisée biaisée, a été introduite par *Scott* et *Terrell* en 1987, l'idée de cette méthode est de trouver la valeur de  $h$  qui minimise un estimateur du *AMISE*.

Nous avons déjà vu que :

$$AMISE = \frac{h^4}{4} \delta_K^4 \int (f^{(2)}(x))^2 dx + \frac{\int K^2(x) dx}{nh}$$

Le paramètre de lissage basé sur la méthode de la validation croisée biaisée est la valeur de  $h$  qui minimise un estimateur du *AMISE*. On peut estimer le *AMISE* si l'on connaît  $\int (f^{(2)}(x))^2 dx$ .

Scott et Terrell (1987) proposent d'estimer  $\int (f^{(2)}(x))^2 dx$  par :

$$\int (\hat{f}^{(2)}(x))^2 dx - \frac{1}{nh^5} \int K^{(2)}(x) dx.$$

où  $\hat{f}^{(2)}(x)$  désigne la dérivé second de l'estimateur à noyau, la forme de l'estimateur de *AMISE* à minimiser (critère de *BCV(h)*), est résumé comme suit :

$$\begin{aligned} BCV(h) &= \frac{(h\delta_K)^4}{4} \left[ \int (\hat{f}^{(2)}(x))^2 dx - \frac{1}{nh^5} \int (K^{(2)}(x))^2 dx \right] + \frac{\int K^2(x) dx}{nh} \\ &= h^4 \frac{\mu_2^2}{4n^2} \sum_i \sum_{j, j \neq i} K^{(2)}(x) K^{(2)}(x) (X_i - X_j). \end{aligned}$$

avec  $R(g) = \int g^2(x) dx$ .

Le paramètre de lissage  $h$  choisi par cette méthode est la valeur de  $h$  qui minimise *BCV(h)*

$$h_{bcv} = \underset{h}{\operatorname{argmin}}(BCV(h)).$$

#### 1.4.4 Validation croisée de la vraisemblance

La plus simple des méthodes de type validation croisée est celle proposée par *Habbema* et al.[1974], appelée validation croisée du maximum de vraisemblance. Elle consiste à maximiser par rapport à  $h$  un estimateur de la vraisemblance donné par :

$$\begin{aligned} LCV(h) &= \prod_{i=1}^n \hat{f}_i(x) \\ &= \frac{1}{(n-1)^n} \prod_{i=1}^n \sum_{j=1, j \neq i}^n K_{x_i, h}(x_j). \end{aligned}$$

Le paramètre de lissage optimal fourni par cette méthode est donné par :

$$h_{lcv} = \operatorname{argmax} LCV(h).$$

Cette méthode est probablement la moins utilisée car elle révèle un certain nombre de faiblesses pour les estimateurs à noyau. Plusieurs études ont mis en avant la mauvaise robustesse de cette méthode ainsi que le risque qu'elle conduise vers une estimée non consistante lorsqu'elle est appliquée à des observations dont la distribution présente des queues. C'est la méthode la moins utilisée pour choisir le paramètre de lissage optimal.

# Chapitre 2

## Le choix de paramètre de lissage par l'approche bayésienne

Dans ce chapitre, nous allons présenter la démarche à suivre pour l'approche bayésienne dans le but d'estimer le paramètre de lissage dans l'estimation à noyau de la fonction de densité. En effet, dans un premier lieu nous allons rappeler quelques notions de base du formalisme bayésien et les estimateurs paramétriques bayésiens ainsi que les méthodes *MCMC* (*Monte Carlo par chaîne de Markov*) qui peuvent être employées lors de l'implémentation pratique de l'approche bayésienne.

Dans un seconde lieu, nous allons présenter une situation du choix du paramètre de lissage par l'inférence bayésienne à savoir : le choix du paramètre de lissage global.

### 2.1 Les principes de la démarche bayésienne

L'idée principale de l'analyse bayésienne repose sur la *loi a posteriori* des paramètres en considérant  $\theta$  comme variable aléatoire. et 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'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.

**Définition 2.1.1** On appelle *modèle statistique bayésien* sur  $(\Theta, \mathcal{F}, \mathcal{P})$  la donnée d'un modèle statistique paramétré  $f(\theta)$ , et d'une loi a priori sur les paramètres  $\pi(\theta)$ .

La loi conditionnelle de  $\theta$  sachant les observations  $x$  est appelée *loi a posteriori*

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int_{\Theta} f(x|\theta)\pi(\theta)d\theta}$$

et on interprète la loi des observations  $f_{\theta}$  comme la loi conditionnelle des observations sachant  $\theta$

$$f(x|\theta) = f_{\theta}(x)$$

et comme la densité  $g(\theta|x) = f(x|\theta)\pi(\theta)$  est appelée *la loi du couple*  $(\theta, x)$ , et la densité  $m(x) = \int_{\Theta} f(x|\theta)\pi(\theta)d\theta$  est une *loi marginale* de  $x$ .

**Exemple 2.1.1** Soit  $X|\theta \rightsquigarrow P(\theta)$  et  $\theta \rightsquigarrow \gamma(2, 1)$ . Donnons la loi de probabilité a posteriori du paramètres  $\theta$  ainsi que la loi marginale de  $X$ .

On a :

$$f(x|\theta) = \frac{\theta^x \exp(-\theta)}{x!} \quad x \in \mathbb{N}$$

$$\pi(\theta) = \theta \exp(-\theta) \quad \theta > 0$$

$$g(\theta|x) = \frac{\theta^{x+1} \exp(-2\theta)}{x!}$$

$$\begin{aligned}
 m(x) &= \int_0^{\infty} f(x|\theta)\pi(\theta)d\theta = \int_0^{\infty} \frac{\theta^{x+1} \exp(-2\theta)}{x!} d\theta \\
 &= \int_0^{\infty} \frac{\Gamma(x+2)}{\Gamma(x+2)} \frac{\theta^{x+1} \exp(-2\theta)}{x!} \frac{2^{x+2}}{2^{x+2}} d\theta \\
 &= \frac{\Gamma(x+2)}{x!2^{x+2}} = \frac{(x+1)!}{x!2^{x+2}} \quad \text{avec } \Gamma(x+2) = (x+1)!
 \end{aligned}$$

d'où la loi marginale de  $X$  :

$$m(x) = \frac{(x+1)}{2^{x+2}}$$

et la loi a posteriori de  $\theta$  :

$$f(\theta|x) = \frac{2^{x+2}\theta^{x+1} \exp(-2\theta)}{(x+1)!}, \quad \theta > 0$$

donc :

$$\theta|x \rightsquigarrow \gamma(x+2, 2)$$

## 2.2 Choix de loi a priori

Le choix de la loi a priori est une étape fondamentale dans l'analyse bayésienne. En théorie cette loi doit représenter notre information a priori sur le paramètre. En pratique, il est rare que l'information a priori soit suffisamment précise pour déterminer exactement la loi a priori du paramètre  $\theta$ .

Dans le cadre de cette étude, on se contente de ne présenter que les types de densités a priori les plus courants : les densités a priori conjuguées et les densités a priori non informatives.

### 2.2.1 Loi a priori conjuguée

Une famille  $F$  de lois sur  $\Theta$  est dite conjuguée ou fermée par échantillonnage si pour tout  $\pi(\theta) \in F$ , la loi a posteriori  $\pi(\theta|x)$  appartient également à  $F$ . *Raiiffa et Schlaifer* (1961), qui sont l'origine des lois conjuguées. Lorsque les observations  $x$  modifient la loi a priori  $\pi(\theta)$  en la loi a posteriori  $\pi(\theta|x)$ , l'information apportée par  $x$  sur  $\theta$  est limitée, donc elle ne doit pas conduire à une remise en cause de la forme de la loi  $\pi(\theta)$ . Mais elle peut seulement changer ses paramètres en fonction de  $x$ .

**Définition 2.2.1** *une famille  $\mathcal{F}$  de distributions de probabilité sur  $\Theta$  est dite conjuguée par une fonction de vraisemblance  $f(x|\theta)$  si pour tout  $\pi \in \mathcal{F}$ , la distribution a posteriori  $\pi(\theta|x)$  appartient également à  $\mathcal{F}$ .*

**Exemple 2.2.1** *tableau ci-dessus représente quelques lois a priori conjuguées naturelles pour quelques familles exponentielles usuelles.*

$f(x \theta)$	$\pi(\theta)$	$\pi(\theta x)$
<i>Binomiale</i> $B(n, \theta)$	<i>Beta</i> $Beta(\alpha, \beta)$	<i>Beta</i> $Beta(\alpha + x, \beta + n - x)$
<i>Normale</i> $N(\theta, \sigma^2)$	$N(\mu, \tau^2)$	$N(\varphi(\sigma^2\mu + \tau^2x), \varphi\sigma^2\tau^2)$ avec $\varphi^{-1} = \sigma^2 + \tau^2$
<i>Poisson</i> $P(\theta)$	<i>Gamma</i> $G(\alpha, \beta)$	<i>Gamma</i> $G(\alpha + x, \beta + 1)$
<i>Gamma</i> $G(v, \theta)$	<i>gamma</i> $G(\alpha, \beta)$	<i>gamma</i> $G(\alpha + v, \beta + x)$

TAB. 2.1 – Lois a priori conjuguées pour quelques familles exponentielles usuelles.

## 2.2.2 Priors non informatives

Dans le cas d'absence d'information précise sur le paramètre à estimer, les bayésiens proposent d'utiliser des priors non informatives et cela le fait que aucune information n'est disponible, il est impossible de bâtir une distribution a priori sur des considérations subjectives. Dans ce cas, on laisse les données conduire l'inférence qui peut être justifié par les résultats théoriques qui montrent que l'inférence dans les deux paradigmes (bayésien et fréquentiste). Voici quelques considérations sur les priors non informatives.

## 2.2.3 A priori Uniforme

La densité a priori non informative la plus simple et la plus communément utilisée est la densité Uniforme. En effet, ce choix repose sur l'équiprobabilité des valeurs possibles du paramètre  $\theta$  sur son domaine de définition, et donc n'apporte aucune information supplémentaire sur  $\theta$ . Ainsi, la densité est définie par :

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

$d$  est une constante positive

## 2.2.4 A priori impropre

La loi a priori peut être impropre i.e.  $\int_{\Theta} \pi(\theta) d\theta = \infty$ .

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.

## 2.3 Estimateur bayésien

Soit  $X = (X_1, \dots, X_n)$  un  $n$  - échantillon de variables aléatoire *iid* de densité  $f(x|\theta)$ , où  $\theta$  est un paramètre inconnu à estimer.

On dit que le  $\theta$  est une valeur d'une variable aléatoire auquel on assigne une loi a priori  $\pi(\theta)$ , puis on détermine la distribution a posteriori  $\pi(\theta|x)$ .

### 2.3.1 Estimation intrinsèque

**Définition 2.3.1** *L'estimateur de Bayes intrinsèque est la fonction des données qui minimise l'espérance postérieure de la perte intrinsèque*

$$\theta^*(x) = \arg \min_{\hat{\theta} \in \Theta} d(\hat{\theta}|x) = \arg \min_{\hat{\theta} \in \Theta} \int_{\Theta} \delta(\hat{\theta}, \theta) \pi(\theta|x) d\theta.$$

*Le cas uni- dimensionnel*

On suppose dans cette section que le paramètre  $\theta$  est réel.

**Proposition 2.3.1** *Si :*

$$l(\hat{\theta}, \theta) = \begin{cases} a & \text{si } \hat{\theta} \neq \theta \quad a > 0 \\ 0 & \text{si } \hat{\theta} = \theta, \end{cases}$$

*l'estimateur de Bayes correspond au mode de la distribution a posteriori  $\pi(\theta|x)$  ; est défini par :*

$$\pi(\hat{\theta}|x) = \sup_{\theta \in \Theta} \pi(\theta|x).$$

$$l(\hat{\theta}, \theta) = a_1(\hat{\theta} - \theta)1_{\{\theta \leq \hat{\theta}\}}(t) + a_2(\theta - \hat{\theta})1_{\{\theta > \hat{\theta}\}}(t),$$

l'estimateur de Bayes  $\hat{\theta}$  est le quantile  $\frac{a_1}{a_1+a_2}$  de la distribution a posteriori  $\pi(\theta|x)$  ; est défini par :

$$P(\Theta < \hat{\theta}|x) = \frac{a_1}{a_1 + a_2}$$

Lorsque  $a_1 = a_2$ , l'estimateur de Bayes est la *médiane a posteriori*  $\pi(\theta|x)$  ; est définie par :

$$\int_{-\infty}^{\hat{\theta}} \pi(\theta|x)d\theta = \int_{\hat{\theta}}^{+\infty} \pi(\theta|x)d\theta.$$

Si  $l(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^2$ , l'estimateur de Bayes est la moyenne a posteriori  $\pi(\theta|x)$  ; est définie par :

$$\hat{\theta}_B(x) = E[\theta|x] = \int_{\Theta} \theta \pi(\theta|x) d\theta = \frac{\int_{\Theta} \theta f(x|\theta) \pi(\theta) d\theta}{\int_{\Theta} f(x|\theta) \pi(\theta) d\theta}$$

**Remarque 2.3.1** Il faut signaler que le mode de la loi a posteriori est parfois retenu comme estimation bayésienne de  $\theta$  si l'on suspecte que la vraisemblance comporte des maximum locaux. En ce cas, on a :

$$\hat{\theta}_B(x) = \arg \max_{\theta \in \Theta} \pi(\theta|x).$$

La moyenne et la médiane sont généralement de meilleurs estimateurs pour  $\theta$  que le mode.

### **Le cas multi-dimensionnel**

Dans un contexte multi-dimensionnel où  $\theta = (\theta_j ; j = 1, \dots, J)$  la moyenne a posteriori  $E[\theta|x]$  est égale au vecteur  $(E[\theta_j|x] ; j = 1, \dots, J)$ , où :

$$E[\theta_j|x] = \int_{\Theta} \theta_j \pi(\theta_j|x) d\theta_j$$

$\pi(\theta_j|x)$  est obtenu en intégrant  $\pi(\theta|x)$  sur toutes les composantes de  $\theta$  autres que  $\theta_j$ .

Le plus souvent, les estimateurs de Bayes des  $\theta_j$  ne peuvent pas être calculés de façon explicite et il faut faire appel aux méthode de simulation de *Monte carlo*. voici cependant une situation dans la quelle le calcul des estimateurs de Bayes ne pose aucun difficulté.

### 2.3.2 Propriétés de l'estimateur de Bayes

- 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 les deux propriétés :
  - 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-à-d pour de grandes valeurs de  $n$ ) approximée par une loi normale  $N(E[\theta|x], Var[\theta|x])$  Cette dernière propriété est particulièrement utile pour construire des intervalles de confiance a posteriori.

## 2.4 Les méthodes de Monte Carlo

Dans un problème statistique, l'approximation de l'intégrale

$$\int_{\Theta} g(\theta) f(x|\theta)\pi(\theta) d\theta$$

doit tirer avantage de la nature particulière, à savoir de fait que  $\pi$  soit densité de probabilité ou plutôt, que  $f(x|\theta)\pi(\theta)$  soit proportionnel à densité. Une conséquence naturelle de cette perspective est d'utiliser la méthode de *Monte Carlo*, introduite par *Metropolis* et *Ulam*(1949) et *von Neumann* (1951).

### 2.4.1 Méthodes de Monte Carlo par chaînes de Markov (MCMC)

Les méthodes de Monte Carlo par Chaînes de Markov (Markov Chain Monte Carlo - *MCMC*) offrent une alternative pour obtenir des échantillons suivant une certaine loi cible  $\pi$  (et par suite un estimateur des moments d'intérêt de  $\pi$ ) sans pour autant tirer

directement suivant  $\pi$ . La méthodologie est cette fois basée sur la construction d'une chaîne de Markov dont la loi stationnaire est  $\pi$ . Nous rappelons brièvement les principes de base des chaînes de Markov et des méthodes *MCMC*, puis présentons deux techniques *MCMC* bien établies : l'algorithme de *Metropolis-Hastings* (*MH*).

### Chaînes de Markov et principe de MCMC

Soit  $(X_n)$  une suite de variables aléatoires associées aux valeurs dans un ensemble  $E$  supposé fini, typiquement  $E = \{1, 2, \dots, M\}$ .  $E$  est appelé l'espace d'états. On dit que  $(X_n)$  est une chaîne de Markov si pour tout  $n \geq 1$  et toute suite  $(i_0, i_1, \dots, i_{n-1}, i, j)$  de  $E$  telle que :

$$P(X_0 = i_0, \dots, X_{n-1} = i_{n-1}, X_n = i) > 0$$

on a l'égalité suivante :

$$P(X_{n+1} = j | X_0 = i_0, \dots, X_{n-1} = i_{n-1}, X_n = i) = P(X_{n+1} = j | X_n = i) = P(X_1 = j | X_0 = i).$$

Dans le contexte des méthodes *MCMC*, nous nous focalisons sur des chaînes de Markov présentant une propriété particulière appelée *stationnarité* ou invariance. Ceci signifie qu'il existe une loi  $\pi$  telle que : si  $x_n \sim \pi$ , alors  $x_{n+1} \sim \pi$ .

Dans les algorithmes de simulation avec indépendance, la convergence des moyennes empiriques vers les moyennes théoriques est assurée par la Loi des Grands Nombres (LGN), mais cette dernière ne fonctionne plus dans le cadre de simulation avec dépendance.

Afin que la simulation via les chaînes de Markov puisse remplacer les méthodes de simulation, indépendantes, la LGN est étendue par le théorème d'ergodicité (voir *Robert and Casella* et *Robert*) qui résout le paradoxe des deux types de convergence. Ainsi ce théorème supprime le besoin de produire des suites de variables aléatoires indépendantes.

Le principe des méthodes *MCMC* consiste ainsi à partir d'une densité cible  $\pi$  donnée à construire un noyau de transition  $K$ , et par suite une chaîne de Markov  $\{x_n\}_{n \geq 0}$ , admettant  $\pi$  comme loi invariante. Nous rappelons ici deux méthodes bien établies permettant d'obtenir un tel noyau  $K$  : l'algorithme de *Metropolis-Hastings*

### Algorithme M-H

L'algorithme *Metropolis-Hastings* (MH) appartient aussi à la famille des méthodes *MCMC*. La première version de l'algorithme a été introduite en 1953 en considérant le cas particulier de la distribution de *Boltzmann*. Cette première version a été beaucoup utilisée pour traiter des problèmes en physique statistique. En 1970, l'algorithme a été généralisé par *Hastings* au cas de n'importe quelle distribution.

La loi cible peut être connue à une constante multiplicative près, ce qui est souvent le cas lorsque l'on calcule la densité a posteriori. Si l'on ne sait pas simuler suivant cette loi, l'algorithme MH propose d'introduire à partir de la loi cible un noyau de transition arbitraire que l'on appelle loi instrumentale ou aussi loi de proposition à partir de laquelle on sait échantillonner.

Le support de la loi de proposition doit inclure le support de la loi cible, c'est-à-dire que la loi de proposition ne doit pas être nulle sur le support de la loi cible. Les étapes de l'algorithme *M-H* sont donnée comme suit :

1. Initialiser  $\theta^{(0)}$ .
2. Pour  $l \in \{1, \dots, N\}$  :
  - a. générer  $\tilde{\theta} \sim q(\tilde{\theta} | \theta^{(k-1)})$ .
  - b. calculer la probabilité d'acceptation  $\alpha = \min \left\{ 1, \frac{\pi(\tilde{\theta}|x) q(\theta^{(k-1)}|\tilde{\theta})}{\pi(\theta^{(k-1)}|x) q(\tilde{\theta}|\theta^{(k-1)})} \right\}$ .
  - c. générer  $u$ ,  $u \sim U_{[0,1]}$  :

$$\theta^{(k+1)} = \begin{cases} \tilde{\theta} & \text{si } u < \alpha \\ \theta^{(k)} & \text{sinon} \end{cases}$$

3. calculer l'estimateur de Bayes

$$\tilde{\theta} = \frac{1}{N - N_0} \sum_{k=N_0+1}^N \theta^{(k)}$$

## 2.5 Étapes de l'approches bayésienne pour la sélection du paramètre de lissage

Dans cette section nous allons présenter les étape de l'approche Bayésienne globale pour la sélection du paramètre de lissage.pour le choix du paramètre de lissage global dans l'estimation à noyau symétrique d'une densité de probabilité.

L'estimation Bayésienne du paramètre  $h$  conditionné aux données se fait par la densité a posteriori.  $\pi(h|données)$ .Concrètement, nous considérons une séquence  $X_1, X_2, \dots, X_n$  de variables aléatoire réelles, indépendantes et identiquement distribuées par une densité de probabilité  $f$  et de réalisations  $x = (x_1, \dots, x_n)$ . L'estimation bayésienne globale du paramètre de lissage est généralement construite par les quatre étapes suivantes :

1. **Donner la forme de l'estimateur de la vraisemblance  $f(x|h)$  de la densité des données sachant le paramètre  $h$ .**

La vraisemblance des données  $x_1, \dots, x_n$  notée  $L(x_1, \dots, x_n)$  est donnée par :

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

l'estimateur à noyau et la technique de validation croisée. de  $f(x_i)$  est donnée par :

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

Ainsi la vraisemblance est approximée par :

$$f(x_1, x_2, \dots, x_n | h) = \frac{1}{(n-1)^n} \prod_{i=1}^n \sum_{j=1, i \neq j}^n K_{x_i, h}(x_j). \quad (2.1)$$

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

## 2 Choisir la loi a priori sur le paramètre de lissage $h$

Rappelons que l'estimation par l'approche bayésienne est caractérisée par le choix d'une loi a priori et notons que le choix n'est pas unique ici. L'utilisation de la loi a priori conjuguée dans ce modèle n'est pas bénéfique, car la vraisemblance de validation croisée donnée par 2.1 s'écrit comme produit de  $n$  termes, et le paramètre de lissage  $h$  intervient dans chacun des termes.

Malgré que d'une manière générale, en pratique, l'estimateur bayésien n'est pas très sensible au choix de la loi a priori, mais dans notre cas, le choix doit être fait d'une manière minutieuse afin que les méthodes *MCMC* fonctionnent correctement.

Nous avons choisi d'utiliser la loi Gamma proposée par *Brewer [1998]* comme loi a priori pour  $h$ . La densité  $\pi(h)$  est alors donnée, a 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 (paramètres de la loi Gamma).

Dans cette partie, nous n'avons pas discuté le choix des hyper-paramètres parce que, le choix est fait empiriquement de sorte que les méthodes de *Monte Carlo* vont fonctionner correctement.

**3 Calculer la distribution a posteriori  $\pi(h|x_1, \dots, x_n)$  du paramètre  $h$  quand on dispose des données  $x_1, \dots, x_n$  en utilisant le théorème de Bayes.**

La loi a posteriori de  $h$  est définie par :

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

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

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

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

$$\pi(h | x_1, x_2, \dots, x_n) \propto \frac{\pi(h|\alpha, \beta)}{(n-1)^n \pi(x_1, x_2, \dots, x_n)} \prod_{i=1}^n \sum_{i=1, j \neq i}^n K_{x_i, h}(x_j)$$

où  $\pi(x_1, x_2, \dots, x_n) = \int f(x_1, \dots, x_n|h)\pi(h|\alpha, \beta)dh$  est la loi marginale de l'échantillon.

**4 MCMC pour estimer le paramètre de lissage**

En 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 a partir d'une variable aléatoire  $\varepsilon$  positive de densité  $q(h | h^{(k)}, \gamma_{(k)})$ , telle que  $h^{(k)}$  est le candidat généré a la  $k^{ème}$  étape,  $\gamma_{(k)}$  est un paramètre de réglage choisi de telle sorte a obtenir un taux d'acceptation optimal  $\tau$ . La variable aléatoire  $\varepsilon$  est généralement choisie comme gaussienne de moyenne  $h^{(k)}$  et de variance  $\gamma_{(k)}^2$ . Ici, nous la remplaçons par une même loi gaussienne tronquée sur

$[0, +\infty[$ , vu la positivité du paramètre de lissage  $h$ . C'est-à-dire  $\varepsilon \sim N(h^{(k)}, \gamma_{(k)}^2, 0, \infty)$ , la densité de l'instrumentale est alors donnée par :

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

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

Garthwaite et *al. brewer (1998)* donnent la formule de mise-à-jour du paramètre :

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

où  $\tau$  est le taux d'acceptation optimal, qui vaut 0.44 dans le cas univarié voir *Gelman et al [1996]*, *Roberts and Rosenthal [2009]* et *Garthwaite et al.[2010]*. Les étapes de l'algorithme *M-H* à marche aléatoire associé au présent problème est donné comme suite :

1. initialiser  $h^{(0)}$  et  $\gamma_{(0)}$
2. pour  $k \in \{1, \dots, N\}$ 
  - 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 :

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

3. Poser  $k = k + 1$  et aller à 2.

4. Calculer l'estimateur de *Bayes* :

$$\tilde{h} = \frac{1}{N - N_0} \sum_{k=N_0+1}^N h^{(k)}.$$

# Chapitre 3

## Simulation et résultats numériques

### 3.1 Application Numérique

#### 3.1.1 Introduction

Le but de cet chapitre est d'identifier l'impact de l'inférence bayésienne utilisé pour le choix de paramètre de lissage  $h_{Bayes}$  lors de l'estimation d'une densité de probabilité à support positif sur la qualité de l'estimateur. Nous comparant les performances des estimateurs associe à  $h_{Bayes}$  avec les performance de ceux associant aux où paramètre de lissage globale et cel sélectionné par la méthode de validation croisée non biaisé.

#### 3.1.2 Etude de Simulation

$h_{UCV}$  : Le paramètre de lissage optimal obtenu par la technique de validation croisée.non bias

$h_{Bayes}$  : est le  $h$  optimale sélectionné par l'inférence bayésienne.

$MSE$  : L'erreur quadratique moyenne.

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

1. Une *loi exponentielle* de paramètre  $\lambda = 5$  :

$$g_1(x) = \lambda \exp(-\lambda x), \quad x \geq 0.$$

2. Une *loi Gamma* de paramètre  $(1, 2)$  :

$$g_2(x) = \frac{\lambda}{\Gamma(p)} (\lambda x)^{p-1} \exp(-\lambda x)$$

3. Une *loi Béta* de paramètre  $(0.5, 1)$  :

$$g_3(x) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1}.$$

**Le noyau  $k$**  : on a opté pour le noyau epanechnikov.

**La loi a priori  $\pi(h)$**  : est une loi gamma.

**La loi instrumentale** : est une loi normale tronquée dont  $\gamma(0) = 1$  et  $h^{(0)} = 1$ .

-Le paramètre de lissage est sélectionné par le biais de deux approches :

- Bayésienne : en utilisant la méthode *MCMC* (Algorithme M-H).
- classique : en utilisant l'algorithme UCV

Le manuel suivant présente les différentes routines, sous le logiciel R, relatives du package Ake. Ce package est disponible sur le site "Comprehensive R Archive Network" (CRAN) via le lien

<http://cran.r-project.org/web/packages/Ake/Ake.pdf>.

	<i>UCV</i>			<i>Bays</i>		
<i>n</i>	$h_{UCV}$	$ Bias $	<i>MSE</i>	$h_{bays}$	$ Bias $	<i>MSE</i>
10	0.1524053	0.03011333	0.1023679	0.6465001	0.0524001	0.2114000
200	0.0135095	0.02636484	0.06662864	0.13116975	0.0287494	0.0942807
1000	0.00231365	0.02540452	0.06600027	0.0118232	0.0261852	0.0580025

TAB. 3.1 – résultats obtenus pour le cas g1

	<i>UCV</i>			<i>Bays</i>		
<i>n</i>	$h_{UCV}$	$ Bias $	<i>MSE</i>	$h_{bays}$	$ Bias $	<i>MSE</i>
10	0.2751791	0.6960137	0.6711991	0.0310445	0.0995259	0.2225669
200	0.0482594	0.5498152	0.3948659	0.0310230	0.0995952	0.2133247
1000	0.0006296	0.4394944	0.0110508	0.0310139	0.0103323	0.2089313

TAB. 3.2 – résultats obtenus pour le cas g2

	<i>UCV</i>			<i>Bays</i>		
<i>n</i>	$h_{UCV}$	$ Bias $	<i>MSE</i>	$h_{bays}$	$ Bias $	<i>MSE</i>
10	0.0319937	0.1260955	0.1718115	0.0299632	0.1917622	0.2643644
200	0.0049698	0.1100550	0.0304692	0.0199027	0.1265470	0.2002220
1000	0.0009914	0.0060955	0.2205212	0.0008996	0.0033330	0.0679331

TAB. 3.3 – résultats obtenus pour le cas g3

# Conclusion

Dans ce mémoire nous avons considéré le choix du paramètre de lissage par l'inférence bayésienne dans l'estimation à noyau d'une densité de probabilité.

En premier lieu, nous avons intéressé à l'estimation classique d'une densité par la méthode du noyau nous avons présenté ses propriétés et choix du noyau  $k$  de cet estimateur. Après on va présenter la techniques classique dans le choix du paramètre de lissage.

En deuxième lieu, nous avons présenté l'approche bayésienne qui est une alternative pour Traitement aux problèmes des techniques classiques de sélection du paramètre de lissage. En effet, après avoir abordé le paradigme bayésien nous avons présenté la démarche à suivre pour la mise en œuvre de cette approche pour le choix du paramètre de lissage globale dans l'estimation à noyau d'une densité de probabilité.

Enfin, nous avons présenté une application numérique, la comparaison de l'approche bayésienne globale (*MCMC*) avec la méthode classique (*UCV*) pour des données simulées montre que les performances de l'approche Bayésienne globale sont meilleures que celle de la méthode *UCV*.

# Bibliographie

- [1] **A.F.M. Smith, G.O. Robert (1993)**. *Bayesian computation via the Gibbs sampler and related Markov Chain Monte Carlo methods*. Journal of the Royal Statistical Society ; vol(55), 3-23, .
- [2] **A. w. Bowman, (1984)**. An alternative methode of cross-validation for the smoothing dendity estimates. *biometrika*, 71, 553-560.
- [3] **C. P. Robert and G, (2004)**. *Casella.monte carlo statistical methods*. Springer-Verlag
- [4] **E. Parzen, (1962)**. On estimation of a probability density function and mod. *Annals of Mathematical statistics*, 33(3), 1056-1076.
- [5] **E. Parent, J. Bernier, (2007)**, *Le raisonnement Bayésien : Modélisation et Inférence*. Springer-Verlag France, Paris. 364 pages.
- [6] **I. Ben Khalifa (2008)** Estimation Non-paramétrique par Noyaux Associés et Données de Panel en Marketing.Projet de Fin d'Etude. Université du 7 Novembre à Carthage.
- [7] **J. Albert, (2009)** *Bayesian computation with R*. Springer science+business media, LLC.
- [8] **J. Dupuis, (2007)** (LSP-UPS) . *Statistique bayésienne et algorithme MCMC*.
- [9] **J. Dupuis, (2007)** (LSP-UPS) . *Statistique bayésienne et algorithme MCMC*.
- [10] **J .J. Boreux, E. Parent, J. Bernier, (2010)**. *Pratique du calcul Bayésien*. Springer-Verlag France, Paris. 333 pages.

- [11] **M. Rosenblatt, (1956)**. Remarks in some nonparametric estimates of a density function. *Ann. Math. Statist.* 27 : 832–837.
- [12] **P. Robert, Christian, (2006)**. *Le Choix Bayésien : Principes et pratique*. Springer-Verlag France, Paris, vol (1), 638 pages.
- [13] **S. Djouder, (2007)** . Méthodes de Monte Carlo dans l'estimation non paramétrique de la densité de probabilité par noyaux associés Mémoire de Magister en Mathématiques Appliquées
- [14] **W. K. Hastings, (1970)**. Monte carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1), 97–109
- [15] **W.E., Wansouwé, S.M., Somé & C.C., Kokonendji(2015c)**. Ake : Associated kernel estimations, URL [http ://cran.r-project.org/web/packages/Ake/](http://cran.r-project.org/web/packages/Ake/).

# Annexe A : Abréviations et Notations

$K(u)$	noyau.
$\propto$	Proportionnel à.
$\theta x$	$\theta$ est conditionne par $x$ .
$\pi(x), \pi(\theta x)$	<i>A priori</i> , <i>.A posteriori</i> .
$f(\theta x)$	Loi de probabilité, indexée par le paramètre $\theta$ .
<i>MCMC</i>	Méthodes de Monte Carlo par chaîne de Markov.
<i>MSE</i>	L'erreur quadratique moyenne.
<i>ISE</i>	L'erreur quadratique intégrée.
<i>AMISE</i>	L'erreur quadratique moyenne intégrée asymptotique.
<i>MISE</i>	L'erreur quadratique moyenne intégrée.
<i>LCV</i>	validation croisée de vraisemblance.
<i>BCV</i>	validation croisée biais.
<i>UCV</i>	validation croisée non biais.
<i>MH</i>	Metropolis-Hastings.

**Résumé :** Dans ce travail, l'approche non-paramétrique par noyaux est présentée pour les fonctions de densités. Pour cela, quelques aspects essentiels des notions d'estimation par noyaux continus (dits classiques) sont d'abord rappelés.

Pour estimer la densité de probabilité des données positives par la méthode du noyau, le choix du noyau et du paramètre de lissage est important. On a utilisé les noyaux symétriques, La sélection du paramètre de lissage est basée sur l'approche classique (UCV) et l'approche Bayésienne globale. La complexité de la loi a posteriori, dans l'approche Bayésienne globale, nécessite l'utilisation des méthodes de Monte Carlo (MCMC).

La comparaison de l'approche Bayésienne globale (MCMC) avec la méthode classique (UCV) pour des données simulées montre que les performances de l'approche Bayésienne globale sont meilleures que celle de la méthode *UCV*.

**Mots clés :** Estimation non paramétrique de densité, noyaux, approche Bayésienne, méthodes de Monte Carlo (MCMC), validation croisée, plug-in, paramètre de lissage.

**Abstract :** This work is about nonparametric approach using kernels for densities. For this, some essential aspects of notions of continuous (classical) kernel estimation are first recalled.

To estimate the probability density of the positive data by the kernel method, the choice of kernel and the smoothing parameter is important. We used asymmetric kernels. The selection of the smoothing parameter is based on the classical approach (*UCV*) and the global Bayesian approach. The complexity of the posterior law, in the global Bayesian approach, requires the use of Monte Carlo methods (*MCMC*).

The comparison of the global Bayesian approach (*MCMC*) with the classical method (*UCV*) for simulated data shows that the performances of the global Bayesian approach are better than that of the *UCV* method.

**Key words :** Nonparametric density estimation, kernels, Bayesian approach, Monte Carlo methods (*MCMC*), cross validation, plug-in, smoothing parameter.