

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

Université A. MIRA de Béjaïa

Faculté des Sciences Exactes

Département de Recherche Opérationnelle



Mémoire de Magister en Mathématiques Appliquées

Option : Modélisation Mathématiques et Techniques de Décision

Thème

MÉTHODES DE MONTE CARLO DANS L'ESTIMATION NON
PARAMÉTRIQUE DE LA DENSITÉ DE PROBABILITÉ PAR
NOYAUX ASSOCIÉS

Présenté par Mr. Djouder Sofiane

Soutenu publiquement le 30/04/2014 devant le jury composé de :

Président	AISSANI DJAMIL	Professeur	U. de Béjaïa
Directeur	ADJABI SMAÏL	Professeur	U. de Béjaïa
Examineurs	MOHDEB ZAHER	Professeur	U. de Constantine
	BOUALEM NÉE BARECHE AÏCHA	M.C.A	U. de Béjaïa
	BOUALEM MOHAMED	M.C.A	U. de Béjaïa
Invité	ZOUGAB NABIL	M.C.B	U. de Tizi-Ouzou

Université de Béjaïa, Avril 2014.

Remerciements

Je tiens à remercier Monsieur ADJABI Smaïl, Professeur à l'Université de Béjaïa, pour la qualité de son encadrement et du soutien qu'il m'a accordé durant ma formation. Ses conseils avisés ont une part plus qu'importante dans la forme et le fond de ce document.

Mes remerciements les plus sincères vont également à Monsieur AISSANI Djamil, Professeur à l'Université de Béjaïa, qui me fait l'honneur de présider ce jury.

Je remercie également Monsieur MOHDEB Zaher, Monsieur BOUALEM Mohamed et Madame BOUALEM Aicha pour avoir accepté de juger ce travail et Monsieur ZOU-GAB Nabil de m'avoir soutenu dans des moments difficiles et aussi de m'avoir honoré en acceptant l'invitation.

Enfin, mes remerciements vont à toutes celles et ceux qui ont contribué de manière directe ou indirecte pour la réalisation de ce travail.

Table des matières

Liste des tableaux	iv
Liste des figures	v
Notations & Abréviations	vi
Introduction générale	1
1 Méthode du noyau associé	5
1.1 Introduction	5
1.2 Construction d'un noyau associé continu	7
1.3 Estimateur à noyau associé	9
1.3.1 Propriétés de l'estimateur à noyau associé continu	9
1.3.2 Réduction de biais	11
1.3.3 Exemples de noyaux associés asymétriques	12
1.3.4 Convergence des estimateurs à noyau associé	16
1.3.5 Normalisation par Monte Carlo	20
1.4 Choix du noyau asymétrique	21
1.5 Conclusion	22
2 Approche bayésienne et méthodes Monte Carlo	24
2.1 Introduction	24
2.2 Lois a priori	25
2.2.1 A priori conjugué	25
2.2.2 A priori non informatif	27

2.2.3	A priori impropre	29
2.3	Loi a posteriori	29
2.4	Marginalisation	30
2.5	Quand la vraisemblance fait le posterior	30
2.6	Inférence Bayésienne	31
2.6.1	Fonctions de coût	31
2.6.2	Estimateurs MMSE et MAP	34
2.6.3	Approches par approximations analytiques	35
2.6.4	Approche numérique	37
2.6.5	Méthodes de simulation Monte Carlo	39
2.6.5.1	Méthodes d'échantillonnage	41
2.6.5.2	Approximation de Monte Carlo	42
2.6.5.3	Échantillonnage préférentiel	43
2.6.5.4	Population Monte Carlo (PMC)	49
2.6.5.5	Monte Carlo par chaînes de Markov (MCMC)	50
2.7	Conclusion	52
3	Approches globales pour la sélection du paramètre de lissage	53
3.1	Introduction	53
3.2	Approche fréquentiste	53
3.2.1	Méthodes plug-in	54
3.2.2	Méthodes de validation croisée	55
3.2.3	Avantages et inconvénients des méthodes classiques	58
3.3	Approche Bayésienne globale	59
3.3.1	Approximation de la vraisemblance	60
3.3.2	Choix de la loi a priori	60
3.3.3	Approximation de la densité a posteriori	61
3.3.4	MCMC pour estimer le paramètre de lissage	62
3.3.5	PMC pour estimer le paramètre de lissage	64
3.4	Conclusion	66

4 Application Numérique	67
4.1 Introduction	67
4.2 Choix des densités cibles	67
4.3 Les méthodes utilisées	70
4.4 Plan de simulation	70
4.5 Résultats de simulation	71
4.6 Discussions des résultats	72
4.7 Application aux données réelles	76
4.8 Conclusion	78
Conclusion générale & Perspectives	102
Bibliographie	105

Liste des tableaux

2.1	Lois a priori conjuguées naturelles pour quelques familles exponentielles usuelles	27
4.1	Paramètres de lissage correspondants aux durées d'éruption	77
4.2	Résultats de simulation pour la sélection du paramètre de lissage par les algorithmes : UCV, MCMC et PMC, cas du modèle D1	80
4.3	Résultats de simulation pour la sélection du paramètre de lissage par les algorithmes : UCV, MCMC et PMC, cas du modèle D2	81
4.4	Résultats de simulation pour la sélection du paramètre de lissage par les algorithmes : UCV, MCMC et PMC, cas du modèle D3	82
4.5	Résultats de simulation pour la sélection du paramètre de lissage par les algorithmes : UCV, MCMC et PMC, cas du modèle D4	83
4.6	Résultats de simulation pour la sélection du paramètre de lissage par les algorithmes : UCV, MCMC et PMC, cas du modèle D5	84
4.7	Résultats de simulation pour la sélection du paramètre de lissage par les algorithmes : UCV, MCMC et PMC, cas du modèle D6	85
4.8	Résultats de simulation pour la sélection du paramètre de lissage par les algorithmes : UCV, MCMC et PMC, cas du modèle D7	86
4.9	Taux d'acceptation et indicateurs SIF moyens fournis par l'algorithme MCMC.	87

Table des figures

1.1	Noyaux Gamma, Gamma modifié, RIG et log-normal pour $h=0.25$	16
1.2	Densité ayant un pôle en $x=0$	17
1.3	Forme du noyau optimal asymétrique	22
4.1	Courbes des densités tests choisies D_i , $i=\overline{1,7}$	69
4.2	Courbes de la densité estimée de la durée d'éruption	77
4.3	Estimateurs à noyaux Gamma et Gamma modifié d'une densité (D1)	88
4.4	Estimateurs à noyaux RIG et log-Normal d'une densité (D1)	89
4.5	Estimateurs à noyaux Gamma et Gamma modifié d'une densité (D2)	90
4.6	Estimateurs à noyaux RIG et log-Normal d'une densité (D2)	91
4.7	Estimateurs à noyaux Gamma et Gamma modifié d'une densité (D3)	92
4.8	Estimateurs à noyaux RIG et log-Normal d'une densité (D3)	93
4.9	Estimateurs à noyaux Gamma et Gamma modifié d'une densité (D4)	94
4.10	Estimateurs à noyaux RIG et log-Normal d'une densité (D4)	95
4.11	Estimateurs à noyaux Gamma et Gamma modifié d'une densité (D5)	96
4.12	Estimateurs à noyaux RIG et log-Normal d'une densité (D5)	97
4.13	Estimateurs à noyaux Gamma et Gamma modifié d'une densité (D6)	98
4.14	Estimateurs à noyaux RIG et log-Normal d'une densité (D6)	99
4.15	Estimateurs à noyaux Gamma et Gamma modifié d'une densité (D7)	100
4.16	Estimateurs à noyaux RIG et log-Normal d'une densité (D7)	101

Notations & Abréviations

Abréviations

AMISE	: Asymptotic Mean Integrate Squared Error
BMSE	: Bath Mean Standard Error
i.i.d.	: indépendants et identiquement distribués
IS	: Importance sampling
ISE	: Integrate Squared Error
LGN	: Loi des Grands Nombres
MAP	: Maximum A Posteriori
MCMC	: Markov chain Monte Carlo
MISE	: Mean Integrated Squared Error
MLE	: Maximum Likelihood Estimator
MMSE	: Minimum Mean Squared Error
MSE	: Mean Squared Error
PMC	: Population Monte Carlo
SIF	: Simulation Inefficiency Factor
SIR	: Sampling/Importance Resampling
v.a.	: variable aléatoire

Notations

A^c	: événement complémentaire de A
$\xrightarrow{\mathbb{P}}$: convergence en probabilité
$\xrightarrow{p.s.}$: convergence presque sûre
$\xrightarrow{L^2}$: convergence en norme L^2
$\xrightarrow{\mathcal{L}}$: convergence en loi
$\mathbf{1}_A$: fonction indicatrice de l'ensemble A
Θ	: l'espace des paramètres
Ω	: l'espace des observations
$\mathcal{N}(\mu, \sigma^2)$: loi Normale de moyenne μ et de variance σ^2
$\mathcal{N}(\mu, \sigma^2, a, b)$: loi Normale de moyenne μ et de variance σ^2 tronquée sur $[a, b]$
$\mathcal{P}(\theta)$: loi de Poisson de paramètre θ
$\mathcal{G}(\alpha, \beta)$: loi Gamma de paramètres α et β
$\mathbf{B}(\alpha, \beta)$: loi Béta de paramètres α et β
$\mathcal{B}(n, \theta)$: loi Binomiale de paramètres n et θ
$\mathcal{B}neg(m, \theta)$: loi Binomiale négative de paramètres n et θ
$\mathbf{E}[\cdot]$: espérance mathématique de l'argument
$\mathbf{V}[\cdot]$: variance de l'argument
$I(\theta)$: information de Fisher sur le paramètre θ
$\pi(x_1, \dots, x_n)$: constante de normalisation (évidence Bayésienne)
θx	: θ est conditionné par x
\mathbb{R}	: ensemble des réels
\mathbb{N}	: ensemble des entiers naturels
dx	: différentielle par rapport à x
x^T	: transposée du vecteur x
$x \sim \pi$: x est simulé suivant π
$P(A)$: probabilité de l'événement A
\mathbb{S}_π	: support de la fonction π

Introduction générale

Dans un premier temps, il est essentiel de noter la dualité entre un échantillon et la distribution dont il est généré. De toute évidence, la densité génère l'échantillon, à l'inverse, étant donné un échantillon, nous pouvons, approximativement recréer la distribution, en particulier la densité.

La densité de probabilité est rarement connue en pratique, son estimation à partir d'un ensemble fini d'observations est un problème fondamental qui a fait l'objet d'une très vaste littérature. Une bonne introduction à cette problématique peut être trouvée dans les livres de [Silverman \[1986\]](#) et [Tsybakov \[2009\]](#).

Deux types d'approches d'estimation de la densité de probabilité sont utilisées et largement discutées dans la littérature : 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. On dit que souvent dans cette approche *les données parlent d'elles mêmes*.

Dans la littérature, plusieurs méthodes ont été dédiées à l'estimation de la densité de probabilité. Nous citons la méthode d'estimation par histogramme dont l'origine est attribué à John Grant au XVII^{ème} siècle. Les propriétés des estimateurs par histogramme ont été initialement étudiées par [Geffroy \[1974\]](#) et [Abou-Jaoude \[1976\]](#). Lorsque des connaissances a priori permettent d'affirmer que f est "lisse", pour l'estimer, plusieurs méthodes ont été proposées et utilisées au cours de ces dernières années. Elle comprennent la méthode du noyau due à [Rosenblatt \[1956\]](#) et [Parzen \[1962\]](#), la méthode d'estimation par les séries orthogonales de densités à supports bornés a été développée à partir des tra-

vaux de [Cencov \[1962\]](#), et étudiée ensuite par plusieurs auteurs dont [Kronmal and Tarter \[1968\]](#), [Wahba \[1981\]](#), [Bosq \[2005\]](#) et récemment [Saadi and Adjabi \[2009\]](#). Les méthodes d'interpolation par les fonctions splines [Wahba \[1975\]](#). Une autre méthode basée sur les développements en ondelettes a été proposée. Les estimateurs par ondelettes surpassent les estimateurs classiques dans la représentation des discontinuités et les oscillations locales. Les premiers résultats sur ces estimateurs sont donnés dans [Walter \[1992\]](#) et [Kerkyacharian and Picard \[1993\]](#).

L'estimation de la densité par la méthode du noyau, qui peut être vue comme une extension de la méthode d'estimation par histogramme, est la plus populaire parmi les multiples méthodes d'estimation non paramétriques de la densité. Cette popularité de l'estimateur à noyau peut s'expliquer par au moins trois raisons : la simplicité de sa forme, ses modes de convergence multiples et sa flexibilité qui s'interprète par la liberté de l'utilisateur dans le choix du noyau K et du paramètre de lissage h . L'estimateur à noyau a été proposé initialement par [Rosenblatt \[1956\]](#) et [Parzen \[1962\]](#) pour estimer des fonctions de densité à support non borné. Dans ce cadre, le choix du noyau K est très peu influant. Les noyaux employés pour ce type de densité (ou de support) sont symétriques (dit aussi classiques). Nous citons comme exemple, le noyau rectangulaire, triangulaire, parabolique ([Epanechnikov \[1969\]](#)), biweight et gaussien. Cependant, lorsqu'on veut estimer des densités à support borné au moins d'un côté, l'estimateur à noyau classique devient non consistant, à cause des effets du bord. Plusieurs solutions ont été proposées dans la littérature pour remédier à cette difficulté, citons *data reflection* de [Schuster \[1985\]](#), *boundary kernels* de [Müller \[1991, 1993\]](#) et *empirical transformation* de [Marron and Ruppert \[1994\]](#). La solution la plus simple, qui est encore d'actualité, est de remplacer le noyau symétrique par un noyau asymétrique, qui n'assigne pas un poids en dehors du support de la densité qu'on veut estimer. Cette idée est due à [Chen \[1999, 2000\]](#), puis d'autres extensions ont été apportées par [Jin and Kawczak \[2003\]](#), [Scaillet \[2004\]](#), [Senga Kiessé \[2008\]](#) et récemment, la notion du noyau K est généralisé en noyau associé $K_{x,h}$ par [Kokonendji and Libengué \[2011\]](#) et [Libengué \[2013\]](#).

En pratique, pour utiliser l'estimateur à noyau associé, il faut choisir le noyau associé

$K_{x,h}$ et le paramètre de lissage h . Pour le noyau, le choix peut être adapté par le support de la fonction inconnue à estimer, et pourtant, en situation non-asymptotique (faible nombre de données), le côté arbitraire du choix du noyau peut avoir une incidence importante sur la qualité de l'estimation. En revanche, le paramètre de lissage est un facteur important et crucial dans l'estimation de la fonction densité par la méthode des noyaux associés. De petites ou de grandes valeurs de h peuvent conduire à une estimation sous ou sur-lissée. Deux catégories de méthodes classiques ont été proposées dans la littérature pour choisir le paramètre de lissage h . 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 laisse guider seulement par les observations. L'inconvénient principal des deux catégories est qu'elles ont tendance à fournir des estimateurs sous ou sur-lissés lorsque les données sont de petite ou moyenne taille, ou encore lorsque on veut estimer des fonctions complexes.

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 de variables aléatoires à valeurs positives par la méthode des noyaux associés (précisément asymétriques), en particulier pour des échantillons de petite ou moyenne taille. Cette alternative Bayésienne est différente des méthodes classiques qui consistent souvent à minimiser ou à maximiser directement un certain critère. La modélisation Bayésienne est caractérisée par la considération du paramètre de lissage h comme une variable aléatoire positive, en lui associant une loi a priori $\pi(\cdot)$. 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 $\mathbf{E}(h|\text{données})$, en calculant la loi a posteriori $\pi(h|\text{données})$ à partir de la règle de Bayes. Les travaux traitant cette thématique sont relativement récents. En ce qui concerne le cas continu où le noyau classique gaussien est souvent utilisé, on peut citer les travaux de [Brewer \[1998\]](#) et [Zhang et al. \[2006\]](#). Récemment, [Zougab et al. \[2013b\]](#) ont généralisé l'approche bayésienne globale pour estimer le paramètre de lissage h en utilisant les noyaux associés dans le cas continu et discret.

Parmi les trois variantes de l'approche Bayésienne, nous proposons l'approche Bayé-

sienne globale pour le choix du paramètre de lissage h dans l'estimation de densité de probabilité à support positif (semi-borné), par la méthode des noyaux asymétriques. Cette proposition est l'extension des travaux de [Brewer \[1998\]](#), [Zhang et al. \[2006\]](#) et [Zougab et al. \[2013b\]](#) proposés pour choisir le paramètre de lissage h dans l'estimation de la densité de probabilité à support non borné, en utilisant, uniquement, le noyau classique gaussien. 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) comme dans [Brewer \[1998\]](#), [Zhang et al. \[2006\]](#) et [Zougab et al. \[2013b\]](#). Outre la méthode MCMC, nous proposons l'utilisation de l'échantillonnage préférentiel adaptatif dont la méthode Population Monte Carlo (PMC), qui est concurrente à la méthode MCMC.

Ce mémoire est composé d'une introduction générale, de deux parties englobant quatre chapitres, d'une conclusion générale et d'une liste de références bibliographiques.

La première partie englobe les deux premiers chapitres et traite les aspects théoriques du sujet. Dans le premier chapitre, nous avons présenté la notion du noyau associé, leur construction et l'estimateur à noyau associé de densité de probabilité en donnant ses propriétés de convergence. Le second chapitre est consacré aux rappels sur le formalisme Bayésien et les méthodes de Monte Carlo dont les méthodes MCMC et les méthodes d'échantillonnage préférentiel (PMC) qu'on utilisera pour simuler suivant la loi a posteriori lorsque celle-ci est complexe.

La deuxième partie est consacrée à l'étude empirique. Elle comporte les deux derniers chapitres. Le troisième chapitre traite le problème de sélection du paramètre de lissage et l'implémentation des deux approches : fréquentiste et Bayésienne globales. Enfin, dans le chapitre 4, nous présentons les résultats des simulations conduites à partir, d'une part, de plusieurs densités cibles connues et ayant différents aspects, d'autre part, des données réelles fréquemment utilisées dans la littérature.

Chapitre 1

Méthode du noyau associé

1.1 Introduction

En raison de leur simplicité, les praticiens ont utilisé les *noyaux classiques* (symétriques), introduits par [Rosenblatt \[1956\]](#) et [Parzen \[1962\]](#) pour l'estimation de la fonction densité de probabilité à support non borné. Citons le noyau gaussien, uniforme, biweight ainsi que le noyau optimal d'[Epanechnikov \[1969\]](#).

L'estimateur à noyau classique d'une densité de probabilité f , étant donné un n-échantillon X_1, \dots, X_n i.i.d. issu d'une v.a. X de fonction densité inconnue f , est de la forme

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right), \quad (1.1)$$

où K est la fonction noyau vérifiant généralement les conditions suivantes :

$$\int_{\mathcal{S}} K(u)du = 1, \int_{\mathcal{S}} uK(u)du = 0 \text{ et } \int_{\mathcal{S}} u^2K(u)du = \sigma_K^2 < \infty,$$

et \mathcal{S} le support de la fonction noyau K .

Dans la littérature, il a été signalé que l'estimateur à noyau classique souffre du problème de biais de bordure, une fois appliqué aux données issues des distributions à supports bornés. Dans ce cadre, beaucoup de travaux ont été réalisés afin de remédier au problème du biais de bordure en proposant d'autres estimateurs, généralement basés sur la modification des noyaux classiques. Citons *data reflection* de [Schuster \[1985\]](#), *boundary*

kernels de Müller [1991, 1993] et *empirical transformation* de Marron and Ruppert [1994].

Chen [1999] a proposé de remplacer les noyaux classiques par les noyaux Beta dans l'estimateur à noyau classique de densité inconnue f , dans le cas des données à support $[0, 1]$. En s'inspirant des noyaux Beta, Chen [2000] a étendu son idée pour l'estimation des densités des données à support semi-borné, en proposant le noyau Gamma.

Dans ce qui suit, nous allons d'abord rappeler la notion unifiée d'un noyau associé de cible x et de paramètre de lissage h . Notons que cette notion a été introduite par Kokonendji et al. [2007] et Senga Kiessé [2008] dans le cas discret, et étendue récemment par Kokonendji and Libengué [2011] et Libengué [2013] pour le cas continu. Par la suite, nous donnerons la définition d'un noyau associé, les techniques récentes utilisées pour leur construction et leurs propriétés. Puis, nous donnerons la définition d'un estimateur à noyau associé d'une fonction de densité inconnue f dans \mathbb{R}^+ , leurs propriétés et leurs modes de convergence, leur normalisation par la méthode Monte Carlo et des exemples récents d'estimateurs à noyaux associés.

Définitions

Les définitions suivantes présentent les notions du noyau associé et de l'estimateur à noyau associé pour la fonction de densité f inconnue sur le support \mathbb{T} .

Définition 1 Soit $x \in \mathbb{T}$ et $h > 0$ avec $\mathbb{T} \subseteq \mathbb{R}$ le support de la densité f à estimer. On appelle noyau associé $K_{x,h}$ toute densité paramétrique de probabilité de support $\mathbb{S}_{x,h}$, vérifiant les trois conditions suivantes :

$$x \in \mathbb{S}_{x,h}, \quad (1.2)$$

$$\mathbf{E}\{\mathcal{K}_{x,h}\} = x + A(x, h), \quad (1.3)$$

$$\mathbf{V}\{\mathcal{K}_{x,h}\} = B(x, h), \quad (1.4)$$

où $\mathcal{K}_{x,h}$ est une variable aléatoire de densité $K_{x,h}$ et les termes $A(x, h)$ et $B(x, h)$ convergent vers 0 quand $h \rightarrow 0$.

Définition 2 Soit X_1, \dots, X_n un n -échantillon aléatoire indépendant et identiquement distribué issu d'une variable aléatoire X de fonction de densité inconnue f sur $\mathbb{T} \subseteq \mathbb{R}$. L'estimateur à noyau associé \hat{f}_h de f utilisant $K_{x,h}$ est défini par

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^n K_{x,h}(X_i), \quad x \in \mathbb{T}, \quad (1.5)$$

où $K_{x,h}$ est le noyau associé continu de cible x et de fenêtre (paramètre de lissage) h sur $\mathbb{S}_{x,h}$. La cible x et le paramètre h interviennent dans la définition intrinsèque du noyau associé $K_{x,h}$.

1.2 Construction d'un noyau associé continu

Les noyaux associés continus englobent les noyaux associés symétriques (dit aussi classiques) et les noyaux associés asymétriques (dit aussi non classiques). Ces noyaux associés peuvent se construire à partir de n'importe quel *type de noyau* qui est une densité de probabilité à deux paramètres ayant un paramètre de dispersion.

En pratique, généralement, on utilise le noyau optimal d'Epanechnikov ou le noyau gaussien en raison de leur simplicité. Cependant, lorsqu'on cherche l'estimateur de la densité de probabilité d'une variable aléatoire bornée ou semi-bornée, l'inconvénient principal de ce type de noyau est qu'il assigne un poids à l'extérieur du support quand le lissage est pris en compte près du bord. Cela cause le problème du biais de bordure et donne un estimateur non consistant. La solution consiste alors à utiliser des noyaux asymétriques continus.

Dans la littérature, une multitude de noyaux associés continus asymétriques liés aux densités de probabilité Beta, Gamma, gaussienne-inverse et gaussienne-inverse-réciproque ont été introduits. En effet, les noyaux associés Beta et Gamma ont été proposés par [Chen \[1999, 2000\]](#), les noyaux BS¹ et log-normal par [Jin and Kawczak \[2003\]](#) et les noyaux gaussien-inverse et gaussien-inverse-réciproque par [Scaillet \[2004\]](#). Notons que [Chen \[1999, 2000\]](#), [Jin and Kawczak \[2003\]](#), [Scaillet \[2004\]](#) et [Hirukaway and Sakudo \[2013\]](#) n'ont jamais montré dans leurs études comment construisent-ils ces noyaux.

1. noyau de Birnbaum-Saunders

Ce sont [Kokonendji and Libengué \[2011\]](#) et [Libengué \[2013\]](#) qui ont proposé récemment une technique de construction des noyaux associés continus dite *mode-dispersion*.

Principe mode-dispersion

La méthode proposée par [Kokonendji and Libengué \[2011\]](#) est basée sur le principe **mode-dispersion** qui consiste à mettre la cible x sur le mode et le paramètre de lissage h sur le paramètre de dispersion. [Libengué \[2013\]](#) applique ce principe de construction de noyaux continus sur les densités Beta, Gamma, gaussien-inverse et gaussien inverse-réciproque ainsi qu'aux autres types de densités tel que Beta étendu, Gamma inverse et Weibull. Pour les premières densités, il retrouve les noyaux correspondants proposés dans la littérature et présentés dans la sous-section (1.3.3).

Dans la définition suivante, on introduit la notion de type de noyau qui fera l'objet de construction de tout noyau associé.

Définition 3 *Le type de noyau continu K est une densité de probabilité à carré intégrable $K = K_\theta$, dépendant du paramètre $\theta \in \Theta \subseteq \mathbb{R}^2$ et de support $\mathbb{S}_K = \mathbb{S}_\theta$.*

Dans ce qui suit, on considère le type de noyau K_θ (Définition 3) unimodal de mode M_θ et ayant un paramètre de dispersion D_θ . Étant donné Θ bidimensionnel, on retient la notation suivante : $\theta = \theta(a, b)$, $M_\theta = M(a, b)$ et $D_\theta = D(a, b)$ avec a et b deux réels positifs.

Ainsi, on est en mesure de construire certains noyaux associés en utilisant toute densité de probabilité satisfaisant les conditions de la Définition 3.

La méthode générale de construction est donnée comme suit

Principe 1 (mode-dispersion) *Soit $K_{\theta(a,b)}$ un type de noyau continu de mode $M(a, b)$ et de mesure de dispersion $D(a, b)$. On construit un noyau associé $K_{\theta(x,h)}$ à partir de $K_{\theta(a,b)}$ en résolvant le système :*

$$\begin{cases} M(a, b) = x, \\ D(a, b) = h. \end{cases} \quad (1.6)$$

Cela conduit à $\theta(x, h) = \theta(a(x, h), b(x, h))$, où $a(x, h)$ et $b(x, h)$ sont solutions de (1.6).

Exemple 1 (Noyau log-normal) En appliquant le principe mode-dispersion sur la densité log-normal $f(x|\mu, \sigma)$ de mode $M(\mu, \sigma) = e^{\mu-\sigma^2}$ et de paramètre de dispersion $D(\mu, \sigma) = \sigma$, on a

$$\begin{cases} e^{\mu-\sigma^2} = x, \\ \sigma = h, \end{cases} \Rightarrow \begin{cases} \mu = \log(x) + h^2, \\ \sigma = h, \end{cases} \quad (1.7)$$

ainsi, le noyau log-normal associé au principe mode-dispersion s'écrit sous la forme :

$$\begin{aligned} K_{\theta(x,h)}(u) &= K_{LN(\log(x)+h^2,h)}(u) \\ &= \frac{1}{uh\sqrt{2\pi}} \exp\left(\frac{-1}{2h^2} [\log(u) - \log(x) - h^2]^2\right). \end{aligned} \quad (1.8)$$

1.3 Estimateur à noyau associé

Soit X_1, \dots, X_n une séquence de variables aléatoires i.i.d., de densité inconnue f sur $\mathbb{T} \subseteq \mathbb{R}$.

L'estimateur à noyau associé \hat{f}_h de f est défini par

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^n K_{\theta(x,h)}(X_i), \quad \forall x \in \mathbb{T}. \quad (1.9)$$

L'estimateur défini dans (1.9) est celui construit par la méthode mode-dispersion.

1.3.1 Propriétés de l'estimateur à noyau associé continu

L'estimateur à noyau associé brut (1.9) doit, du moins approximativement avoir les mêmes propriétés qu'une fonction densité de probabilité.

Proposition 1 (Kokonendji and Libengué [2011]) Soit \hat{f}_h un estimateur à noyau associé (1.9) de f . Pour $x \in \mathbb{T}$ et $h > 0$, on a :

$$\mathbf{E}[\hat{f}_h(x)] = \mathbf{E}[f(\mathcal{K}_{\theta(x,h)})], \quad (1.10)$$

$$\hat{f}_h(x) \geq 0, \quad (1.11)$$

$$\int_{\mathbb{T}} \hat{f}_h(x) dx = \Lambda(n, h, K), \quad (1.12)$$

avec $\Lambda(n, h, K)$ est une quantité dépendant de l'échantillon, du type de noyau et du paramètre de lissage sachant qu'elle peut être différente de 1.

Proposition 2 (Kokonendji and Libengué [2011]) Soit \hat{f}_h un estimateur à noyau associé (1.9) de f . Pour $x \in \mathbb{T}$ et $h > 0$, on a :

$$\mathbf{Biais}[\hat{f}_h(x)] = A_{\theta(x,h)}f'(x) + \frac{1}{2}\{A_{\theta(x,h)}^2 + B_{\theta(x,h)}\}f''(x) + o(h^2). \quad (1.13)$$

en outre, si f est bornée en \mathbb{T} alors

$$\mathbf{V}[\hat{f}_h(x)] = \frac{1}{n}f(x)R(K) + o\left(\frac{1}{nh^r}\right). \quad (1.14)$$

où r est le plus grand nombre réel positif tel que : $R(K) = \int_{\mathbb{S}_{x,h} \cap \mathbb{T}} K_{\theta(x,h)}^2(u)du \leq C(x)h^{-r}$ et $0 < C(x) < \infty$.

Critères d'erreurs

L'évaluation de la similarité entre l'estimateur à noyau associé \hat{f}_h et la vraie densité f à estimer, nécessite des critères d'erreur. La mesure la plus naturelle utilisée est la moyenne intégrée des erreurs quadratiques. Ainsi, on définit d'abord la moyenne des erreurs quadratiques (MSE) par

$$\mathbf{MSE}[\hat{f}_h(x), f(x)] = \mathbf{V}[\hat{f}_h(x)] + \mathbf{Biais}^2[\hat{f}_h(x)]. \quad (1.15)$$

où le biais et la variance de l'estimateur sont définis dans (1.13) et (1.14).

La forme intégrale du MSE en \mathbb{T} et son approximation sont données respectivement par

$$\begin{aligned} \mathbf{MISE}[\hat{f}_h, f] &= \int_{\mathbb{T}} \left\{ \mathbf{V}[\hat{f}_h(x)] + \mathbf{Biais}^2[\hat{f}_h(x)] \right\} dx \\ &= \int_{\mathbb{T}} \left(\left[A_{\theta(x,h)}f'(x) + \frac{1}{2}\{A_{\theta(x,h)}^2 + B_{\theta(x,h)}\}f''(x) \right]^2 + \frac{1}{n}R(K)f(x) \right) dx \\ &\quad + o\left(h^4 + \frac{1}{nh^r}\right) \end{aligned} \quad (1.16)$$

et

$$\mathbf{AMISE}[\hat{f}_h, f] = \int_{\mathbb{T}} \left(\left[A_{\theta(x,h)}f'(x) + \frac{1}{2}\{A_{\theta(x,h)}^2 + B_{\theta(x,h)}\}f''(x) \right]^2 + \frac{1}{n}R(K)f(x) \right) dx. \quad (1.17)$$

La proposition suivante donne le taux de convergence au sens du *AMISE*

Proposition 3 (Libengué [2013]) *Supposons que $f \in \mathcal{C}^2(\mathbb{T})$, de dérivées première et seconde bornées. Alors, le paramètre de lissage optimal minimisant le AMISE est*

$$h_{AMISE}^* = C(x)n^{-\frac{1}{r+2}}$$

avec $r = r(K_{\theta(x,h)})$ et $C(x)$ constante dépendant de x .

Notons qu'il est plus pratique d'utiliser l'intégrale des erreurs quadratiques (ISE) pour l'évaluation du MISE. Le ISE est défini par

$$\text{ISE}[\hat{f}_h, f] = \int_{\mathbb{T}} [\hat{f}_h(x) - f(x)]^2 dx. \quad (1.18)$$

1.3.2 Réduction de biais

La présence du terme non nul $A_{\theta(x,h)}$ dans (1.13) augmente le biais de $\hat{f}_h(x)$. Ainsi, Libengué [2013] s'inspirent des travaux de Chen [1999, 2000] et proposent un algorithme général de réduction de biais de l'estimateur à noyau associé brut. En effet, à travers cet algorithme on peut définir des versions modifiées de tout estimateur à noyau associé préalablement défini. Autant que le principe *mode-dispersion*, la technique de *réduction de biais* peut être considérée comme méthode de construction d'un noyau associé Chen [1999, 2000], Libengué [2013] et Malec and Schienle [2014].

L'algorithme est de deux étapes. La première consiste à diviser le support en deux régions, le bord et l'intérieur et la seconde dans la modification du noyau associé.

Dans ce qui suit, on retiendra l'adaptation de l'algorithme pour les noyaux à supports $\mathbb{T} = [0, \infty[$. Ces noyaux feront l'objet principal de notre présente étude.

Étape I : diviser le support $\mathbb{T} = [0, \infty[$ en deux régions d'ordre $\alpha(h) > 0$ avec $\alpha(h) \rightarrow 0$ quand $h \rightarrow 0$,

(i) L'intérieur (la région la plus large contenant au moins 95% des observations) notée $\mathbb{T}_{\alpha(h),+1}$ et définie par

$$\mathbb{T}_{\alpha(h),+1} =]\alpha(h), \infty[, \quad (1.19)$$

(ii) Le bord (complémentaire de $\mathbb{T}_{\alpha(h),+1}$, peut être vide) noté par l'intervalle $\mathbb{T}_{\alpha(h),-1}$

$$\mathbb{T}_{\alpha(h),-1} = [0, \alpha(h)], \quad (1.20)$$

Étape II : modifier le noyau associé $K_{\theta(x,h)}$ par une autre fonction noyau notée $K_{\tilde{\theta}(x,h)}$ tel que $\forall h$ fixé

$$\tilde{\theta}(x, h) = \begin{cases} \tilde{\theta}_{-1}(x, h) & \text{si } x \in \mathbb{T}_{\alpha(h), -1}, \\ \tilde{\theta}_{+1}(x, h) & \text{si } x \in \mathbb{T}_{\alpha(h), +1}, \end{cases} \quad (1.21)$$

doit être continue sur \mathbb{T} .

Proposition 4 (Libengué [2013]) *La fonction $K_{\tilde{\theta}(x,h)}$ obtenue dans (1.21) est un noyau associé (asymétrique).*

1.3.3 Exemples de noyaux associés asymétriques

Noyau gamma

Le noyau associé Gamma a été introduit par Chen [2000] pour l'estimation des densités à support $\mathbb{T} = [0, \infty[$. Il a utilisé la loi Gamma pour construire des noyaux associés continus asymétriques. Deux classes de noyaux ont été proposées. Le noyau de première classe est alors de la forme :

$$K_{Gam(\frac{x}{h}+1, h)}(u) = \frac{u^{\frac{x}{h}} e^{-\frac{u}{h}}}{b^{\frac{x}{h}+1} \Gamma(\frac{x}{h} + 1)}, \quad (1.22)$$

où $\Gamma(\alpha) = \int_0^\infty \exp(-t)t^{\alpha-1}dt$, $\alpha > 0$ est la fonction Gamma et h est le paramètre de lissage satisfaisant les conditions $h \rightarrow 0$ et $nh \rightarrow \infty$ quand $n \rightarrow \infty$. Le premier estimateur à noyau Gamma est donné par

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^n K_{Gam(\frac{x}{h}+1, h)}(X_i). \quad (1.23)$$

Le paramètre de lissage optimal minimisant le *MISE* induit par l'estimateur (1.23) est

$$h_1^* = \frac{1}{4^{\frac{2}{5}}} \left[\frac{1}{2\sqrt{\pi}} \int_0^\infty x^{-\frac{1}{2}} f(x) dx \right]^{\frac{2}{5}} \left[\int_0^\infty \left\{ x f'(x) + \frac{1}{2} x f''(x) \right\}^2 dx \right]^{-\frac{2}{5}} n^{-\frac{2}{5}}. \quad (1.24)$$

Le *MISE* optimal est alors :

$$MISE_1^* = \frac{5}{4^{\frac{4}{5}}} \left[\frac{1}{2\sqrt{\pi}} \int_0^\infty x^{-\frac{1}{2}} f(x) dx \right]^{\frac{4}{5}} \left[\int_0^\infty \left\{ x f'(x) + \frac{1}{2} x f''(x) \right\}^2 dx \right]^{\frac{1}{5}} n^{-\frac{4}{5}}. \quad (1.25)$$

La deuxième classe a été proposée par le même auteur pour améliorer les performances (réduire le biais) de $\hat{f}_h(x)$. La forme du noyau Gamma modifié est donnée par :

$$K_{Gam^m(\rho_h(x), h)}(u) = \frac{u^{\frac{x}{h}} e^{-\frac{u}{h}}}{h^{\rho_h(x)} \Gamma(\rho_h(x))}, \quad (1.26)$$

où

$$\rho_h(x) = \begin{cases} \frac{1}{4}\left(\frac{x}{h}\right)^2 + 1 & \text{si } 0 \leq x < 2h, \\ \frac{x}{h} & \text{si } x \geq 2h. \end{cases} \quad (1.27)$$

L'estimateur à noyau Gamma modifié s'écrit :

$$\tilde{f}_h(x) = \frac{1}{n} \sum_{i=1}^n K_{\text{Gam}^m(\rho_h(x), h)}(X_i). \quad (1.28)$$

Le paramètre de lissage optimal pour (1.28) est donné par

$$h_2^* = \left[\frac{1}{2\sqrt{\pi}} \int_0^\infty x^{-\frac{1}{2}} f(x) dx \right]^{\frac{2}{5}} \left[\int_0^\infty \{x f''(x)\}^2 dx \right]^{-\frac{2}{5}} n^{-\frac{2}{5}}. \quad (1.29)$$

Le MISE optimal est

$$\text{MISE}_2^* = \frac{5}{4^{\frac{4}{5}}} \left[\frac{1}{2\sqrt{\pi}} \int_0^\infty x^{-\frac{1}{2}} f(x) dx \right]^{\frac{4}{5}} \left[\int_0^\infty \{x f''(x)\}^2 dx \right]^{\frac{1}{5}} n^{-\frac{4}{5}}. \quad (1.30)$$

Deux autres classes de noyaux Gamma, dites Gamma modifiées raffinées, ont été proposés par [Malec and Schienle \[2014\]](#) afin de remédier au problème des pôles au bord, là où les performances du noyau Gamma modifié (1.28) sont médiocres.

Les deux versions de Gamma modifié raffiné se distinguent de Gamma modifié (1.28) uniquement par leur paramétrisation $\rho_h(x)$. La première version de Gamma modifié raffiné est paramétrée par

$$\rho_h^{vI}(x) = \begin{cases} \left[\frac{1}{4}\left(\frac{x}{hc}\right)^2 + 1 \right] [c + 2h(1 - c)] & \text{si } 0 \leq x < 2hc, \\ \frac{x}{hc}(c + 2h - x) & \text{si } 2hc \leq x < 2h, \\ \frac{x}{h} & \text{si } x \geq 2h, \end{cases} \quad (1.31)$$

la seconde version par

$$\rho_h^{vII}(x) = \begin{cases} \frac{1}{4}\left(\frac{x}{hc}\right)^2 + 1 & \text{si } 0 \leq x < 2hc, \\ \frac{x}{hc} & \text{si } x \geq 2hc, \end{cases} \quad (1.32)$$

où $c \in]0, 1]$. Si $c = 1$, on retrouve la paramétrisation originelle (1.27) de [Chen \[2000\]](#) dans les deux cas.

Récemment, [Hirukaway and Sakudo \[2013\]](#) ont unifié l'approche des noyaux asymétriques en famille de noyaux Gamma généralisée.

La forme du noyau Gamma généralisé est donnée par :

$$K_{GG(\alpha, \beta\Gamma(\frac{\alpha}{\gamma})/\Gamma(\frac{\alpha+1}{\gamma}), \gamma)}(u; x, h) = \frac{\gamma u^{\alpha-1} \exp \left[- \left\{ \frac{u}{\beta\Gamma(\frac{\alpha}{\gamma})/\Gamma(\frac{\alpha+1}{\gamma})} \right\}^{\gamma} \right]}{\left\{ \beta\Gamma(\frac{\alpha}{\gamma})/\Gamma(\frac{\alpha+1}{\gamma}) \right\}^{\alpha} \Gamma(\frac{\alpha}{\gamma})}. \quad (1.33)$$

Afin que cette densité soit une famille de noyaux Gamma généralisée, elle doit satisfaire cinq conditions données dans [Hirukaway and Sakudo \[2013\]](#). Certaines conditions ont été inspirées par la construction du noyau Gamma modifié de [Chen \[2000\]](#) et d'autres pour la validation des approximations des termes de biais et de variance de l'estimateur, donné par

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^n K_{GG}(X_i; x, h), \quad (1.34)$$

où K_{GG} est le noyau Gamma généralisé, donné dans (1.33).

Noyaux gaussien-inverse et gaussien-inverse-réciproque

[Scaillet \[2004\]](#) a introduit les noyaux associés gaussien-inverse et gaussien-inverse-réciproque définis sur $]0, \infty[$. Les noyaux gaussien-inverse et gaussien-inverse-réciproque sont donnés respectivement comme suit

$$K_{IG(x, \frac{1}{h})}(u) = \frac{1}{\sqrt{2\pi h u^3}} \exp \left(-\frac{1}{2hx} \left[\frac{u}{x} - 2 + \frac{x}{u} \right] \right) \quad (1.35)$$

et

$$K_{RIG(\frac{1}{x-h}, \frac{1}{h})}(u) = \frac{1}{\sqrt{2\pi h u}} \exp \left(-\frac{x-h}{2h} \left[\frac{u}{x-h} - 2 + \frac{x-h}{u} \right] \right), \quad (1.36)$$

où h est le paramètre de lissage satisfaisant la condition $h + \frac{1}{nh} \rightarrow 0$ quand $n \rightarrow \infty$. Les estimateurs de la densité de probabilité inconnue f en utilisant les noyaux gaussien-inverse et gaussien-inverse-réciproque sont donnés respectivement par

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^n K_{IG(x, \frac{1}{h})}(X_i) \quad (1.37)$$

et

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^n K_{RIG(\frac{1}{x-h}, \frac{1}{h})}(X_i). \quad (1.38)$$

Les paramètres de lissage optimaux au sens du *MISE* sont

$$h_{IG}^* = \left[\frac{1}{2\sqrt{\pi}} \int_0^{\infty} x^{-\frac{3}{2}} f(x) dx \right]^{\frac{2}{5}} \left[\int_0^{\infty} \{x^3 f''(x)\}^2 dx \right]^{-\frac{2}{5}} n^{-\frac{2}{5}} \quad (1.39)$$

et

$$h_{RIG}^* = \left[\frac{1}{2\sqrt{\pi}} \int_0^\infty x^{-\frac{1}{2}} f(x) dx \right]^{\frac{2}{5}} \left[\int_0^\infty \{x f''(x)\}^2 dx \right]^{-\frac{2}{5}} n^{-\frac{2}{5}}, \quad (1.40)$$

induisant les *MISE* optimaux suivants

$$MISE_{IG}^* = \frac{5}{4} \left[\frac{1}{2\sqrt{\pi}} \int_0^\infty x^{-\frac{3}{2}} f(x) dx \right]^{\frac{4}{5}} \left[\int_0^\infty \{x^3 f''(x)\}^2 dx \right]^{1/5} n^{-\frac{4}{5}} \quad (1.41)$$

et

$$MISE_{RIG}^* = \frac{5}{4} \left[\frac{1}{2\sqrt{\pi}} \int_0^\infty x^{-\frac{1}{2}} f(x) dx \right]^{\frac{4}{5}} \left[\int_0^\infty \{x f''(x)\}^2 dx \right]^{\frac{1}{5}} n^{-\frac{4}{5}}. \quad (1.42)$$

Noyaux Log-normaux

[Jin and Kawczak \[2003\]](#) ont été les premiers à avoir introduit le noyau log-normal ainsi que le noyau Birnbaum-Saunders (BS) qui n'est qu'un mélange, avec poids égaux des noyaux $IG(x, \frac{x}{h})$ et $RIG(x, \frac{h}{x})$. Comme [Chen \[2000\]](#), les auteurs ne montrent pas la manière avec laquelle ils construisent leurs noyaux. Ici, on se limitera au noyau log-normal dont la forme est donnée par

$$K_{LN(\log(x), 4\log(1+h))}(u) = \frac{1}{u\sqrt{8\pi\log(1+h)}} \exp\left(-\frac{[\log(u) - \log(x)]^2}{8\log(1+h)}\right), \quad (1.43)$$

le paramètre de lissage h satisfait les conditions $h \rightarrow 0$ et $nh \rightarrow \infty$ quand $n \rightarrow \infty$.

L'estimateur à noyau LN correspondant, de la densité inconnue f est

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^n K_{LN(\log(x), 4\log(1+h))}(X_i). \quad (1.44)$$

En utilisant la technique mode-dispersion, [Libengué \[2013\]](#) construisent un noyau log-normal différent de celui proposé par [Jin and Kawczak \[2003\]](#). Sa forme est

$$K_{LN(\log(x)+h^2, h)}(u) = \frac{1}{uh\sqrt{2\pi}} \exp\left(-\frac{1}{2h^2} [\log(u) - \log(x) - h^2]^2\right). \quad (1.45)$$

En utilisant ce noyau, la densité inconnue f est estimée par

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^n K_{LN(\log(x)+h^2, h)}(X_i). \quad (1.46)$$

[Libengué \[2013\]](#) proposent une modification de leur noyau afin de l'adapter aux problèmes d'estimation au bord. Ils ont utilisé $\alpha(h)$, une fonction de h , pour séparer le support de f en deux régions précisées dans (1.48).

$$K_{LN(\log(\theta_h(x)), h)}(u) = \frac{1}{uh\sqrt{2\pi}} \exp\left(-\frac{1}{2h^2} [\log(u) - \log(\theta_h(x))]^2\right), \quad (1.47)$$

où

$$\theta_h(x) = \begin{cases} \frac{\alpha^3(h)}{x^2} \exp\left(-\frac{3h^2}{2}\right) & \text{si } 0 \leq x \leq \alpha(h), \\ x \exp\left(-\frac{3h^2}{2}\right) & \text{si } x > \alpha(h). \end{cases} \quad (1.48)$$

En posant $\alpha(h) = 2h$, on obtient la modification adéquate (au sens du séparateur de régions) de celle de [Chen \[2000\]](#) donnée dans (1.27).

La figure (1.1) présente les formes de certains noyaux pour différentes cibles x et pour un paramètre de lissage h fixé.

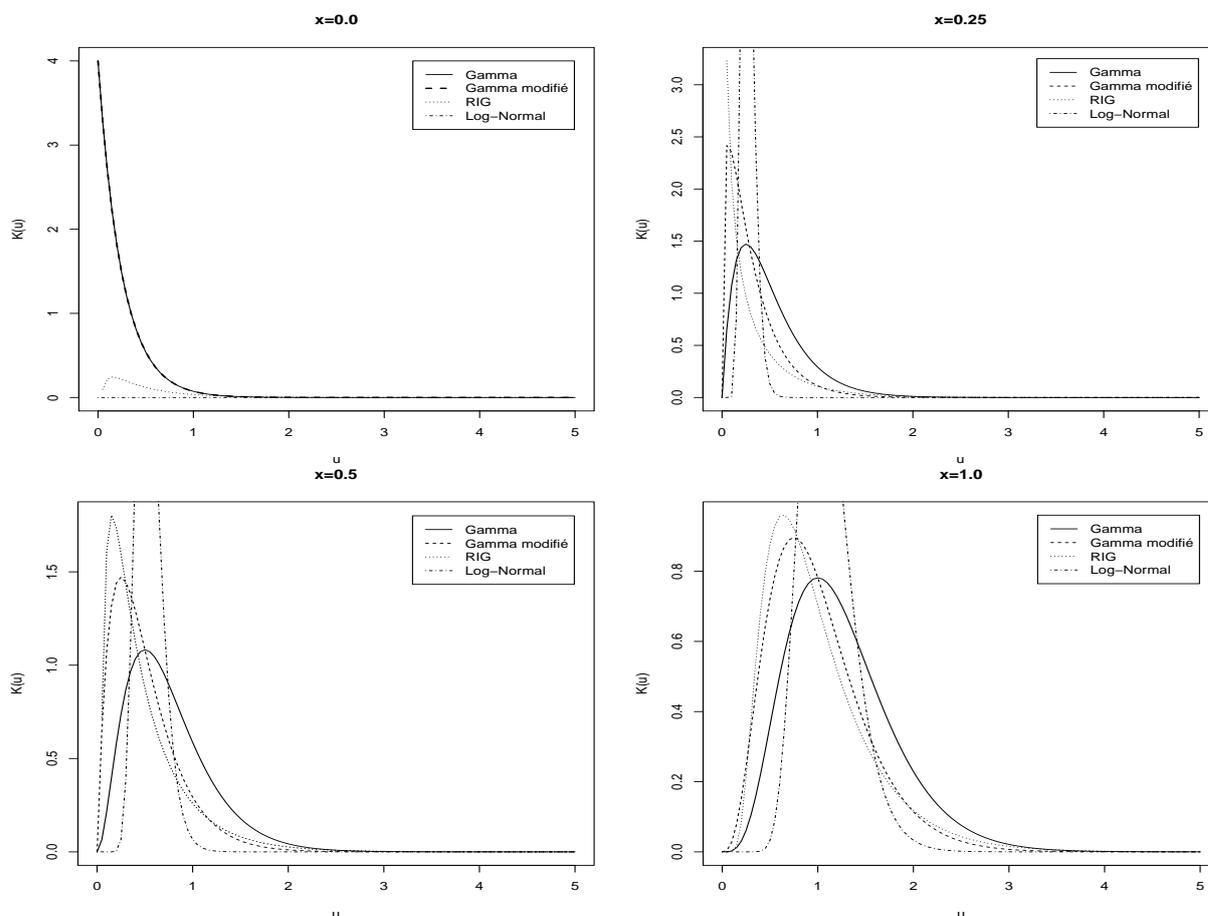


FIGURE 1.1: Noyaux Gamma, Gamma modifié, RIG et log-normal pour $h=0.25$

1.3.4 Convergence des estimateurs à noyau associé

Nous présentons d'abord les résultats ponctuels puis les consistances globales.

Résultats ponctuels

Une abondante littérature a été particulièrement dédiée aux problèmes d'estimation au bord du support par les estimateurs à noyaux asymétriques. [Bouezmarni and Scaillet \[2005\]](#) traitent, dans le résultat suivant, le comportement de l'estimateur à noyau asymétrique au point $x = 0$ dans le cas des densités ayant un pôle en $x = 0$ comme dans la Figure (1.2).

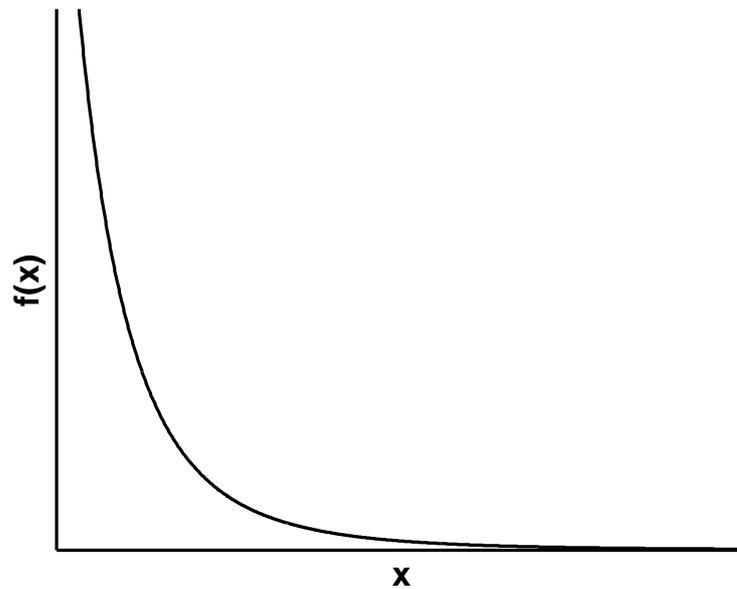


FIGURE 1.2: Densité ayant un pôle en $x=0$

Théorème 1 ([Bouezmarni and Scaillet \[2005\]](#)) Soit f une densité de probabilité de support $[0, +\infty[$, non bornée en $x = 0$ et \hat{f}_h son estimateur à noyau associé asymétrique.

si $\lim_{n \rightarrow +\infty} h = 0$ et $\lim_{n \rightarrow +\infty} nh^{2a} = +\infty$ ($a > 0$) alors

$$\hat{f}_h(0) \xrightarrow{\mathbb{P}} +\infty, \quad \text{quand } n \rightarrow +\infty,$$

si

$$\text{pour tout } \delta > 0, \quad \int_0^\delta K_{\theta(0,h)}(u) du \longrightarrow 1, \quad \text{quand } h \rightarrow 0,$$

où " $\xrightarrow{\mathbb{P}}$ " désigne la "convergence en probabilité".

La condition supplémentaire du précédent théorème est vérifiée pour l'estimateur à noyau Gamma. En effet, pour $\delta > 0$ et $x = 0$ on a

$$\int_0^{\delta} K_{Gam(1,h)}(u) du = 1 - \exp\left(-\frac{\delta}{h}\right) \rightarrow 1, \quad \text{quand } h \rightarrow 0.$$

Par conséquent, l'estimateur à noyau Gamma attribue d'importants poids aux points du bord pour de petits paramètres de lissage. Ceci est dû à la propriété particulière du noyau Gamma en $x = 0$.

Ainsi, [Bouezmarni and Scaillet \[2005\]](#) déconseillent d'utiliser d'autres noyaux asymétriques ne vérifiant pas la dernière propriété pour l'estimation des densités ayant un pôle en $x = 0$, tel que les noyaux gaussien-inverse et gaussien-inverse-réciproque. De même, [Malec and Schienle \[2014\]](#) ont conseillé l'utilisation du noyau Gamma pour l'estimation de ce type de densité.

Le théorème suivant traite les conditions ponctuelles des consistances (faible et forte) de l'estimateur à noyau associé.

Théorème 2 ([Kokonendji et al. \[2012\]](#)) *Soit $f \in \mathcal{C}^2(\mathbb{T})$ une densité de probabilité et \hat{f}_h son estimateur à noyau associé pour un type de noyau K donné. Pour tout $x \in \mathbb{T}$ fixé et $h = h(n)$, on suppose qu'il existe un réel positif $r = r(K, x)$ tel que $nh^r \rightarrow +\infty$. Alors*

$$\hat{f}_h(x) \xrightarrow{\mathbb{L}^2 \ \& \ p.s.} f(x), \quad \text{quand } n \rightarrow +\infty,$$

où " $\xrightarrow{\mathbb{L}^2 \ \& \ p.s.}$ " désigne les deux convergences "en probabilité dans \mathbb{L}^2 (par rapport à la mesure de Lebesgue) et presque sûre"

Puisque la convergence en \mathbb{L}^2 implique la convergence en probabilité, la faible consistance de \hat{f}_h en découle.

La normalité asymptotique est assuré par le théorème ci-dessous

Théorème 3 ([Kokonendji et al. \[2012\]](#)) *Soit f une densité de probabilité de support \mathbb{T} et \hat{f}_h son estimateur à noyau associé pour un type de noyau K donné. Pour tout $x \in \mathbb{T}$*

fixé tel que $f(x) > 0$, on a

$$\frac{\hat{f}_h(x) - \mathbf{E} [\hat{f}_h(x)]}{\sqrt{\mathbf{V} [\hat{f}_h(x)]}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1), \quad \text{quand } n \rightarrow +\infty,$$

où " $\xrightarrow{\mathcal{L}}$ " désigne la "convergence en loi".

Résultats globaux

Ces résultats ont été initialement étudiés par Bouezmarni and Scaillet [2005] pour le cas des noyaux asymétriques puis généralisé par Kokonendji et al. [2012] pour les noyaux associés.

La proposition suivante nous assure la convergence uniforme du biais de l'estimateur à noyau associé.

Proposition 5 (Kokonendji et al. [2012] - Convergence uniforme du biais) *Soit f une densité de probabilité bornée et continue sur son support \mathbb{T} , et \hat{f}_h son estimateur à noyau associé. Pour tout compact $I \subseteq \mathbb{T}$, on a*

$$\sup_{x \in I} \left| \mathbf{E} [\hat{f}_h(x)] - f(x) \right| \rightarrow 0, \quad \text{quand } n \rightarrow +\infty.$$

Théorème 4 (Kokonendji et al. [2012] - Convergence uniforme faible et forte)

Soit f une densité de probabilité bornée et continue sur son support \mathbb{T} , et \hat{f}_h son estimateur à noyau associé pour un type de noyau K donné. On suppose qu'il existe un réel $r_0 = r_0(K)$ tel que, pour tout $x \in \mathbb{T}$, $\int_{\mathbb{S}_{x,h}} K_{x,h}(u) du \leq c_0(x)h^{-r_0}$ avec $c_0(x)$ positif et borné sur tout compact I contenant x . Pour tout compact $I \subseteq \mathbb{T}$, on a

- si $\lim_{n \rightarrow +\infty} nh^{2r_0} = +\infty$, alors $\sup_{x \in I} \left| \hat{f}_h(x) - f(x) \right| \xrightarrow{\mathbb{P}} 0$, quand $n \rightarrow +\infty$.
- si $\lim_{n \rightarrow +\infty} \frac{nh^{2r_0}}{\log(n)} = +\infty$, alors $\sup_{x \in I} \left| \hat{f}_h(x) - f(x) \right| \xrightarrow{p.s.} 0$, quand $n \rightarrow +\infty$.

Bouezmarni and Scaillet [2005] précisent la valeur de r_0 pour que la convergence uniforme faible ou forte ait lieu. En effet, $r_0 = 1$ pour le noyau Gamma et $r_0 = \frac{5}{2}$ pour les noyaux gaussien-inverse et gaussien-inverse-réciproque.

Proposition 6 (Kokonendji et al. [2012] - Convergence en \mathbb{L}^1 du biais) Soit f une densité de probabilité de support \mathbb{T} et \hat{f}_h son estimateur à noyau associé. Alors

$$\int_{\mathbb{T}} \left| \mathbf{E} \left[\hat{f}_h(x) \right] - f(x) \right| dx \rightarrow 0, \quad \text{quand } n \rightarrow +\infty.$$

Théorème 5 (Kokonendji et al. [2012] - Convergence en \mathbb{L}^1 faible et forte) Soit f une densité de probabilité bornée et continue sur son support \mathbb{T} , et \hat{f}_h son estimateur à noyau associé pour un type de noyau K donné. On suppose qu'il existe un réel $r_1 = r_1(K)$ tel que, pour tout $x \in \mathbb{T}$, $\int_{\mathbb{S}_{x,h}} K_{x,h}(u) du \leq c_1(x) h^{-r_1}$ avec $c_1(x)$ positif et borné sur tout compact I contenant x . Pour tout compact $I \subseteq \mathbb{T}$, on a

- si $\lim_{n \rightarrow +\infty} nh^{2r_1} = +\infty$, alors $\int_{\mathbb{T}} \left| \hat{f}_h(x) - f(x) \right| dx \xrightarrow{\mathbb{P}} 0$, quand $n \rightarrow +\infty$.
- si $\lim_{n \rightarrow +\infty} \frac{nh^{2r_1}}{\log(n)} = +\infty$, alors $\int_{\mathbb{T}} \left| \hat{f}_h(x) - f(x) \right| dx \xrightarrow{p.s.} 0$, quand $n \rightarrow +\infty$.

1.3.5 Normalisation par Monte Carlo

Comme il est signalé dans (1.12), l'estimateur (1.9) est défini à une constante près. Pour cela, nous généralisons la normalisation traitée par Gouriéroux and Monfort [2006] dans le cas du noyau Beta pour la normalisation de l'estimateur à noyau associé. De plus, nous proposons une normalisation par les techniques Monte Carlo qui seront présentées dans le chapitre suivant.

Il existe deux manières pour corriger le manque de normalisation de l'estimateur à noyau associé standard donné dans (1.9) :

- Dans la *macro-approche*, la correction est effectuée globalement. L'estimateur à noyau associé \hat{f}_h est remplacé par

$$\hat{f}_h^1(x) = \frac{\hat{f}_h(x)}{\int_{\mathbb{S}_{x,h}} \hat{f}_h(x) dx}. \quad (1.49)$$

- Dans la *micro-approche*, la correction est effectuée au niveau du noyau associé. L'estimateur normalisé est défini par

$$\hat{f}_h^2(x) = \frac{1}{n} \sum_{i=1}^n \frac{K_{\theta(x,h)}(X_i)}{\int_{\mathbb{S}_{x,h}} K_{\theta(x,h)}(X_i) dx}. \quad (1.50)$$

Pour les noyaux asymétriques de support $\mathbb{S}_{x,h} = [0, +\infty[$, les approximations respectives des estimateurs à noyau associé normalisés par la méthode Monte Carlo standard sont données respectivement par :

$$\hat{f}_h^1(x) \approx \frac{\sum_{i=1}^n K_{\theta(x,h)}(X_i)}{\frac{1}{m} \sum_{j=1}^m \frac{1}{y_j^2} \sum_{i=1}^n K_{\theta(\frac{1-y_j}{y_j},h)}(X_i)}. \quad (1.51)$$

$$\hat{f}_h^2(x) \approx \frac{1}{n} \sum_{i=1}^n \frac{K_{\theta(x,h)}(X_i)}{\frac{1}{m} \sum_{j=1}^m \frac{1}{y_j^2} K_{\theta(\frac{1-y_j}{y_j},h)}(X_i)}. \quad (1.52)$$

où les y_j , $j = 1, \dots, n$ sont des réalisations d'une v.a. distribuée suivant une uniforme sur $[0, 1]$.

Notons bien que la normalisation de l'estimateur \hat{f}_h n'est pas toujours nécessaire, mais elle dépend du domaine d'application étudié.

1.4 Choix du noyau asymétrique

En pratique, le choix du noyau associé dépend du support de la distribution des données dont on se propose d'estimer la densité. Il est évident que ce dernier critère de choix est peu informatif, sinon on aura plus besoin de proposer d'autres types de noyaux après ceux proposés par [Chen \[2000\]](#) dans le cas de support $[0, +\infty[$. En effet, le type de données peut fortement influencer sur le choix du noyau, et cela s'interprète par des queues, des pôles et d'autres spécificités des distributions de ces données. Ainsi [Bouezmarni and Scaillet \[2005\]](#) et [Malec and Schienle \[2014\]](#) conseillent l'utilisation du noyau Gamma pour les distributions ayant des pôles au bord, et [Jin and Kawczak \[2003\]](#) l'utilisation du noyau BS, qui est approximativement similaire aux noyaux RIG et log-normal pour des données des distributions à queue.

Comme [Epanechnikov \[1969\]](#) dans le cas d'estimation de densités à supports non bornés, [Abadir and Lawford \[2004\]](#) ont proposé un noyau asymétrique de forme cubique et

optimal au sens du $MISE$ pour l'estimation des densités à support $[0, +\infty[$. Ce noyau s'écrit sous la forme

$$K_c(t) = \frac{12}{625}(t+2)(t-3)^2 \mathbf{1}_{t \in]-2, 3[}. \quad (1.53)$$

Abadir and Lawford [2004] ont vérifié l'optimalité de leur noyau par simulation, pour la localisation du mode et les quantiles de certaines densités asymétriques.

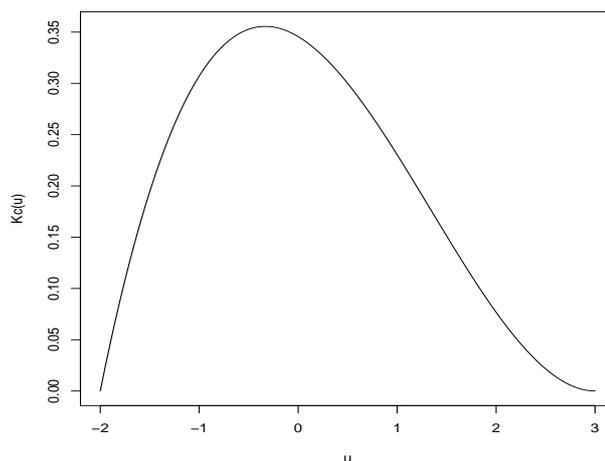


FIGURE 1.3: Forme du noyau optimal asymétrique

L'estimateur à noyau cubique de la densité inconnue f est donné par

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K_c \left(\frac{x - X_i}{h} \right). \quad (1.54)$$

1.5 Conclusion

Dans ce chapitre, nous avons présenté la méthode du noyau associé pour l'estimation non paramétrique de la densité de probabilité. Les techniques de construction du noyaux associé ont été établies, ainsi que la convergence et la consistance de la méthode.

Nous nous sommes focalisé sur l'estimation de densités à support $[0, +\infty[$, étant donné que l'estimation de densités à supports non bornés ont été traités dans Zougab [2007] et Zougab [2013]. Nous avons donné la forme des noyaux non classiques cités dans la littérature, ainsi que les propriétés des estimateurs basés sur ces noyaux.

Lors de l'estimation de la densité par la méthode du noyau associé, l'estimateur doit être normalisé. Nous avons donné deux techniques de normalisation ainsi que leurs approximations par les méthodes Monte Carlo. Ces méthodes d'approximation feront l'objet du chapitre suivant.

L'estimateur à noyau associé dépend du noyau $K_{x,h}$ et du paramètre de lissage h . Le choix du noyau dans l'estimation des densités asymétriques est d'importance capitale contrairement au cas d'estimation de densités symétriques. En fait, même le noyau optimal asymétrique d'[Abadir and Lawford \[2004\]](#) ne pallie pas au problème d'estimation au bord.

Chapitre 2

Approche bayésienne et méthodes Monte Carlo

2.1 Introduction

L'objectif de ce chapitre est d'une part, rappeler quelques notions de base du formalisme Bayésien et les estimateurs paramétriques Bayésiens, utiles pour des considérations majeures dans le cadre de cette étude et d'autre part, passer en revue les principales méthodes de calcul Bayésien qui peuvent être employées lors de l'implémentation pratique de l'approche Bayésienne ([Robert and Casella \[2004\]](#), [Robert \[2006\]](#), [Parent and Bernier \[2007\]](#), [Marin and Robert \[2007\]](#) et [Boreux et al. \[2010\]](#)).

L'inférence Bayésienne mobilise deux sources d'information : d'une part les données, via la vraisemblance et d'autre part le savoir de l'expert via la distribution a priori sur les paramètres et via les hypothèses structurelles sur lesquelles repose le modèle utilisé.

Le concept fondamental de l'approche Bayésienne est la distribution a posteriori $\pi(\theta|x)$. En effet cette distribution opère de façon conditionnelle sur les observations et met donc en oeuvre automatiquement l'inversion des probabilités, tout en incluant le principe de vraisemblance. La distribution a posteriori représente l'actualisation de l'information disponible sur les paramètres θ , au vu de l'information contenue dans la vraisemblance $\pi(x|\theta)$, tandis que la distribution a priori $\pi(\theta)$ représente l'information disponible préala-

blement à l'observation de x .

Définition 4 *Un modèle statistique bayésien est constitué d'un modèle statistique paramétrique, $f(x|\theta)$, et d'une distribution a priori pour les paramètres, $\pi(\theta)$.*

L'impact du théorème de Bayes provient de la décision audacieuse de mettre les causes et effets sur le même niveau conceptuel, puisque les deux sont aléatoires¹.

2.2 Lois a priori

Le point le plus critiquable et le plus critiqué de l'analyse Bayésienne est le choix de la loi a priori. Cependant, une fois cette dernière connue, l'inférence peut être conduite d'une façon quasi mécanique.

Dans une certaine mesure, c'est aussi la plus difficile à déterminer. Dans la pratique, il est rare que l'information a priori soit suffisamment précise pour conduire à une détermination exacte de la loi a priori, au sens où plusieurs lois de probabilité peuvent être compatibles avec cette information. Il y a plusieurs raisons pour cela : le décideur, le client ou le statisticien n'ont pas forcément le temps, les ressources ou souvent la volonté de chercher à construire un a priori exact et doivent compléter l'information partielle qu'ils ont rassemblée à l'aide de données subjectives afin d'obtenir une loi a priori.

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 A priori conjugué

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 prises en compte. En particulier, [Raiffa and Schlaifer \[1961\]](#) ont proposé l'utilisation des densités a priori dites conjuguées afin de faciliter l'analyse Bayésienne.

1. historiquement, l'idée que les paramètres sont aléatoires peut être perçue comme allant à l'encontre du déterminisme athée de Laplace, ainsi que des conceptions religieuses de Bayes

Définition 5 Une famille \mathfrak{F} de distribution de probabilité sur Θ est dite conjuguée (ou fermée par échantillonnage) par une fonction de vraisemblance $f(x|\theta)$ si, pour tout $\pi \in \mathfrak{F}$, la distribution a posteriori $\pi(\cdot|x)$ appartient également à \mathfrak{F} .

Cette définition est formellement vague puisque si l'on choisit \mathfrak{F} comme la famille de toutes les densités de probabilité, la famille de la densité a priori \mathfrak{F} est toujours conjuguée par la fonction de vraisemblance. Il est plus intéressant d'étudier des familles de densités a priori conjuguées tel que celles-ci aient la même forme fonctionnelle que la fonction de vraisemblance. Dans ce cas, le passage de la fonction de vraisemblance à la densité a posteriori se réduit à un changement de paramètres et non à une modification de la forme fonctionnelle de la famille correspondante. Les densités a posteriori sont donc toujours calculables et l'évaluation des statistiques a posteriori est grandement simplifiée.

Considérons l'exemple des familles de densités exponentielles :

Définition 6 Soient Θ l'espace des paramètres et Ω l'espace des observations. On définit C et h respectivement fonctions de Θ et Ω dans \mathbb{R}^+ et R , et T fonctions de Θ et Ω dans \mathbb{R}^k .

La famille \mathfrak{F} tel que les densités soient de la forme :

$$f(x|\theta) = h(x)C(\theta) \exp(R(\theta)^t T(x)), \quad (2.1)$$

est dite famille exponentielle de dimension k .

Dans le cas particulier où $\Theta \subset \mathbb{R}^k$, $\Omega \subset \mathbb{R}^k$ et

$$f(x|\theta) = h(x)C(\theta) \exp(\theta^t x), \quad (2.2)$$

la famille est dite naturelle.

La fonction de vraisemblance du vecteur x tel que les x_i , $i = \overline{1, N}$, sont i.i.d. est définie par :

$$f(x_1, \dots, x_n|\theta) = \left[\prod_{i=1}^N h(x_i) \right] C(\theta)^N \exp \left(R(\theta)^t \sum_{i=1}^N T(x_i) \right). \quad (2.3)$$

Supposons que la densité a priori conjuguée s'écrive de la manière suivante :

$$\pi(\theta) \propto {}^2 C(\theta)^\eta \exp(R(\theta)^T \nu), \quad (2.4)$$

2. Le symbole de proportionnalité, \propto , s'entend en termes de fonction de θ (et non de x).

$f(x \theta)$	$\pi(\theta)$	$\pi(\theta x)$
$\mathcal{N}(\theta, \sigma^2)$	$\mathcal{N}(\mu, \tau^2)$	$\mathcal{N}(\varrho(\mu\sigma^2 + \tau^2x), \varrho\sigma^2\tau^2)$ $\varrho^{-1} = \sigma^2 + \tau^2$
$\mathcal{N}(\mu, 1/\theta)$	$\mathcal{G}(\alpha, \beta)$	$\mathcal{G}(\alpha + 1/2, \beta + (\mu - x)^2/2)$
$\mathcal{P}(\theta)$	$\mathcal{G}(\alpha, \beta)$	$\mathcal{G}(\alpha + x, \beta + 1)$
$\mathcal{G}(\nu, \theta)$	$\mathcal{G}(\alpha, \beta)$	$\mathcal{G}(\alpha + \nu, \beta + x)$
$\mathcal{B}(n, \theta)$	$\mathbf{B}(\alpha, \beta)$	$\mathbf{B}(\alpha + \mathbf{x}, \beta + \mathbf{n} - \mathbf{x})$
$\mathcal{B}neg(m, \theta)$	$\mathbf{B}(\alpha, \beta)$	$\mathbf{B}(\alpha + \mathbf{m}, \beta + \mathbf{x})$

TABLE 2.1: Lois a priori conjuguées naturelles pour quelques familles exponentielles usuelles

où les deux paramètres (ou hyperparamètres) η et ν doivent être spécifiés pour mener à son terme l'analyse Bayésienne. La densité a posteriori appartient, donc, à la famille exponentielle :

$$\pi(\theta|x) \propto C(\theta)^{\eta+N} \exp \left(R(\theta)^T \left(\nu + \sum_{i=1}^N T(x_i) \right) \right). \quad (2.5)$$

Le choix d'un a priori conjugué est toujours un choix particulier et influence donc, dans une certaine mesure, l'inférence résultante. De plus, il peut obliger à ignorer une partie de l'information a priori si cette dernière n'est pas complètement compatible avec la structure de la loi a priori conjuguée.

2.2.2 A priori non informatif

A priori de Laplace

Historiquement, Laplace³ 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 θ est définie par

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

où k est une constante.

3. Nous devons envisager l'état présent de l'Univers comme un effet de l'état antérieur et comme la cause de l'état suivant, 1795.

Trois critiques ont été avancées sur ce choix :

- Les lois résultantes sont impropres quand l'espace des paramètres n'est pas compact, ces dernières mènent à des difficultés comme le paradoxe de marginalisation.
- Le principe des événements équiprobables de Laplace n'est pas cohérent en termes de Partitionnement.
- Une autre critique concerne le problème de l'invariance par reparamétrisation, si on passe de $\theta \in \Theta$ à $\eta = g(\theta)$ par une transformation bijective g l'information a priori reste totalement inexistante et ne devrait pas être modifiée.

Cependant, si $\pi(\theta) = 1$, la loi a priori sur η est $\pi(\eta) = \left| \frac{d}{d\eta} g^{-1}(\eta) \right|$ par le changement de variable. Donc $\pi(\eta)$ est le plus souvent non constante.

Exemple 2 Si p , une proportion, suit une loi uniforme sur $[0, 1]$, le paramètre de rapport des chances $\varrho = \frac{p}{1-p}$ suit une loi a priori $\pi(\varrho) = \frac{1}{(1+\varrho)^2}$, qui est donc non constante.

A priori de Jeffreys

Afin d'éviter le besoin de prendre en compte une structure d'invariance potentielle, Jeffreys (1946, 1961) propose des lois a priori non informatives fondées sur l'information de Fisher, donnée par

$$I(\theta) = E_{\theta} \left[\left(\frac{\delta \log f(x|\theta)}{\delta \theta} \right)^2 \right], \quad (2.6)$$

Dans le cas unidimensionnel, sous certaines conditions de régularité, cette information est aussi égale à

$$I(\theta) = -E_{\theta} \left[\frac{\delta^2 \log f(x|\theta)}{\delta \theta^2} \right]. \quad (2.7)$$

La loi a priori de Jeffreys est

$$\pi(\theta) \propto \sqrt{I(\theta)},$$

elle définit un coefficient de normalisation près quand π est propre.

Ce choix, dépendant de l'information de Fisher se justifie par le fait que $I(\theta)$ est largement accepté comme un indicateur de la quantité d'information apportée par le modèle sur θ , (Fisher 1956).

2.2.3 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 Loi a posteriori

L'analyse statistique se ramène fondamentalement à une inversion⁴, car elle doit déterminer les causes réduites aux paramètres du mécanisme probabiliste générateur, à partir des effets résumés par les observations.

Une description générale de l'inversion des probabilités est donnée par le théorème de Bayes :

Théorème 6 (Bayes) *Si A et E sont deux événements tels que $P(E) \neq 0$, $P(A|E)$ et $P(E|A)$ sont reliés par*

$$\begin{aligned} P(A|E) &= \frac{P(E|A)P(A)}{P(E|A)P(A) + P(E|A^c)P(A^c)} \\ &= \frac{P(E|A)P(A)}{P(E)}. \end{aligned}$$

Bayes donne en réalité une version continue de ces résultats, à savoir, pour deux v.a. X et Y de distributions conditionnelle $f(x|y)$ et marginale $g(y)$, la distribution conditionnelle de y sachant x est

$$g(y|x) = \frac{f(x|y)g(y)}{\int f(x|y)g(y)dy}.$$

Bayes et Laplace ont considéré que l'incertitude sur le paramètre θ d'un modèle peut être décrite par une distribution de probabilité π sur Θ , appelée distribution a priori. L'inférence est alors fondée sur la distribution de θ conditionnelle à X , $\pi(\theta|x)$, appelée distribution **a posteriori** et définie par

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int f(x|\theta)\pi(\theta)d\theta}. \quad (2.8)$$

4. À l'époque de Bayes et de Laplace, c'est-à-dire à la fin du XVIII ème siècle, la Statistique était souvent appelée Probabilités inverses, à cause de cette perspective.

2.4 Marginalisation

Une caractéristique de l'approche bayésienne est la possibilité d'éliminer les paramètres qui ne nous intéressent plus, par marginalisation.

Supposons que le vecteur des paramètres θ puisse se décomposer en deux vecteurs tel que $\theta = (\theta_1, \theta_2) \in \Theta_1 \times \Theta_2$ et que seul θ_1 nous intéresse. La densité a posteriori marginale de θ_1 , $\pi(\theta_1|x)$ est déduite par intégration de la densité a posteriori de $\pi(\theta|x)$ par rapport à θ_2 sur son domaine de définition, c'est-à-dire :

$$\pi(\theta_1|x) = \int_{\Theta_2} \pi(\theta|x) d\theta_2.$$

Cette technique implique donc une réduction de la dimension de l'espace des paramètres à estimer qui peut être très intéressante lors de l'analyse Bayésienne.

2.5 Quand la vraisemblance fait le posterior

L'inférence Bayésienne mobilise deux sources d'information : d'une part les données via la vraisemblance et d'autre part le savoir de l'expert via la distribution a priori sur les paramètres et via les hypothèses structurelles sur lesquelles repose le modèle utilisé. Les approches Bayésienne et fréquentiste sont équivalentes dans deux cas :

I- Grande taille de données ($n \rightarrow \infty$) : L'influence du prior s'estompe et c'est la vraisemblance qui fait le posterior.

Exemple 3 (Modèle exponentiel) Soit x un n -échantillon de v.a. i.i.d. par une exponentielle paramétrée par θ dont la vraisemblance est donnée par :

$$f(x|\theta) = \theta^n \exp(-n\bar{x}\theta),$$

le prior conjugué est une distribution Gamma de paramètres a et b :

$$\pi(\theta|a, b) = \frac{b^a}{\Gamma(a)} \theta^{a-1} \exp(-b\theta),$$

la règle de Bayes donne la distribution a posteriori de θ suivante

$$\pi(\theta|x, a, b) \propto \theta^{n+a-1} \exp(-(n\bar{x} + b)\theta).$$

On reconnaît la forme analytique d'une nouvelle distribution Gamma.

Si la taille de l'échantillon ($n \rightarrow \infty$) alors

$$n + a - 1 \approx n \quad \text{et} \quad n\bar{x} + b \approx n\bar{x},$$

Ainsi, le posterior et la vraisemblance ont la même forme analytique.

II- Prior vague : La vraisemblance fait aussi le posterior quand l'état de connaissance disponible n'autorise qu'un prior vague (peu informatif).

Exemple 4 (Suite) Dans l'exemple précédent, un prior vague est obtenu pour $(a, b) \rightarrow (0, 0)$.

La forme analytique de la densité a posteriori devient

$$\pi(\theta|x) \propto \theta^{n-1} \exp(-(n\bar{x})\theta),$$

Quand n est assez grand, n et $n - 1$ possèdent le même ordre de grandeur. De même, le posterior et la vraisemblance ont la même forme analytique.

2.6 Inférence Bayésienne

Dans cette section, nous présentons les estimateurs ponctuels Bayésiens, et leurs propriétés. Nous la débutons par une discussion sur le concept de fonction de risque Bayésien dont la minimisation conduit aux différents estimateurs paramétriques Bayésiens. En particulier, nous présentons, en détail, les estimateurs par maximum a posteriori et par moyenne a posteriori.

2.6.1 Fonctions de coût

Étant donné $\hat{\theta}$ l'estimateur du paramètre θ , on définit une fonction de coût non négative $C(\epsilon)$ telle que $\epsilon = \theta - \hat{\theta}$ est l'erreur d'estimation pour un vecteur d'observations x donné. L'objectif est de déterminer l'estimateur $\hat{\theta}$ qui minimise le coût moyen $E[C(\epsilon)]$ nommé le risque de Bayes "R" défini par :

$$R = E[C(\epsilon)] = \int_{\Theta} \left[\int_{\Omega} C(\epsilon) \pi(x|\theta) dx \right] \pi(\theta) d\theta. \quad (2.9)$$

Sous certaines conditions de régularité (2.9) peut s'écrire sous la forme :

$$R = \int_{\Omega} \left[\int_{\Theta} C(\epsilon) \pi(\theta|x) d\theta \right] \pi(\theta) m(x) dx. \quad (2.10)$$

Pour x fixé, l'estimateur qui minimise R est l'estimateur qui minimise le coût moyen sur θ , définit par :

$$\mathbb{E} \left[C(\theta - \hat{\theta}(x)) | x \right] = \int_{\Theta} C(\theta - \hat{\theta}(x)) \pi(\theta|x) d\theta.$$

On présente les estimateurs Bayésiens correspondant aux fonctions de coûts classiques.

Coût quadratique

Considérons le cas où $C(\epsilon)$ est quadratique

$$C(\theta - \hat{\theta}(x)) = (\theta - \hat{\theta}(x))^2.$$

L'estimateur de Bayes qui minimise le risque de Bayes étant donné cette fonction de coût, doit minimiser :

$$\begin{aligned} \mathbb{E} \left[(\theta - \hat{\theta}(x))^2 | x \right] &= \hat{\theta}^2(x) - 2\hat{\theta}(x)\mathbb{E}[\theta|x] + \mathbb{E}[\theta^2|x] \\ &= \hat{\theta}^2(x) - 2\hat{\theta}(x)\mathbb{E}[\theta|x] - \mathbb{E}^2[\theta|x] + \mathbb{E}^2[\theta|x] + \mathbb{E}[\theta^2|x] \\ &= \mathbb{V}[\theta|x] + (\hat{\theta}(x) - \mathbb{E}[\theta|x])^2. \end{aligned} \quad (2.11)$$

L'égalité qui correspond au minimum du risque quadratique est atteinte pour

$$\hat{\theta}_{MMSE}(x) = \mathbb{E}[\theta|x]. \quad (2.12)$$

Par conséquent, pour une fonction de coût quadratique, l'estimateur de risque de Bayes minimum est la moyenne a posteriori ou en notation anglo-saxonne *Minimum Mean Squared Error* (*MMSE*).

Coût absolu

Considérons le cas où $C(\epsilon)$ est la fonction valeur absolue,

$$C(\theta - \hat{\theta}(x)) = |\theta - \hat{\theta}(x)|,$$

L'estimateur de Bayes qui minimise le risque de Bayes étant donnée cette fonction de coût, doit minimiser :

$$\mathbb{E} \left[C(\theta - \hat{\theta}(x)) | x \right] = \int_{\Theta} |\theta - \hat{\theta}(x)| \pi(\theta | x) d\theta. \quad (2.13)$$

Si $\theta \in \mathbb{R}$, on peut écrire :

$$\mathbb{E} \left[|\theta - \hat{\theta}(x)| | x \right] = \int_{-\infty}^{\hat{\theta}(x)} (\hat{\theta}(x) - \theta) \pi(\theta | x) + \int_{\hat{\theta}(x)}^{+\infty} (\theta - \hat{\theta}(x)) \pi(\theta | x) d\theta \quad (2.14)$$

$$= \int_{-\infty}^{\hat{\theta}} p^{\pi}(\theta < y | x) dy + \int_{\hat{\theta}}^{+\infty} p^{\pi}(\theta > y | x) dy. \quad (2.15)$$

Dérivant en $\hat{\theta}(x)$, on obtient

$$p^{\pi}(\theta < \hat{\theta}(x)) - p^{\pi}(\theta > \hat{\theta}(x)) = 0,$$

soit encore

$$p^{\pi}(\theta < \hat{\theta}(x)) = \frac{1}{2}.$$

Dans le cas du coût absolu, l'estimateur de Bayes est la médiane a posteriori, qui est l'estimateur obtenu par Laplace.

Coût uniforme

Considérons le cas où $C(\epsilon)$ est défini par :

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

ou encore

$$C(\theta - \hat{\theta}(x)) = 1 - \mathbf{1}_{[\hat{\theta} - \frac{\delta}{2}, \hat{\theta} + \frac{\delta}{2}]}(\theta),$$

tel que δ est la longueur maximale du domaine de l'erreur d'estimation.

Étant donné x , l'estimateur qui minimise le risque de Bayes associé à la fonction de coût uniforme, doit minimiser

$$\begin{aligned} \mathbb{E} \left[C(\theta - \hat{\theta}) | x \right] &= \int_{\Theta} \left[1 - \mathbf{1}_{[\hat{\theta} - \frac{\delta}{2}, \hat{\theta} + \frac{\delta}{2}]}(\theta) \right] \pi(\theta | x) d\theta \\ &= 1 - \int_{\hat{\theta} - \frac{\delta}{2}}^{\hat{\theta} + \frac{\delta}{2}} \pi(\theta | x) d\theta. \end{aligned} \quad (2.16)$$

Minimiser (2.16) revient à maximiser l'intégrale ci-dessus. Si δ est arbitrairement petit, l'intégrale est maximisée lorsque $\hat{\theta}$ correspond au mode, c-à-d à l'argument du maximum de la densité a posteriori.

Par conséquent, pour une fonction de coût uniforme, l'estimateur qui minimise le risque de Bayes est le mode a posteriori, appelé aussi l'estimateur du maximum a posteriori (*MAP*).

2.6.2 Estimateurs MMSE et MAP

Estimateur MMSE

L'estimateur MMSE est déterminé par la moyenne de la densité a posteriori considérée. Étant donné, un vecteur de paramètres θ et un vecteur d'observations x , l'estimation de θ_{MMSE} est définie par

$$\hat{\theta}_{MMSE} = \int_{\Theta} \theta \pi(\theta|x) d\theta. \quad (2.17)$$

- Une integration est requise pour estimer la moyenne de la densité a posteriori $\pi(\theta|x)$.
- Le principe d'invariance⁵ n'est pas vérifié pour l'estimateur *MMSE*.

Estimateur MAP

L'estimateur MAP est déterminé par le mode de la densité a posteriori. Il est défini par :

$$\hat{\theta}_{MAP} = \arg \max_{\theta} \pi(\theta|x). \quad (2.18)$$

- L'évaluation de $\hat{\theta}_{MAP}$ nécessite uniquement la résolution d'un problème d'optimisation.
- Trouver l'argument du maximum de $\pi(\theta|x)$ est équivalent à trouver celui de $f(x|\theta)\pi(\theta)$, tel que le calcul de la constante de normalisation $m(x)$ n'est pas nécessaire.
- Lorsque $\pi(\theta|x) \propto f(x|\theta)$, c-à-d $\pi(\theta)$ est uniforme sur Θ , l'estimation MAP coïncide avec celle du maximum de vraisemblance (MLE).
- L'estimateur MAP vérifie le principe d'invariance.

5. Un estimateur est invariant si, pour une reparamétrisation $\phi = f(\theta)$ du modèle, les estimateurs $\hat{\theta}$ et $\hat{\phi}$ sont telles que $\hat{\phi} = f(\hat{\theta})$.

- Les estimations MMSE et MAP sont identiques lorsque la densité a posteriori est unimodale et symétrique.

À l'exception des cas académiques, l'inférence Bayésienne se heurte très rapidement à des problèmes d'intégration et/ou d'optimisation. En effet, l'évaluation des intégrales suivantes intervient souvent dans l'analyse Bayésienne :

- Calcul de la constante de normalisation

$$m(y) = \int_{\Theta} f(y|\theta)\pi(\theta)d\theta. \quad (2.19)$$

- Evaluation d'une densité marginale a posteriori

Soit $\theta = (\theta_1, \theta_2) \in \Theta_1 \times \Theta_2$

$$\pi(\theta_1|y) = \int_{\Theta_2} \pi(\theta_1, \theta_2|y)d\theta_2. \quad (2.20)$$

- Calcul de moment a posteriori

$$E[h(\theta)|y] = \int_{\Theta} h(\theta)\pi(\theta|y)d\theta. \quad (2.21)$$

- On peut également se heurter à des problèmes d'optimisation quand, par exemple, on a besoin de déterminer le maximum de densité a posteriori

$$\hat{\theta}_{MAP} = \arg \max_{\theta \in \Theta} \pi(\theta|y). \quad (2.22)$$

Pour résoudre ces problèmes, trois approches sont envisageables. Dans un premier temps, nous allons présenter brièvement l'approche par approximation analytique, l'approche numérique puis nous abordons, plus en détails l'approche par simulation, particulièrement les méthodes de Monte Carlo qui sont l'objet de ce chapitre.

2.6.3 Approches par approximations analytiques

Si la densité a posteriori de θ , étant donné le vecteur des observations y , $\pi(\theta|y)$, satisfait certaines conditions de régularité, des méthodes d'approximation analytique peuvent être utilisées pour faciliter l'analyse Bayésienne. Elles sont en général fondées sur des considérations asymptotiques. Nous allons présenter les deux méthodes analytiques les plus courantes pour approximer les inférences a posteriori :

Approximation normale

La densité a posteriori $\pi(\theta|y)$ peut être approximée asymptotiquement par une densité normale $\mathcal{N}(\hat{\theta}, \hat{\Sigma})$ lorsque le nombre d'observations N tend vers l'infini où $\hat{\theta}$ est l'estimation par maximum de vraisemblance de θ et $\hat{\Sigma}$ est l'inverse de la matrice hessienne évaluée pour $\hat{\theta}$, Walker [1969]. Cette méthode est donc basée sur le principe de normalité asymptotique. Le grand avantage de cette approximation est que toutes les inférences a posteriori peuvent, alors, être facilement déterminées. Bien que cette méthode nous donne uniquement le comportement asymptotique de la densité a posteriori, elle peut être également intéressante dans le cas où le vecteur y est de taille modérée et même, lorsque cette dernière est petite et que l'approximation normale n'est pas justifiée. Cette approximation peut être un point de départ à partir duquel les densités a posteriori non Gaussiennes peuvent être analysées.

Approximation de Laplace

Tierney and Kadane [1986] ont proposé une méthode analytique faisant appel à la méthode de Laplace et qui n'est valide qu'asymptotiquement.

Supposons que l'on désire évaluer l'espérance a posteriori suivante :

$$E[h(\theta)|y] = \frac{\int_{\Theta} h(\theta)\ell(\theta)\pi(\theta)d\theta}{\int_{\Theta} \ell(\theta)\pi(\theta)d\theta}, \quad (2.23)$$

Où ℓ est la fonction vraisemblance

L'idée de base de cette méthode est d'appliquer séparément la méthode de Laplace⁶ sur le numérateur et le dénominateur de (2.23).

$$\begin{aligned} \ell_N &= \frac{1}{n}(\ln h + \ln \pi + \ln \ell) \quad \text{et} \\ \ell_D &= \frac{1}{n}(\ln \pi + \ln \ell). \end{aligned}$$

La moyenne a posteriori de $h(\theta)$ est alors donnée par :

$$E[h(\theta)|y] = \frac{\int_{\Theta} e^{n\ell_N} d\theta}{\int_{\Theta} e^{n\ell_D} d\theta}. \quad (2.24)$$

6. la méthode de Laplace (méthode du col ou du point selle), due à Pierre-Simon de Laplace, est une méthode pour l'évaluation numérique d'intégrales de la forme : $\int_a^b e^{Mf(x)} dx$

Soit $\hat{\theta}_N$ le mode de ℓ_N , et $\sigma_N^2 = -1/\ell_N''(\hat{\theta}_N)$. La méthode d'approximation de Laplace nous donne :

$$\begin{aligned} \int e^{nL_N(\theta)} d\theta &\approx \int e^{n[\ell_N(\hat{\theta}_N) - \frac{(\theta - \hat{\theta}_N)^2}{2\sigma_N^2}]} d\theta \\ &= \sqrt{2\pi\sigma_N} n^{-1/2} e^{n\ell_N(\hat{\theta}_N)}. \end{aligned}$$

D'une manière similaire, si $\hat{\theta}_D$ le mode de ℓ_D , et $\sigma_D^2 = -1/\ell_D''(\hat{\theta}_D)$, l'approximation de Laplace du dénominateur est $\sqrt{2\pi\sigma_D} n^{-1/2} e^{n\ell_D(\hat{\theta}_D)}$.

Enfin la moyenne a posteriori de $h(\theta)$ est approximée par

$$E[h(\theta)|y] = \frac{\sigma_N}{\sigma_D} e^{n[\ell_N(\hat{\theta}_N) - \ell_D(\hat{\theta}_D)]}. \quad (2.25)$$

Cette technique peut être utilisée dans le cas multidimensionnel pour l'évaluation des densités marginales et l'approximation des moments a posteriori comme le soulignent Lindely [1980] et Tierney and Kadane [1986].

Remarques 1

- $h(\theta)$ doit être une fonction positive. Des approximations alternatives ont été proposées pour contourner cette difficulté par Tierney and Kadane [1986].
- Cette méthode d'approximation analytique n'est intéressante que si les estimateurs par Maximum de Vraisemblance nécessaires sont faciles à évaluer.

2.6.4 Approche numérique

De nombreuses méthodes numériques d'intégration et d'optimisation ont été développées en Analyse Numérique (voir par exemple Jedrzejewski [2005]). Certaines d'entre elles peuvent être intéressantes pour résoudre les problèmes d'implémentation pratique de l'analyse Bayésienne. Nous ne présenterons qu'une méthode d'intégration numérique, la quadrature numérique, et une méthode d'optimisation, l'algorithme de Newton-Raphson.

Quadrature numérique

Supposons que l'on désire calculer une intégrale de la forme suivante :

$$I = \int h(\theta) f(y|\theta) \pi(\theta) d\theta. \quad (2.26)$$

La quadrature numérique est basée sur une approximation de la fonction intégrée par le produit d'une densité normale et d'un polynôme, tel que :

$$g(\theta) = h(\theta)f(y|\theta)\pi(\theta) \approx P(\theta)\phi(\theta; \mu, \sigma^2), \quad (2.27)$$

où $\phi(\theta; \mu, \sigma^2)$ est la densité Normale de moyenne μ et de variance σ^2 au point θ et $P(\theta)$ est un polynôme de degré $2M$.

Ainsi $g(\theta)$ peut être intégrée efficacement en utilisant la quadrature de Gauss-Hermite telle que :

$$\int P(\theta)\phi(\theta; \mu, \sigma^2)d\theta \approx \sum_{i=1}^M \omega_i P(\theta_i) e^{\theta_i^2}, \quad (2.28)$$

où

$$\begin{aligned} \omega_i &= 2^{M-1} M \sqrt{\pi} H_{M-1}^2(x_i) \text{ et} \\ \theta_i &= \sqrt{2}\sigma x_i + \mu, \end{aligned}$$

où x_i est la $i^{\text{ème}}$ racine du polynôme de Hermite $H_M(x)$ de degré M .

Remarques 2

- Contrairement à la méthode de Simpson⁷, les racines x_i ne sont pas uniformément espacés, ce qui permet une évaluation très efficace de l'intégrale (2.26).
- D'autres bases orthogonales peuvent être employées pour approcher les intégrales mais ces méthodes exigent que la fonction g soit suffisamment régulière.
- Quelle que soient les méthodes d'intégration numérique utilisées, elles deviennent très instables et peu fiables lorsque la dimension de θ augmente.

Méthode de Newton-Raphson

Cette méthode est une approche itérative basée sur une approximation quadratique du développement en série de Taylor du logarithme de la densité a posteriori $L(\theta) = \ln \pi(\theta|y)$ où cette densité peut ne pas être normalisée. Cette approximation quadratique de $L(\theta)$ est, généralement, suffisamment précise quand le nombre des observations est grand par rapport au nombre de paramètres.

7. Dû à Thomas Simpson, est une technique de calcul numérique d'une intégrale $\int_a^b f(x)dx$ en utilisant l'approximation de f par un polynôme quadratique d'ordre 2

La procédure de recherche du mode est alors la suivante :

Algorithme 1 (Newton-Raphson)

1. choix d'une valeur initiale $\theta^{(0)}$
2. pour $k=1, 2, \dots$
 - calcul de $L'(\theta^{(k-1)})$ et $L''(\theta^{(k-1)})$
 - évaluation de $\theta^{(k)}$ à partir de la relation suivante :
$$\theta^{(k)} = \theta^{(k-1)} - \frac{L'(\theta^{(k-1)})}{L''(\theta^{(k-1)})}.$$

Remarques 3

- Cette méthode nécessite la détermination analytique des deux premières dérivées de $L(\theta)$.
- La valeur initiale joue un rôle très important. En effet, la convergence vers le mode n'est pas garantie pour toutes les valeurs initiales, en particulier, lorsqu'elles sont issues des régions où $-L''(\theta)$ n'est pas définie positive.
- L'avantage de cette méthode numérique est que la convergence vers le mode est très rapide une fois que l'on est proche de la solution. Par contre s'il n'y a pas convergence, il faut essayer avec une autre valeur initiale.

Une des approches possibles pour résoudre des problèmes d'intégration et d'optimisation rencontrés en inférence Bayésienne est d'utiliser des méthodes de simulation de densités de probabilité. Elles sont basées sur la génération de variables aléatoires distribuées suivant une loi à simuler.

2.6.5 Méthodes de simulation Monte Carlo

A l'origine, l'objectif des méthodes de simulation de Monte Carlo était de simuler des échantillons de réplicats indépendants de distributions de probabilités. La simulation Monte Carlo permet d'estimer les moments de tout ordre de distributions de variables aléatoires, trop complexes pour être étudiées de façon analytique, en utilisant des algorithmes numériques itératifs de génération.

Méthode d'acceptation-rejet

La première technique de simulation générale (échantillonnage indirecte), introduite par J. Von Neuman dans les années 40 puis publiée en 1956, est la méthode d'acceptation-rejet.

Soit f une densité de probabilité de support \mathbb{S}_f difficile à simuler et soit g une autre densité de même support que f mais aisément simulable.

Soit $M > 0$ un nombre réel tel que :

$$\forall x \in \mathbb{S}_f : f(x) \leq Mg(x),$$

alors les étapes de simulation d'une réalisation de f sont résumées dans l'algorithme suivant :

Algorithme 2 *Acceptation-Rejet*

– *Données* : les densités $f(\cdot)$ et $g(\cdot)$ et la constante M ,

1. générer θ_g à partir de g ,
2. générer u à partir d'une distribution uniforme en $[0, 1]$,
3. **Si** $u \leq \frac{f(\theta_g)}{Mg(\theta_g)}$ alors
accepter la réalisation θ_g ,

Si non

répéter les étapes 1, 2, 3.

Fin Algorithme

Tout θ_g accepté dans l'algorithme 2 n'est qu'une réalisation d'une variable aléatoire ayant $\frac{f(\theta)}{\int f(\theta)d\theta}$ comme distribution de probabilité. Sans perte de généralités, on se contente de présenter un bref résumé de la preuve heuristique donnée par [Smith and Gelfand \[1992\]](#).

Leur construction se base sur deux événements S et S_0 donnés comme suit :

$$S = \{(\theta, u) : \theta \leq \theta_0, u \leq f(\theta_g)/Mg(\theta_g)\} \text{ et}$$

$$S_0 = \{(\theta, u) : u \leq f(\theta_g)/Mg(\theta_g)\}.$$

La fonction de répartition de θ (considéré réel) accepté selon la procédure précédente est alors donnée par

$$P(\theta \leq \theta_0 | \theta \text{ accept}) = \frac{P(\theta \leq \theta_0, \theta \text{ accept})}{P(\theta \text{ accept})} \quad (2.29)$$

$$= \frac{\int_{-\infty}^{\theta_0} f(\theta) d\theta}{\int_{-\infty}^{+\infty} f(\theta) d\theta}. \quad (2.30)$$

Il s'ensuit que le paramètre θ accepté ait une densité proportionnelle à la cible f .

Choix de la constante M

La probabilité qu'une réalisation θ soit acceptée est donnée par

$$\begin{aligned} P(\theta \text{ accept}) &= \int \int \mathbf{1}_S(u, \theta) g(\theta) du d\theta \\ &= \frac{1}{M} \int_{-\infty}^{+\infty} f(\theta) d\theta. \end{aligned} \quad (2.31)$$

Dans l'équation (2.31), le terme $\frac{1}{M} \int_{-\infty}^{+\infty} f(\theta) d\theta$ est bien une probabilité et cela est vérifié pour

$$M \geq \int_{-\infty}^{+\infty} f(\theta) d\theta.$$

En pratique, il est conseillé de choisir $M = [\int_{\mathbb{S}_f} f(x) dx] + 1$ ou $M = \int_{\mathbb{S}_f} f(x) dx$ dans la mesure où la primitive de f soit facilement calculable. Sinon il est nécessaire de faire appel à d'autres techniques qu'on présentera par la suite.

Remarque 1 *La méthode d'acceptation-rejet a le défaut de nécessiter la connaissance a priori de la constante M qui n'est pas toujours disponible.*

En effet, dans le cadre bayésien cette dernière est (en partie) l'objet de recours aux techniques de simulation.

2.6.5.1 Méthodes d'échantillonnage

Les méthodes d'échantillonnage sont utilisées pour l'approximation des espérances dans le cadre de l'estimation statistique ainsi que la visualisation des distributions complexes. L'historique de ces méthodes est disponible dans [Metropolis \[1987\]](#).

Le choix des espérances comme base des méthodes d'échantillonnage est motivé par leur adaptativité et représentativité des mesures et grandeurs statistiques, comme indiqué ci-dessous :

1. Toute probabilité peut être exprimée sous forme d'espérance :

Soit X une v.a et A un événement probabiliste donné, alors

$$P(x \in A) = E[\mathbf{1}_{x \in A}].$$

2. Certaines intégrales et/ou sommations complexes nécessitent des approximations. Or que ces dernières peuvent toujours être représentées autant qu'espérance sous certaine distribution de probabilité (en général, l'uniforme).

Avantages des méthodes de Monte Carlo

- faciles à comprendre et à implémenter,
- domaine d'applicabilité général.

Inconvénients des méthodes de Monte Carlo

- leur facilité cause leur faiblesse lors des utilisations inappropriées,
- vitesse de convergence lente comparées aux méthodes d'approximation déterministes,
- difficulté de génération des échantillons représentatifs des cibles,
- difficulté de contrôle de leur convergence.

2.6.5.2 Approximation de Monte Carlo

Le but est d'approximer des espérances $E[h(X)]$ motivées par leur difficulté d'évaluation.

Définition 7 Si $X_1, X_2, \dots, X_n \stackrel{iid}{\sim} f$ alors

$\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^n h(X_i)$ (moyenne empirique) est l'estimateur Monte Carlo standard de $E[h(X)]$.

Remarques 4

1. $E[\hat{\mu}_n] = E[h(X)]$,
 \Rightarrow l'estimateur standard Monte Carlo est non biaisé.
2. $\hat{\mu}_n \xrightarrow{\mathbb{P}} E[h(X)]$ (LGN) pour $n \rightarrow \infty$ et $\sigma^2[h(X)] < \infty$,
 $\Rightarrow \hat{\mu}_n$ est consistant.
3. $\sigma^2[\hat{\mu}_n] = \frac{1}{n}\sigma^2[h(X)]$,
 \Rightarrow le taux de convergence de $\hat{\mu}_n$ est de $\frac{1}{\sqrt{n}}$.
4. $MSE(\hat{\mu}_n) = \frac{1}{n}\sigma^2[h(X)] \rightarrow 0$,
 $\Rightarrow \hat{\mu}_n$ est convergent en moyenne quadratique.

2.6.5.3 Échantillonnage préférentiel

La méthode que nous présentons ici est appelée *échantillonnage préférentiel* (*pondéré*), car elle est basée sur les fonctions de préférence (ou, aussi appelées, fonctions de pondération).

L'échantillonnage préférentiel est une technique, initialement introduit par Marshall en 1956 Liu [1996]. Elle contourne la simulation suivant la cible (généralement complexe ou inconnue) par un tirage suivant une loi de proposition (dite aussi loi d'importance ou loi instrumentale) et associe à chaque particule issue de ce tirage un poids d'importance. Le résultat de cette technique est un système de particules pondérées dont la loi empirique approche la cible.

On présente quelques généralités de l'échantillonnage préférentiel.

Échantillonnage préférentiel basique

L'échantillonnage préférentiel basique est la technique brute décrite par Marshall. Étant donnée une loi cible Π de densité π , l'échantillonnage préférentiel vise à approximer des intégrales du type

$$\mathfrak{J}(h) = E_{\pi}[h] = \int h(x)\pi(x)dx, \quad (2.32)$$

pour une certaine fonction test h . Pour cela on échantillonne une loi de proposition Q de densité q , de la façon suivante :

$$\begin{aligned} E_\pi[h] &= \int h(x)\pi(x)dx \\ &= \int_{\mathbb{S}_\pi} \frac{h(x)\pi(x)}{q(x)}q(x)dx, \quad \forall q(\text{pdf}) : q(x) = 0 \Rightarrow \pi(x) = 0 \\ &= \int_{\mathbb{S}_q} \frac{h(x)\pi(x)}{q(x)}q(x)dx, \quad \mathbb{S}_\pi \subseteq \mathbb{S}_q \end{aligned} \quad (2.33)$$

$$\begin{aligned} &= E_q \left[h(x) \frac{\pi(x)}{q(x)} \right] \\ &\approx \frac{1}{N} \sum_{i=1}^N h(\theta_i) \frac{\pi(\theta_i)}{q(\theta_i)}, \quad \theta_i \sim q \end{aligned} \quad (2.34)$$

$$= \frac{1}{N} \sum_{i=1}^N h(\theta_i)w(\theta_i), \quad w(\theta_i) = \frac{\pi(\theta_i)}{q(\theta_i)}. \quad (2.35)$$

L'intégrale (2.32) est approchée par l'estimateur (2.36) donné par

$$\widehat{\mathfrak{J}}_N^{IS}(h) = \frac{1}{N} \sum_{i=1}^N w(\theta_i)h(\theta_i), \quad (2.36)$$

avec θ un N-échantillon issu de Q et w est la fonction d'importance définie comme le rapport

$$w(\theta_i) = \frac{\pi(\theta_i)}{q(\theta_i)}. \quad (2.37)$$

L'estimateur (2.36) converge vers (2.32) pour les mêmes raisons que l'estimateur Monte Carlo standard pour tout choix de la distribution Q telle que $\mathbb{S}_q \supseteq \mathbb{S}_\pi$, dans ce cas $\widehat{\mathfrak{J}}_N^{IS}(h)$ est sans biais (Robert and Casella [2004]). En effet, nous avons

$$\widehat{\mathfrak{J}}_N^{IS}(h) \xrightarrow{p.s.} \mathfrak{J}(h),$$

quand $N \rightarrow \infty$ et lorsque la fonction h est telle que

$$\int \frac{h^2(x)\pi^2(x)}{q(x)}dx < +\infty.$$

En appliquant le théorème de la limite centrale (TCL), nous obtenons

$$\sqrt{N} \left\{ \widehat{\mathfrak{J}}_N^{IS}(h) - \mathfrak{J}(h) \right\} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_{IS}^2(h)) \quad \text{quand } N \rightarrow \infty. \quad (2.38)$$

La variance asymptotique est donnée par :

$$\sigma_{IS}^2(h) = \int \frac{h^2(x)\pi^2(x)}{q(x)} - \mathfrak{J}^2(h). \quad (2.39)$$

- **Choix de l'instrumentale**

D'après les développements présentés ci-dessus (voir les equations (2.33)), le choix de l'instrumentale est guidé, uniquement, par son support qui doit englober celui de la cible (le produit du prior et de la vraisemblance dans le cadre bayésien).

Théorème 7 (Robert and Casella [2004]) *Le choix de q qui minimise la variance de l'estimateur 2.39 est :*

$$q^*(x) = \frac{|h(x)|\pi(x)}{\int |h(z)|\pi(z)dz}. \quad (2.40)$$

La variance de l'estimateur est nulle, en prenant pour loi de proposition $q^* \propto |h|f$. Ce choix n'est évidemment pas possible en pratique puisqu'il requiert la connaissance de la quantité d'intérêt f . Néanmoins, l'expression de q^* signifie que le choix de q doit être guidé par un critère de proximité au produit $|h|f$.

Échantillonnage pondéré auto-normalisé

Dans le domaine statistique bayésien, la forme même des distributions a posteriori (connues seulement à une constante près) n'as pas permis pendant longtemps une application aisée des algorithmes de simulation hormis celui de l'échantillonnage préférentiel Parent and Bernier [2007].

En absence de connaissance de constantes de normalisation (paradoxe des constantes de normalisation), l'échantillonnage préférentiel basique ne peut être appliqué d'où la proposition d'une adaptation et généralisation de ce dernier.

$$\begin{aligned} E_\pi[h] &= \int h(\theta)\pi(\theta|x)d\theta \\ &= \int_{\mathbb{S}_\pi} \frac{h(\theta)\pi(\theta|x)}{q(\theta)}q(\theta)d\theta, \quad \forall q(pdf) : q(\theta) = 0 \Rightarrow \pi(\theta|x) = 0 \\ &= \int_{\mathbb{S}_q} \frac{h(\theta)\pi(\theta|x)}{q(\theta)}q(\theta)d\theta, \quad \mathbb{S}_\pi \subseteq \mathbb{S}_q \\ &= \int_{\mathbb{S}_q} \frac{h(\theta)f(x|\theta)\pi(\theta)}{m(x)q(\theta)}q(\theta)d\theta, \quad q(\theta) = 0 \Rightarrow f(x|\theta)\pi(\theta) = 0 \end{aligned} \quad (2.41)$$

$$\begin{aligned}
E_\pi[h] &= \int_{\mathbb{S}_q} \frac{h(\theta)w(\theta)q(\theta)d\theta}{\int_{\mathbb{S}_\pi} f(x|\theta)\pi(\theta)d\theta}, \quad w(\theta) = \frac{f(x|\theta)\pi(\theta)}{q(\theta)} \quad (2.42) \\
&= \int_{\mathbb{S}_q} \frac{h(\theta)w(\theta)q(\theta)d\theta}{\int_{\mathbb{S}_q} \frac{f(x|\theta)\pi(\theta)}{q(\theta)}q(\theta)d\theta} \\
&= \int_{\mathbb{S}_q} \frac{h(\theta)w(\theta)q(\theta)d\theta}{\int_{\mathbb{S}_q} w(\theta)q(\theta)d\theta} \\
&= \frac{\int_{\mathbb{S}_q} h(\theta)w(\theta)q(\theta)d\theta}{\int_{\mathbb{S}_q} w(\theta)q(\theta)d\theta} \\
&= \frac{E_q[h(\cdot)w(\cdot)]}{E_q[w(\cdot)]} \\
&\approx \frac{\frac{1}{N} \sum_{i=1}^N h(\theta_i)w(\theta_i)}{\frac{1}{N} \sum_{j=1}^N w(\theta_j)}, \quad \theta \sim q \\
&= \sum_{i=1}^N h(\theta_i)\hat{w}(\theta_i), \quad \hat{w}(\theta_i) = \frac{w(\theta_i)}{\sum_{j=1}^N w(\theta_j)}. \quad (2.43)
\end{aligned}$$

NB : même dans le cas où l'instrumentale q soit connue à une constante près, l'échantillonnage préférentiel auto-normalisé demeure applicable.

Soient $\pi \propto \tilde{\pi}$ et $q \propto \tilde{q}$ ou encore $\pi = \frac{1}{q_\pi} \tilde{\pi}$ et $q = \frac{1}{q_q} \tilde{q}$ avec q_π et q_q sont des constantes de normalisation respectives.

$$\begin{aligned}
E_\pi[h] &= \int h(x)\pi(x)dx \\
&= \int_{\mathbb{S}_\pi} \frac{h(x)\pi(x)}{q(x)}q(x)dx \\
&= \int_{\mathbb{S}_q} h(x) \frac{\tilde{\pi}(x)}{\tilde{q}(x)} \frac{q_q}{q_\pi} q(x)dx \\
&= \frac{q_q}{q_\pi} \int_{\mathbb{S}_q} h(x) \frac{\tilde{\pi}(x)}{\tilde{q}(x)} q(x)dx \\
&= \left(\frac{q_\pi}{q_q}\right)^{-1} E_q \left[h(x) \frac{\tilde{\pi}(x)}{\tilde{q}(x)} \right] \quad (2.44)
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}_\pi[h] &= \left(\int \frac{1}{q_q} \tilde{\pi}(x) dx \right)^{-1} \mathbb{E}_q \left[h(x) \frac{\tilde{\pi}(x)}{\tilde{q}(x)} \right] \\
&= \left(\int \frac{\tilde{\pi}(x)}{\tilde{q}(x)} q(x) dx \right)^{-1} \mathbb{E}_q \left[h(x) \frac{\tilde{\pi}(x)}{\tilde{q}(x)} \right] \\
&= \frac{\mathbb{E}_q[h(\cdot)\tilde{w}(\cdot)]}{\mathbb{E}_q[\tilde{w}(\cdot)]}, \quad \tilde{w}(\cdot) = \frac{\tilde{\pi}(\cdot)}{\tilde{q}(\cdot)} \\
&\approx \frac{\frac{1}{N} \sum_{i=1}^N h(\theta_i) \tilde{w}(\theta_i)}{\frac{1}{N} \sum_{j=1}^N \tilde{w}(\theta_j)}, \quad \theta \sim q \\
&= \sum_{i=1}^N h(\theta_i) \hat{w}(\theta_i), \quad \hat{w}(\theta_i) = \frac{\tilde{w}(\theta_i)}{\sum_{j=1}^N \tilde{w}(\theta_j)}. \tag{2.45}
\end{aligned}$$

En utilisant le résultat (2.45), l'algorithme général de l'échantillonnage préférentiel est donné comme suit :

Algorithme 3 Échantillonnage préférentiel

- Données : un échantillon x , $x \sim f$
- Initialiser $(h(\cdot), \pi(\cdot), q(\cdot), N)$,
- Simuler un N -échantillon θ , $(\theta_1, \theta_2, \dots, \theta_N) \sim q$

1. calculer la vraisemblance $f(x|\theta)$.

2. évaluer les poids $w(\theta_i)$ selon l'Eq. 2.42 :

$$w(\theta_i) = \frac{f(x|\theta_i)\pi(\theta_i)}{q(\theta_i)}.$$

3. normaliser les $w(\theta_i)$ suivant l'Eq. 2.43

$$\tilde{w}(\theta_i) = \frac{w(\theta_i)}{\sum_{j=1}^N w(\theta_j)}.$$

4. évaluer l'estimateur de Monte Carlo

$$\bar{h}_\theta = \sum_{i=1}^N h(\theta_i) \tilde{w}(\theta_i).$$

Fin Algorithme

Ré-échantillonnage

Rubin [1987] a proposé une technique dite ré-échantillonnage afin de pallier, dans une certaine mesure, aux difficultés évoquées précédemment. On l'appelle dans la littérature anglo-saxonne *sampling importance resampling (SIR)*. L'algorithme du Ré-échantillonnage est donné par :

Algorithme 4 Ré-échantillonnage

- Données : un échantillon x , $x \sim f$
- Initialiser ($h(\cdot)$, $\pi(\cdot)$, $q(\cdot)$, N),
- Simuler un N -échantillon θ , $(\theta_1, \theta_2, \dots, \theta_N) \sim q$

1. calculer la vraisemblance $f(x|\theta)$.

2. évaluer les poids $w(\theta_i)$ selon l'Eq. 2.42 :

$$w_i = \frac{f(x|\theta_i)\pi(\theta_i)}{q(\theta_i)}.$$

3. normaliser les w_i suivant l'Eq. 2.43

$$\tilde{w}_i = \frac{w_i}{\sum_{j=1}^N w_j}.$$

4. Tirer un échantillon avec remise de la distribution finie sur $\theta^1, \theta^2, \dots, \theta^N$ affectant les poids \tilde{w}_i à chaque θ_i

5. évaluer l'estimateur de Monte Carlo

$$\bar{h}_\theta = \frac{1}{N} \sum_{i=1}^N h(\theta^i).$$

Fin Algorithme

L'avantage de cette méthode est d'obtenir un véritable échantillon (tout les poids générés retrouvent le même poids). Son principal défaut est de produire un grand nombre de doublons, ce qui appauvrit les possibilités d'exploration du support de la cible.

2.6.5.4 Population Monte Carlo (PMC)

Les algorithmes de Monte Carlo dépendent de paramètres d'implémentation dont le choix est crucial pour la convergence de l'algorithme et son efficacité. Par exemple, dans les algorithmes de type *échantillonnage pondéré*, les tirages se font sous une loi instrumentale puis corrigés par la suite pour approcher une loi cible.

Choisir des paramètres appropriés est une étape difficile, qui demande une certaine expertise de l'utilisateur. En conséquence, les nouveaux algorithmes de simulation s'orientent vers des procédures itératives au cours desquelles l'algorithme corrige la valeur des paramètres d'implémentation en fonction de son comportement passé. Cette procédure d'adaptation est guidée par la recherche des paramètres d'implémentation qui optimisent un critère d'efficacité de l'échantillonneur.

En combinant la méthode d'échantillonnage préférentiel et celle du ré-échantillonnage, Cappé et al. [2004] ont proposé l'approche PMC pour traiter le problème de reconstitution de signal (observé bruité) dans un canal ionique.

Les étapes de l'algorithme PMC général (sans adaptation à un problème bien spécifique) sont résumées dans l'algorithme suivant :

Algorithme 5 Schéma PMC général

Données : un échantillon x ,

Initialiser ($h(\cdot)$, $f(\cdot)$, $\pi(\cdot)$, $q(\cdot)$, N , T),

• **étape 0 :** *Choix de $(\theta_0^1, \theta_0^2, \dots, \theta_0^N)$;*

• **étape t :** ($t = 1, \dots, T$) :

1. **pour** $i=1, \dots, N$

générer $\theta_t^i \sim q_{it}(\theta)$ et calculer

$$r_t^i = \frac{f(x|\theta_t^i)\pi(\theta_t^i)}{q_{it}(\theta_t^i)};$$

2. *calculer $w_t^i = r_t^i / \sum_{k=1}^N r_t^k$ et ré-échantillonner les θ_t^i en utilisant les poids w_t^i ;*

3. *construire $q_{i(t+1)}$ à partir de l'échantillon courant.*

Fin Algorithme

2.6.5.5 Monte Carlo par chaînes de Markov (MCMC)

La génération des échantillons à partir d'une densité de probabilité π , connue à une constante multiplicative près et qui est suffisamment complexe de sorte que l'application des méthodes usuelles de génération ([Robert and Casella \[2004\]](#)) telle que la méthode d'acceptation-rejet n'est pas possible est un problème et sujet d'actualité.

L'idée de base des méthodes MCMC est de construire une chaîne de Markov $\{\theta^{(t)}\}$, à partir d'un noyau de transition qui garantit la convergence en loi vers la loi cible π , et ce, en partant d'un état initial arbitraire $\{\theta^{(0)}\}$. Au bout d'une période N_0 , appelée période de chauffage, suffisamment grande, on estime que les particules générées (simulées) sont représentatives de la loi cible π .

L'idée d'utiliser le comportement limite d'une chaîne de Markov donnée dans [Metropolis et al. \[1953\]](#) apparaît à la même époque que la technique de Monte Carlo originelle, mais elle nécessite une puissance de calcul qui n'était alors pas suffisamment grande pour qu'elle soit appréciée dans sa globalité.

Algorithmes de Métropolis-Hastings

La mise en oeuvre du principe d'utilisation d'une chaîne de Markov de loi stationnaire π nécessite la construction d'un mécanisme de génération pour produire de telles chaînes. En effet, [Metropolis et al. \[1953\]](#) ont développé un algorithme quasi universel satisfaisant cette contrainte, au départ pour la physique particulaire (et la bombe H), puis généralisé par [Hastings \[1970\]](#) dans un cadre statistique (et plus pacifique). Cet algorithme est connu par ses applications variées car sa principale restriction est que la loi d'intérêt soit connue à une constante près.

Asymptotiquement, une simulation par chaînes de Markov doit assurer deux convergences :

- convergence de la chaîne vers sa loi stationnaire. De plus, cette dernière doit être aussi proche que possible de la cible π ,
- convergence des moyennes empiriques vers l'espérance correspondante $E_\pi[h(\cdot)]$

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 ergodique (voir [Robert and Casella \[2004\]](#) et [Robert \[2006\]](#)) qui résout le paradoxe des deux types de convergence. Ainsi ce théorème supprime le besoin de produire des suites de variables aléatoires indépendantes.

Contrairement à l'algorithme d'acceptation-rejet, l'algorithme M-H ne dépend que des quotients $\pi(y)/\pi(x)$ et $q(x|y)/q(y|x)$ tel que x et y sont deux particules générées successivement. En effet, cette dépendance est la motivation de leur non besoin de connaissance de constantes de normalisation que ce soit pour la cible ou l'instrumentale. L'algorithme est donné comme suit :

Algorithme 6 M-H

– *Données* : les densités $\pi(\cdot)$ et $q(\cdot)$, la taille et la période de chauffe N , N_0 et le vecteur x ,

1. initialiser θ^0 ,

2. Pour $t \in \{1, \dots, N\}$ faire

(a) générer $\tilde{\theta}$, $\tilde{\theta} \sim q(\theta|\theta^{t-1})$,

(b) calculer la probabilité d'acceptation $\rho = \min\{1, \frac{\pi(\tilde{\theta}|x)q(\theta^{t-1}|\tilde{\theta})}{\pi(\theta^{t-1}|x)q(\tilde{\theta}|\theta^{t-1})}\}$,

(c) générer u , $u \sim U_{[0, 1]}$,

$$\theta^{t+1} = \begin{cases} \tilde{\theta}, & \text{si } u < \rho, \\ \theta^t, & \text{sinon.} \end{cases}$$

3. évaluer l'estimateur de Monte Carlo, $\bar{\theta} = \frac{1}{N - N_0} \sum_{t=N_0+1}^N h(\theta^t)$.

Fin Algorithme

Il existe plusieurs formes particulières de l'algorithme M-H induites par le procédé de génération de candidats via l'instrumentale. On cite les algorithmes M-H les plus utilisés et cités en littérature : M-H indépendant, M-H à marche aléatoire et M-H dit une variable à la fois.

2.7 Conclusion

L'objet principal de ce chapitre est la présentation des algorithmes basés sur les différentes méthodes et techniques de Monte Carlo.

Nous avons suivi l'ordre chronologique des événements dans l'évolution et l'introduction des différentes approches d'approximation des intégrales, soit analytiquement, soit numériquement ou par simulation. Les critiques faites aux deux premières approches nous orientent sur l'approche simulation dont les méthodes d'échantillonnage préférentiel adaptatif et les méthodes MCMC l'emportent sur les autres techniques que nous avons présentées.

Une partie du chapitre suivant est consacrée à l'adaptation et implémentation des algorithmes MCMC et PMC pour l'estimation de la moyenne a posteriori de la loi du paramètre de lissage.

Chapitre 3

Approches globales pour la sélection du paramètre de lissage

3.1 Introduction

Le choix du paramètre de lissage h est d'une importance capitale dans l'estimation de la fonction densité par la méthode des noyaux associés.

Dans ce chapitre, nous présentons deux approches différentes pour la sélection du paramètre de lissage. La première approche est basée sur les méthodes classiques (fréquentistes) dont les méthodes de ré-injection (plug-in) et les méthodes validation croisée, la seconde est bayésienne et englobe trois approches : globale, adaptative et locale (voir par exemple les travaux récents de [Zougab et al. \[2012\]](#), [Zougab et al. \[2013a\]](#) et [Zougab et al. \[2013b\]](#)). Nous présentons et utiliserons par la suite uniquement l'approche bayésienne globale à cause de sa similitude avec les méthodes classiques.

3.2 Approche fréquentiste

Plusieurs méthodes (dites classiques) pour choisir le paramètre de lissage ont été proposées dans la littérature et quelques études comparatives intéressantes ont été effectuées (cas des noyaux symétriques, voir [Zougab \[2007\]](#)) notamment les méthodes reposant sur

la validation croisée, dont le principal intérêt est leur caractère direct. Une autre classe de méthodes dite de ré-injection (ou plug-in) a également été introduite. Son principe repose sur l'estimation d'une (ou des) quantité(s) inconnue(s), donnant un paramètre de lissage dépendant des observations.

Nous présentons brièvement le principe de ces méthodes pour le choix du paramètre de lissage dans le cas de données positives.

3.2.1 Méthodes plug-in

Dans le cas d'utilisation des noyaux associés asymétriques, le paramètre de lissage idéal au sens de AMISE dépend généralement de trois quantités inconnues exprimées à travers f , f' et f'' . Ceci rend plus difficile le choix du paramètre de lissage par les méthodes plug-in. La seule méthode qui a été proposée dans ce cas est la méthode de règle de référence par analogie avec le cas des noyaux symétriques. [Scaillet \[2004\]](#) suggère d'estimer les quantités inconnues en remplaçant f par un modèle de référence log-normal de paramètres μ_f et σ_f^2 . Par exemple, les paramètres de lissage obtenus par [Scaillet \[2004\]](#) pour la règle de référence en utilisant le noyau gaussien-inverse et le noyau gaussien-inverse-réciproque sont donnés respectivement par

$$h_{IG}^{rdr} = \left[\frac{16\sigma_f^5 \exp\left\{\frac{1}{8}(7\sigma_f^2 - 20\mu_f)\right\}}{12 + 68\sigma_f^2 + 225\sigma_f^4} \right]^{\frac{2}{5}} n^{-\frac{2}{5}}, \quad (3.1)$$

$$h_{RIG}^{rdr} = \left[\frac{16\sigma_f^5 \exp\left\{\frac{1}{8}(-17\sigma_f^2 + 20\mu_f)\right\}}{12 + 4\sigma_f^2 + \sigma_f^4} \right]^{\frac{2}{5}} n^{-\frac{2}{5}}. \quad (3.2)$$

En pratique, les deux paramètres μ_f et σ_f^2 sont estimés à l'aide des observations X_1, \dots, X_n par la moyenne et la variance empiriques. Cette méthode a tendance à fournir des valeurs très petites pour le paramètre de lissage, ce qui conduit au phénomène de sous-lissage. Une étude de simulation conduite par [Scaillet \[2004\]](#) montre que le choix de paramètre de lissage par la règle de référence en utilisant les noyaux gaussien-inverse et gaussien-inverse-réciproque confirme le défaut de cette approche.

La méthode la plus répandue lorsque le noyau utilisé est asymétrique est la méthode de validation croisée dont nous allons présenter deux de ses variantes :

3.2.2 Méthodes de validation croisée

Validation croisée non biaisée

La validation croisée non biaisée pour la sélection du paramètre de lissage dans l'estimation à noyau de densité a été initialement introduite par [Rudemo \[1982\]](#) et [Bowman \[1984\]](#). La méthode consiste à choisir le paramètre de lissage qui minimise un estimateur optimal de

$$ISE(h) = \int (\hat{f}_h(x) - f(x))^2 dx = \int \hat{f}_h^2(x) dx - 2 \int \hat{f}_h(x) f(x) dx + \int f^2(x) dx. \quad (3.3)$$

Étant donné que $\int f^2(x) dx$ ne dépend pas du paramètre de lissage h . Le paramètre h est choisi de sorte à ce qu'il minimise un estimateur de

$$UCV(h) = \int \hat{f}_h^2(x) dx - 2 \int \hat{f}_h(x) f(x) dx. \quad (3.4)$$

La relation donnant le terme $UCV(h)$ est composée de deux termes $I_1(h)$ et $I_2(h)$ avec $UCV(h) = I_1(h) - 2I_2(h)$

En s'inspirant des techniques de Monte Carlo, le terme $\int \hat{f}_h(x) f(x) dx$ peut être estimé par $\frac{1}{n} \sum_{i=1}^n \hat{f}_h(X_i)$ mais la technique décrite dans [Scott and Terrell \[1987\]](#) nécessite d'estimer

le second terme par un estimateur jackknife¹ $\frac{1}{n} \sum_{i=1}^n \hat{f}_{h,i}(X_i)$ afin d'assurer l'indépendance des variables, où $\hat{f}_{h,i}(X_i)$ est l'estimateur de la densité calculé à partir de l'échantillon privé de l'observation X_i , soit $\hat{f}_{h,i}(X_i) = \frac{1}{n-1} \sum_{j=1, j \neq i}^n K_{X_i, h}(X_j)$.

Le second terme est estimé par $\hat{I}_2(h)$ tel que

$$\hat{I}_2(h) = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j=1, j \neq i}^n K_{X_i, h}(X_j). \quad (3.5)$$

L'évaluation du premier terme $\int_0^\infty \hat{f}_h^2(x) dx$, est obtenue en lui appliquant une approximation Monte Carlo comme suit :

1. Une méthode de ré-échantillonnage qui tire son nom de couteau suisse. À partir des années 70, cette méthode de ré-échantillonnage a été "remplacée" par une méthode plus sophistiquée, le bootstrap.

$$\begin{aligned}
I_1(h) &= \int_0^\infty \hat{f}_h^2(x) dx = \int_0^1 \hat{f}_h^2\left(\frac{1-y}{y}\right) \frac{1}{y^2} dy, \quad \text{avec } x = \frac{1-y}{y} \\
&= \mathbb{E}_{U_{[0,1]}} \left[\hat{f}_h^2\left(\frac{1-y}{y}\right) \frac{1}{y^2} \right] \\
&\approx \frac{1}{m} \sum_{i=1}^m \frac{1}{y_i^2} \hat{f}_h^2\left(\frac{1-y_i}{y_i}\right), \quad y_i \sim U_{[0,1]} \\
&= \frac{1}{m} \sum_{i=1}^m \frac{1}{y_i^2} \left\{ \frac{1}{n} \sum_{j=1}^n K_{\frac{1-y_i}{y_i}, h}(X_j) \right\}^2 \\
&= \frac{1}{mn^2} \sum_{i=1}^m \frac{1}{y_i^2} \left\{ \sum_{j=1}^n K_{\frac{1-y_i}{y_i}, h}(X_j) \right\}^2, \tag{3.6}
\end{aligned}$$

où m est la taille d'une séquence uniforme : y_1, \dots, y_m .

Le critère à optimiser devient alors

$$\text{UCV}(h) = \frac{1}{mn^2} \sum_{i=1}^m \frac{1}{y_i^2} \left\{ \sum_{j=1}^n K_{\frac{1-y_i}{y_i}, h}(X_j) \right\}^2 - \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{j=1, j \neq i}^n K_{X_i, h}(X_j). \tag{3.7}$$

Reste à montrer que $\text{UCV}(h)$ est un estimateur sans biais de $\text{MISE}(h) - \int f^2(x) dx$.

On a :

$$\begin{aligned}
\text{MISE}(h) - \int f^2(x) dx &= \mathbb{E} \left[\int \left\{ \hat{f}_h(x) - f(x) \right\}^2 dx \right] - \int f^2(x) dx \\
&= \mathbb{E} \int \hat{f}_h^2(x) dx - 2\mathbb{E} \int \hat{f}_h(x) f(x) dx. \tag{3.8}
\end{aligned}$$

Or $\int \hat{f}_h^2(x) dx$ est approximée par un estimateur Monte Carlo standard qui est un estimateur approximativement non biaisé. Il reste donc à montrer que $\frac{1}{n} \sum_{i=1}^n \hat{f}_{h,i}(X_i)$ est un estimateur sans biais de $\mathbb{E} \int \hat{f}_h(x) f(x) dx$.

D'une part, on a

$$\begin{aligned}
\mathbb{E} \left[\frac{1}{n} \sum_{i=1}^n \hat{f}_{h,i}(X_i) \right] &= \frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[\frac{1}{n-1} \sum_{j=1, j \neq i}^n K_{X_i, h}(X_j) \right], \\
&= \frac{1}{n} \sum_{i=1}^n \int K_{X_i, h}(z) f(z) dz, \\
&= \int f(x) \int K_{x, h}(z) f(z) dz dx. \tag{3.9}
\end{aligned}$$

D'autre part,

$$\begin{aligned}
 \mathbb{E} \left[\int \hat{f}_h(x) f(x) dx \right] &= \mathbb{E} \left[\int \frac{1}{n} \sum_{i=1}^n \int K_{x,h}(X_i) f(x) dx \right], \\
 &= \frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[\int K_{x,h}(X_i) f(x) dx \right], \\
 &= \int f(x) \int K_{x,h}(z) f(z) dz dx. \tag{3.10}
 \end{aligned}$$

$\frac{1}{n} \sum_{i=1}^n \hat{f}_{h,i}(X_i)$ est donc un estimateur non biaisé de $\mathbb{E} \int \hat{f}_h(x) f(x) dx$.

Algorithme de la méthode validation croisée non biaisée

La sélection du paramètre de lissage optimal, noté h_{ucv} par validation croisée non biaisée, est donnée par la minimisation numérique de la fonctionnelle $UCV(h)$ donnée dans (3.7). Pour cela, on propose un algorithme général pour le calcul du paramètre de lissage optimal d'un noyau associé asymétrique de support \mathbb{R}^+ . Les étapes de l'algorithme sont résumées comme suit :

Algorithme 7 *Validation croisée non biaisée (UCV)* _____

1. Données : un n -échantillon x ,
2. générer un m -échantillon y , $y \sim U_{[0, 1]}$
 $Terme1(h) \leftarrow 0$, $Terme2(h) \leftarrow 0$
3. **pour** $i \leftarrow 1$ **à** m
 $Terme1(h) \leftarrow Terme1(h) + \frac{1}{y_i^2} \left\{ \sum_{j=1}^n K_{\theta(\frac{1-y_i}{y_i}, h)}(x_j) \right\}^2$
4. **pour** $i \leftarrow 1$ **à** n
 $Terme2(h) \leftarrow Terme2(h) + \sum_{j=1, j \neq i}^n K_{\theta(x_i, h)}(x_j)$
5. $UCV(h) \leftarrow \frac{1}{mn^2} Terme1(h) - \frac{2}{n(n-1)} Terme2(h)$
6. $h_{ucv} \leftarrow \arg \min_h UCV(h)$

Fin Algorithme _____

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

$$\begin{aligned} \text{LCV}(h) &= \prod_{i=1}^n \hat{f}_{h,i}(x_i) \\ &= \frac{1}{(n-1)^n} \prod_{i=1}^n \sum_{j=1, j \neq i}^n K_{x_i, h}(x_j). \end{aligned} \quad (3.11)$$

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

$$h_{lcv} = \arg \max_h \text{LCV}(h).$$

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.

3.2.3 Avantages et inconvénients des méthodes classiques

Pour les méthodes plug-in, plusieurs points peuvent être soulignés. D'abord la technique de plug-in est satisfaisante théoriquement dans le cas où le paramètre de lissage optimal h^* obtenu en minimisant le critère AMISE s'exprime d'une manière simple. Par exemple, dans le cas d'utilisation des noyaux associés symétriques continus, la forme de h^* est donné dans [Zougab \[2013\]](#). Cependant, plusieurs difficultés se posent en pratique pour cette méthode. En effet, pour calculer l'optimal h^* , il faut estimer une ou plusieurs quantités inconnues. Ceci s'avère techniquement très difficile selon le type du noyau utilisé et des restrictions sont aussi imposées sur la densité à estimer f souvent très difficiles à vérifier dans la pratique. De plus, le paramètre de lissage obtenu par plug-in n'est valable qu'asymptotiquement.

Le principal intérêt des méthodes validation croisée est qu'elles n'exigent pas des hypothèses sur la densité à estimer f comme c'est le cas des méthodes plug-in. Pour être applicable en pratique, ces méthodes sont guidées entièrement par les observations. Cependant cette approche présente quelques inconvénients majeurs : le paramètre de lissage estimé par les techniques de validation croisée présente une grande variabilité, c'est à dire que les valeurs des paramètres de lissage peuvent être très différentes après plusieurs applications de la méthode à un même échantillon. Ceci relève du problème d'automatisation de la méthode validation croisée.

3.3 Approche Bayésienne globale

Dans cette section, nous présentons 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, \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 le biais de ces étapes :

1. Donner la forme de l'estimateur de la vraisemblance $f(x_1, \dots, x_n|h)$, 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 $\pi(h|x_1, \dots, x_n)$ du paramètre de lissage h et les données x_1, \dots, x_n en utilisant le théorème de Bayes.
4. Estimer le paramètre de lissage h par la moyenne, le mode ou la médiane a posteriori en utilisant les méthodes de Monte Carlo.

3.3.1 Approximation de la vraisemblance

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 de la vraisemblance est obtenu en utilisant l'estimateur à noyau associé donné dans (1.9) et la technique de jackknife, parfois appelée technique de validation croisée.

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

La technique du jackknife consiste à estimer $f(x_i)$ à partir de l'ensemble des points sauf le point x_i en utilisant l'estimateur à noyau associé. L'estimateur à noyau associé et jackknife de $f(x_i)$ est donné par

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

La vraisemblance est alors approximée par

$$f(x_1, \dots, x_n | h) \approx \frac{1}{(n-1)^n} \prod_{i=1}^n \sum_{j=1, j \neq i}^n K_{x_i, h}(x_j). \quad (3.12)$$

L'estimateur (3.12) de la vraisemblance sera utilisé pour le calcul de la densité a posteriori.

3.3.2 Choix de la loi a priori

L'inférence Bayésienne nécessite le choix de la loi a priori sur le(s) paramètre(s) à estimer. Dans le cas du paramètre de lissage, ce choix n'est pas unique mais il est fait de sorte à ce que les méthodes Monte Carlo fonctionnent correctement.

Zhang et al. [2006] ont utilisé la loi a priori $\pi(h)$ dont la densité est donnée, à une constante près par

$$\pi(h) \propto \frac{1}{1+h^2}$$

L'avantage de ce prior est qu'il ne présente pas d'hyper-paramètres. cet avantage fait aussi son défaut, vu la difficulté de son adaptation à l'estimation du paramètre de lissage par la méthode du noyau associé.

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, à une constante près, par

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

où α et β sont des hyper-paramètres.

Notons aussi que le choix de la loi a priori peut se faire dans le but d'avoir la conjugalité avec la vraisemblance.

3.3.3 Approximation de la densité a posteriori

En utilisant le théorème de Bayes, la densité a posteriori de h est estimée par :

$$\begin{aligned} \hat{\pi}(h|\alpha, \beta, x_1, \dots, x_n) &= \frac{f(x_1, \dots, x_n|h)\pi(h|\alpha, \beta)}{\pi(x_1, \dots, x_n)} \approx \frac{\pi(h|\alpha, \beta) \prod_{i=1}^n \hat{f}_{h,i}(x_i)}{\pi(x_1, \dots, x_n)} \\ &\approx \frac{\pi(h|\alpha, \beta) \prod_{i=1}^n \sum_{j=1, j \neq i}^n K_{x_i, h}(x_j)}{\int \pi(h|\alpha, \beta) \prod_{i=1}^n \sum_{j=1, j \neq i}^n K_{x_i, h}(x_j) dh} \end{aligned} \quad (3.13)$$

où $\pi(x_1, \dots, x_n) = \int \prod_{i=1}^n \hat{f}_{h,i}(x_i) \pi(h|\alpha, \beta) dh$ est la loi marginale de l'échantillon.

Vu la complexité de la loi a posteriori (3.13) il est difficile d'obtenir directement l'estimateur Bayésien de h . Pour cela, On utilise les méthodes Monte Carlo présentées dans le chapitre précédent pour estimer le paramètre de lissage h .

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

$$\pi(h|\alpha, \beta, x_1, \dots, x_n) \propto h^{\alpha-1} \exp\left(-\frac{h}{\beta}\right) \prod_{i=1}^n \sum_{j=1, j \neq i}^n K_{x_i, h}(x_j). \quad (3.14)$$

Nous présenterons, ci-dessous, l'adaptation numérique des méthodes Monte Carlo (MCMC et PMC) pour la sélection du paramètre de lissage h .

3.3.4 MCMC pour estimer le paramètre de lissage

Parmi les algorithmes se basant sur la méthode MCMC, nous avons choisi l'algorithme M-H [Hastings \[1970\]](#) à marche aléatoire pour estimer le paramètre de lissage h par la moyenne a posteriori. Cet algorithme est basé sur l'utilisation d'une loi génératrice de candidats de la forme $\mathcal{Q}(\cdot|h^{(k)})$. Le candidat \tilde{h} est généré à partir d'une variable aléatoire ε positive de densité $q(h|h^{(k)}, \gamma_{(k)})$, tels que $h^{(k)}$ est le candidat généré à la $k^{\text{ème}}$ étape, $\gamma_{(k)}$ est un paramètre de réglage choisi de telle sorte à obtenir un taux d'acceptation optimal τ . ε est généralement choisie comme gaussienne de moyenne $h^{(k)}$ et de variance $\gamma_{(k)}^2$ (voir [Zougab \[2013\]](#) et [Li et al. \[2013\]](#)). 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 $\varepsilon \sim \mathcal{N}(h^{(k)}, \gamma_{(k)}^2, 0, \infty)$. La densité de l'instrumentale est alors donnée par

$$q(h|h^{(k)}, \gamma_{(k)}) = \frac{2}{\gamma_{(k)}\sqrt{2\pi}\Phi\left(\frac{h^{(k)}}{\gamma_{(k)}}\right)} \exp\left(-\frac{(h - h^{(k)})^2}{2\gamma_{(k)}^2}\right), \quad (3.15)$$

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

[Garthwaite et al. \[2010\]](#) donnent la formule de mise-à-jour du paramètre γ à l'étape k par

$$\gamma_{(k)} = \begin{cases} \gamma_{(k-1)} + \frac{\gamma_{(k-1)}}{k\tau}, & \text{si } h^{(k)} \text{ est accepté,} \\ \gamma_{(k-1)} - \frac{\gamma_{(k-1)}}{k(1-\tau)}, & \text{si } h^{(k)} \text{ est rejeté,} \end{cases} \quad (3.16)$$

où τ est le taux d'acceptation optimal, il est égal à 0.234 dans le cas multivarié et 0.44 dans l'univarié, voir [Gelman et al. \[1996\]](#), [Roberts and Rosenthal \[2009\]](#) et [Garthwaite et al. \[2010\]](#).

Une des difficultés rencontrées dans l'implémentation des méthodes MCMC est le diagnostic de convergence de la chaîne construite. Plusieurs critères et indicateurs ont été présentés dans la littérature pour vérifier la convergence des méthodes MCMC. Nous citons par exemple, le critère *Intern-Intra-chaîne* proposé par [Gelman et al. \[1996\]](#) et les indicateurs *Bath Mean Standard Error (BMSE)*, aussi appelée erreur standard de Monte Carlo, et *Simulation Inefficiency Factor (SIF)* proposé par [Meyer and Yu \[2000\]](#). Ces critères et indicateurs ont été utilisés pour vérifier la convergence des méthodes MCMC

par plusieurs auteurs, par exemple, Meyer and Yu [2000], Tse et al. [2004], Zhang et al. [2006] et Roberts and Rosenthal [2009].

L'algorithme M-H consiste à construire une chaîne de longueur N avec valeur initiale $h^{(0)}$ puis à évaluer les estimations Monte Carlo désirées. Afin de réduire le biais de l'estimateur Monte Carlo qui est dû à l'effet des valeurs initiales. Les N_0 premières iterations de départ, appelée phase de chauffe, sont ignorées.

Nous présentons les étapes de calcul des indicateurs BMSE et SIF. Ces critères consistent à subdiviser la chaîne $(h^{(k)})_{0 \leq k \leq N}$ en m chaînes de longueur n . Le BMSE est donné par

$$BMSE = \sqrt{\frac{\hat{\sigma}^2}{N}}, \quad (3.17)$$

$$\text{où } \hat{\sigma}^2 = \frac{n}{m-1} \sum_{k=1}^m (\bar{h}_k - \bar{h})^2 \quad \text{avec} \quad \bar{h}_k = \frac{1}{n} \sum_{i=(k-1)n+1}^{kn} h^{(i)}, \quad k = \overline{1, m} \quad \text{et}$$

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

Le SIF est donné par

$$SIF = \frac{\hat{\sigma}^2}{\tilde{\sigma}^2}, \quad (3.18)$$

où $\tilde{\sigma}$ est l'estimation de la déviation standard donnée par

$$\tilde{\sigma} = \left[\frac{1}{N - N_0 - 1} \sum_{i=N_0+1}^N (h^{(i)} - \bar{h})^2 \right]^{\frac{1}{2}}.$$

De petites valeurs de ces indicateurs révèlent la bonne convergence de l'échantionneur. Nous utiliserons ces indicateurs pour contrôler la convergence des Méthodes MCMC.

L'objectif de l'algorithme M-H est de simuler un échantillon selon la densité a posteriori $\pi(h|\alpha, \beta, x_1, \dots, x_n)$ connue à une constante multiplicative près, à partir de la loi de proposition $\mathcal{Q}(h|\tilde{h})$ facilement simulable. Les étapes de l'algorithme M-H sont données comme suit

Étape 1- initialiser $h^{(0)}$ et $\gamma_{(0)}$

Étape 2- pour $k \in \{1, \dots, N\}$

(a) Générer $\tilde{h} \sim q(h|h^{(k-1)}, \gamma_{(k-1)})$

(b) Calculer la probabilité d'acceptation

$$\rho = \min \left\{ 1, \frac{\pi(\tilde{h}|x_1, \dots, x_n)}{\pi(h^{(k-1)}|x_1, \dots, x_n)} \times \frac{q(h^{(k-1)}|\tilde{h}, \gamma_{(k-1)})}{q(\tilde{h}|h^{(k-1)}, \gamma_{(k-1)})} \right\}$$

(c) Calculer l'état de la chaîne

$$h^{(k)} = \begin{cases} \tilde{h}, & \text{si } u < \rho, \quad u \sim U_{[0,1]} \\ h^{(k-1)}, & \text{sinon} \end{cases}$$

(d) Mettre à jour le paramètre de réglage

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

Étape 3- $k = k + 1$ et aller à 2.

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

où $U_{[0,1]}$ désigne la loi uniforme dans $[0, 1]$.

Pour éviter les erreurs de calcul dans la pratique, la probabilité d'acceptation est modifiée comme suit :

$$\hat{\rho} = \min \left\{ 0, \log(\rho) = \log \left(\pi(\tilde{h}|x_1, \dots, x_n) + q(h^{(k-1)}|\tilde{h}, \gamma_{(k-1)}) \right) - \log \left(\pi(h^{(k-1)}|x_1, \dots, x_n) + q(\tilde{h}|h^{(k-1)}, \gamma_{(k-1)}) \right) \right\}.$$

Puisque la fonction logarithme est strictement monotone, la règle de décision devient :

$$h^{(k)} = \begin{cases} \tilde{h}, & \text{si } \log(u) < \hat{\rho}, \quad u \sim U_{[0,1]} \\ h^{(k-1)}, & \text{sinon} \end{cases}$$

3.3.5 PMC pour estimer le paramètre de lissage

Lorsque le calcul explicite de la loi a posteriori du paramètre h est impossible, les méthodes de Monte Carlo par chaînes de Markov (MCMC) permettent de fournir un échan-

tillon approximativement distribué suivant celle loi. Nous avons alors utilisé ce dernier afin d'estimer l'espérance a posteriori du paramètre h . Le recours excessif à l'utilisation des méthodes MCMC a peu à peu marginalisé l'utilisation des méthodes d'échantillonnage préférentiel qui leurs sont ultérieurs. En effet, le schéma d'échantillonnage préférentiel nécessite son adaptativité en lui introduisant une dimension temporelle pour fonctionner correctement.

Dans l'algorithme M-H présenté ci-dessus, la vraisemblance est doublement évaluée dans chaque étape lors du calcul de la probabilité d'acceptation. Cela cause la lenteur de l'algorithme M-H. Cette motivation est largement suffisante pour l'utilisation des méthodes d'échantillonnage préférentiel, précisément l'algorithme Population Monte Carlo (PMC) qui nécessite une seule évaluation de la vraisemblance dans chaque étape (ou pour chaque particule).

L'algorithme PMC est basé sur l'utilisation d'une loi génératrice de particules de la forme $\mathcal{Q}_t(h|h^{(k)})$ où $h_t^{(k)}$ est la $k^{\text{ème}}$ particule générée à l'iteration t . En choisissant comme instrumentale la loi normale tronquée sur $[0, \infty[$, sa densité s'écrit sous la forme

$$q(h|h_{t-1}^{(k)}, \sigma) = \frac{2}{\sigma\sqrt{2\pi}\Phi\left(\frac{h_{t-1}^{(k)}}{\sigma}\right)} \exp\left(-\frac{(h - t - 1^{(k)})^2}{2\sigma^2}\right), \quad (3.19)$$

où σ est un hyper-paramètre fixé ($\sigma = 0.3$ dans notre cas) et $h_0^{(1)}, \dots, h_0^{(M)}$ sont M particules initialisées au départ. Les particules initiales peuvent être générées à partir de la loi a priori.

Le diagnostic de convergence de l'approche PMC ou d'une manière générale l'échantillonnage préférentiel itératif est encore un sujet ouvert [Guillin et al. \[2005\]](#). Par ailleurs, les estimations faites à travers les échantillons générés sont approximativement non biaisées.

L'algorithme PMC consiste en l'itération de deux étapes : tout d'abord, on simule une population de M particules suivant la loi instrumentale choisie puis on calcule les poids de chaque particule et on retire un échantillon dans la population générée proportionnellement aux poids calculés et normalisés. La représentation algorithmique suivante correspond à ces deux étapes :

• **étape 0** : Choix de $h_0^{(1)}, \dots, h_0^{(M)}$;

•• **étape t** : ($t = 1, \dots, T$) :

(a) **pour** $i = 1, \dots, M$

1. générer $h_t^{(i)} \sim q_{it}(h|h_{t-1}^{(i)})$ et calculer $r_t^{(i)} = \frac{\pi(x_1, \dots, x_n | h_t^{(i)}) \pi(h_t^{(i)})}{q_{it}(h_t^{(i)} | h_{t-1}^{(i)})}$

2. normaliser les poids $r_t^{(i)}$, $w_t^{(i)} = r_t^{(i)} / \sum_{k=1}^M r_t^{(k)}$;

(b) ré-échantillonner les particules $h_t^{(i)}$ en utilisant les poids $w_t^{(i)}$;

(c) $t=t+1$ et aller à (••) ;

(d) Calculer l'estimateur de Bayes $\hat{h} = \frac{1}{MT} \sum_{t=1}^T \sum_{i=1}^M w_t^{(i)} \times h_t^{(i)}$

La paramétrisation de l'a priori est liée au paramètre de lissage h . Pour ce faire, nous avons à satisfaire deux contraintes : $h \rightarrow 0$ quand $n \rightarrow \infty$ et la vraisemblance doit faire le posterior quand $n \rightarrow \infty$. Nous avons utilisé la méthode de validation croisée de la vraisemblance pour vérifier la deuxième contrainte.

3.4 Conclusion

L'objectif majeur de ce chapitre est l'utilisation des méthodes de Monte Carlo dans la sélection du paramètre de lissage h , qui est aussi la motivation du choix de l'approche bayésienne globale.

Dans l'approche classique, nous avons présenté les méthodes : plug-in, validation croisée non biaisée et validation croisée de la vraisemblance. Quant à l'approche bayésienne globale, nous avons présenté l'algorithme M-H basé sur les méthodes MCMC et l'algorithme PMC basé sur les méthodes d'échantillonnage préférentiel.

Dans l'approche bayésienne globale, nous n'avons pas discuté le choix des hyperparamètres précisément les paramètres de la loi a priori. En effet, le choix est fait empiriquement de sorte que les méthodes de Monte Carlo fonctionnent correctement.

Chapitre 4

Application Numérique

4.1 Introduction

Nous présentons dans ce chapitre le travail de simulation effectué pour étudier, d'une part, l'influence du noyau dans l'estimation de la densité de probabilité, d'autre part, les performances des méthodes MCMC et PMC 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 sept densités tests présentant différentes formes.

4.2 Choix des densités cibles

Densités définies sur $]0, +\infty[$

- Les densités **D1** et **D2** sont unimodales et avec queues, de plus **D2** est à queue lourde.
- La densité **D4** est unimodale ayant un pôle au voisinage de $x = 0$ et sans maximum local.
- La densité **D7**, combinaison linéaire de deux densités de sorte qu'elle soit bimodale.

Densités définies sur $[0, +\infty[$

- La densité **D3** est unimodale, ayant un maximum en $x = 0$.

- Les densités **D5** et **D6** sont unimodales présentant deux particularités. Les points de forte densité sont concentrés dans le voisinage de $x = 0$ (bord) pour le cas du modèle **D5** et ailleurs pour le modèle **D6**.

Densités des différents modèles :

- **D1** Modèle log-normale de paramètres $\mu = 0$ et $\sigma = 1$:

$$f_1(x) = \frac{1}{x\sqrt{2\pi}} \exp\left(-\frac{1}{2} [\log(x)]^2\right) \mathbf{1}_{]0,+\infty[}(x),$$

- **D2** Modèle log-normale de paramètres $\mu = 2$ et $\sigma = 1$:

$$f_2(x) = \frac{1}{x\sqrt{2\pi}} \exp\left(-\frac{1}{2} [\log(x) - 2]^2\right) \mathbf{1}_{]0,+\infty[}(x),$$

- **D3** Modèle normale tronquée de paramètres $\mu = 0$ et $\sigma = 1$:

$$f_3(x) = \frac{2}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) \mathbf{1}_{]0,+\infty[}(x),$$

- **D4** Modèle Weibull de paramètre de forme $\alpha = 0.91$ et de paramètre d'échelle $\beta = 1$:

$$f_4(x) = 0.91x^{-0.09} \exp(-x^{0.91}) \mathbf{1}_{]0,+\infty[}(x),$$

- **D5** Modèle normale tronquée de paramètres $\mu = \frac{1}{10}$ et $\sigma = \frac{1}{10}$:

$$f_5(x) = \frac{10}{\sqrt{2\pi}\Phi(1)} \exp\left(-50 \left[x - \frac{1}{10}\right]^2\right) \mathbf{1}_{]0,+\infty[}(x),$$

- **D6** Modèle normale tronquée de paramètres $\mu = 1$ et $\sigma = \frac{1}{10}$:

$$f_6(x) = \frac{10}{\sqrt{2\pi}\Phi\left(\frac{1}{10}\right)} \exp(-50[x - 1]^2) \mathbf{1}_{]0,+\infty[}(x),$$

- **D7** Le mélange de deux densités de loi log-normale de paramètres $(\mu_1 = 1, \sigma_1 = \frac{1}{10})$ et $(\mu_2 = 0, \sigma_2 = 1)$ pondérées par les poids $(p_1 = \frac{1}{4}, p_2 = \frac{3}{4})$:

$$f_7(x) = \left[\frac{1}{4} \frac{10}{x\sqrt{2\pi}} \exp(-50[\log(x) - 1]^2) + \frac{3}{4} \frac{1}{x\sqrt{2\pi}} \exp\left(-\frac{1}{2} [\log(x)]^2\right) \right] \mathbf{1}_{]0,+\infty[}(x),$$

Les courbes de ces densités tests sont données dans la figure (4.1) suivante :

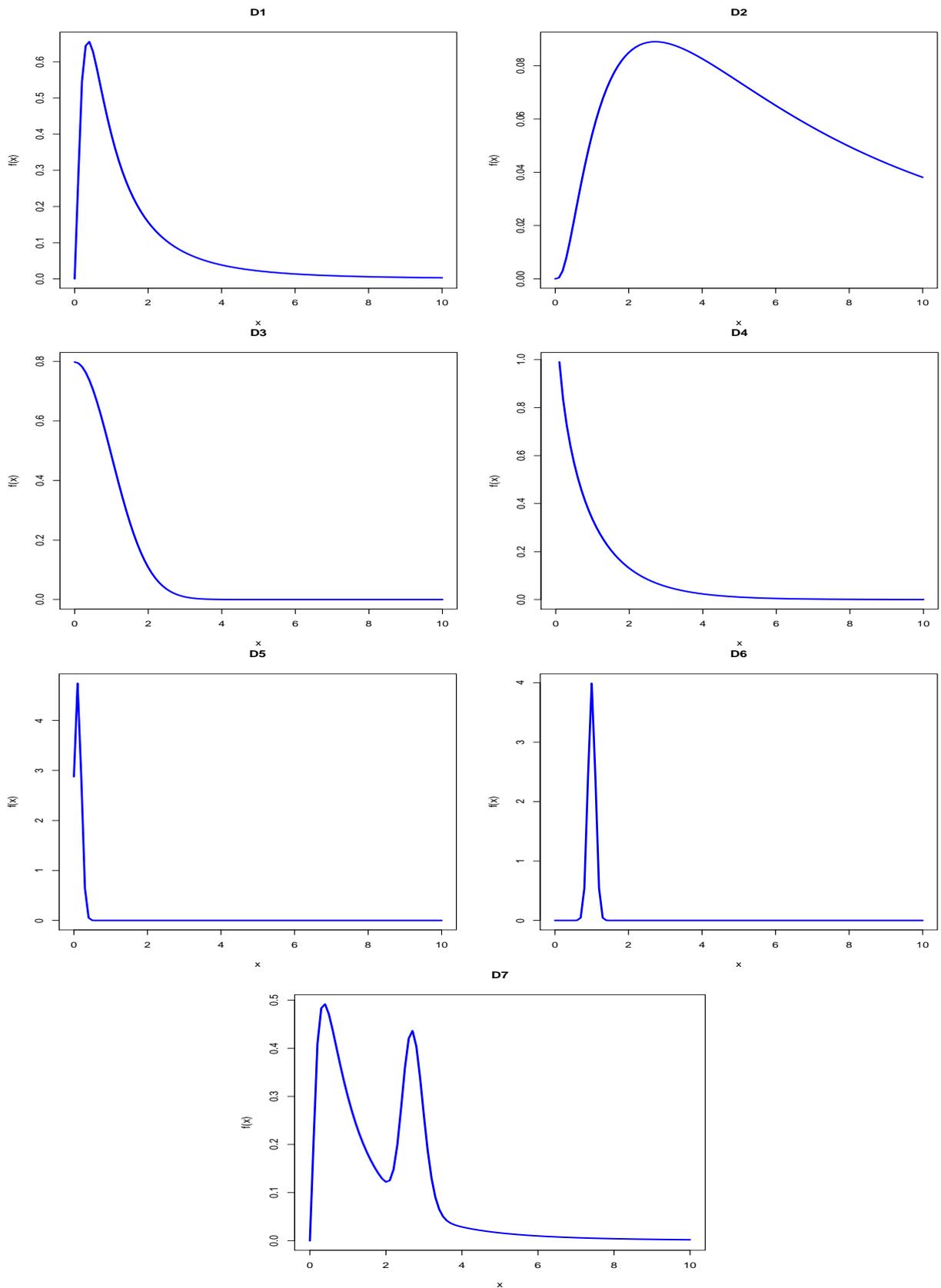


FIGURE 4.1: Courbes des densités tests choisies D_i , $i=\overline{1,7}$

4.3 Les méthodes utilisées

Parmi les méthodes et critères présentés dans les chapitres précédents, nous utilisons :

- **La fonction noyau** : afin de mettre en évidence l'influence du noyau, nous avons utilisé quatre noyaux : Gamma, Gamma modifié, gaussien-inverse-réciproque et log-normal.
- **Le paramètre de lissage** : le paramètre de lissage est sélectionné par le biais de deux approches :
- ★ bayésienne : en utilisant la méthode MCMC (Algorithme M-H) et l'algorithme PMC ;
- ★ classique : en utilisant l'algorithme UCV ;
- **Les critères d'erreur** : nous avons choisi d'évaluer l'erreur d'estimation par le ISE et la MISE, présentés ci-dessous :

Pour une fonction de densité de probabilité f à support positif, nous utilisons l'approximation de Monte Carlo pour l'évaluation de l'erreur quadratique intégrée ainsi que sa moyenne. Le ISE est donné par :

$$\begin{aligned} ISE(h) &= \int_0^\infty \left\{ \hat{f}_h(x) - f(x) \right\}^2 dx \\ &\approx \frac{1}{m} \sum_{i=1}^m \frac{1}{y_i^2} \left\{ \frac{1}{n} \sum_{j=1}^n K_{\frac{1-y_i}{y_i}, h}(x_j) - f\left(\frac{1-y_i}{y_i}\right) \right\}^2, \quad y_i \sim U_{[0,1]}, \end{aligned} \quad (4.1)$$

la \overline{ISE} , la moyenne empirique du $ISE(h)$ est alors donnée par

$$\begin{aligned} \overline{ISE}(h) &= \mathbb{E} \left[\int_0^\infty \left\{ \hat{f}_h(x) - f(x) \right\}^2 dx \right] \\ &\approx \frac{1}{m N_{sim}} \sum_{r=1}^{N_{sim}} \sum_{i=1}^m \frac{1}{y_i^2} \left\{ \frac{1}{n} \sum_{j=1}^n K_{\frac{1-y_i}{y_i}, h}^r(x_j) - f\left(\frac{1-y_i}{y_i}\right) \right\}^2. \end{aligned} \quad (4.2)$$

4.4 Plan de simulation

Nous nous contenterons de faire des simulations et d'observer le comportement de l'estimateur à noyau calculé à partir des échantillons simulés pour les densités tests f_i , $i=\overline{1,7}$. Nous utilisons, pour les simulations, des échantillons de lois **Di**, $i=\overline{1,7}$ et de tailles $n \in \{25, 50, 100, 200, 400\}$. Pour chaque type de densité, pour chaque type de noyau, pour chaque méthode de sélection et pour chaque taille d'échantillon, $N_{sim} = 50$ répétitions d'expériences ont été conduites afin d'approcher le plus possible la valeur du *MISE*.

Algorithme

Étant donnée une distribution paramétrique \mathbf{D}_i , l'algorithme de simulation que nous avons utilisé, pour un seul jeu de données, comporte six phases :

- Simuler un échantillon de taille n à partir d'une des densités paramétriques cibles \mathbf{D}_i , $i=\overline{1,7}$,
- Choisir un noyau asymétrique : Gamma, Gamma modifié, gaussien-inverse-réciproque ou log-normal,
- Calculer le paramètre de lissage optimal (h_{UCV}, h_{MCMC} ou h_{PMC}) en utilisant l'un des algorithmes de sélection (UCV, M-H ou PMC),
- Construire l'estimateur par la méthode du noyau à partir des observations et le paramètre de lissage optimal calculé,
- Calculer l'erreur quadratique intégrée commise lors de l'estimation de f ,
- Tracer les deux courbes : la densité test (théorique) et la densité estimée.

Le calcul du \overline{ISE} est donné après plusieurs exécutions de ce dernier algorithme.

La programmation, les simulations et les graphiques ont été réalisés à l'aide du logiciel \mathcal{R} ([R Development Core Team \[2013\]](#)) version 3.0.0.

4.5 Résultats de simulation

Les résultats de simulation sont donnés sous forme de tableaux et de graphiques. Les tableaux (4.2) à (4.9) contiennent les résultats suivant :

1. h^* : le paramètre de lissage théorique optimal calculé à partir des équations (1.24, 1.29 et 1.40) ;
2. \bar{h}_{UCV} : le paramètre de lissage moyen obtenu par la méthode validation croisée non biaisée ;
3. \bar{h}_{MCMC} : le paramètre de lissage moyen obtenu par l'algorithme de Metropolis-Hasting ;
4. \bar{h}_{PMC} : le paramètre de lissage moyen obtenu par l'algorithme PMC ;

5. σ_h : l'écart type du paramètre de lissage pour chaque taille d'échantillon ;
6. $MISE^*$: l'erreur quadratique intégrée moyenne théorique et optimale, calculée à partir des equations (1.25, 1.30 et 1.42) ;
7. \overline{ISE}_{UCV} : le ISE moyen empirique calculée par la relation (4.2) avec le paramètre de lissage h_{UCV} ;
8. \overline{ISE}_{MCMC} : le ISE moyen empirique calculée par la relation (4.2) avec le paramètre de lissage h_{MCMC} ;
9. \overline{ISE}_{PMC} : le ISE moyen empirique calculée par la relation (4.2) avec le paramètre de lissage h_{PMC} ;
10. σ_{ISE} : l'écart type du ISE pour chaque taille d'échantillon ;
11. τ_{ac} : le taux d'acceptation empirique de l'algorithme M-H ;
12. SIF : le facteur d'inefficacité de simulation empirique, utilisé pour diagnostiquer la convergence de la méthode MCMC ;

Les tableaux (4.2) à (4.9) nous permettent de comparer les résultats obtenus par la méthode de validation croisée non biaisée, les méthodes bayésiennes et les quatre noyaux pour les différentes densités tests \mathbf{D}_i , $i=\overline{1,7}$.

Les figures (4.3) à (4.16) présentent la qualité de lissage des densités \mathbf{D}_i , $i=\overline{1,7}$ à travers les noyaux Gamma, Gamma modifié, RIG et LN en choisissant les fenêtres h_{UCV} , h_{MCMC} et h_{PMC} . Pour la construction des figures, la taille des jeux de données simulées est fixée à $n = 400$.

4.6 Discussions des résultats

Modèle D1

La visualisation des courbes estimées et théoriques présentées dans les figures (4.3) et (4.4) et la lecture du tableau (4.2) montrent que :

- Les valeurs de h_{UCV} sont les plus proches du paramètre de lissage optimal h^* particulièrement pour des échantillons de tailles $n \in \{100, 200, 400\}$.

- h_{MCMC} et h_{PMC} sont proches sauf dans le cas du noyau RIG où h_{PMC} est trop faible, il a tendance à sous-lisser la densité estimée. De plus, la convergence de h_{MCMC} et h_{PMC} vers h^* est très lente.
- Le \overline{ISE} empirique est proche de l'optimal $MISE^*$ associé et ceci quelque soit le noyau choisi.
- Les courbes estimées comparées aux courbes théoriques présentées dans les figures (4.3) et (4.4) confirment ces résultats. La qualité de lissage est satisfaisante, elle est presque identique quelque soit les noyaux et les paramètres de lissage, sauf dans le cas du noyau log-Normal pour les paramètres de lissage sélectionnés par la méthode Bayésienne où les courbes sont sujet à un phénomène de sous-lissage.

Modèle D2

La lecture du tableau (4.3) nous permet de remarquer que :

- Toutes les méthodes de sélection fournissent des valeurs \bar{h} très voisines des valeurs h^* dans le cas des noyaux Gamma modifié et RIG. Notons que la dispersion du paramètre de lissage est importante dans tous les cas, surtout pour la méthode Bayésienne (MCMC et PMC).
- Les performances (selon le critère ISE moyen) de ces méthodes sont très proches, et ceci quelque soit le noyau choisi. Cependant, le noyau RIG associé au paramètre de lissage h_{PMC} est meilleur pour des échantillons de tailles $n \in \{25, 50, 100\}$ et $n \in \{200, 400\}$ pour une combinaison avec h_{UCV} .
- Ces résultats sont confirmés graphiquement (voir les figures 4.5 et 4.6). Enfin, la qualité de lissage est presque identique pour tous les noyaux. On constate que les courbes estimées ne sont pas suffisamment lisses, à l'exception du noyau RIG combiné avec le paramètre de lissage h_{PMC} .

Modèle D3

Les résultats donnés dans le tableau (4.4) nous permettent de remarquer que :

- Les paramètres de lissage obtenus par les différentes approches diffèrent des paramètres de lissage optimaux dans toutes les situations. La même remarque est

constatée pour les \overline{ISE} s et $MISE^*$ s associés.

- En terme du \overline{ISE} , l’approche Bayésienne l’emporte sur la validation croisée pour les tailles d’échantillon $n \in \{25, 50, 100\}$. Les \overline{ISE} minimales sont associées au noyau Gamma modifié, qui sont proches de celles associées au noyau Gamma.
- Les figures (4.7) et (4.8) montrent les similitudes des courbes estimées par les estimateurs à noyaux Gamma et Gamma modifié de celles des courbes de la densité théorique. Les estimations obtenues par les noyaux RIG ou log-Normal divergent au voisinage de 0.

Modèle D4

La première lecture du tableau (4.5) nous permet de remarquer que :

- Le paramètre de lissage \bar{h}_{UCV} associé au noyau Gamma modifié est le seul qui est proche de l’optimum h^* .
- Les erreurs minimales sont enregistrées dans le cas de l’estimateur à noyau Gamma pour le paramètre de lissage h_{UCV} . Notons que ces dernières sont proches des erreurs induites par le même estimateur pour les paramètres de lissage h_{MCMC} et h_{PMC} . Les performances des noyaux Gamma et Gamma modifié sont similaires.
- Les figures (4.9) et (4.10) confirment ces dernières remarques. En effet, la qualité du lissage obtenue par l’estimateur à noyau Gamma est meilleure. Le noyau Gamma est avantageux au voisinage de $x = 0$, ce qui confirme les résultats de [Bouezmarni and Scaillet \[2005\]](#) et [Malec and Schienle \[2014\]](#). Les estimations obtenues par les noyaux RIG et log-Normal divergent au voisinage de 0 et convergent lentement ailleurs.

Modèle D5

La lecture des résultats donnés dans le tableau (4.6) nous permet de constater que :

- Les paramètres de lissage \bar{h}_{UCV} , \bar{h}_{MH} et \bar{h}_{PMC} sont proches et convergent vers h^* et ceci quelque soit le noyau.
- Les erreurs moyennes \overline{ISE} sont minimales et approximativement proches dans le cas des noyaux Gamma et Gamma modifié. De plus, l’approche Bayésienne est meilleure

pour $n \in \{25, 50, 100\}$.

- Les estimateurs à noyaux RIG et log-Normal ne convergent pas vers la densité théorique cible au voisinage de $x = 0$.
- Les figures (4.11) et (4.12) confirment ces dernières remarques. Les courbes estimées par les estimateurs à noyaux Gamma et Gamma modifié sont suffisamment lisses.

Modèle D6

La lecture du tableau (4.7) nous permet de remarquer que :

- Le paramètre de lissage h_{ucv} est constant ($h_{ucv} = 0.00311$) pour tout échantillon de taille $n \geq 200$, et cela pour les noyaux Gamma, Gamma modifié et RIG. Les trois méthodes de sélection convergent vers le même paramètre de lissage dans le cas du noyau log-Normal.
- Les erreurs (selon le critère ISE) sont toutes dans le voisinage des $MISE^*$ associées. Les erreurs minimales sont associées aux paramètres de lissage h_{MCMC} et h_{PMC} pour $n \in \{200, 400\}$.
- Les figures (4.13) et (4.14) confirment ces dernières remarques. La qualité de lissage est presque parfaite dans tout les cas. Contrairement à la validation croisée, l'approche Bayésienne est compétitive dans l'estimation au voisinage du mode de la densité cible.

Modèle D7

La lecture du tableau (4.8) nous permet de remarquer que :

- Les performances de la méthode UCV sont meilleures que celles de l'approche Bayésienne pour $n \in \{200, 400\}$. Pour $n \in \{25, 50, 100\}$ les performances sont proches.
- Les estimateurs à noyaux Gamma et Gamma modifié fonctionnent correctement. En effet, les courbes estimées présentent un effet de sous-lissage engendré par h_{UCV} et une sur-estimation causée par les fenêtres h_{MCMC} et h_{PMC} (Voir la figure (4.15)).
- L'estimateur à noyau RIG fournit des estimations proches de la densité théorique avec un effet de sous-lissage pour la première constituante du mélange (voisinage de $x = 1$) causé par h_{UCV} et h_{PMC} , le lissage est acceptable ailleurs. De plus, on

constate un effet de sur-lissage causé par h_{MCMC} (Voir figure 4.16). Par contre, des sur-estimations de la densité du mélange sont données par l'estimateur à noyau log-Normal (Voir figure 4.16).

Remarque : D'une manière générale, que ce soit pour les noyaux utilisés, ou pour les méthodes de sélection du paramètre de lissage ou les modèles choisis, on peut constater quelques points :

- La taille de l'échantillon n influe sur les valeurs de h^* , \bar{h} , $MISE^*$ et \overline{ISE} . L'augmentation de n entraîne leur décroissance.
- Le tableau (4.9) affiche de petites valeurs du paramètre SIF et des taux d'acceptation proches de la valeur optimal. Ces résultats confirment la bonne convergence de l'algorithme M-H. D'ailleurs, cette remarque est valable pour l'algorithme PMC vu que les deux algorithmes convergent vers des valeurs fortement proches.
- Contrairement à ce qu'a été cité dans la littérature, le choix des particules initiales dans l'algorithme PMC peut influencer dans certains cas sur sa convergence (cas du noyau RIG, voir les tableaux (4.2, 4.4, 4.5 et 4.8)).
- En terme de qualité de lissage, on peut admettre que l'approche Bayésienne fournit des courbes suffisamment lisses comparée aux méthodes classiques (UCV).

4.7 Application aux données réelles

Nous considérons les durées (en minutes) de 299 éruptions du geyser¹ d'Old Faithful dans le parc national de Yellowstone, aux États-Unis d'Amérique. Ces données sont disponibles dans les bases de données du package "MASS" du logiciel \mathcal{R} . Elles sont aussi utilisées par Silverman [1986] et Zougab et al. [2013b] pour l'estimation non paramétrique de densité de probabilité.

En utilisant les différents estimateurs à noyau (Gamma, Gamma modifié, RIG et Log-Normal) dont les paramètres de lissage correspondants sont calculés par les méthodes UCV, MCMC et PMC et présentés dans le tableau (4.1), nous estimons la densité de la

1. Type particulier de source d'eau chaude qui jaillit par intermittence en projetant de l'eau et de la vapeur.

variable continue à support positif "durée d'une éruption". Les différentes courbes estimées sont présentées dans la figure (4.2).

	Gamma	Gamma modifié	RIG	Log-Normal
h_{UCV}	0.00311	0.00311	0.00311	0.00311
h_{MCMC}	0.01092	0.01086	0.00999	0.08219
h_{PMC}	0.01085	0.01133	0.01105	0.08213

TABLE 4.1: Paramètres de lissage correspondants aux durées d'éruption

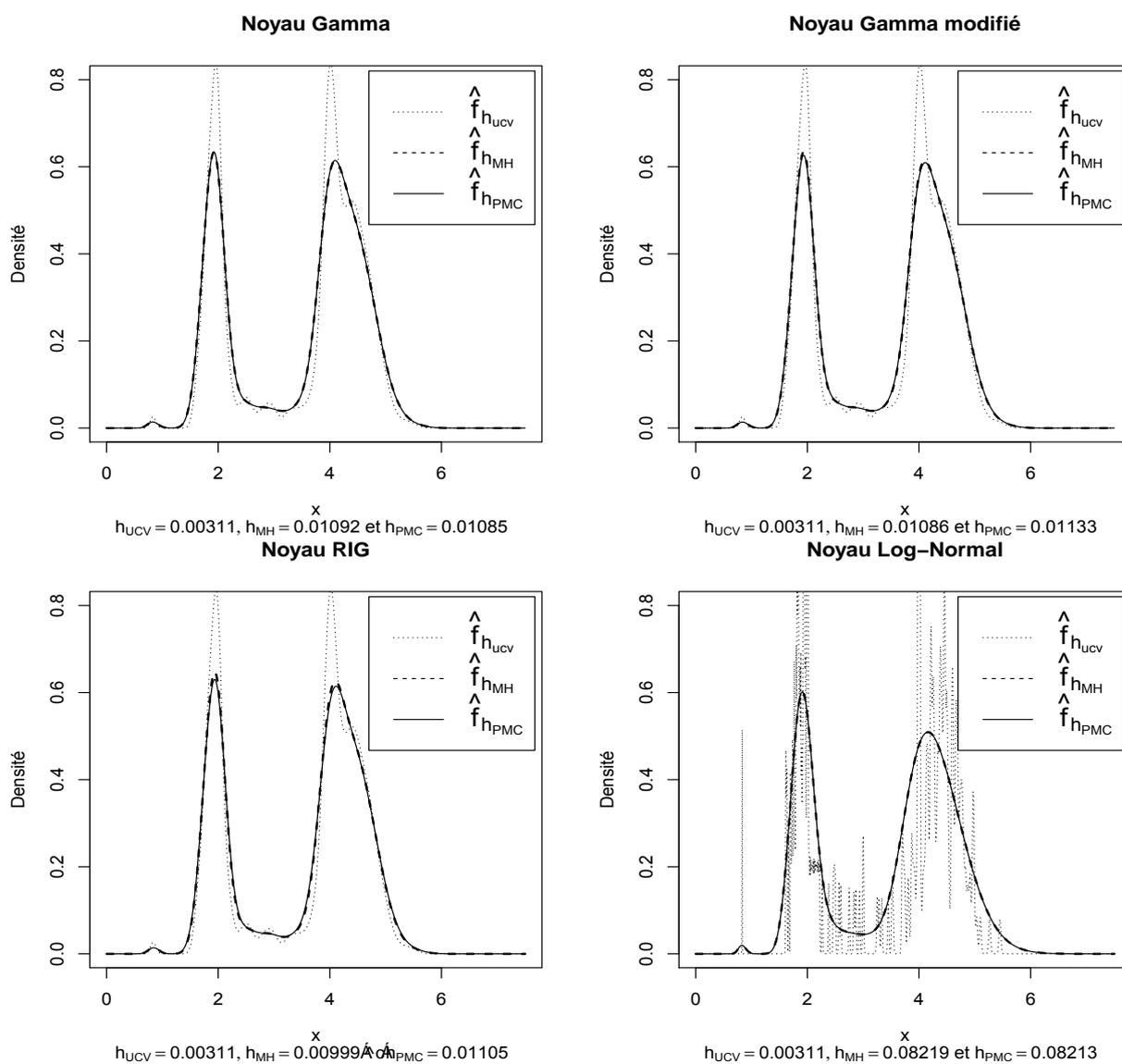


FIGURE 4.2: Courbes de la densité estimée de la durée d'éruption

La méthode UCV fournit un paramètre de lissage identique ($h_{UCV} = 0.00311$) pour tous les noyaux. Quant aux paramètres de lissage h_{MCMC} et h_{PMC} , donnés par la méthode Bayésienne, ils sont approximativement proches mais présentent d'importants écarts comparés à h_{UCV} .

La visualisation des courbes estimées données dans la figure (4.2) permet de remarquer que :

- Le paramètre de lissage h_{UCV} engendre le phénomène de sous-lissage des courbes de densités estimées et ceci quelque soit le noyau. Par contre, les courbes lissées en utilisant les paramètres de lissage h_{MCMC} et h_{PMC} sont suffisamment lisses quelque soit le noyau utilisé.
- Les courbes estimées présentent deux modes. Leurs niveaux sont différents dans le cas du noyau Log-Normal et sont approximativement proches pour les autres noyaux.
- Les estimateurs à noyau Gamma, Gamma modifié et RIG fournissent des estimations similaires de la densité des durées d'éruption, soit par le paramètre de lissage h_{UCV} ou par les paramètres sélectionnés par l'approche Bayésienne.

4.8 Conclusion

Dans ce chapitre, il a été présenté l'étude numérique basée sur des données simulées et des données réelles. L'étude est motivée par l'illustration des résultats théoriques sur l'apport du choix adéquat du noyau et du paramètre de lissage dans l'estimation de densité de probabilité de données positives.

Dans cette étude de simulation, on a observé l'impact sur la qualité des estimations, d'une part, de la variation des estimateurs par le biais du noyau et des méthodes de sélection du paramètre de lissage et d'autre part, des données (simulées ou réelles) par leurs types (variation des modèles pour les données simulées) ainsi que la taille des échantillons.

Il est clair que le type de données influe sur le choix du noyau. Cela est constaté lors d'utilisation des noyaux RIG et log-Normal pour des données dont la plupart sont concentrées au voisinage de $x = 0$. Par contre les noyaux Gamma et Gamma modifié sont suffisamment flexibles à l'estimation de la densité de données positives.

Les deux algorithmes MCMC et PMC convergent tous les deux vers des quantités très proches, mis à part quelques situations où l'algorithme PMC diverge. Cette divergence est causée par le choix inapproprié des particules initiales ou par le nombre insuffisant d'itérations. Si les particules initiales sont bien choisies, l'algorithme PMC est meilleur de l'algorithme MCMC du point de vue temps d'exécution.

La méthode de validation croisée présente certaines limites. Les paramètres fournis par cette méthode ont tendance à reproduire les données, ce qui cause, généralement, le phénomène de sous-lissage. Par contre, l'approche Bayésienne (MCMC et PMC) est compétitive au lissage de la densité estimée. Cependant, les paramètres donnés par cette approche ont parfois tendance au sur-lissage des densités estimées.

Les algorithmes de sélection basés sur la méthode de validation croisée ou l'approche Bayésienne ont parfois tendance à diverger, en fournissant des paramètres non appropriés. Cela relève, d'une part, du défaut de la méthode validation croisée (la fonction score $UCV(h)$, à minimiser, présente plusieurs minima locaux) et d'autre part, du choix inapproprié des paramètres d'implémentation dont les hyper-paramètres de l'a priori dans l'approche Bayésienne.

f	n	$Noyau$	h^*	$(\bar{h}_{ucv}, \sigma_h)$	$(\bar{h}_{mcmc}, \sigma_h)$	$(\bar{h}_{pmc}, \sigma_h)$	$MISE^*$	$(\overline{ISE}_{ucv}, \sigma_{ISE})$	$(\overline{ISE}_{mcmc}, \sigma_{ISE})$	$(\overline{ISE}_{pmc}, \sigma_{ISE})$
D1	25	Gam	0.12327	(0.29719, 0.15411)	(0.19744, 0.18060)	(0.18563, 0.18131)	0.04552	(0.04564, 0.02291)	(0.03971, 0.02186)	(0.04037, 0.02287)
		Gam^m	0.11512	(0.43765, 0.24695)	(0.28209, 0.19316)	(0.27000, 0.19917)	0.06216	(0.05003, 0.02542)	(0.04266, 0.02326)	(0.04248, 0.02385)
		RIG	0.11512	(0.20657, 0.11859)	(0.20358, 0.17851)	(0.04047, 0.00298)	0.04711	(0.05001, 0.03847)	(0.04915, 0.03746)	(0.05118, 0.03357)
		LN		(0.52330, 0.14008)	(0.35284, 0.10110)	(0.35030, 0.10115)		(0.04964, 0.03728)	(0.04515, 0.03652)	(0.04542, 0.03619)
	50	Gam	0.09342	(0.14672, 0.10300)	(0.18664, 0.21159)	(0.18126, 0.21178)	0.02615	(0.02710, 0.01330)	(0.02585, 0.01339)	(0.02601, 0.01364)
		Gam^m	0.08724	(0.19452, 0.17450)	(0.26011, 0.22575)	(0.25373, 0.22697)	0.03570	(0.02903, 0.01343)	(0.02895, 0.01373)	(0.02891, 0.01381)
		RIG	0.08724	(0.11326, 0.05543)	(0.22579, 0.34735)	(0.02099, 0.00185)	0.02706	(0.02481, 0.01405)	(0.03876, 0.05501)	(0.03649, 0.01475)
		LN		(0.41832, 0.10525)	(0.28613, 0.08395)	(0.28444, 0.08432)		(0.02590, 0.02042)	(0.02956, 0.02227)	(0.02974, 0.02247)
	100	Gam	0.07080	(0.07435, 0.05089)	(0.15332, 0.14775)	(0.15127, 0.14835)	0.01502	(0.01910, 0.01499)	(0.01935, 0.00986)	(0.01936, 0.00987)
		Gam^m	0.06612	(0.08633, 0.06837)	(0.22187, 0.15902)	(0.21917, 0.15979)	0.02050	(0.01904, 0.01581)	(0.02282, 0.01057)	(0.02269, 0.01061)
		RIG	0.06612	(0.07379, 0.03855)	(0.18699, 0.25803)	(0.01044, 0.00144)	0.01554	(0.01858, 0.01676)	(0.02825, 0.04315)	(0.02926, 0.01378)
		LN		(0.35447, 0.08997)	(0.25612, 0.06474)	(0.25564, 0.06511)		(0.01767, 0.01453)	(0.01748, 0.01138)	(0.01750, 0.01136)
200	Gam	0.05366	(0.04912, 0.02276)	(0.09521, 0.08604)	(0.09401, 0.08647)	0.00862	(0.01024, 0.00598)	(0.01218, 0.00718)	(0.01218, 0.00720)	
	Gam^m	0.05011	(0.05622, 0.02621)	(0.13810, 0.11143)	(0.13693, 0.11210)	0.01178	(0.00872, 0.00536)	(0.01378, 0.00851)	(0.01373, 0.00854)	
	RIG	0.05011	(0.05703, 0.02260)	(0.11064, 0.11260)	(0.00518, 0.00099)	0.00893	(0.00835, 0.00523)	(0.01441, 0.01901)	(0.01978, 0.00665)	
	LN		(0.30487, 0.07523)	(0.19494, 0.04983)	(0.19440, 0.05050)		(0.00951, 0.00712)	(0.01081, 0.00699)	(0.01091, 0.00715)	
400	Gam	0.04066	(0.02971, 0.01222)	(0.09321, 0.14937)	(0.09283, 0.14950)	0.00495	(0.00659, 0.00442)	(0.00953, 0.01009)	(0.00953, 0.01008)	
	Gam^m	0.03797	(0.03584, 0.01350)	(0.11631, 0.14450)	(0.11600, 0.14476)	0.00676	(0.00530, 0.00459)	(0.00933, 0.00882)	(0.00933, 0.00883)	
	RIG	0.03797	(0.03821, 0.01522)	(0.12241, 0.23440)	(0.00286, 0.00068)	0.00513	(0.00515, 0.00442)	(0.01550, 0.04074)	(0.01423, 0.00485)	
	LN		(0.26373, 0.06583)	(0.17275, 0.03362)	(0.17250, 0.03371)		(0.00638, 0.00539)	(0.00614, 0.00281)	(0.00615, 0.00282)	

TABLE 4.2: Résultats de simulation pour la sélection du paramètre de lissage par les algorithmes : UCV, MCMC et PMC, cas du modèle **D1**.

f	n	Noyau	h^*	$(\bar{h}_{ucv}, \sigma_h)$	$(\bar{h}_{mcmc}, \sigma_h)$	$(\bar{h}_{pmc}, \sigma_h)$	$MISE^*$	$(\overline{ISE}_{ucv}, \sigma_{ISE})$	$(\overline{ISE}_{mcmc}, \sigma_{ISE})$	$(\overline{ISE}_{pmc}, \sigma_{ISE})$
D2	25	Gam	0.26841	(0.77416, 0.32130)	(0.84078, 0.64353)	(0.82806, 0.64196)	0.01135	(0.00730, 0.00798)	(0.00576, 0.00436)	(0.00586, 0.00446)
		Gam^m	0.85061	(0.75872, 0.30414)	(0.89648, 0.67856)	(0.82806, 0.64196)	0.00841	(0.00658, 0.00643)	(0.00548, 0.00363)	(0.00586, 0.00446)
		RIG	0.85061	(0.74199, 0.31767)	(0.97739, 0.85979)	(0.77551, 0.13072)	0.00638	(0.00663, 0.00683)	(0.00572, 0.00378)	(0.00421, 0.00265)
		LN		(0.47679, 0.16220)	(0.36292, 0.08274)	(0.36027, 0.08467)		(0.00838, 0.01020)	(0.00660, 0.00654)	(0.00667, 0.00655)
	50	Gam	0.20342	(0.67818, 0.32080)	(0.57904, 0.89670)	(0.57416, 0.89938)	0.00652	(0.00360, 0.00283)	(0.00357, 0.00225)	(0.00364, 0.00239)
		Gam^m	0.64464	(0.66538, 0.31443)	(0.59531, 0.86063)	(0.57416, 0.89938)	0.00483	(0.00368, 0.00287)	(0.00363, 0.00252)	(0.00364, 0.00239)
		RIG	0.64464	(0.66535, 0.31357)	(0.51641, 0.45134)	(0.71872, 0.14977)	0.00366	(0.00363, 0.00297)	(0.00367, 0.00267)	(0.00252, 0.00200)
		LN		(0.40417, 0.11022)	(0.28916, 0.07975)	(0.28748, 0.08150)		(0.00382, 0.00385)	(0.00355, 0.00269)	(0.00361, 0.00276)
	100	Gam	0.15416	(0.55049, 0.27387)	(0.44225, 0.68768)	(0.44208, 0.68970)	0.00374	(0.00221, 0.00142)	(0.00212, 0.00108)	(0.00212, 0.00107)
		Gam^m	0.48855	(0.56302, 0.27393)	(0.45037, 0.66369)	(0.44208, 0.68970)	0.00277	(0.00225, 0.00140)	(0.00212, 0.00124)	(0.00212, 0.00107)
		RIG	0.48855	(0.57876, 0.26918)	(0.54752, 1.10102)	(0.72330, 0.13121)	0.00210	(0.00221, 0.00148)	(0.00265, 0.00388)	(0.00176, 0.00106)
		LN		(0.35238, 0.08519)	(0.23799, 0.05083)	(0.23727, 0.05061)		(0.00202, 0.00154)	(0.00220, 0.00130)	(0.00220, 0.00128)
200	Gam	0.11683	(0.33808, 0.15862)	(0.36282, 0.51159)	(0.36370, 0.51371)	0.00215	(0.00145, 0.00095)	(0.00148, 0.00077)	(0.00149, 0.00077)	
	Gam^m	0.37025	(0.37344, 0.17960)	(0.36421, 0.49451)	(0.36370, 0.51371)	0.00159	(0.00132, 0.00102)	(0.00135, 0.00085)	(0.00149, 0.00077)	
	RIG	0.37025	(0.40910, 0.17688)	(0.44246, 0.77721)	(0.69896, 0.11496)	0.00121	(0.00127, 0.00108)	(0.00169, 0.00246)	(0.00118, 0.00079)	
	LN		(0.29240, 0.07035)	(0.19134, 0.03921)	(0.19116, 0.03935)		(0.00139, 0.00127)	(0.00155, 0.00082)	(0.00155, 0.00082)	
400	Gam	0.08854	(0.23838, 0.09512)	(0.29108, 0.37339)	(0.29251, 0.37535)	0.00123	(0.00082, 0.00047)	(0.00097, 0.00052)	(0.00098, 0.00053)	
	Gam^m	0.28060	(0.28399, 0.10646)	(0.29025, 0.36461)	(0.29251, 0.37535)	0.00092	(0.00073, 0.00049)	(0.00094, 0.00057)	(0.00098, 0.00053)	
	RIG	0.28060	(0.30843, 0.11224)	(0.34369, 0.48403)	(0.69718, 0.13946)	0.00069	(0.00068, 0.00045)	(0.00111, 0.00104)	(0.00103, 0.00053)	
	LN		(0.26538, 0.05676)	(0.16153, 0.04273)	(0.16151, 0.04290)		(0.00073, 0.00043)	(0.00090, 0.00056)	(0.00090, 0.00056)	

TABLE 4.3: Résultats de simulation pour la sélection du paramètre de lissage par les algorithmes : UCV, MCMC et PMC, cas du modèle D2.

f	n	Noyau	h^*	$(\bar{h}_{ucv}, \sigma_h)$	$(\bar{h}_{mcmc}, \sigma_h)$	$(\bar{h}_{pmc}, \sigma_h)$	MISE*	$(\overline{ISE}_{ucv}, \sigma_{ISE})$	$(\overline{ISE}_{mcmc}, \sigma_{ISE})$	$(\overline{ISE}_{pmc}, \sigma_{ISE})$
D3	25	Gam	0.37311	(0.23089, 0.10572)	(0.08846, 0.03057)	(0.07728, 0.03048)	0.00993	(0.04391, 0.05417)	(0.04393, 0.03939)	(0.04915, 0.04213)
		Gam^m	0.47715	(0.31966, 0.13931)	(0.16412, 0.04868)	(0.15236, 0.04802)	0.01159	(0.04026, 0.03774)	(0.03825, 0.03815)	(0.04056, 0.04040)
		RIG	0.47715	(0.19013, 0.10422)	(0.09187, 0.03516)	(0.02225, 0.00410)	0.00878	(0.08626, 0.04439)	(0.07514, 0.04401)	(0.10889, 0.06471)
		LN		(0.58973, 0.23314)	(0.43200, 0.16620)	(0.58973, 0.23314)		(0.16081, 0.21487)	(0.12977, 0.11703)	(0.16081, 0.21487)
	50	Gam	0.28277	(0.17351, 0.07963)	(0.05600, 0.02001)	(0.05013, 0.01922)	0.00570	(0.03009, 0.02845)	(0.03173, 0.02412)	(0.03444, 0.02607)
		Gam^m	0.36161	(0.25034, 0.11188)	(0.11829, 0.03574)	(0.11101, 0.03536)	0.00665	(0.02729, 0.02488)	(0.02287, 0.01895)	(0.02420, 0.01988)
		RIG	0.36161	(0.14355, 0.08930)	(0.05504, 0.02344)	(0.01239, 0.00274)	0.00504	(0.06500, 0.02948)	(0.04973, 0.02429)	(0.08058, 0.03816)
		LN		(0.51273, 0.16586)	(0.36035, 0.14066)	(0.51273, 0.16586)		(0.08167, 0.05025)	(0.09573, 0.07055)	(0.08167, 0.05025)
	100	Gam	0.21430	(0.11428, 0.05270)	(0.02952, 0.01121)	(0.02678, 0.01070)	0.00328	(0.01769, 0.02029)	(0.02525, 0.01804)	(0.02725, 0.01928)
		Gam^m	0.27405	(0.16985, 0.08089)	(0.07433, 0.02278)	(0.06963, 0.02272)	0.00382	(0.01722, 0.01987)	(0.01562, 0.01245)	(0.01643, 0.01284)
		RIG	0.27405	(0.08471, 0.06523)	(0.03075, 0.01232)	(0.00667, 0.00135)	0.00290	(0.04100, 0.02026)	(0.03407, 0.01599)	(0.05958, 0.02537)
		LN		(0.44350, 0.15606)	(0.28081, 0.10626)	(0.44350, 0.15606)		(0.06154, 0.06171)	(0.06885, 0.05091)	(0.06154, 0.06171)
200	Gam	0.16241	(0.09195, 0.03501)	(0.01901, 0.00759)	(0.01732, 0.00745)	0.00188	(0.00954, 0.00911)	(0.01605, 0.00973)	(0.01718, 0.01049)	
	Gam^m	0.20769	(0.13435, 0.05588)	(0.05451, 0.02084)	(0.05121, 0.02000)	0.00220	(0.00936, 0.00930)	(0.01036, 0.00664)	(0.01083, 0.00699)	
	RIG	0.20769	(0.05844, 0.04421)	(0.01982, 0.00892)	(0.00380, 0.00088)	0.00166	(0.02636, 0.01201)	(0.02201, 0.00927)	(0.03863, 0.01368)	
	LN		(0.36259, 0.10237)	(0.24879, 0.09426)	(0.36259, 0.10237)		(0.04015, 0.05426)	(0.04377, 0.03533)	(0.04015, 0.05426)	
400	Gam	0.12308	(0.07122, 0.02266)	(0.01186, 0.00406)	(0.01116, 0.00416)	0.00108	(0.00518, 0.00639)	(0.00916, 0.00404)	(0.00962, 0.00428)	
	Gam^m	0.15740	(0.10682, 0.03454)	(0.03428, 0.01293)	(0.03233, 0.01275)	0.00126	(0.00439, 0.00455)	(0.00548, 0.00354)	(0.00572, 0.00363)	
	RIG	0.15740	(0.04316, 0.03430)	(0.01163, 0.00536)	(0.00205, 0.00073)	0.00096	(0.01710, 0.00942)	(0.01373, 0.00615)	(0.02689, 0.00899)	
	LN		(0.30733, 0.08182)	(0.20658, 0.08161)	(0.30733, 0.08182)		(0.02248, 0.01984)	(0.02789, 0.01428)	(0.02248, 0.01984)	

TABLE 4.4: Résultats de simulation pour la sélection du paramètre de lissage par les algorithmes : UCV, MCMC et PMC, cas du modèle **D3**.

f	n	Noyau	h*	$(\bar{h}_{ucv}, \sigma_h)$	$(\bar{h}_{mcmc}, \sigma_h)$	$(\bar{h}_{pmc}, \sigma_h)$	MISE*	$(\overline{ISE}_{ucv}, \sigma_{ISE})$	$(\overline{ISE}_{mcmc}, \sigma_{ISE})$	$(\overline{ISE}_{pmc}, \sigma_{ISE})$
D4	25	Gam	0.38835	(0.24211, 0.10522)	(0.08654, 0.03457)	(0.08443, 0.03703)	0.04205	(0.07792, 0.07127)	(0.08262, 0.05232)	(0.08429, 0.05327)
		<i>Gam^m</i>	0.33134	(0.35073, 0.15927)	(0.17517, 0.05100)	(0.16262, 0.05060)	0.06007	(0.09813, 0.06012)	(0.08432, 0.05194)	(0.08628, 0.05389)
		RIG	0.33134	(0.22832, 0.15291)	(0.09784, 0.04685)	(0.02371, 0.00365)	0.04553	(0.19481, 0.06043)	(0.15420, 0.05497)	(0.17348, 0.07356)
		LN		(0.58973, 0.23314)	(0.39835, 0.11334)	(0.39733, 0.11369)		(0.16081, 0.21487)	(0.13565, 0.07018)	(0.13616, 0.07090)
	50	Gam	0.29432	(0.15798, 0.08154)	(0.04702, 0.01645)	(0.04560, 0.01755)	0.02415	(0.05505, 0.03178)	(0.07068, 0.03727)	(0.07213, 0.03809)
		<i>Gam^m</i>	0.25111	(0.24563, 0.10791)	(0.11453, 0.03545)	(0.10759, 0.03484)	0.03450	(0.06876, 0.03173)	(0.06583, 0.03671)	(0.06667, 0.03718)
		RIG	0.25111	(0.14649, 0.12368)	(0.05237, 0.02020)	(0.01259, 0.00263)	0.02615	(0.14912, 0.05120)	(0.11541, 0.03594)	(0.13480, 0.04504)
		LN		(0.51273, 0.16586)	(0.33543, 0.11858)	(0.33328, 0.11892)		(0.08167, 0.05025)	(0.12518, 0.07160)	(0.12624, 0.07259)
	100	Gam	0.22305	(0.12021, 0.05287)	(0.02802, 0.01006)	(0.02806, 0.01093)	0.01387	(0.04752, 0.03086)	(0.06436, 0.03005)	(0.06467, 0.03010)
		<i>Gam^m</i>	0.19031	(0.17619, 0.08442)	(0.07227, 0.02372)	(0.06701, 0.02274)	0.01982	(0.06049, 0.03152)	(0.05810, 0.02926)	(0.05890, 0.02983)
		RIG	0.19031	(0.09839, 0.07069)	(0.03232, 0.01304)	(0.00702, 0.00157)	0.01502	(0.11899, 0.04139)	(0.09505, 0.03080)	(0.11505, 0.06607)
		LN		(0.44350, 0.15606)	(0.30851, 0.10152)	(0.30892, 0.10141)		(0.06154, 0.06171)	(0.09409, 0.06436)	(0.09391, 0.06383)
200	Gam	0.16904	(0.09059, 0.03497)	(0.01791, 0.00627)	(0.01833, 0.00694)	0.00797	(0.04109, 0.01585)	(0.05506, 0.01787)	(0.05497, 0.01802)	
	<i>Gam^m</i>	0.14423	(0.14373, 0.05936)	(0.05199, 0.01840)	(0.04919, 0.01815)	0.01138	(0.05129, 0.01769)	(0.05109, 0.01934)	(0.05149, 0.01969)	
	RIG	0.14423	(0.06213, 0.05585)	(0.02044, 0.00882)	(0.00362, 0.00119)	0.00863	(0.09250, 0.03119)	(0.07546, 0.01960)	(0.08433, 0.01961)	
	LN		(0.36259, 0.10237)	(0.24373, 0.08561)	(0.24335, 0.08481)		(0.04015, 0.05426)	(0.07301, 0.02770)	(0.07308, 0.02790)	
400	Gam	0.12811	(0.06904, 0.02827)	(0.01167, 0.00410)	(0.01223, 0.00456)	0.00458	(0.04561, 0.01804)	(0.05655, 0.01660)	(0.05628, 0.01660)	
	<i>Gam^m</i>	0.10930	(0.10045, 0.04715)	(0.03492, 0.01147)	(0.03335, 0.01152)	0.00654	(0.05438, 0.01757)	(0.05378, 0.01668)	(0.05398, 0.01682)	
	RIG	0.10930	(0.03973, 0.02645)	(0.01306, 0.00553)	(0.00180, 0.00064)	0.00495	(0.08254, 0.02201)	(0.07005, 0.01647)	(0.07994, 0.02092)	
	LN		(0.30733, 0.08182)	(0.22065, 0.07658)	(0.22072, 0.07677)		(0.02248, 0.01984)	(0.06495, 0.02005)	(0.06497, 0.02012)	

TABLE 4.5: Résultats de simulation pour la sélection du paramètre de lissage par les algorithmes : UCV, MCMC et PMC, cas du modèle **D4**.

f	n	Noyau	h^*	$(\bar{h}_{ucv}, \sigma_h)$	$(\bar{h}_{mcmc}, \sigma_h)$	$(\bar{h}_{pmc}, \sigma_h)$	$MISE^*$	$(\overline{ISE}_{ucv}, \sigma_{ISE})$	$(\overline{ISE}_{mcmc}, \sigma_{ISE})$	$(\overline{ISE}_{pmc}, \sigma_{ISE})$
25		Gam	0.02116	(0.02801, 0.01756)	(0.01639, 0.00577)	(0.01371, 0.00523)	0.12072	(0.35146, 0.25137)	(0.27255, 0.23217)	(0.30866, 0.27246)
		Gam^m	0.02052	(0.03976, 0.02594)	(0.03871, 0.01417)	(0.03532, 0.01356)	0.16174	(0.35923, 0.24482)	(0.27400, 0.15701)	(0.27359, 0.16332)
		RIG	0.02052	(0.02908, 0.02181)	(0.01609 0.00639)	(0.01284, 0.00733)	0.12258	(0.45900, 0.28346)	(0.40732 0.40728)	(0.48460, 0.48785)
		LN		(0.44693, 0.18895)	(0.36723, 0.14096)	(0.36568, 0.14235)		(0.77678, 0.67123)	(0.73932, 0.89151)	(0.74182, 0.88436)
50		Gam	0.01604	(0.01966, 0.01061)	(0.01107, 0.00355)	(0.00933, 0.00342)	0.06934	(0.22487, 0.18123)	(0.16856, 0.11910)	(0.18702, 0.13371)
		Gam^m	0.01555	(0.02643, 0.01619)	(0.03180, 0.01109)	(0.02934, 0.01088)	0.09290	(0.20805, 0.16865)	(0.16659, 0.12152)	(0.16302, 0.11995)
		RIG	0.01555	(0.02070, 0.01483)	(0.01096 0.00412)	(0.00940, 0.00499)	0.07040	(0.26970, 0.19772)	(0.21525 0.13986)	(0.25555, 0.18361)
		LN		(0.38761, 0.11746)	(0.32966, 0.14721)	(0.32886, 0.14783)		(0.46152, 0.29419)	(0.46954, 0.32091)	(0.47199, 0.32429)
D5		Gam	0.01215	(0.01304, 0.00700)	(0.00697, 0.00189)	(0.00590, 0.00184)	0.03982	(0.11840, 0.07307)	(0.10682, 0.06958)	(0.11779, 0.07517)
		Gam^m	0.01179	(0.01664, 0.00973)	(0.02170, 0.00877)	(0.02042, 0.00861)	0.05336	(0.12652, 0.08151)	(0.10851, 0.06914)	(0.10792, 0.07088)
		RIG	0.01179	(0.01497, 0.01055)	(0.00727 0.00252)	(0.00640, 0.00312)	0.04044	(0.16685, 0.10126)	(0.13702 0.07015)	(0.15714, 0.09195)
		LN		(0.31984, 0.09129)	(0.29352, 0.11844)	(0.29396, 0.11846)		(0.27435, 0.14373)	(0.30011, 0.16390)	(0.29979, 0.16377)
200		Gam	0.00921	(0.00986, 0.00397)	(0.00454, 0.00122)	(0.00392, 0.00117)	0.02287	(0.06227, 0.03453)	(0.06594, 0.03418)	(0.07090, 0.03527)
		Gam^m	0.00893	(0.01143, 0.00522)	(0.01635, 0.00637)	(0.01545, 0.00624)	0.03064	(0.06682, 0.03756)	(0.06570, 0.03544)	(0.06482, 0.03555)
		RIG	0.00893	(0.01005, 0.00546)	(0.00465 0.00146)	(0.00409, 0.00183)	0.02322	(0.09824, 0.03079)	(0.08620 0.03333)	(0.09419, 0.03991)
		LN		(0.28281, 0.06854)	(0.25210, 0.10069)	(0.25214, 0.10097)		(0.16069, 0.07810)	(0.17571, 0.08214)	(0.17574, 0.08233)
400		Gam	0.00698	(0.00669, 0.00252)	(0.00284, 0.00095)	(0.00252, 0.00089)	0.01314	(0.03654, 0.02207)	(0.04467, 0.02442)	(0.04768, 0.02526)
		Gam^m	0.00677	(0.00789, 0.00315)	(0.01171, 0.00502)	(0.01102, 0.00492)	0.01760	(0.03735, 0.02125)	(0.03930, 0.02216)	(0.03872, 0.02195)
		RIG	0.00677	(0.00613, 0.00376)	(0.00281 0.00103)	(0.00273, 0.00124)	0.01334	(0.05774, 0.02263)	(0.05708 0.02444)	(0.05970, 0.02770)
		LN		(0.22967, 0.05481)	(0.20239, 0.10100)	(0.20237, 0.10106)		(0.10274, 0.05056)	(0.12784, 0.08240)	(0.12797, 0.08282)

TABLE 4.6: Résultats de simulation pour la sélection du paramètre de lissage par les algorithmes : UCV, MCMC et PMC, cas du modèle **D5**.

f	n	Noyau	h^*	$(\bar{h}_{ucv}, \sigma_h)$	$(\bar{h}_{mcmc}, \sigma_h)$	$(\bar{h}_{pmc}, \sigma_h)$	MISE*	$(\overline{ISE}_{ucv}, \sigma_{ISE})$	$(\overline{ISE}_{mcmc}, \sigma_{ISE})$	$(\overline{ISE}_{pmc}, \sigma_{ISE})$
D6	25	Gam	0.00234	(0.00500, 0.00205)	(0.00496, 0.00224)	(0.00416, 0.00196)	0.10323	(0.17120, 0.09243)	(0.20971, 0.18479)	(0.19264, 0.14605)
		Gam^m	0.00234	(0.00500, 0.00205)	(0.00507, 0.00228)	(0.00466, 0.00206)	0.13621	(0.16760, 0.08914)	(0.20555, 0.17605)	(0.19141, 0.15617)
		RIG	0.00234	(0.00500, 0.00205)	(0.00493, 0.00219)	(0.00324, 0.00183)	0.10323	(0.16763, 0.08931)	(0.20575, 0.18373)	(0.20946, 0.18748)
		LN		(0.06051, 0.02427)	(0.05862, 0.01718)	(0.05640, 0.01766)		(0.49056, 1.50495)	(0.21638, 0.21494)	(0.22238, 0.22429)
	50	Gam	0.00178	(0.00357, 0.00083)	(0.00350, 0.00114)	(0.00308, 0.00111)	0.05929	(0.09061, 0.06457)	(0.09799, 0.06257)	(0.09653, 0.06433)
		Gam^m	0.00178	(0.00357, 0.00083)	(0.00347, 0.00116)	(0.00331, 0.00114)	0.07823	(0.09183, 0.06335)	(0.09938, 0.06117)	(0.09948, 0.06209)
		RIG	0.00178	(0.00357, 0.00083)	(0.00348, 0.00113)	(0.00271, 0.00110)	0.05929	(0.09213, 0.06344)	(0.09907, 0.06141)	(0.10372, 0.07335)
		LN		(0.05100, 0.01529)	(0.05228, 0.01010)	(0.05120, 0.01036)		(0.13725, 0.10944)	(0.09847, 0.06668)	(0.09895, 0.06700)
	100	Gam	0.00135	(0.00318, 0.00038)	(0.00240, 0.00087)	(0.00222, 0.00080)	0.03405	(0.07138, 0.05072)	(0.07105, 0.04812)	(0.07025, 0.04883)
		Gam^m	0.00135	(0.00318, 0.00038)	(0.00239, 0.00087)	(0.00230, 0.00088)	0.04493	(0.07039, 0.04907)	(0.06978, 0.04724)	(0.07076, 0.04704)
		RIG	0.00135	(0.00318, 0.00038)	(0.00237, 0.00086)	(0.00198, 0.00083)	0.03405	(0.07050, 0.04903)	(0.07011, 0.04716)	(0.07442, 0.05562)
		LN		(0.04406, 0.01328)	(0.04398, 0.01021)	(0.04348, 0.01015)		(0.09653, 0.10262)	(0.07238, 0.04771)	(0.07223, 0.04779)
200	Gam	0.00102	(0.00311, 0.00000)	(0.00165, 0.00043)	(0.00155, 0.00046)	0.01956	(0.04405, 0.02129)	(0.03175, 0.01793)	(0.03226, 0.01910)	
	Gam^m	0.00102	(0.00311, 0.00000)	(0.00164, 0.00044)	(0.00164, 0.00042)	0.02581	(0.04415, 0.02049)	(0.03203, 0.01723)	(0.03160, 0.01793)	
	RIG	0.00102	(0.00311, 0.00000)	(0.00165, 0.00043)	(0.00145, 0.00047)	0.01956	(0.04431, 0.02054)	(0.03189, 0.01697)	(0.03254, 0.01785)	
	LN		(0.03687, 0.00777)	(0.03695, 0.00642)	(0.03655, 0.00648)		(0.04064, 0.03539)	(0.03313, 0.01809)	(0.03335, 0.01818)	
400	Gam	0.00077	(0.00311, 0.00000)	(0.00122, 0.00039)	(0.00123, 0.00042)	0.01123	(0.03922, 0.01429)	(0.01968, 0.01281)	(0.02018, 0.01335)	
	Gam^m	0.00077	(0.00311, 0.00000)	(0.00121, 0.00037)	(0.00129, 0.00038)	0.01482	(0.03970, 0.01469)	(0.01989, 0.01333)	(0.01971, 0.01304)	
	RIG	0.00077	(0.00311, 0.00000)	(0.00121, 0.00038)	(0.00118, 0.00046)	0.01123	(0.03989, 0.01481)	(0.02001, 0.01333)	(0.02061, 0.01321)	
	LN		(0.03206, 0.00584)	(0.03219, 0.00638)	(0.03194, 0.00658)		(0.02209, 0.01345)	(0.02061, 0.01363)	(0.02081, 0.01378)	

TABLE 4.7: Résultats de simulation pour la sélection du paramètre de lissage par les algorithmes : UCV, MCMC et PMC, cas du modèle **D6**.

f	n	Noyau	h^*	$(\bar{h}_{ucv}, \sigma_h)$	$(\bar{h}_{mcmc}, \sigma_h)$	$(\bar{h}_{pmc}, \sigma_h)$	MISE*	$(\overline{ISE}_{ucv}, \sigma_{ISE})$	$(\overline{ISE}_{mcmc}, \sigma_{ISE})$	$(\overline{ISE}_{pmc}, \sigma_{ISE})$
D7	25	Gam	0.01083	(0.20959, 0.22113)	(0.21567, 0.24312)	(0.20447, 0.24161)	0.10011	(0.06426, 0.03800)	(0.05643, 0.02276)	(0.05707, 0.02422)
		Gam^m	0.01134	(0.34081, 0.32881)	(0.26934, 0.27903)	(0.26113, 0.27728)	0.12909	(0.06298, 0.02915)	(0.05898, 0.02326)	(0.05899, 0.02359)
		RIG	0.01134	(0.18788, 0.18682)	(0.24471, 0.34755)	(0.02351, 0.00413)	0.09783	(0.06488, 0.03404)	(0.06636, 0.03690)	(0.07191, 0.03482)
		LN		(0.38310, 0.18751)	(0.29914, 0.10166)	(0.29738, 0.10186)		(0.08026, 0.03119)	(0.07858, 0.03244)	(0.07882, 0.03262)
	50	Gam	0.00821	(0.09054, 0.09882)	(0.18020, 0.15666)	(0.17487, 0.15669)	0.05750	(0.03736, 0.01767)	(0.04057, 0.01505)	(0.04027, 0.01515)
		Gam^m	0.00860	(0.16837, 0.22834)	(0.22920, 0.19326)	(0.22565, 0.19486)	0.07414	(0.03987, 0.01847)	(0.04238, 0.01518)	(0.04216, 0.01530)
		RIG	0.00860	(0.08836, 0.09255)	(0.19699, 0.18227)	(0.01267, 0.00278)	0.05619	(0.03701, 0.01869)	(0.04598, 0.02593)	(0.04355, 0.01859)
		LN		(0.29736, 0.15236)	(0.25993, 0.07806)	(0.25911, 0.07865)		(0.06099, 0.03171)	(0.05016, 0.01464)	(0.05014, 0.01458)
	100	Gam	0.00622	(0.03271, 0.01733)	(0.12367, 0.09850)	(0.12127, 0.09772)	0.03302	(0.02432, 0.01371)	(0.03013, 0.01145)	(0.02987, 0.01149)
		Gam^m	0.00652	(0.03483, 0.01847)	(0.15552, 0.12474)	(0.15441, 0.12607)	0.04258	(0.02359, 0.01359)	(0.03174, 0.01310)	(0.03158, 0.01318)
		RIG	0.00652	(0.03611, 0.02027)	(0.13856, 0.12366)	(0.00665, 0.00193)	0.03227	(0.02348, 0.01328)	(0.03246, 0.01886)	(0.03494, 0.01758)
		LN		(0.18981, 0.08102)	(0.20496, 0.04642)	(0.20459, 0.04624)		(0.04023, 0.01355)	(0.03698, 0.01050)	(0.03699, 0.01051)
200	Gam	0.00472	(0.02357, 0.01002)	(0.12564, 0.22012)	(0.12446, 0.22036)	0.01897	(0.01447, 0.00504)	(0.02536, 0.01398)	(0.02517, 0.01404)	
	Gam^m	0.00494	(0.02463, 0.01076)	(0.14431, 0.21521)	(0.14385, 0.21545)	0.02446	(0.01413, 0.00478)	(0.02733, 0.01385)	(0.02726, 0.01388)	
	RIG	0.00494	(0.02498, 0.01181)	(0.15828, 0.34325)	(0.00476, 0.00383)	0.01854	(0.01422, 0.00490)	(0.03029, 0.02996)	(0.02211, 0.00840)	
	LN		(0.12958, 0.04148)	(0.16585, 0.05092)	(0.16574, 0.05115)		(0.02616, 0.00911)	(0.02651, 0.00663)	(0.02658, 0.00663)	
400	Gam	0.00357	(0.01407, 0.00523)	(0.07982, 0.13243)	(0.07967, 0.13262)	0.01089	(0.00879, 0.00446)	(0.01638, 0.01293)	(0.01635, 0.01294)	
	Gam^m	0.00374	(0.01470, 0.00573)	(0.08398, 0.12876)	(0.08396, 0.12913)	0.01405	(0.00869, 0.00456)	(0.01685, 0.01338)	(0.01684, 0.01338)	
	RIG	0.00374	(0.01467, 0.00602)	(0.10172, 0.20178)	(0.00266, 0.00184)	0.01065	(0.00878, 0.00455)	(0.02054, 0.02661)	(0.01520, 0.00662)	
	LN		(0.09601, 0.01844)	(0.12786, 0.03386)	(0.12807, 0.03392)		(0.01487, 0.00548)	(0.01659, 0.00538)	(0.01661, 0.00540)	

TABLE 4.8: Résultats de simulation pour la sélection du paramètre de lissage par les algorithmes : UCV, MCMC et PMC, cas du modèle **D7**.

n	Noyau	τ_{ac}, SIF									
		D1	D2	D3	D4	D5	D6	D7			
25	Gam	0.43774, 5.00725	0.43625, 4.27948	0.43203, 5.49637	0.42945, 5.42120	0.40291, 6.04239	0.36402, 6.56765	0.43420, 5.37467			
	Gam^m	0.44399, 5.06916	0.43985, 4.72455	0.44449, 5.72292	0.44495, 5.23959	0.43218, 5.67782	0.39556, 6.72336	0.43908, 5.72487			
	RIG	0.43947, 4.60716	0.43877, 4.66810	0.43055, 5.65979	0.43058, 5.62617	0.39984, 5.93202	0.36523, 6.50212	0.43156, 5.65562			
	LN	0.44725, 4.66927	0.44795, 4.33185	0.44416, 4.88348	0.44133, 4.79561	0.44116, 4.52277	0.44116, 4.52277	0.43857, 4.53529			
50	Gam	0.42840, 4.71853	0.42805, 4.80355	0.42463, 5.02608	0.42335, 5.54976	0.39399, 5.47575	0.35269, 5.64591	0.43178, 4.65842			
	Gam^m	0.44141, 4.98396	0.42639, 4.56180	0.44501, 5.20326	0.44533, 4.61616	0.43165, 5.48608	0.38421, 5.50712	0.43935, 4.98024			
	RIG	0.43185, 4.90063	0.43287, 4.81889	0.41976, 5.68349	0.42427, 5.84001	0.38641, 5.66253	0.35375, 5.23693	0.43177, 4.75431			
	LN	0.44053, 4.35612	0.44651, 4.50316	0.44243, 4.16383	0.44565, 4.52387	0.44279, 4.22484	0.44279, 4.22484	0.44045, 4.50662			
100	Gam	0.42741, 4.64027	0.41563, 4.71201	0.41183, 5.40881	0.41329, 4.85667	0.38294, 5.27883	0.32615, 5.08130	0.42347, 4.51077			
	Gam^m	0.43978, 4.71698	0.41909, 4.85344	0.44143, 5.09104	0.44396, 5.50520	0.43038, 5.23305	0.37378, 4.93078	0.43333, 4.88330			
	RIG	0.42949, 4.50881	0.42203, 4.86224	0.41317, 5.74656	0.41327, 5.29215	0.37567, 5.01641	0.32910, 5.21446	0.42462, 4.64421			
	LN	0.44613, 4.20248	0.44447, 4.19342	0.44139, 3.97312	0.44181, 4.32875	0.43955, 4.33699	0.43955, 4.33699	0.43687, 4.53325			
200	Gam	0.41511, 4.85389	0.39825, 5.52316	0.40435, 5.05252	0.40270, 4.50276	0.36757, 4.90375	0.31313, 5.04108	0.41290, 5.03267			
	Gam^m	0.43849, 4.59169	0.39831, 5.02698	0.44689, 4.60751	0.44003, 5.06508	0.42891, 5.07951	0.35807, 4.44106	0.41608, 4.41705			
	RIG	0.41831, 4.59191	0.39328, 5.33027	0.39615, 5.50733	0.39913, 4.81335	0.35683, 5.84815	0.31056, 5.00061	0.41833, 5.12762			
	LN	0.44053, 4.27397	0.44885, 4.05337	0.44541, 4.26241	0.45082, 4.03351	0.43917, 4.35439	0.43917, 4.35439	0.42964, 4.21321			
400	Gam	0.40351, 4.76602	0.36104, 5.58352	0.38601, 4.81317	0.38092, 4.83781	0.34221, 5.25426	0.28675, 5.21685	0.39111, 5.05685			
	Gam^m	0.43357, 5.05453	0.37459, 5.57508	0.43582, 4.94816	0.44183, 4.67116	0.42076, 4.87436	0.34533, 4.49156	0.39859, 4.76226			
	RIG	0.40542, 4.94442	0.36677, 5.74929	0.37927, 5.20741	0.38307, 5.44978	0.32925, 5.13336	0.28871, 4.79530	0.38921, 5.10896			
	LN	0.44001, 4.35766	0.44064, 4.22273	0.43010, 4.57920	0.44049, 4.39594	0.43389, 4.69103	0.43389, 4.69103	0.43301, 4.35657			

TABLE 4.9: Taux d'acceptation et indicateurs SIF moyens fournis par l'algorithme MCMC.

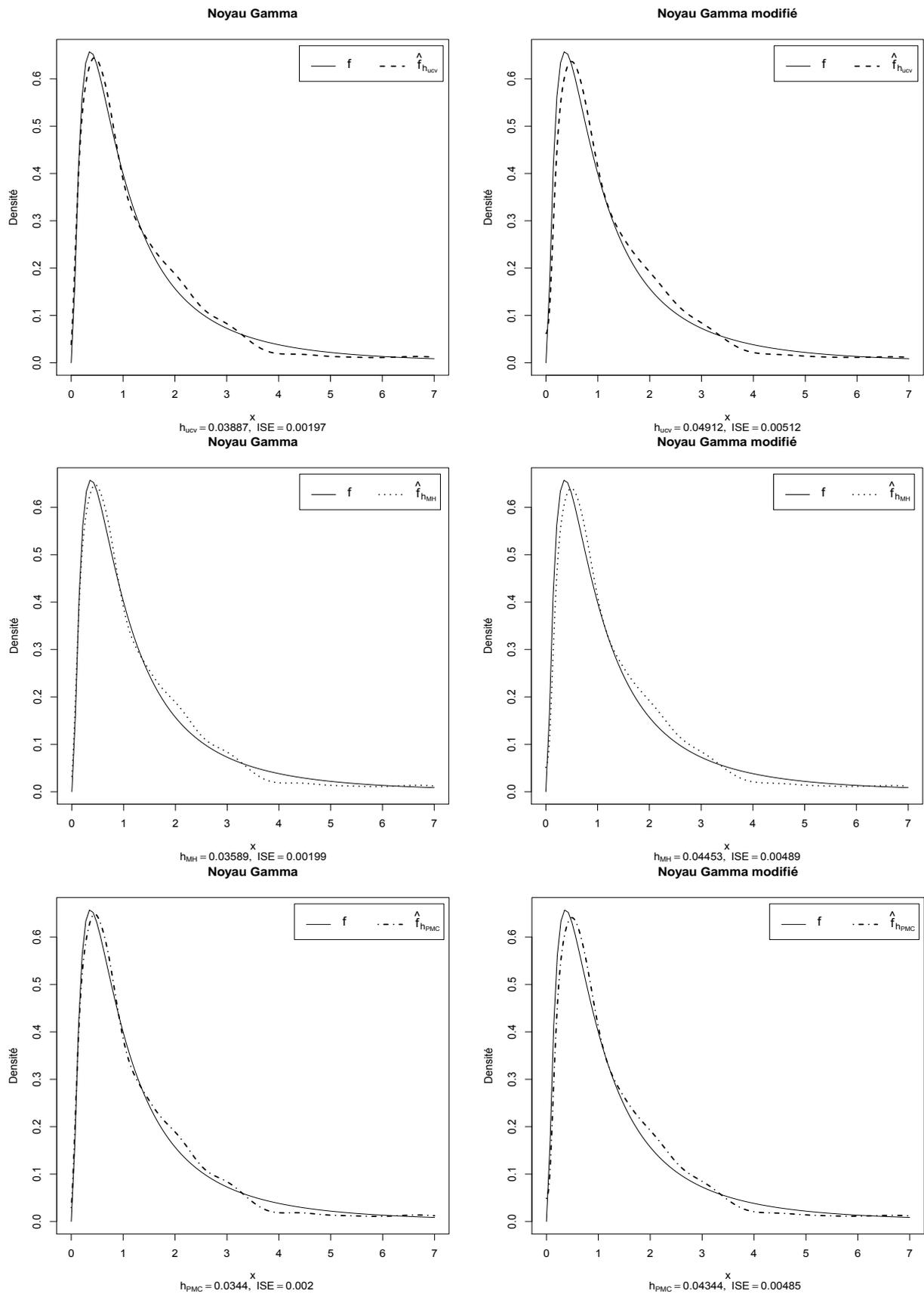


FIGURE 4.3: Estimateurs à noyaux Gamma et Gamma modifié d'une densité (D1)

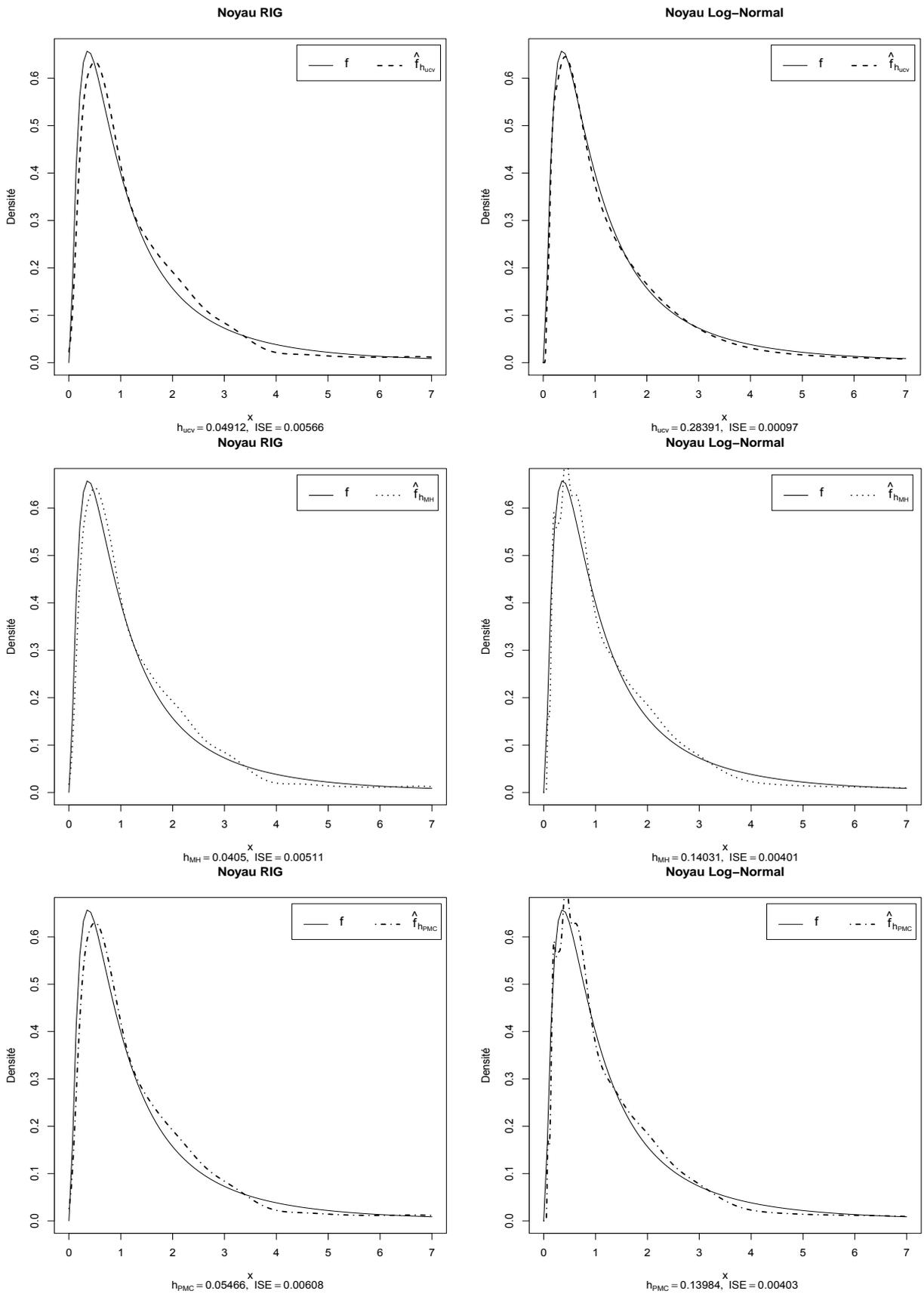


FIGURE 4.4: Estimateurs à noyaux RIG et log-Normal d'une densité (D1)

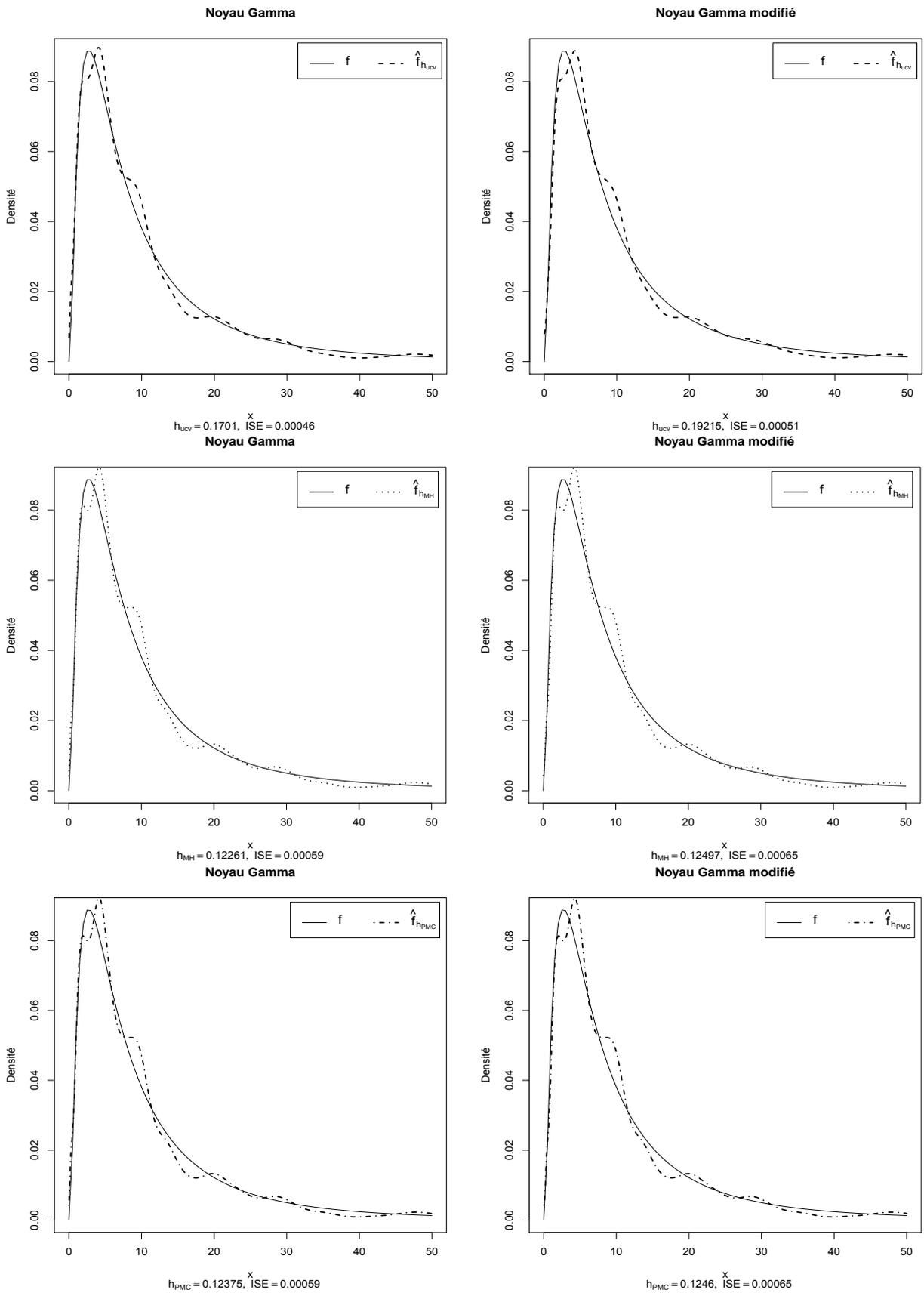


FIGURE 4.5: Estimateurs à noyaux Gamma et Gamma modifié d'une densité (D2)

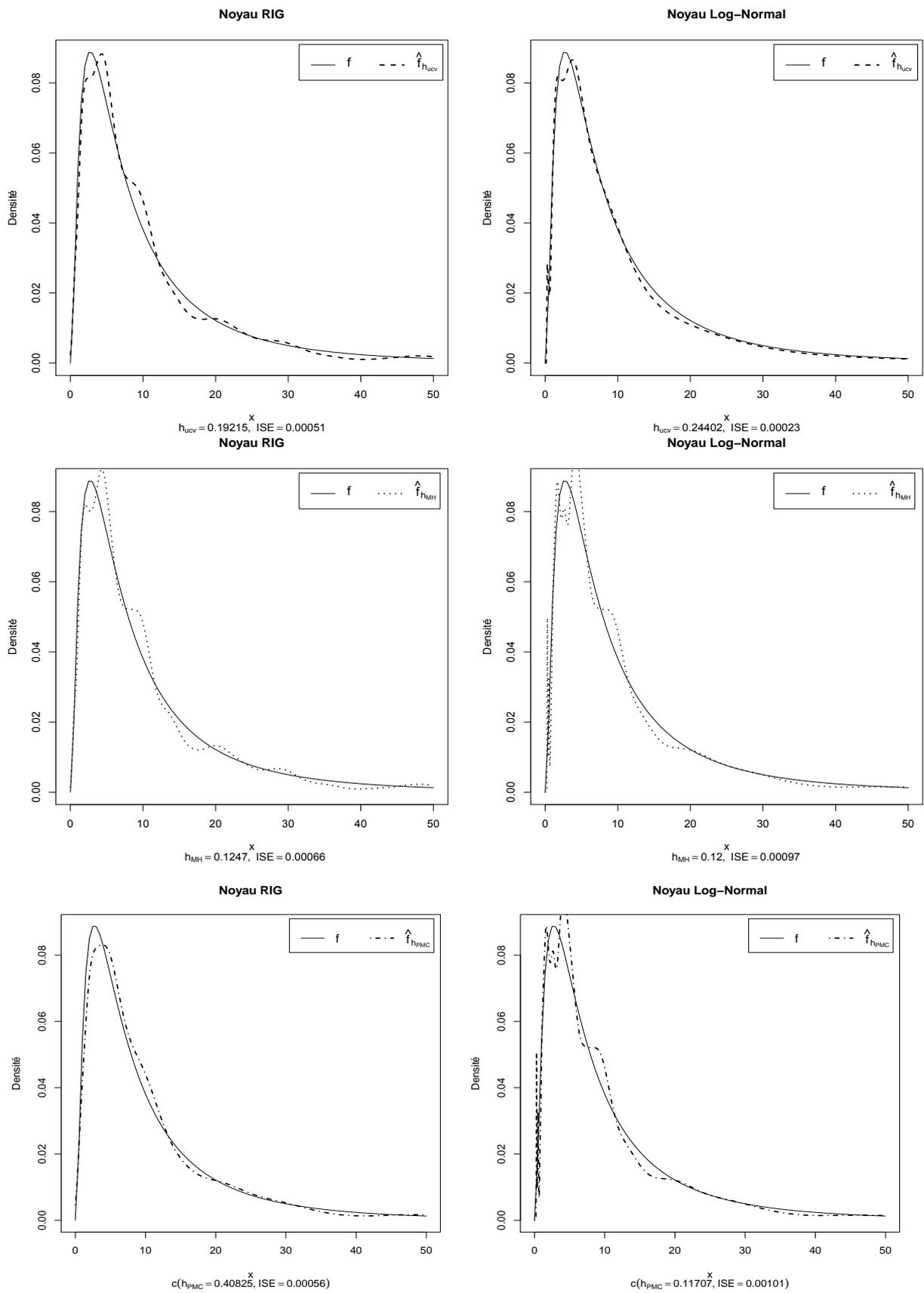


FIGURE 4.6: Estimateurs à noyaux RIG et log-Normal d'une densité (D2)

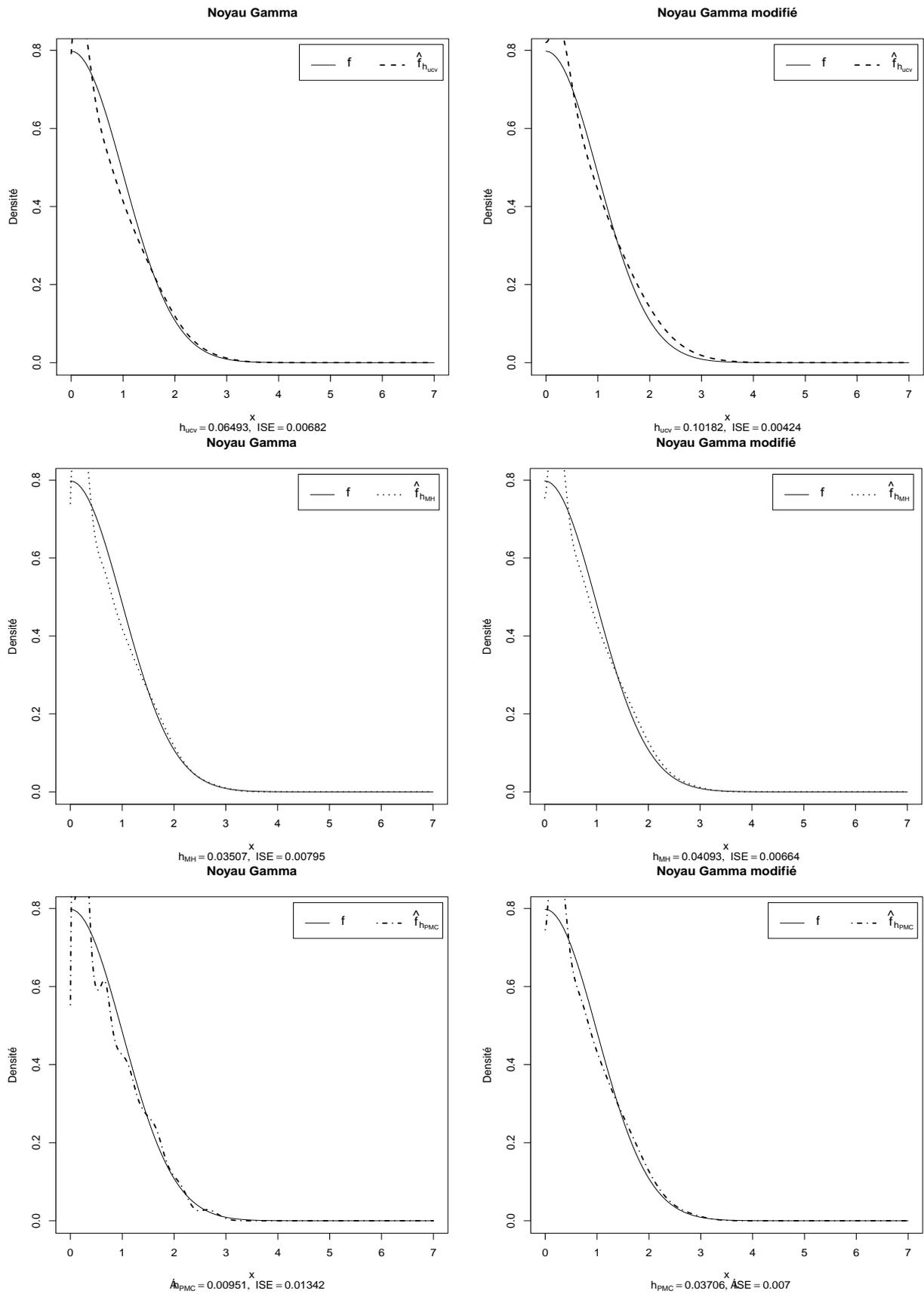


FIGURE 4.7: Estimateurs à noyaux Gamma et Gamma modifié d'une densité (D3)

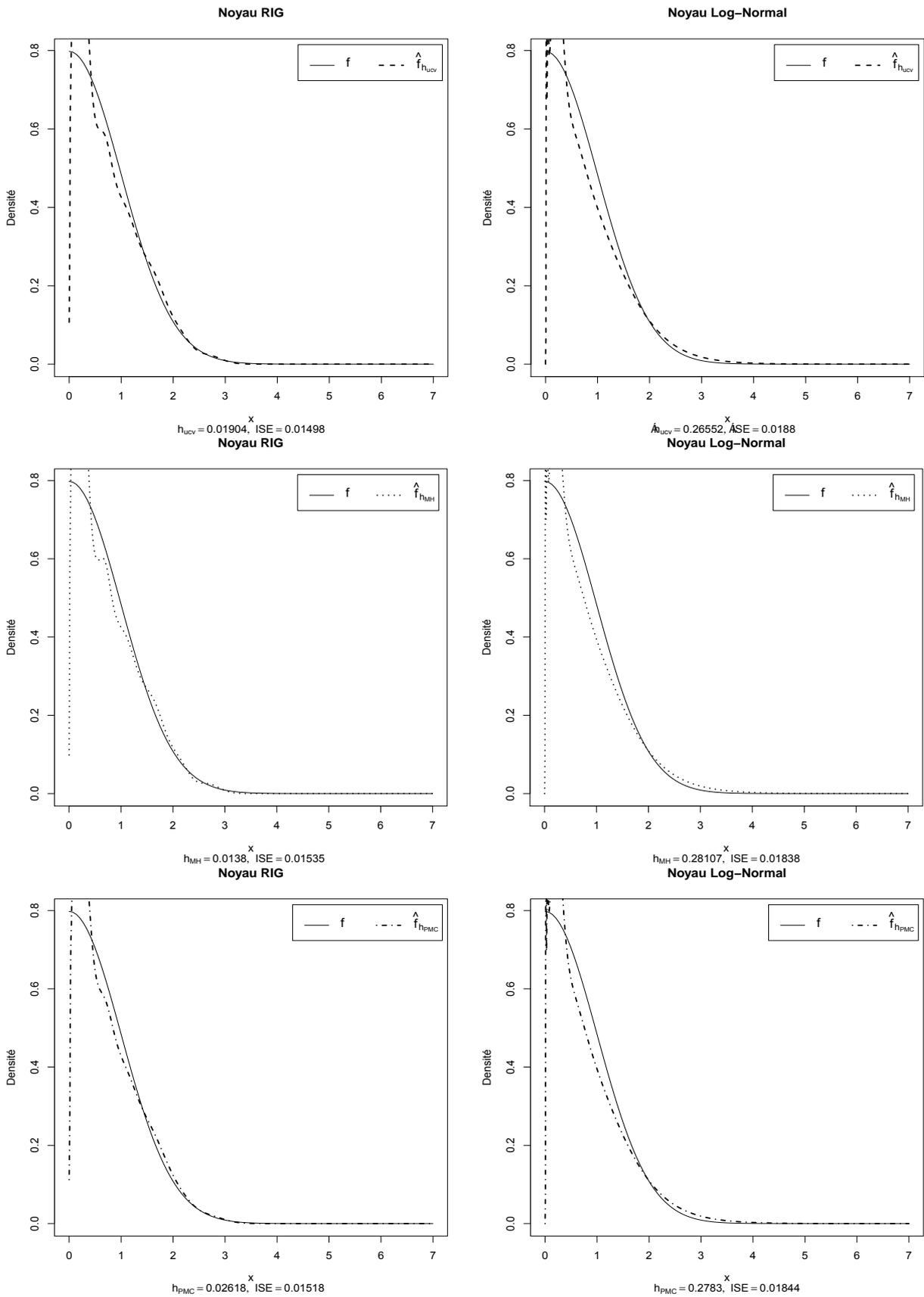


FIGURE 4.8: Estimateurs à noyaux RIG et log-Normal d'une densité (D3)

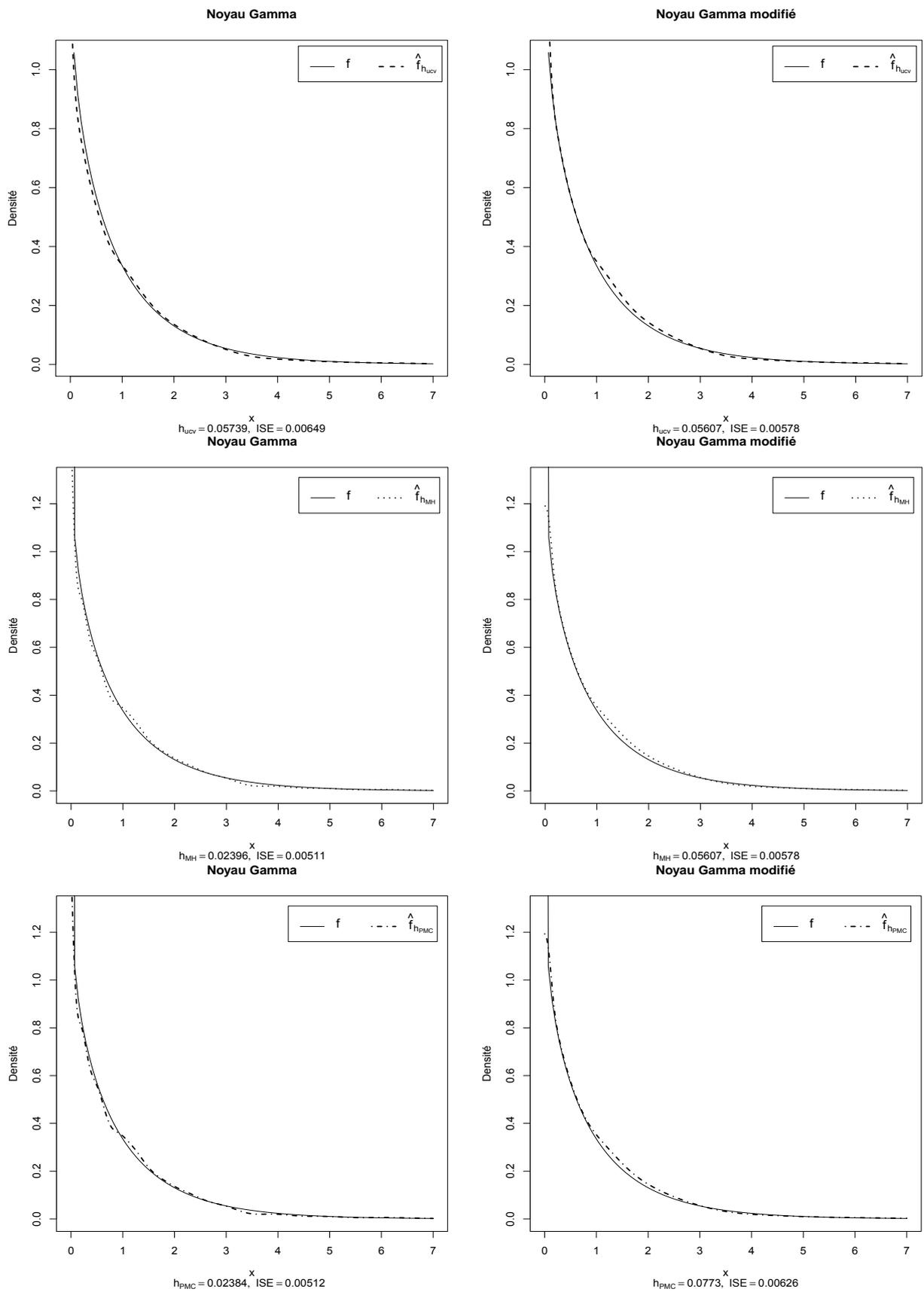


FIGURE 4.9: Estimateurs à noyaux Gamma et Gamma modifié d'une densité (D4)

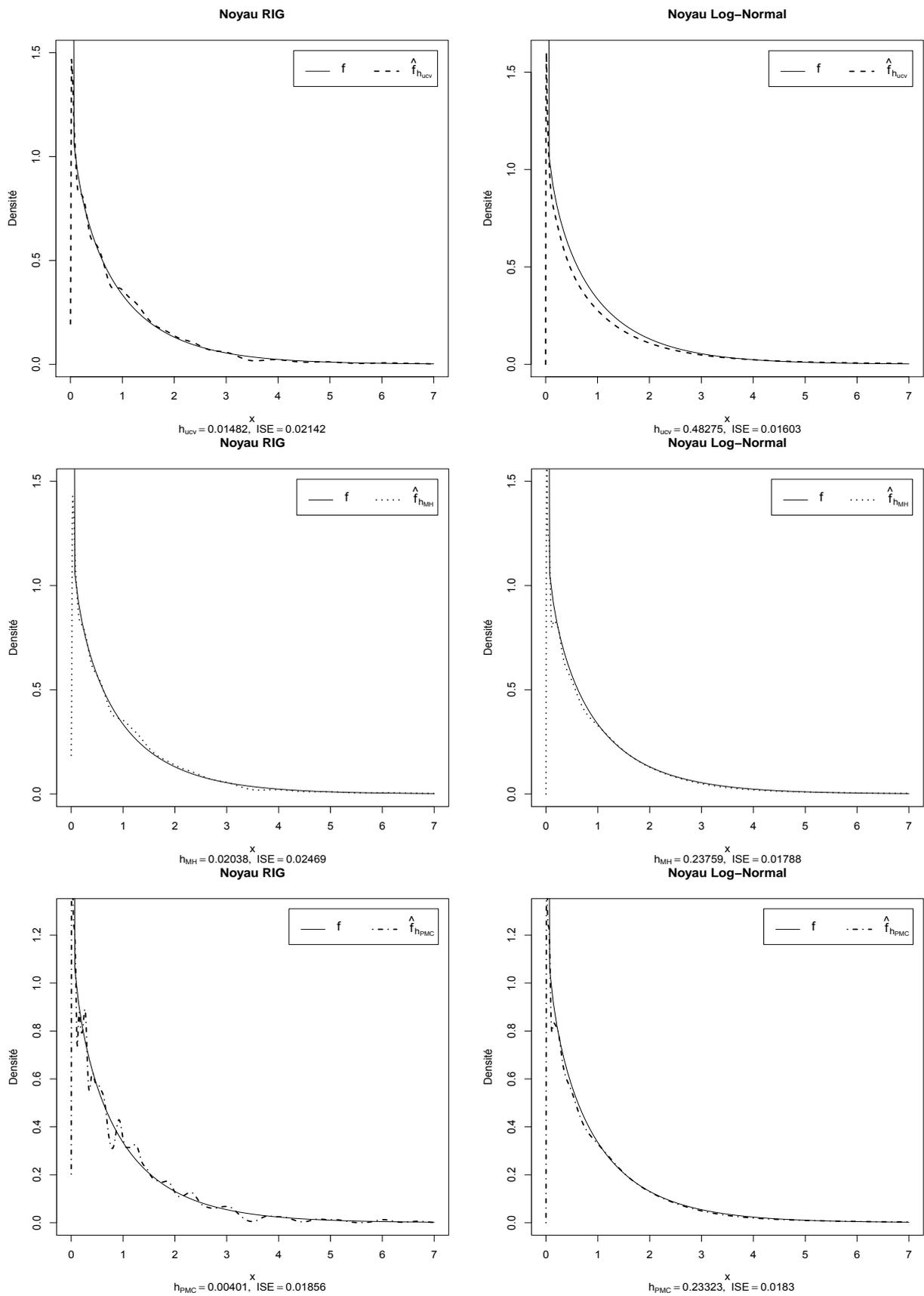


FIGURE 4.10: Estimateurs à noyaux RIG et log-Normal d'une densité (D4)

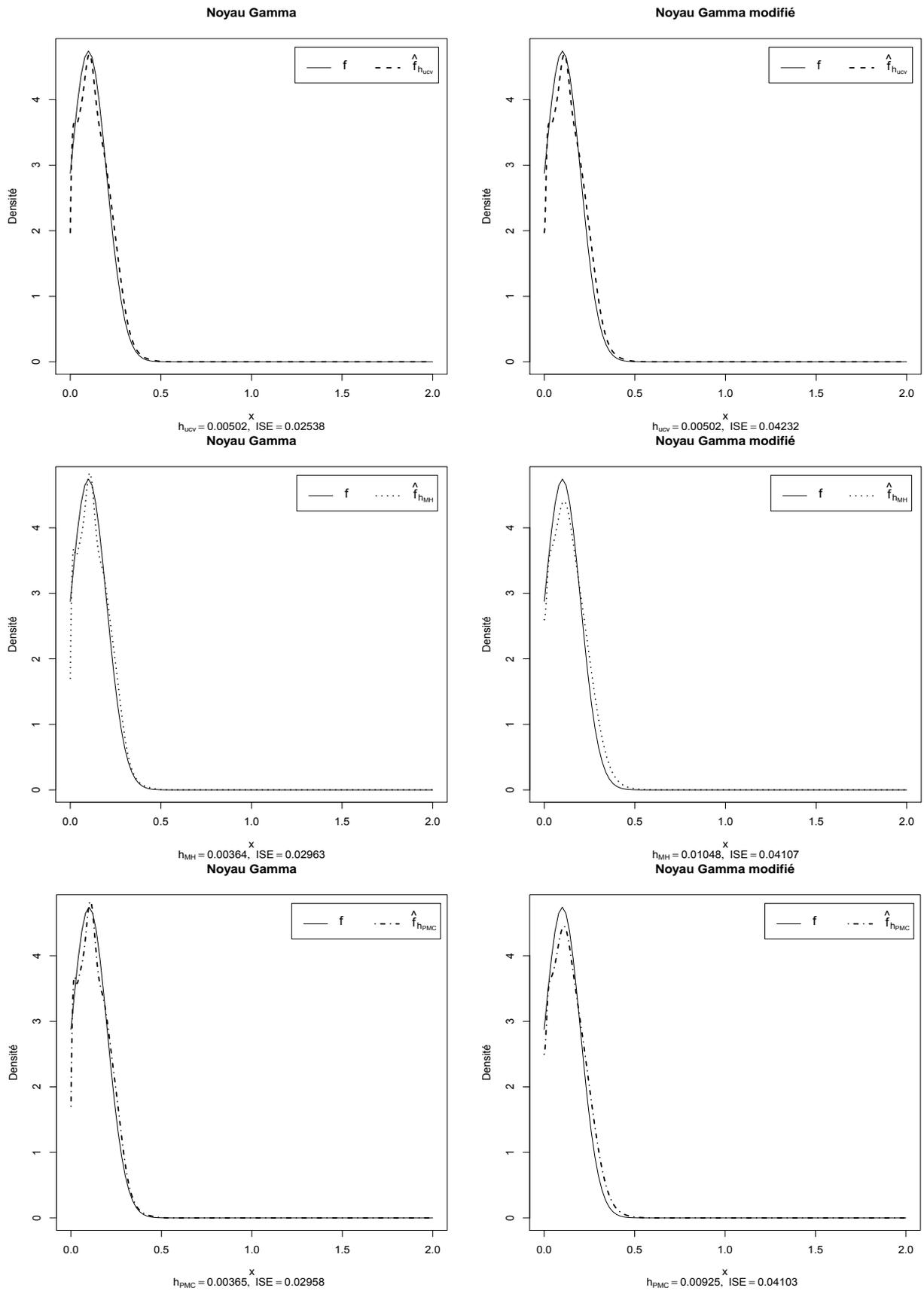


FIGURE 4.11: Estimateurs à noyaux Gamma et Gamma modifié d'une densité (D5)

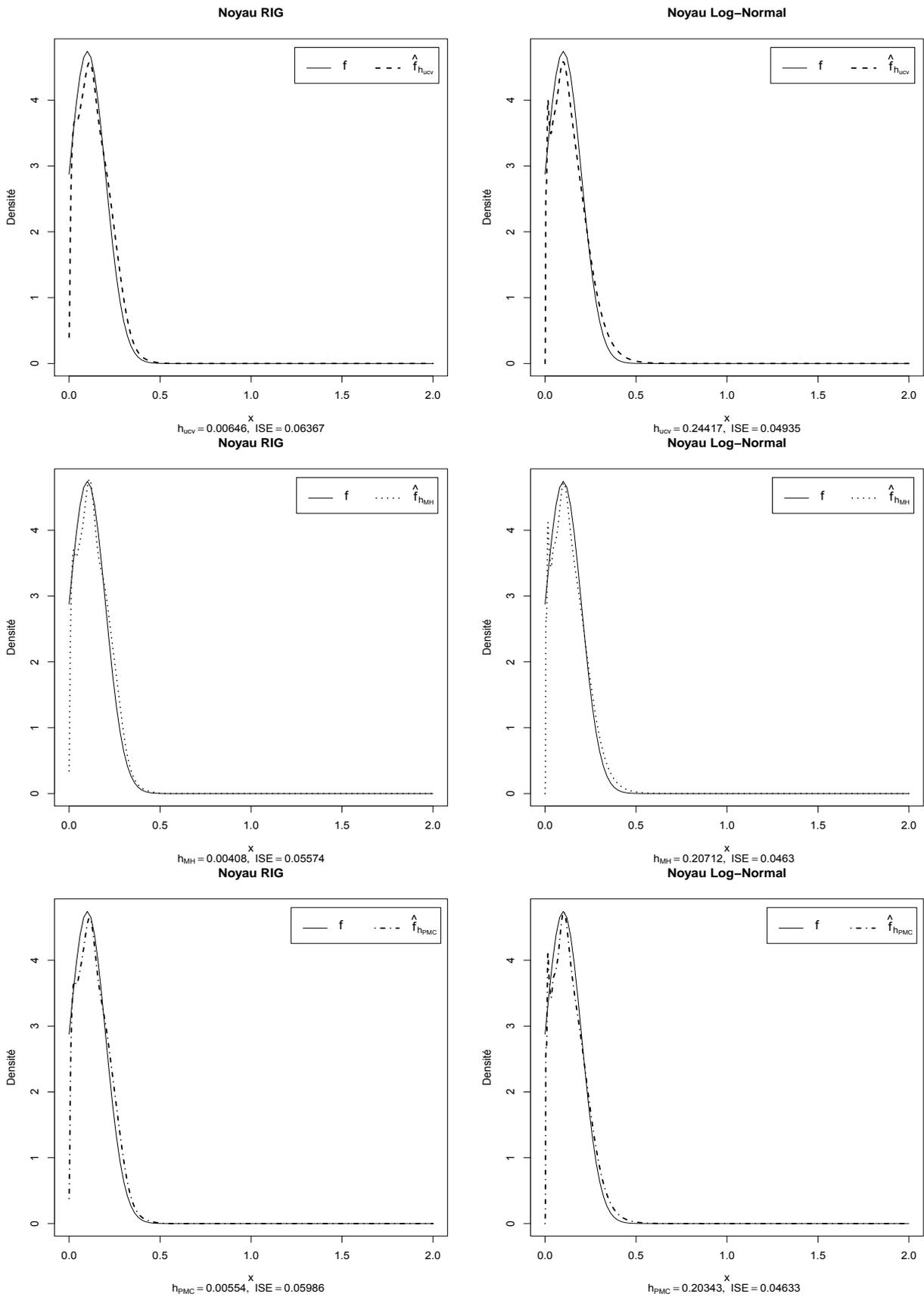


FIGURE 4.12: Estimateurs à noyaux RIG et log-Normal d'une densité (D5)

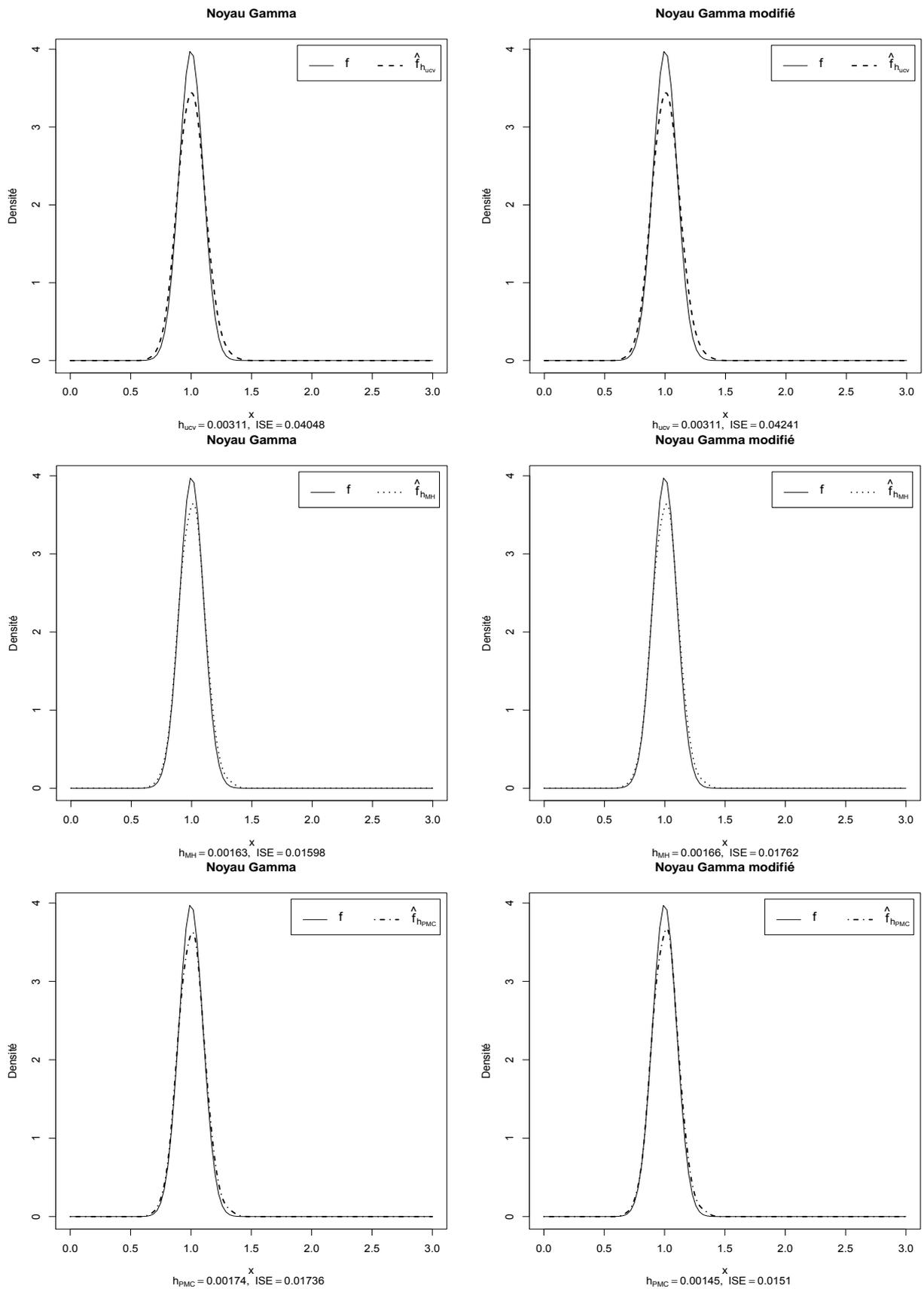


FIGURE 4.13: Estimateurs à noyaux Gamma et Gamma modifié d'une densité (D6)

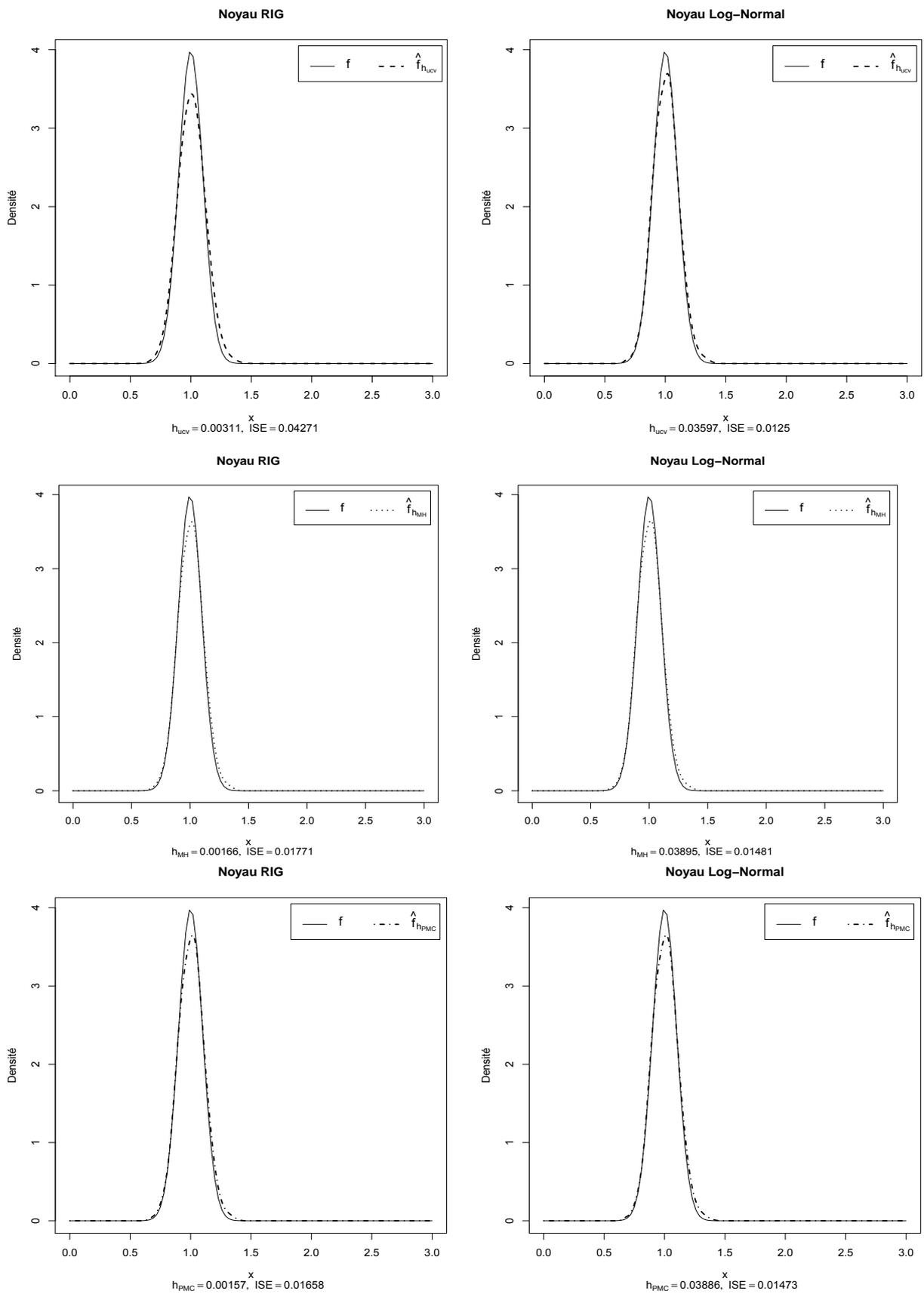


FIGURE 4.14: Estimateurs à noyaux RIG et log-Normal d'une densité (D6)

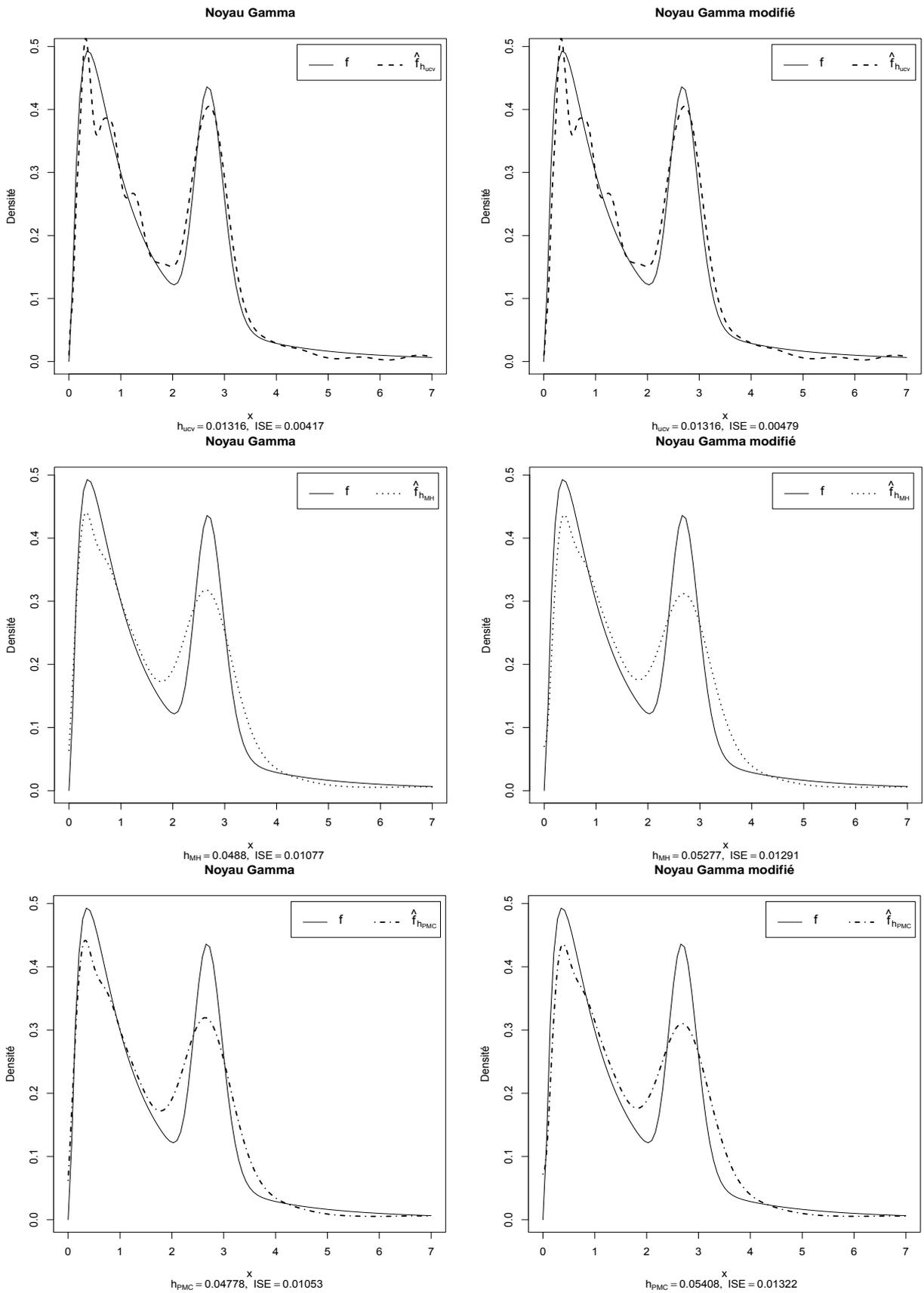


FIGURE 4.15: Estimateurs à noyaux Gamma et Gamma modifié d'une densité (D7)

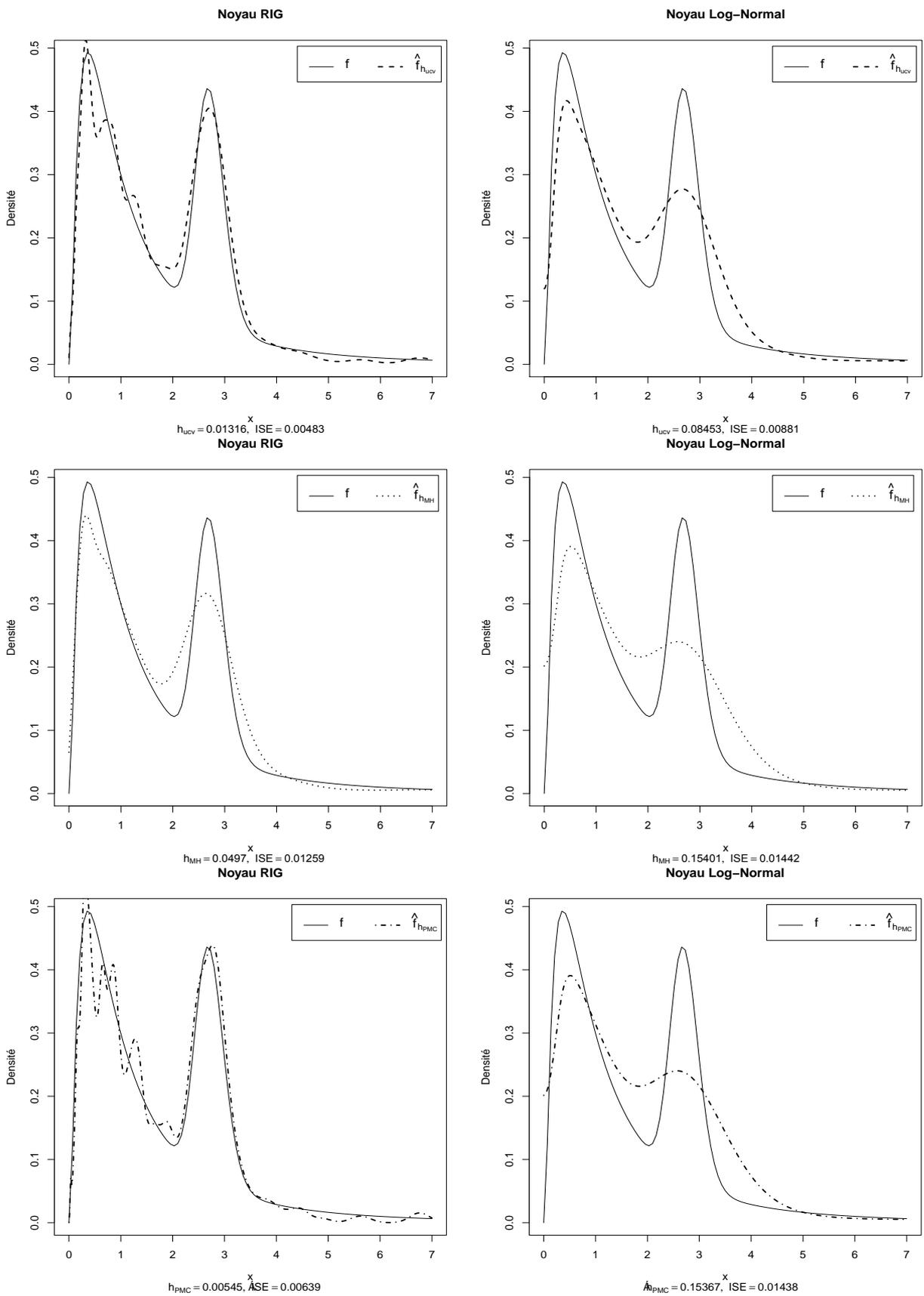


FIGURE 4.16: Estimateurs à noyaux RIG et log-Normal d'une densité (D7)

Conclusion générale & Perspectives

Le travail présenté dans ce mémoire traite le problème de l'estimation de la densité de probabilité de variables aléatoires réelles à valeurs positives par la méthode du noyau associé. L'applicabilité de cette méthode nécessite, préalablement, le choix du noyau et du paramètre de lissage.

Dans un premier temps, nous avons rappelé la notion de l'estimateur à noyau associé récemment introduite par [Kokonendji and Libengué \[2011\]](#) et [Libengué \[2013\]](#). Cette notion donne la définition unifiée d'un noyau associé $K_{x,h}$ de cible x et du paramètre de lissage h , à partir d'une quelconque loi de probabilité vérifiant certaines conditions de régularité. En pratique, pour estimer une densité de probabilité f de support \mathbb{S} , il est utile de choisir un noyau associé construit à partir d'une distribution de même support \mathbb{S} . Cette technique a permis de remédier au problème des effets de bord rencontré souvent en pratique lorsqu'on utilise par exemple un noyau continu symétrique comme le noyau gaussien, pour lisser une densité dont le support est borné d'un ou des deux cotés.

Pour estimer la densité de variables aléatoires à valeurs positives, nous avons utilisé les noyaux asymétriques : Gamma et Gamma modifié de [Chen \[2000\]](#), Gaussien-Inverse-Réciproque de [Scaillet \[2004\]](#) et log-Normal de [Libengué \[2013\]](#). Les noyaux Gamma et Gamma modifié sont les plus adaptés pour ce type de données, les résultats obtenus coïncident avec ceux de [Bouezmarni and Scaillet \[2005\]](#) et [Malec and Schienle \[2014\]](#). Cependant, les autres noyaux (voire Gaussien-Inverse-Réciproque et Log-Normal) conduisent à des estimateurs divergeant au voisinage de 0.

Dans un second temps, nous avons traité le problème du choix du paramètre de lissage h qui est d'importance capitale dans l'estimation de la fonction densité par la méthode du noyau associé. Deux différentes approches ont été présentées : l'approche classique

(fréquentiste) et l'approche Bayésienne globale. L'approche fréquentiste est basée sur les méthodes classiques qui sont regroupées en deux classes. La première classe repose sur la minimisation du critère de l'erreur quadratique moyenne intégrée MISE. Cependant cette classe a tendance à fournir des paramètres de lissage de très faibles valeurs qui conduisent à des estimations sous lissées, comme l'a souligné Scaillet [2004]. La seconde classe englobe les méthodes de validation croisée, guidée seulement par les observations. L'intérêt de cette catégorie de méthodes est dans les applications. Cependant, elle ne garantit pas l'existence d'un seul minimum local. L'inconvénient principal de cette classe est qu'elle a tendance à produire des estimateurs sous ou sur-lissés lorsque la densité à estimer est complexe ou quand la taille de l'échantillon est petite ou modérée. L'approche Bayésienne globale est caractérisée par la considération du paramètre de lissage comme une variable aléatoire de loi a priori $\pi(\cdot)$. La loi a priori sert à compenser le manque d'information lorsque la taille de l'échantillon est petite ou modérée. L'estimateur Bayésien de h est obtenu par la moyenne a posteriori sous la perte quadratique Bayésienne. Le calcul explicite de la moyenne a posteriori a été impossible vu la complexité de la loi a posteriori. Cette complexité a été résolue par les méthodes de Monte Carlo par chaînes de Markov (MCMC) et Population Monte Carlo (PMC), qui permettent de générer des échantillons suivant la loi a posteriori connue à une constante près.

Les résultats numériques obtenus sur des données simulées et sur des données réelles, en utilisant les noyaux asymétriques, montrent que les performances de l'approche Bayésienne globale, lorsque la taille de l'échantillon est petite ou modérée, sont légèrement meilleures que celles obtenues par la méthode UCV. Pour les échantillons de grande taille, les performances des méthodes de sélection sont équivalentes. En revanche, la qualité de lissage a été très satisfaisante par l'approche Bayésienne. Dans l'approche Bayésienne, les deux méthodes (MCMC et PMC) sont comparées et toutes les deux convergent vers des quantités très proches.

Le travail réalisé dans ce mémoire ouvre de nombreuses perspectives intéressantes :

- Dans ce travail, on s'est limité à des estimateurs unidimensionnels à noyaux asymétriques : un travail important reste à mener sur les estimateurs multidimensionnels,

là où les performances des deux approches (Bayésienne et classique) peuvent se distinguer considérablement.

- Utiliser l’approche Bayésienne adaptative et locale pour le lissage des densités ayant des queues lourdes ou des pôles au voisinage de $x=0$.
- Généralisation et automatisation des algorithmes de sélection du paramètre de lissage pour un noyau associé donné.
- Réaliser un Package sous \mathcal{R} traitant le problème de l’estimation de la densité de probabilité de variables aléatoires positives, où sera traité le problème du choix du noyau et du paramètre de lissage par l’approche classique et les approches Bayésiennes.

Bibliographie

- K. M. Abadir and S. Lawford. Optimal asymmetric kernels. *Economics Letters*, 83(1) : 61–68, 2004.
- S. Abou-Jaoude. Sur une condition nécessaire et suffisante de l^1 -convergence presque complète de l'estimateur de la partition fixe pour une densité. *CR Acad. Sci. Paris Sér. AB*, 283 :16, 1976.
- J. J. Boreux, E. Parent, and J. Bernier. *Pratique du calcul bayésien*. Springer-Verlag, 2010.
- D. Bosq. Estimation suroptimale de la densité par projection. *Canadian Journal of Statistics*, 33(1) :21–37, 2005.
- T. Bouezmarni and O. Scaillet. Consistency of asymmetric kernel density estimators and smoothed histograms with application to income data. *Econometric Theory*, 21 : 390–412, 2005.
- A. W. Bowman. An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, 71(2) :353–360, 1984.
- M. J. Brewer. A modelling approach for bandwidth selection in kernel density estimation. *In : Proceedings of COMPSTAT Physica Verlag, Heidelberg*, pages 203–208, 1998.
- O. Cappé, A. Guillin, J. M. Marin, and C. P. Robert. Population monte carlo. *Journal of computational and graphical statistics*, 13(4) :907–929, 2004.
- N. N. Cencov. Evaluation of an unknown distribution density from observations. *Soviet Mathematics*, 3 :1559–1562, 1962.
- S. X. Chen. Beta kernels estimators for density functions. *Computational Statistics and Data Analysis*, 31 :131–145, 1999.
- S. X. Chen. Gamma kernel estimators for density functions. *Annals of the Institute of Statistical Mathematics*, 52 :471–480, 2000.

- E. A. Epanechnikov. Nonparametric estimation of a multidimensional probability density. *Theory of Probability and its Applications*, 14 :153–158, 1969.
- C. Gaetan and X. Guyon. *Modélisation et statistique spatiales*. Springer, 2008.
- P. H. Garthwaite, Y. Fan, and S. A. Sisson. Adaptive optimal scaling of metropolis-hastings algorithms using the robbins-monro process. *Working paper*, 2010.
- J. Geffroy. Sur l'estimation d'une densité dans un espace métrique. *CR Acad. Sci. Paris Sér. A*, 278 :1449–1452, 1974.
- A. Gelman, G. O. Roberts, and W. R. Gilks. Efficient metropolis jumping rules. *Bayesian Statistics*, 5 :599–608, 1996.
- C. Gouriéroux and A. Monfort. (non) consistency of the beta kernel estimator for recovery rate distribution. *CREST Discussion Paper, N° 31*, 2006.
- A. Guillin, J. M. Marin, and C. P. Robert. Estimation bayésienne approximative par échantillonnage préférentiel. *Revue de statistique appliquée*, 53(1) :79–96, 2005.
- J. D. F. Habbema, J. Hermans, and K. Vandebroek. A stepwise discriminant analysis program using density estimation. *in Compstat, ed. G. Bruckmann, Physica-Verlag, Wien*, pages 101–110, 1974.
- J.M. Hammersley and D. C. Handscomb. *Monte Carlo methods*. Oxford University, 1967.
- W. K. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1) :97–109, 1970.
- M. Hirukaway and M. Sakudo. Family of generalized gamma kernels : A unified approach to the asymptotics on asymmetric kernels. 2013.
- F. Hua, X. Shen, Z. Chen, F. Yang, and J. Gu. Bayesian DOA estimation method using population monte carlo. In *Signal Processing, Communication and Computing (ICSPCC), 2012 IEEE International Conference on*, pages 158–161. IEEE, 2012.
- Y. Iba. Population monte carlo algorithms. *Transactions of the Japanese Society for Artificial Intelligence*, 16 :279–286, 2001.

- F. Jedrzejewski. *Introduction aux méthodes numériques*. Springer, 2005.
- J. Jiang. *Large Sample Techniques for Statistics*. Springer, 2010.
- X. Jin and J. Kawczak. Birnbaum-saunders and lognormal kernel estimators for modelling durations in high frequency financial data. *Annals of Economics and Finance*, 4 :103–124, 2003.
- G. Kerkycharian and D. Picard. Density estimation by kernel and wavelets methods : optimality of besov spaces. *Statistics and Probability Letters*, 18(4) :327–336, 1993.
- C. C. Kokonendji and F. G. Libengué. Méthode des noyaux associés continus et estimation de densité. *Journées de Statistique de la SFdS, 6 pages, Tunis*, 2011.
- C. C. Kokonendji, T. Senga Kiessé, and S. S. Zocchi. Discrete triangular distributions and non-parametric estimation for probability mass function. *Journal of Nonparametric Statistics*, 19 :241–257, 2007.
- C. C. Kokonendji, F. G. Libengué, and D. Varron. Convergences des estimateurs à noyaux associés de densité. *Journées de Statistique de la SFdS, 5 pages, Bruxelles*, 2012.
- R. Kronmal and M. Tarter. The estimation of probability densities and cumulatives by fourier series methods. *Journal of the American Statistical Association*, pages 925–952, 1968.
- S. Li, M. J. Silvapulle, P. Silvapulle, and X. Zhang. Bayesian approaches to non-parametric estimation of densities on the unit interval. *Econometric Reviews*, 2013.
- F. G. Libengué. *Méthodes non paramétriques des noyaux associés mixtes et applications*. PhD thesis, Université de Franche-Comté (France) et Université de Ouagadougou (Burkina-Faso), 2013.
- D. V. Lindely. *Approximate bayesian methods*. University College London, 1980.
- J. S. Liu. Metropolized independent sampling with comparaison to rejection sampling and importance sampling. *Statistics and Computing*, 6(2) :113–119, 1996.

- P. Malec and M. Schienle. Nonparametric kernel density estimation near the boundary. volume 72, pages 57–76, 2014.
- J. M. Marin and C. P. Robert. *Bayesian Core : A practical approach to computational bayesian statistics*. Springer, 2007.
- J. S. Marron and D. Ruppert. Transformations to reduce boundary bias in kernel density estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 653–671, 1994.
- N. Metropolis. The Begining of the Monte Carlo Method. *Los Alamos Science*, 15(584) : 125–130, 1987.
- N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equations of state calculations by fast computing machines. *Journal of Chemical physics*, (21) : 1087–91, 1953.
- R. Meyer and J. Yu. Bugs for a bayesian analysis of stochastic volatility models. *The Econometrics Journal*, 3(2) :198–215, 2000.
- P. Del Moral, A. Doucet, and A. Jasra. Sequential monte carlo samplers. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 68(3) :411–436, 2006.
- H. G. Müller. Smooth optimum kernel estimators near endpoints. *Biometrika*, 78(3) : 521–530, 1991.
- H. G. Müller. On the boundary kernel method for non-parametric curve estimation near endpoints. *Scandinavian Journal of Statistics*, pages 313–328, 1993.
- E. Parent and J. Bernier. *Le raisonnement bayésien - Modélisation et inférence*. Springer-Verlag, 2007.
- E. Parzen. On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 33 :1065–1076, 1962.

- R Development Core Team. *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013. URL <http://www.R-project.org>. ISBN 3-900051-07-0.
- H. Raiffa and R. Schlaifer. *Applied statistical decision theory*. Boston, Massachusetts : Harvard Business School, 1961.
- C. P. Robert. *Le choix bayésien - Principes et pratique*. Springer-Verlag France, Paris, 2006.
- C. P. Robert and G. Casella. *Monte Carlo statistical methods*. Springer-Verlag, 2004.
- G. O. Roberts and J. S. Rosenthal. Examples of adaptive MCMC. *Journal of Computational and Graphical Statistics*, 18(2) :349–367, 2009.
- M. Rosenblatt. Remarks in some nonparametric estimates of a density function. *Annals of Mathematical Statistics*, 27 :832–837, 1956.
- D. B. Rubin. A noniterative sampling/importance resampling alternative to the data augmentation algorithm for creating a few imputations when fraction of missing information are modest : The sir algorithm. *American Statistical Association, Theory and methods*, 82(398) :543–546, 1987.
- R. Y. Rubinstein and D. P. Kroese. *Simulation and the Monte Carlo method*. John Wiley & Sons, 2008.
- M. Rudemo. Empirical choice of histograms and kernel density estimators. *Scandinavian Journal of Statistics*, pages 65–78, 1982.
- N. Saadi and S. Adjabi. On the estimation of the probability density by trigonometric series. *Communications in Statistics Theory and Methods*, 38(19) :3583–3595, 2009.
- O. Scaillet. Density estimation using inverse and reciprocal inverse gaussian kernels. *Journal of Nonparametric Statistics*, 16 :217–226, 2004.
- E. F. Schuster. Incorporating support constraints into nonparametric estimators of densities. *Communications in Statistics-Theory and Methods*, 14(5) :1123–1136, 1985.

- S. C. Schwartz. Estimation of probability density by an orthogonal series. *The Annals of Mathematical Statistics*, pages 1261–1265, 1967.
- D. W. Scott and G. R. Terrell. Biased and unbiased cross-validation in density estimation. *Journal of the American Statistical Association*, 82 :1131–1146, 1987.
- T. Senga Kiessé. *Approche non-paramétrique par noyaux associés discrets des données de dénombrement*. Thèse de Doctorat, Université de Pau, France, 2008.
- B. W. Silverman. *Density estimation*. Chapman and Hall, London, 1986.
- A. F. M. Smith and A. E. Gelfand. Bayesian statistics without tears : A sampling-resampling perspective. *The American Statistician*, 46(2) :84–88, 1992.
- L. Tierney and J. B. Kadane. Accurate approximation for posterior moments and marginal functions. *Journal of the American Statistician Association*, 81 :82–86, 1986.
- Y. K. Tse, X. Zhang, and J. Yu. Estimation of hyperbolic diffusion using the markov chain monte carlo method. *Quantitative Finance*, 4 :158–169, 2004.
- A. B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer, New York, 2009.
- M. Vieira. *Estimation Bayésienne par des méthodes MCMC - Application à la surveillance des moteurs asynchrones*. PhD thesis, Université de Nice - Sophia-Antipolis, 1999.
- G. Wahba. Optimal convergence properties of variable knot, kernel, and orthogonal series methods for density estimation. *The Annals of Statistics*, pages 15–29, 1975.
- G. Wahba. Data-based optimal smoothing of orthogonal series density estimates. *The Annals of Statistics*, pages 146–156, 1981.
- A. M. Walker. On the asymptotic behaviour of posterior distributions. *Journal of the Royal Statistical Society*, 31 :80–88, 1969.
- G. G. Walter. Approximation of the delta function by wavelets. *Journal of Approximation Theory*, 71(3) :329–343, 1992.

- X. Zhang, M. L. King, and R. J. Hyndman. A bayesian approach to bandwidth selection for multivariate kernel density estimation. *Computational Statistics and Data Analysis*, 50 :3009–3031, 2006.
- X. Zhang, M. L. King, and H. L. Shang. Bayesian bandwidth selection for a nonparametric regression model with mixed types of regressors. *Submitted paper*, 2013.
- X. Zhang, M. L. King, and H. L. Shang. Bayesian estimation of bandwidths for a nonparametric regression model with an unknown error density. *presented at Econometric Society North American Summer Meeting*, 9-12 June 2011, St. Louis, MO, USA.
- N. Zougab. Etude comparative des méthodes de sélection du paramètre de lissage dans l'estimation de la densité de probabilité par la méthode du noyau. In *Mémoire de Magister*. Université de Béjaia, Algérie, 2007.
- N. Zougab. *Approche Bayésienne dans l'estimation non paramétrique de la densité de probabilité et la courbe de régression de la moyenne*. PhD thesis, Université de Béjaia, Algérie, 2013.
- N. Zougab, S. Adjabi, and C. C. Kokonendji. Binomial kernel and Bayes local bandwidth in discrete functions estimation. *Journal of Nonparametric Statistics*, 24 :3 :783–795, 2012.
- N. Zougab, S. Adjabi, and C. C. Kokonendji. Adaptive smoothing in associated kernel discrete functions estimation using Bayesian approach. *Journal of Statistical Computation and Simulation*, 83(12) :2219–2231, 2013a.
- N. Zougab, S. Adjabi, and C. C. Kokonendji. A bayesian approach to bandwidth selection in univariate associate kernel estimation. *Journal of Statistical Theory and Practice*, 7 (1) :8–23, 2013b.

Résumé : Pour estimer la densité de probabilité des données positives par la méthode du noyau associé, le choix du noyau et du paramètre de lissage est important. On a utilisé les noyaux asymétriques du type Gamma, Gamma modifié, Gaussien-Inverse-Réciproque et Log-Normal. 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 et PMC). La comparaison de l'approche Bayésienne globale (MCMC et PMC) avec la méthode classique (UCV) pour des données simulées et des données réelles montre que les performances de l'approche Bayésienne globale sont meilleures que celle de la méthode UCV pour des données de petites ou moyennes tailles. Les noyaux Gamma sont plus performants que les autres noyaux.

Mots clés : estimation non paramétrique de densité, noyaux associés, approche Bayésienne, méthodes de Monte Carlo, validation croisée, plug-in.

Abstract : To estimate the probability density of positive data by the associated kernels method, the choice of kernel and smoothing parameter is important. We used asymmetric kernels of the type Gamma, modified Gamma, Reciprocal Inverse Gaussian and Log-Normal. The selection of the smoothing parameter is based on the classical approach (UCV) and the Bayesian global approach. The complexity of the posterior distribution in the Bayesian global approach required the use of Monte Carlo methods (MCMC and PMC). The performance of the Bayesian global approach (MCMC and PMC) are compared with the classical method (UCV) through simulated data and real data. The results obtained show that the performance of the Bayesian global approach is better than that of the UCV method for small or moderate size data. Furthermore, the performance of the Gamma kernels are better than those of the other kernels.

Key words : non-parametric density estimation, associated kernels, Bayesian approach, Monte Carlo methods, cross-validation, plug-in.

مُلخّص : لتقدير دالة الكثافة البيّنات الإيجابيّة بواسطة طريقتي النوى الغير المتماثلة فإن تقدير وسيط
المس (النّافذة) مهمّ جدّا . تمّ إختيار وسيط المس بإستعمال الطرق الكلاسيكية و الطريقة البايزية الكتيّة
تعقيد توزيع الخلفي في الطريقة البايزية الكتيّة يتطلّب إستخدام أساليب مونتّي كارلو .
أخيرًا، قمنا بمقارنة أداء الطريقة البايزية الكتيّة و الطرق الكلاسيكية بإستعمال معطيات مصنوعة و حقيقية

كلمات المفتاح : الطريقة البايزية الكتيّة، النّواة، دالة الكثافة، المتوسط الخلفي، مونتّي كارلو .