

**REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE**

Ministère de l'enseignement Supérieur et de la Recherche Scientifique

**UNIVERSITE Mohamed Seddik Ben Yahia – Jijel**

**Faculté des Sciences Exactes et Informatique**

**Département de Mathématiques**



## **Mémoire**

Pour l'obtention du diplôme de : **Master**

**Spécialité : Mathématique**

**Option : Probabilités et Statistiques**

**Thème**

***La Simulation Stochastique Par MCMC***

**Présenté par :**

**Deffas Elaram**

**Fedsi Sihem**

**Devant le jury :**

**Président : Mme Laouir Dalila**

**Professeur**

**Encadreur : Mme Abdi Zeyneb**

**M. A. A**

**Examineur : Mme Ghouil Djaweyda**

**M. A. A**

**Examineur : Mlle Sellami Nawel**

**M. A. A**

**Promotion 2016/2017**



# *Remerciements*

*Nous remercions en premier lieu, notre dieu qui à bien voulu nous donnée la force effectuer le présent travail.*

*Nous remercions très vivement ABDI ZINEB notre promoteur de nous avoir encadré, orienté, conseillé et corrigé tour au long de notre travail.*

*Nous remercions également le professeur : GHERDA. M et laouir. D , ghouil. D , sellami . N*

*Nous tenons également à remercier l'ensemble des enseignants et les étudiants de mathématique de l'université de JJJEL.*

*A tous, un grand merci.*

**SJHEM ETELARAM**

*Merci*



# Dédicaces

*Rien n'est aussi beau à offrir que le fruit d'un labeur qu'on dédie du fond du cœur à ceux qu'on aime et qu'on remercie en exprimant la gratitude et la reconnaissance durant toute notre existence.*

*Je dédie ce modeste travail :*

*A ma **mère** et mon **père** qui ont toujours été présent pour me soutenir, me conseiller, sacrifient pour moi et m'avoir permis d'arriver à ce stade d'éducation.*

*A mes frères, et mes sœurs, en particulier KARIMA.*

*A tous mes amis en particulier INES et DJAHIDA.*

*A tous mes enseignants.*

*A tous la promotion 2017.*

*A mon binom DEFFAS ELARAM et sa famille.*

*إنّ الذين أحبهم وأعزهم مكانتهم ليست بين الأسطر و الصفحات لأن مقامهم  
أجل وأعلى فالقلب سكناهم والذكرى نكراهم والفؤاد لن ينساهم*

# Dédicaces

*Rien n'est aussi beau à offrir que le fruit d'un labeur qu'on dédie du fond du cœur à ceux qu'on aime et qu'on remercie en exprimant la gratitude et la reconnaissance durant toute notre existence.*

*Je dédie ce modeste travail :*

*A ma **mère** et mon **père** qui ont toujours été présent pour me soutenir, me conseiller, sacrifient pour moi et m'avoir permis d'arriver à ce stade d'éducation.*

*A mes frères et ma sœurs NESSMA, en particulier : SALAH-EDIN, ZAKARIYA.*

*A tous mes amis en particulier : SARRA YELLAS.*

*A tous mes enseignants.*

*A tous la promotion 2017.*

*A mon binom FEDSI SIHEM et sa famille.*

*إنّ الذين أحبهم وأعزهم مكاتتهم ليست بين الأسطر و الصفحات لأن مقامهم  
أجل وأعلى فالقلب سكناهم والذكرى نكراهم والفؤاد لن ينساهم*

# Table des matières

<b>Introduction</b>	<b>3</b>
<b>1 Le paradigme Bayésienne</b>	<b>5</b>
1.1 Introduction . . . . .	5
1.2 Aperçu à le paradigme bayésien . . . . .	5
1.2.1 Notations et Définitions . . . . .	5
1.2.2 La loi a priori impropre . . . . .	10
1.2.3 La loi a priori informative . . . . .	10
1.2.4 La loi conjuguées . . . . .	10
1.2.5 La loi a priori non informative . . . . .	11
1.2.6 Estimation Bayésienne . . . . .	13
1.3 Les chaînes de Markov . . . . .	15
1.3.1 Irréductibilité . . . . .	16
1.3.2 A périodicité . . . . .	16
1.3.3 Réversibilité . . . . .	16
1.3.4 Chaîne récurrence . . . . .	16
<b>2 Méthodes de simulations avec MCMC</b>	<b>18</b>
2.1 Introduction . . . . .	18
2.2 L'approche indépendante . . . . .	19
2.3 Méthodes d'approximations . . . . .	20
2.3.1 Intégrale numérique . . . . .	20
2.3.2 Méthode de Monte Carlo (MC) . . . . .	21
2.4 Méthodes de Monte Carlo par Chaînes de Markov (MCMC) . . . . .	22
2.4.1 Principes des méthodes MCMC . . . . .	23

2.4.2	L'algorithme de Metropolis-Hasting(MH) . . . . .	23
2.4.3	L'algorithme de Gibbs : . . . . .	27
2.4.4	Comparaison d'échantillonnage de Gibbs avec Metropolis-Hasting	30
2.5	MCMC et convergence . . . . .	30
2.5.1	Convergence de Monte-Carlo . . . . .	31
2.5.2	Convergence des algorithmes de MH et Gibbs . . . . .	32
<b>3</b>	<b>Applications</b>	<b>33</b>
3.1	Introduction . . . . .	33
3.2	Applications de Monte-Carlo (MC) . . . . .	33
3.2.1	Simulation de MC avec une distribution normale . . . . .	33
3.2.2	Simulation de MC avec une distribution binomiale . . . . .	35
3.3	Applications de MCMC . . . . .	37
3.3.1	La fonction de vraisemblance pour le modèle . . . . .	38
3.3.2	Comment définir l'a priori ? . . . . .	39
3.3.3	Comment définir l'a posteriori ? . . . . .	39
3.3.4	Application de M-H . . . . .	40
	<b>Conclusion générale</b>	<b>43</b>
	<b>Résumé</b>	<b>44</b>
	<b>Annexe</b>	<b>46</b>
	<b>Bibliographie</b>	<b>53</b>

# Introduction

Les méthodes MCMC sont apparues il y a 50 ans pour la physique statistique par **Nicholas Metropolis** dans les années 1953, en 1970, **W.Keith Hastings** (1930) a étendu l'algorithme au cas de n'importe quelle distribution, et en 1984 **Geman** étendre l'échantillonneur de Gibbs. Depuis le début des années 90, les méthodes MCMC ont été très utilisées et ont permis des avancées notables dans le traitement bayésien de nouveaux modèles : c'est notamment le cas de nombreux modèles à données manquantes. L'algorithme de Metropolis-Hasting et l'échantillonneur de Gibbs sont les schémas de construction des chaînes ergodiques les plus répandus [**Robert and Casella** 1999].

Les méthodes de Monte-Carlo par chaînes de Markov sont des méthodes d'intégration numérique qui utilisent l'aléatoire. Elles permettent de résoudre de nombreux problèmes autrement insolubles comme par exemple l'évaluation d'intégrales sur des domaines complexes et/ou de grande dimension. Les algorithmes de Monte-Carlo dépendent de paramètres d'implémentation dont le choix est crucial pour la convergence de l'algorithme et pour son efficacité. Dans les algorithmes de type Monte-Carlo par chaînes de Markov (MCMC), il s'agit de construire une chaîne de Markov stationnaire, de mesure invariante égale à la loi cible.

Dans notre travail nous exposons l'essentiel dans le cadre de la recherche sur les méthodes de Monte-Carlo par chaînes de Markov, la description de nos travaux relatifs aux méthodes de simulation est organisée en trois chapitres.

Le premier chapitre représente d'abord aperçu à le paradigme bayésienne avec des définitions des lois et d'estimation bayésienne, ensuite nous savons les chaînes de Markov du choix du noyau de transition dépend les propriétés d'irréductibilité, apériodicité, réversibilité et la chaîne réccurrence, des théorèmes d'ergodicité et de convergence.

Le deuxième chapitre décrit quelques méthodes de simulation qui sont résumés dans

la question suivante : pourquoi et comment simuler ? première méthode stochastique est la méthode de Monte-Carlo pour le calcul d'intégrale qui remplace par des sommes (section 2-3). MCMC par l'approche de Metropolis-Hasting et par l'algorithme de Gibbs et la comparaison entre eux (section 2-4). A la fin du chapitre, nous parlons des règles de la convergence des façons dont nous avons parlé précédemment.

Le troisième chapitre présente quelques applications numériques, en envisageant la simulation par MC et MCMC, nous avons employé le logiciel R et les résultats apparaissent sous la forme des graphes.



# Chapitre 1

## Le paradigme Bayésienne

### 1.1 Introduction

La statistique bayésienne est une théorie concurrente à la statistique dite classique en ce sens que chacune d'elles propose vis-à-vis d'un même problème une approche et une résolution complètement différentes. Donc nous présentons cette théorie dans une première partie (aperçu à le paradigme bayésien). Et la deuxième partie traite les chaînes de Markov avec des propriétés et des théories de base.

### 1.2 Aperçu à le paradigme bayésien

En premier lieu, l'approche bayésienne considère les paramètres comme des variables aléatoires. Ainsi ces paramètres ont toutes une distribution dite a priori de densité  $\pi(\theta)$ . Ensuite, on peut utiliser les règles de probabilité (i.e le théorème de bayes) pour déduire une distribution a posteriori.

#### 1.2.1 Notations et Définitions

L'ensemble de  $\theta$  paramètre et  $x$  observations. Dans ce chapitre  $x = (x_1, \dots, x_i, \dots, x_n)$ . Autrement dit, on dispose d'un échantillon de taille  $n$ . Le cadre statistique de ce chapitre étant celui de la statistique inférentielle, les observations  $x_i, i = \overline{1, n}$  sont donc considérées comme des réalisations de variables aléatoires, notées  $x_i$ .

**Définition 1.1 (Loi a priori)**

La loi a priori d'un paramètre  $\theta$  est une loi informative ou non informative qui résume les informations sur  $\theta$ , que nous avons considéré, sa densité est notée  $\pi(\theta)$ .

**Définition 1.2 (La loi conditionnelle)**

On appelle loi des observations, la loi conditionnelle de  $x$  sachant  $\theta$  ( $x$  discrète ou continue), sa densité est notée  $f(x/\theta)$ .

Dans ce chapitre on fera systématiquement que sachant  $\theta$ , les variables aléatoires  $x_i$  sont indépendantes. Autrement dit on aura :

$$f(x/\theta) = \prod_{i=1}^n f(x_i/\theta)$$

Nous indiquons maintenant les autres lois de probabilité qui interviennent en statistique bayésienne.

**Définition 1.3 (Loi a posteriori)**

C'est la loi conditionnelle de  $\theta$  sachant  $x$ , sa densité est notée  $\pi(\theta/x)$ . En vertu de la formule de Bayes on a :

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

**Exemple 1.1 (Situation simple de la loi a posteriori)**

On dispose d'un vecteur d'observations  $x = (x_1, \dots, x_i, \dots, x_n)$ , et on considère le modèle bayésien suivant :  $x_i/\theta \sim \text{Bernoulli}(\theta)$  et  $\theta \sim \text{Beta}(a, b)$  on a :

$$\begin{aligned} f(x/\theta) &= \prod_{i=1}^n P(x = x_i/\theta) \\ &= \theta^s (1 - \theta)^{n-s} \text{ où } : s = \sum_{i=1}^n x_i \end{aligned}$$

Comme  $\theta \sim \text{Beta}(a, b)$  on a :

$$\pi(\theta) = \frac{1}{\beta(a, b)} \theta^{a-1} (1 - \theta)^{b-1} 1_{[0,1]}(\theta)$$

D'autre part on a :

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

Il est facile de vérifier que :

$$\int_{\Theta} f(x/\theta) \pi(\theta) d\theta = \beta(\alpha, \beta)$$

Où:  $\alpha = a + s$  et  $\beta = b + n - s$

D'où :

$$\pi(\theta/x) = \frac{\theta^{\alpha-1} (1-\theta)^{\beta-1}}{\beta(\alpha, \beta)} 1_{[0,1]}(\theta)$$

Par conséquent :  $\theta/x \sim \text{Beta}(a + \sum_{i=1}^n x_i, b + n - \sum_{i=1}^n x_i)$

**Définition 1.4 (La loi couple  $(\theta, x)$ )**

Sa densité est notée  $h(\theta, x)$ , on a donc :

$$h(\theta, x) = f(x/\theta) \pi(\theta)$$

**Définition 1.5 (La loi marginale de  $x$ )**

Sa densité est notée  $m(x)$ , on a donc :

$$m(x) = \int_{\Theta} f(x/\theta) \pi(\theta) d\theta$$

## Le raisonnement proportionnelle de la loi a posteriori

Soit deux fonctions réelles  $f$  et  $g$  définies sur le même espace  $Y$  on dit que  $f$  et  $g$  sont proportionnelles, ce qu'on note  $f \propto g$ , s'il existe une constante  $a$  telle que  $f(y) = a g(y)$  pour tout  $y \in Y$ .

Dans un contexte bayésien on a :  $\pi(\theta/x) \propto f(x/\theta) \pi(\theta)$ . En tant que fonction de  $\theta$ , les deux expressions  $\pi(\theta/x)$  et  $f(x/\theta)$  sont effectivement proportionnelles, la constante  $a$  est égale ici à  $\frac{1}{m(x)}$ ; à noter que cette quantité est bien que constante, au sens où elle ne dépend pas de  $\theta$  l'écriture  $\pi(\theta/x) \propto f(x/\theta) \pi(\theta)$  est souvent reformulée de la façon suivante :  $\pi(\theta/x) \propto L(\theta, x) \pi(\theta)$  où  $L(\theta, x)$  désigne la vraisemblance rappelons que  $L(\theta, x) = f(x/\theta)$ .

### Exemple 1.2

Reprenons le modèle statistique bayésien précédent où :

$$x_i/\theta \sim \text{Bernoulli}(\theta) \text{ et } \theta \sim \text{Beta}(a, b)$$

On a :

$$\pi(\theta/x) \propto L(\theta, x) \pi(\theta)$$

Soit :

$$\pi(\theta/x) \propto \theta^{\sum_{i=1}^n x_i} (1-\theta)^{n-\sum_{i=1}^n x_i} \theta^{a-1} (1-\theta)^{b-1}$$

Par conséquent :

$$\pi(\theta/x) \propto \theta^{a+\sum_{i=1}^n x_i-1} (1-\theta)^{b+n-\sum_{i=1}^n x_i-1} 1_{[0,1]}(\theta)$$

Dans le membre de gauche de densité une loi beta de paramètre  $(a + \sum_{i=1}^n x_i, b + n - \sum_{i=1}^n x_i)$

D'après Le raisonnement proportionnelle, on déduit que

$$\theta/x \sim \text{Beta}\left(a + \sum_{i=1}^n x_i, b + n - \sum_{i=1}^n x_i\right)$$

### Définition 1.6 (Famille exponentielle)

On appelle famille exponentielle à  $n$  paramètres tout famille de loi de distribution  $\{p_\theta\}$  dont la densité a la forme suivante :

$$f(x/\theta) = \exp\left[\sum_{i=1}^n \eta_i(\theta) T_i(x) + B(\theta)\right] h(x)$$

Où :  $\eta_i(\cdot)$  et  $B(\cdot)$  sont des fonctions du paramètre  $\theta$  et  $T_i(\cdot)$  sont des statistiques exhaustives.

Le tableau suivant présente quelques lois a priori pour quelques familles exponentielles usuelles

$f(x/\theta)$	$\pi(\theta)$	$\pi(\theta/x)$
Normale $N(\theta, \delta^2)$	Normale $N(\mu, T^2)$	Normale $N\left(\frac{\delta^2\mu + T^2x}{\delta^2 + T^2}, \frac{\delta^2T^2}{\delta^2 + T^2}\right)$
Gamma $G(\alpha, \beta)$	$G(\alpha, \beta)$	$G(\alpha + n, \beta + x)$
Binomiale $B(n, \theta)$	$Beta(\alpha, \beta)$	$Beta(\alpha + x, \beta + n - x)$
Poisson $P(\theta)$	$G(\alpha, \beta)$	$G(\alpha + x, \beta + 1)$
Multinomiale $M(\theta_1, \dots, \theta_n)$	$Dirichlet D(\alpha_1, \dots, \alpha_n)$	$D(\alpha_1 + x_1, \dots, \alpha_n + x_n)$

Tableau 1.1 : Quelques lois a priori pour quelques familles exponentielles

### Exemple 1.3

Soit la densité de la loi exponentielle qui définie par :

$$f(x/\theta) = \frac{1}{\theta} \exp\left(\frac{-x}{\theta}\right) 1_{[0, +\infty[}(x)$$

Ici :  $n = 1$ ;  $\eta_1(\theta) = \frac{1}{\theta}$ ;  $T_1(x) = x$ ;  $B(\theta) = \log(\theta)$  et  $h(x) = 1_{[0, +\infty[}(x)$

La loi binomiale :

$$\begin{aligned} P(X = x/\theta) &= C_n^x \theta^x (1 - \theta)^{n-x} \\ &= C_n^x \exp\{x \log \theta + (n - x) \log(1 - \theta)\} \\ &= C_n^x \exp\left\{x \log \left[\frac{\theta}{1 - \theta}\right] + n \log(1 - \theta)\right\} \end{aligned}$$

On a :  $n = 1$ ;  $\eta_1(\theta) = \log\left(\frac{\theta}{1 - \theta}\right)$ ;  $T_1(x) = x$ ,  $B(\theta) = n \log(1 - \theta)$  et  $h(x) = C_n^x$

La loi Gamma :

$$f(x/\alpha, \beta) = \left(\frac{\beta^\alpha}{\Gamma(\alpha)}\right) x^{\alpha-1} \exp(-\beta x)$$

En remarquant que :

$$x^{\alpha-1} \exp(-\beta x) = \exp\{-\beta x + (1 - \alpha) \log x\}$$

Et  $n = 2$  on peut écrire :  $\eta_1(\alpha, \beta) = -\beta$ ;  $\eta_2(\alpha, \beta) = \alpha - 1$ ;  $T_1(x) = x$ ;  $T_2(x) = \log x$ ,  
 $B(\alpha, \beta) = \log\left(\frac{\beta^\alpha}{\Gamma(\alpha)}\right)$

### 1.2.2 La loi 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. Nous verrons par la suite que la construction des lois non informatives peut conduire à des lois a priori de ce type

#### Exemple 1.4

Considérons la loi uniforme sur  $\mathbb{R}^+$  notée  $1_{\mathbb{R}^+}(x)$ . Supposons

$$f(x/\lambda) = \lambda \exp\{-\lambda x\}$$

La loi a posteriori est :

$$\pi(\lambda/x) = \frac{\lambda \exp\{-\lambda x\}}{\int_0^{+\infty} \lambda \exp\{-\lambda x\} dx}$$

Le dénominateur est égal à :  $\Gamma(2)$  d'où :  $\pi(\lambda/x) = \frac{1}{\Gamma(2)} \lambda \exp\{-\lambda x\}$  une loi gamma de paramètres  $(2, 1)$

### 1.2.3 La loi a priori informative

La modélisation a priori informative est le point le plus délicat de l'analyse bayésienne. Il existe plusieurs procédés pour obtenir des lois informatives, nous donnons la description de l'un des plus intéressants des procédés qui est celui des familles naturelles conjuguées.

### 1.2.4 La loi conjuguées

Une des difficultés de l'approche bayésienne est le calcul de la loi a posteriori. Ce calcul est facile lorsque la loi a priori et la loi a posteriori ont la même forme. Dans ce cas, on parle de loi a priori conjuguée.

#### Définition 1.7

Une famille  $F$  de loi sur  $\Theta$  est dite conjuguée si pour tout  $\pi$  appartenant à cette famille, la loi  $\pi(\theta/X)$  appartient également à celle-ci

## 1.2.5 La loi a priori non informative

Ces distributions sont conçues dans le but de faire de l'analyse bayésienne lorsqu'il y a absence de l'information a priori sur le paramètre d'intérêt ou dans le cas où c'est difficile de traduire en terme de loi a priori l'information disponible sur les paramètres par une loi de probabilité.

### La règle de Jeffreys

Une méthode de Jeffreys (1961) permet de fabriquer des lois a priori non informatives, cette méthode utilise l'information de fisher  $I(\theta)$ , la règle de Jeffreys consiste donc à considérer des lois a priori de forme :

$$\pi(\theta) = \sqrt{I(\theta)} \text{ où } I(\theta) = E \left[ \frac{-\partial^2}{\partial \theta^2} \log f(x/\theta) \right]$$

### L'information de fisher

Soit  $n$  échantillons  $(X_1, \dots, X_n)$  de la loi de  $f(x/\theta)$  on a le résultat suivant : sous certaines conditions de régularité l'EMV  $\hat{\theta}_n$  de  $\theta$  est tel que, pour  $n$  assez grand,  $\sqrt{n}(\hat{\theta}_n - \theta)$  suit une loi normale centrée de variance  $\frac{1}{I(\theta)}$ .

Autre résultat important : l'inégalité de Cramer-Rao soit  $t_n$  un estimateur de  $g(\theta)$  alors :

$$E(t_n - g(\theta))^2 \leq \frac{[g'(\theta) + b'_n(\theta)]^2}{nI(\theta)}$$

Où  $b_n(\theta) = E(t_n) - g(\theta)$

### Remarque 1.1

Si  $g(\theta) = \theta$  et si l'estimateur est sans biais, on a :  $\text{var}(t_n) \geq \frac{1}{nI(\theta)}$

Ces résultats illustrent l'idée que l'information de fisher se compare à une variance.

### Remarque 1.2

Dans le cas multidimensionnelle  $\theta = (\theta_1, \dots, \theta_q)$  on peut considérer des lois de la forme :

$$\pi(\theta) = [\det I(\theta)]^{\frac{1}{2}}$$

Où :  $I(\theta)$  la matrice d'information de fisher :

$$I(\theta) |_{i,j} = E \left[ -\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log f(x/\theta) \right]; 1 \leq i, j \leq q$$

### Exemple 1.5

Dans le cas de la loi de gauss de paramètre  $(\theta, \delta^2)$  :

$$f(x/\theta, \delta^2) \propto \frac{1}{\delta} \exp \left\{ -\frac{(x-\theta)^2}{2\delta^2} \right\}$$

$$\log f(x/\theta, \delta^2) \propto \frac{-1}{2} \log \delta^2 - \frac{(x-\theta)^2}{2\delta^2}$$

$$\frac{\partial}{\partial \theta} \log f(x/\theta, \delta^2) = \frac{x-\theta}{\delta^2}$$

Cas

$$\frac{\partial^2}{\partial \theta \partial \delta^2} \log f(x/\theta, \delta^2) = -\frac{(x-\theta)}{2(\delta^2)^2}$$

$$\frac{\partial^2}{\partial \theta^2} \log f(x/\theta, \delta^2) = -\frac{1}{\delta^2}$$

$$\frac{\partial^2}{\partial (\delta^2)^2} \log f(x/\theta, \delta^2) = \frac{1}{2(\delta^2)^2} - \frac{(x-\theta)^2}{(\delta^2)^3}$$

La matrice d'information de fisher s'obtient comme suit :

$$\begin{aligned} E(x-\theta) &= 0 \\ E[(x-\theta)^2] &= \delta^2 \end{aligned}$$

On a donc :

$$I(\theta, \delta^2) = \begin{pmatrix} \frac{1}{\delta^2} & 0 \\ 0 & \frac{1}{2\delta^4} \end{pmatrix}$$

Une loi informative est donc de la forme :  $\pi(\theta) \propto \frac{1}{\delta^2}$



## 1.2.6 Estimation Bayésienne

L'estimation bayésienne de paramètre  $\theta$  est la moyenne de la loi a posteriori, elle est définie par deux cas :

### Le cas uni-dimensionnel

On suppose dans cette section que le paramètre  $\theta$  est réel. Rappelons que  $\pi(\theta/x)$  s'interprète comme résumé de l'information disponible sur  $\theta$ , une fois  $x$  observée d'un point de vue bayésien l'idée serait que le statisticien, communiqué à son interlocuteur (le médecin, l'économiste, l'ingénieur, etc) la loi a posteriori de façon à ne pas perdre d'information. Quand on souhaite cependant disposer d'une estimation pour  $\theta$ , on retient le plus souvent la moyenne de la loi a posteriori.

Dans ce cas l'estimateur de Bayes est noté  $E(\theta/x)$ . Et on a

$$\begin{aligned} E(\theta/x) &= \int_{\Theta} \theta \pi(\theta/x) d\theta \\ &= \frac{\int_{\Theta} \theta f(x/\theta) \pi(\theta) d\theta}{\int_{\Theta} f(x/\theta) \pi(\theta) d\theta} \end{aligned}$$

### Notation 1.1

L'estimateur de Bayes de  $\theta$  est noté  $\hat{\theta}_B$ . Il est donc défini par :

$$\hat{\theta}_B(x) = E(\theta/x)$$

### Exemple 1.6

On considère le modèle bayésien suivant :

$$x_i/\theta \sim \text{Bernoulli}(\theta) \text{ et } \theta \sim \text{beta}(a, b)$$

Rappelons que :  $\theta/x \sim \text{beta}(\alpha, \beta)$  où :  $\alpha = a + s$  et  $\beta = b + n - s$  et  $s = \sum_{i=1}^n x_i$ .

D'où :

$$E(\theta/x) = \frac{a + \sum_{i=1}^n x_i}{a + n + b}$$

## Le cas multi-dimensionnel

Dans un contexte multi-dimensionnel où  $\theta = (\theta_j, j = 1, \dots, J)$  la moyenne a posteriori  $E(\theta/x)$  est égale au vecteur  $(E(\theta_j/x), j = \overline{1, J})$  où :

$$E(\theta_j/x) = \int_{\Theta_j} \pi(\theta_j/x) d\theta_j$$

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

## Le risque de Bayes

La recherche d'estimateur de Bayes peut se faire dans le cadre de la théorie de la décision. Alors il est nécessaire de fixer une règle de préférence  $R$

### Définition 1.8

Le risque de Bayes de  $\hat{\theta}$  est noté  $R'(\hat{\theta})$ , il est défini comme suit :

$$\begin{aligned} R'(\hat{\theta}) &= E[R(\hat{\theta})] \\ &= \int_{\Theta} R(\hat{\theta}) \pi(\theta) d\theta \end{aligned}$$

Si  $R(\hat{\theta}_1) < R(\hat{\theta}_2)$  alors  $\hat{\theta}_1$  est le meilleur que  $\hat{\theta}_2$  au sens du risque de Bayes

$R'(\hat{\theta})$  est le risque bayésien

$R(\hat{\theta})$  le risque fréquentielle

### Définition 1.9

Rappelons qu'en statistique classique la règle de préférence repose (le plus souvent) sur le risque quadratique, noté  $R(\hat{\theta})$  est défini comme suit :

$$R(\hat{\theta}) = \text{var}(\hat{\theta}) + [\text{biais}(\hat{\theta}_B)]^2$$

## 1.3 Les chaînes de Markov

Dans cette section nous définirons ce qu'est une chaîne de Markov et nous donnerons ses principales propriétés.

### Définition 1.10

-Une chaîne de Markov est une suite de variables aléatoires, notées  $\theta_t$  la valeur d'une variable aléatoire au temps  $t$ . Ces variables aléatoires peuvent être vues comme changement dans le temps, avec une probabilité de transition dépendante de l'état présent dans la chaîne.

-On peut définir un noyau de transition (où transition kernel) comme une fonction qui détermine les transitions. Cette fonction est notée  $K$  et définie sur  $\Theta$  où  $\Theta$  est l'espace des états (donc tous les états possibles).

-Si  $\Theta$  est discret, alors le noyau de transition est simplement une matrice où chaque élément est :

$$\begin{aligned} K(\theta, A) &= P(\theta_{t+1} \in A/\theta_t, \theta_j, j < t), \theta \in \Theta, A \subset \Theta \\ &= P(\theta_{t+1} \in A/\theta_t) \end{aligned}$$

-Si toutefois  $\Theta$  est continue alors le noyau de transition est défini comme  $K(\theta, \theta')$ , où  $\theta$  est l'ensemble de l'état présent, et  $\theta'$  est l'ensemble de l'état vers lequel on transite. Alors, étant donné un noyau de transition  $K$ , une séquence  $\{\theta_n\}_{n \geq 0}$  de variables aléatoires est une chaîne de Markov si pour tout  $k$ , on a que :

$$\begin{aligned} P(\theta_{k+1} \in A/\theta_0, \theta_1, \dots, \theta_k) &= P(\theta_{k+1} \in A/\theta_k) \\ &= \int_A K(\theta_k, d\theta) \end{aligned}$$

Cette propriété est appelée la propriété de Markov.

► Nous présentons maintenant les propriétés fondamentales que doivent vérifier les paramètres des chaînes de Markov produites par les algorithmes MCMC présentés prochainement pour assurer leurs convergence.

### 1.3.1 Irréductibilité

On dit qu'une chaîne de Markov, ou sa noyau de transition est irréductible si pour tous  $\theta, \theta' \in \Theta$ , la probabilité partant de  $\theta$  d'atteindre  $\theta'$  est strictement positive c-à-d : tous les ensembles de probabilité non nulle peuvent être atteints à partir de tout point de départ.

### 1.3.2 A périodicité

La période, notée  $d(\theta)$ , d'un état  $i$  est le plus grand nombre qui divise tout  $n$  tel que la probabilité d'aller de  $\theta$  à  $\theta$  en  $n$  pas est strictement positive. En d'autres mots :

$$d(\theta) = \text{PGCD} \{n : K^n(\theta, \theta) > 0, n \geq 1\}$$

Où le PGCD est le plus grand commun diviseur, c à d aucun noyau n'induit un comportement périodique des trajectoires.

### 1.3.3 Réversibilité

Une chaîne de Markov est réversible par rapport à  $f$  si la probabilité d'aller de  $\theta$  à  $\theta'$  en  $n$  pas la même que d'aller de  $\theta'$  à  $\theta$ . Si on considère le cas discret, on a que :

$$f(\theta) K(\theta, \theta') = f(\theta') K(\theta', \theta)$$

Où  $f(\theta)$  est la valeur de la distribution stationnaire à l'état  $\theta$ . Alors, ces propriétés nous permettent de construire notre chaîne de Markov dont la loi stationnaire converge vers la distribution que nous souhaitons.

### 1.3.4 Chaîne récurrence

On dit qu'une chaîne de Markov est récurrence si les trajectoires  $(\theta_t)_t$  passent a une infinité de fois dans tout ensemble de probabilité non nulle.

**Théorème 1.1 (Théorème ergodique)**

Soit  $\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(T)}$ ,  $T$  valeurs d'une chaîne de Markov ergodique de distribution invariante "stationnaire" et telque  $E^\pi [|h(\theta)|] < +\infty$

$$\frac{1}{T} \sum_{t=1}^T h(\theta^{(t)}) \xrightarrow{T \rightarrow +\infty} \int_{\Theta} h(\theta) \pi(\theta/x) d\theta = E^\pi [h(\theta)]$$

Où  $\pi$  est la distribution stationnaire. Le théorème ergodique répond aux problèmes dans la simulation par chaîne de Markov, il étand la loi de grand nombre à des suites dépendantes de variable aléatoire est supprime de construire un échantillon *i.i.d.*

**Théorème 1.2 (Théorème de convergence)**

On considère une chaîne de Markov irréductible et apériodique, admettant un état récurrent positif. Alors tous les états sont recurrent positifs et il existe une unique mesure invariante. De plus, quelque soit la mesure initiale  $\pi_0$  la suite des lois  $\pi_n$  des  $\theta_n$  converge vers  $\pi$ .

# Chapitre 2

## Méthodes de simulations avec MCMC

### 2.1 Introduction

Dans ce chapitre le but est de simuler selon  $\pi$  et l'idée de base est de construire une chaîne de Markov ergodique de la loi stationnaire  $\pi$ . Au début, nous adresses l'approche indépendante.

Nous parlons de la fondation de mémoire "Méthodes d'approximations" qui sont résumés dans la question suivante : pourquoi et comment simuler ? premier algorithme stochastique, et la réponse à cette question est de calculer l'intégrale numérique et les méthodes de Monte-Carlo peut remplaçant les intégrales par des sommes. En suite, il se compose d'un trois sections lorsque la loi a posteriori n'est pas aisément semblable par des algorithmes usueles comme la méthode d'acceptation rejet, il existe une technique forte générale, intitulée méthode de Metropolis-Hastings d'après les travaux de Metropolis et al [11] et de Hastings [11], qui permet d'obtenir des réalisations approximatives de la loi a posteriori et des estimateurs convergents de quantités a posteriori.

Enfin, nous avons discuté de la convergence des méthodes de Monte-Carlo par chaîne de Markov.

## 2.2 L'approche indépendante

De nombreux modèles utilisés reposent sur l'approche bayésienne dans laquelle les observations  $x$  sont supposées suivre un modèle probabiliste  $l(x/\theta)$  dépendant d'un paramètre  $\theta$  lui-même muni d'une loi a priori  $\pi(\theta)$ . Dans ces conditions, l'inférence sur le paramètre  $\theta$  est liée à la détermination des propriétés (moyenne, variance, quantiles, mode, etc.) de sa loi a posteriori définie par la formule de Bayes :

$$\pi(\theta/x) = \frac{l(x/\theta)\pi(\theta)}{\int l(x/\theta)\pi(\theta)d\theta}$$

Malgré la simplicité apparente de l'équation ci-dessus il est bien connu que l'inférence bayésienne ne conduit à des expressions explicites que dans un nombre réduit de modèles relativement simples. Dans le cas général, le calcul de l'évidence  $\int l(x/\theta)\pi(\theta)d\theta$  est impossible de même que celui des moments (ou autres caractéristiques numériques) de la loi a posteriori  $\pi(\theta/x)$ . La seule connaissance généralement exploitable est le fait que la loi a posteriori est donnée à une constante de proportionnalité inconnue près par  $l(x/\theta)\pi(\theta)$ . Si  $\theta \in \Theta$  de grande dimension les approches les plus simples reposent sur une exploration systématique de  $\theta$  (grilles, etc) sont vouées à l'échec.

Dans cette partie, l'approche la plus générique et la plus utilisée à ce jour est celle dite de Monte Carlo Par Chaîne de Markov (MCMC). Nous nous intéressons ici à l'inférence de paramètres de modèles cosmologiques à partir de mesures d'observables. Dans ce cadre, l'approche MCMC est très largement utilisée du fait du caractère particulièrement complexe de la fonction de vraisemblance  $l$ .

### **Théorème 2.1 (Loi forte de grand nombres)**

soit  $(\theta_n)_{n \geq 0}$  une suite de variable aléatoire i.i.d à valeur dans  $\mathbb{R}^d$  ( $d \in \mathbb{N}^*$ ) en suppose que  $E(|\theta_1|) < +\infty$

$$\frac{\theta_1 + \dots + \theta_n}{n} \xrightarrow{ps} E(\theta) \text{ tq } n \rightarrow +\infty$$

### **Théorème 2.2 (Théorème de la limite centrale)**

soit  $(\theta_n)_{n \geq 0}$  une suite de variable aléatoire i.i.d à valeur dans  $\mathbb{R}$ , on suppose que  $E(\theta_1^2) > +\infty$  soit  $\delta^2 = V(\theta)$  (la variance de  $\theta$ ) alors :

$$\frac{\sqrt{n}}{\delta} \left( \frac{\theta_1 + \dots + \theta_n}{n} - E(\theta_1) \right) \xrightarrow{loi} N(0, 1) \text{ tq } n \rightarrow +\infty$$

## 2.3 Méthodes d'approximations

Nous présentons dans cette section les techniques classiques de calcul bayésien, avant d'introduire dans la section suivante une méthode qui nous semble plus adaptée au contexte bayésien. Pour une revue plus détaillée, le lecteur pourra consulter Smith et al [1985], Ripley [1987], Tanner [1991], Robert [1996 b], Gamerman [1997], ou Robert et Casella [1999].

### 2.3.1 Intégrale numérique

Nous considérons un vecteur aléatoire  $x$  ayant pour densité de probabilité  $f(x/\theta)$  où  $\theta$  est le vecteur des paramètres inconnus. En statistique bayésienne,  $\theta$  est considéré comme un vecteur aléatoire de densité  $\pi(\theta)$ . L'estimation bayésienne est basée sur le calcul de la loi a posteriori de  $\theta$ , loi conditionnelle de  $\theta$  sachant  $x$  dont la densité est notée  $\pi(\theta/x)$ . La majorité des estimateurs bayésiens sont de la forme  $E_\pi(h(\theta))$  où  $E_\pi$  est l'opérateur d'espérance mathématique pour la loi  $\pi(\theta/x)$  et  $h(\theta)$  est une fonction réelle.

Dans de nombreux cas, le calcul explicite de  $\pi(\theta/x)$  est impossible, et celui de  $E_\pi(h(\theta))$ , même si  $\pi(\theta/x)$  est connue. Ainsi, de nombreuses méthodes d'approximations sont proposées dans la littérature. Nous nous intéressons ici aux méthodes stochastiques. Il s'agit de déterminer une approximation de :

$$I_h = E_\pi(h(\theta)) = \int_{\Theta} h(\theta) \pi(\theta/x) d\theta \quad (2.1)$$

Où  $\Theta = (\theta_1, \dots, \theta_n)$

Soit à résoudre l'intégrale (2.1), si on considère une suite de variables aléatoires indépendantes  $(\theta_1, \dots, \theta_n, \dots)$  est distribuées selon la loi a posteriori de  $\theta$ , on obtient un échantillon de nombres réels en tirant au hasard une valeur dans chacune d'entre elles la moyenne arithmétique de leur image par fonction  $h$  soit  $\frac{1}{n} \sum_{i=1}^n h(\theta_i)$  converge (presque sûrement) vers la cible quand  $n \rightarrow \infty$  (loi des grands nombres), ce qui justifie l'approximation :

$$\frac{1}{n} \sum_{i=1}^n h(\theta_i) \simeq E_\pi(h(\theta)) = \int_{\Theta} h(\theta) \pi(\theta/x) d\theta \quad (2.2)$$

#### Remarque 2.1



Il importe de remarquer que la formule (2.2) s'applique tout aussi bien au calcul d'une probabilité  $P(A) = \text{prob}(\theta \in A)$  par le biais d'une fonction indicatrice

$$h(\theta) = 1_A(\theta) = \begin{cases} 0 & \text{si } \theta \notin A \\ 1 & \text{si } \theta \in A \end{cases}$$

Puisque :

$$\text{prob}(\theta \in A) = \int_{\Theta} 1_A(\theta \in A) [\theta/y] d\theta$$

### 2.3.2 Méthode de Monte Carlo (MC)

La méthode de Monte-Carlo est une méthode d'approximation au sens statistique du terme, il n'ya pas de définition précise de ce qu'est une technique de type de Monte-Carlo, mais la discription la plus habituelle consiste à dire que les méthodes de ce type se caractérisent par l'utilisation du hasard pour résoudre des problèmes centrées sur le calcul d'une valeur numérique.

#### Principe

Les méthodes de Monte-Carlo sont des techniques d'échantillonnages, c'est à dire des simulations de séquences de réalisation de la distribution a posteriori, connue à un facteur de proportionnalité près :  $[\theta/x] \propto [x/\theta] [\theta]$  on doit cependant distinguer les techniques de calcul d'espérances telle que MCMC qui reposent généralement sur un échantillonnage complémentaire. Alors les méthodes MC sont des techniques de simule avec indépendance qui utilisent une loi de probabilité auxiliaire  $h(\theta)$ , dite instrumentale, facile à simuler.

#### Description de la méthode de MC

Dans un problème statistique l'approximation  $I_h$  doit tirer aavantage de la nature particulier de (2.1) à savoir que  $\pi$  soit une densité de probabilité (en supposant qu'il existe une a priori propre) ou plutôt, que  $\pi(\theta/x)$  soit proportionnel à une densité, une conséquence naturelle de cette perspective est d'utiliser la méthode de Monte-Carlo introduite par Métropolis et Ulam (1949) et Von Neumann (1951).

Lorsque  $\pi(\theta/x)$  est connue, la méthode de Monte-Carlo standard consiste à estimer  $I_h$

par :

$$\hat{I}_h = \frac{1}{n} \sum_{i=1}^n h(\theta_i) \quad (2.3)$$

Où  $\theta_1, \dots, \theta_n \stackrel{i.i.d}{\sim} \pi(\theta/x)$ . De plus, si la variance a posteriori  $V(h(\theta)/x)$  est finie, le théorème centrale limite s'applique à la moyenne (2.2), qui est alors asymptotiquement normale, de variance  $V(h(\theta)/x)/n$ , des régions de confiance peuvent alors se construire à partir décisive, il découple aussi du théorème centrale limite que l'ordre de grandeur de l'erreur est  $\frac{1}{\sqrt{n}}$  quelle que soit la dimension de problème, au contraire des méthodes numériques.

La mise en oeuvre de cette méthode nécessite la production d'une suite i.i.d  $\theta_i$  par ordinateur, reposant sur un générateur pseudo-aléatoire déterministe imitant la génération de  $\pi(\theta)$  ou de  $\pi(\theta/x)$  comme suit un échantillon i.i.d d'une loi uniforme  $U([0, 1])$  est générer puis transformé en variable de la loi d'intérêt.

### Remarque 2.2

*Les techniques statistiques standards peuvent aussi être utilisée pour déterminer l'erreur d'approximation de (2.1) par la moyenne :*

$$\frac{1}{n} \sum_{i=1}^n h(\theta_i) f(x/\theta_i)$$

*en réalité la méthode de Monte-Carlo s'applique dans un cadre beaucoup plus général que pour la simulation de  $\pi$ , comme dans le cas ci-dessus par exemple (2.1) peut se représenter de plusieurs manières, il n'est pas nécessaire de simuler les loi  $\pi(. / x)$  ou  $\pi$  obtenir une bonne approximation.*

## 2.4 Méthodes de Monte Carlo par Chaînes de Markov (MCMC)

Nous considérons dans cette section une méthode de Monte-Carlo plus générale, permettant d'approcher la génération de variables aléatoires, lorsque la loi a posteriori  $\pi(\theta/x)$  est inconnue ou trop complexe pour être simulée directement, les méthodes de Monte-Carlo par Chaînes de Markov (MCMC) proposant d'estimer  $I_h$  par :

$$\hat{I}_h = \frac{1}{T} \sum_{t=1}^T h(\theta^{(t)}) \quad (2.4)$$

Où  $\theta^{(t)}$  est une Chaîne de Markov ayant pour distribution stationnaire  $\pi(\theta/x)$  la convergence de  $\hat{I}_h$  vers  $I_h$  est assurée par l'ergodicité de la Chaîne de Markov  $\theta^{(t)}$ .

### 2.4.1 Principes des méthodes MCMC

- On connaît la loi cible  $\pi$  à une constante multiplicative près.
- On définit une loi de proposition (appelée aussi loi instrumentale)  $q(x/\theta)$ .
- Initialisation : choix de  $\theta^{(0)}$ .
- A partir de  $\theta^{(t)}$ , on génère  $x_t$  à l'aide de la loi de proposition et on accepte ou rejette cette valeur de  $x_t$  à l'aide d'une procédure d'acceptation-rejet. La valeur retenue est notée :  $\theta^{(t+1)}$ .

### 2.4.2 L'algorithme de Metropolis-Hasting(MH)

Nous présentons d'abord l'algorithme de Metropolis-Hasting dont la dynamique sera à la base de nos méthodes adaptatives.

#### L'algorithme générale de Metropolis-Hasting

Cet algorithme très général nécessite la connaissance de la densité  $f$  de la loi cible  $\pi$  à une constante multiplicative près, et donc tout à fait adapté à la reconstruction des lois a posteriori des modèles bayésiens. Chaque déplacement est basé sur la génération, à partir de la position courante  $\theta$ , d'une valeur candidate  $x$  à partir d'une loi instrumentale qui est une densité conditionnelle  $q(x/\theta)$  "presque quelconque" (mais facile à simuler) Un mécanisme d'acceptation-rejet fait que la chaîne se déplace ou reste sur place avec probabilité positive.

Le pas  $\theta^{(t)} \rightarrow \theta^{(t+1)}$  est donné par :

1- Simuler

$$x_t \sim q(x/\theta^{(t)})$$

2-Calculer

$$\rho(\theta^{(t)}, x) = \min \left\{ \frac{f(x) q(\theta^{(t)}/x)}{f(\theta^{(t)}) q(x/\theta^{(t)})}, 1 \right\}$$

3-Prendre

$$\theta^{(t+1)} = \begin{cases} x & \text{avec probabilité } \rho(\theta^{(t)}, x) \\ \theta^{(t)} & \text{avec la probabilité } 1 - \rho(\theta^{(t)}, x) \end{cases}$$

### Exemple 2.1

On veut simuler une loi de khi-deux inverse :

$$f(x) = x^{-\frac{n}{2}} \exp\left(\frac{-a}{2x}\right)$$

On peut prendre une loi uniforme sur (1, 100)

$N=500$

$n=3$

$a=4$

$f=function(x)\{(x^{-(n/2)})*exp(-a/(2*x))\}$

$x=numeric(N)$

$x[1]=1$

$for(i in 2 :N)\{y<-runif(1,0,100)$

$rho=min(1,f(y)/f(x[i-1]))$

$u=runif(1)$

$if (u<rho)$

$\{x[i]=y\}$

$else$

$\{x[i]=x[i-1]\}\}$

$mean(x)$

$[1] 15.15144$

$x[i]$  produit par cet algorithme MH construisant une Chaîne de Markov ergodique de loi stationnaire  $f$ , on a la convergence vers l'espérance (la moyenne a posteriori d'après (2.1)).

## L'algorithme de Metropolis-Hasting à marche aléatoire

C'est la version de l'algorithme de Metropolis-Hasting probablement la plus employée en raison de sa simplicité de mise en œuvre. La Chaîne de Markov associée à la loi proposition  $q$  est telle que :  $x_t = \theta^{(t)} + \varepsilon_t$ , où :  $\varepsilon_t$  indépendant de  $\theta^{(t)}$  ; i.e :  $q(x/\theta) = q(x - \theta)$

si  $q$  est symétrique on obtient l'algorithme suivant :

Etant donné  $\theta^{(t)}$  :

1-Générer

$$x_t \sim q(x - \theta^{(t)})$$

2-Acceptation-rejet

$$\theta^{(t+1)} = \begin{cases} x_t & \text{avec proba } \min \left\{ \frac{f(x_t)}{f(\theta^{(t)})}, 1 \right\} \\ \theta^{(t)} & \text{sinon} \end{cases}$$

### Exemple 2.2 (Loi normale)

Simulation de données suivant la loi normale  $N(0, 1)$

-Metropolis-Hasting indépendant : avec  $q(x) \sim U[-3, 3]$

-Algorithme Metropolis-Hasting-marche aléatoire avec

$$q(\varepsilon_t) \sim U[-\delta, \delta] \text{ (hasting1970)}$$

-Probabilité d'acceptation

$$\min \left\{ \exp \left\{ \theta_{(t)}^2 - \frac{x_t^2}{2} \right\}, 1 \right\}$$

## L'algorithme de Metropolis-Hasting indépendant

C'est l'autre version de l'algorithme de Metropolis-Hasting très utilisée. Ici, la loi de proposition  $q(x/\theta^{(t)})$  est indépendante de  $\theta^{(t)}$ .

Etant donné  $\theta^{(t)}$  :

1-Simuler

$$x_t \sim q(x)$$

## 2-Acceptation-rejet

$$\theta^{(t+1)} = \begin{cases} x_t & \text{avec proba } \min \left\{ \frac{f(x_t)q(\theta^{(t)})}{f(\theta^{(t)})q(x_t)}, 1 \right\} \\ \theta^{(t)} & \text{sinon} \end{cases}$$

### Exemple 2.3 (Loi gamma)

Soit  $f$  la densité d'une loi gamma  $G(\alpha, \beta)$  calcul de

$$I = \int_{-\infty}^{+\infty} \theta^2 f(\theta) d\theta$$

Acceptation-rejet avec

$$q(\theta) \sim G\left([\alpha], \frac{[\alpha]}{\alpha}\right), f(\theta) < Mq(\theta)$$

$Tq$

$$M = \exp\{\alpha(\ln(\alpha) - 1) - \alpha([\alpha] - 1)\}$$

Algorithme de Metropolis-Hasting avec :  $q(\theta) \sim G\left([\alpha], \frac{[\alpha]}{\alpha}\right)$

$$\rho(\theta^{(t)}, x_t) = \min \left\{ \left( \frac{x_t}{\theta^{(t)}} \exp \left[ \frac{\theta^{(t)} - x_t}{\alpha} \right] \right)^{\alpha - [\alpha]}, 1 \right\}$$

### Propriétés.

-Cas symétrique

$$\rho(\theta^{(t)}, x) = \min \left\{ \frac{f(x_t)}{f(\theta^{(t)})}, 1 \right\}$$

-On accepte toujours les valeurs de  $x_t$  augmentant la "vraisemblance"

-La loi cible  $\pi$  peut être connue à une constante multiplicative près

-La chaîne  $(\theta^{(t)})_t$  peut prendre plusieurs fois la même valeur  $\Rightarrow$  échantillon non i.i.d ■

### 2.4.3 L'algorithme de Gibbs :

Cette section traite des échantillonneurs de gibbs à deux est à plusieurs étapes. Ici est un cas particulier du second l'échantillonneur à deux étapes de meilleure propriétés de convergence et il s'applique naturellement à un grand nombre de modèles statistiques pour lesquels la généralisation de gibbs à plusieurs étapes possède de nombreuses propriétés d'optimalité.

#### L'échantillonneur de Gibbs à deux étapes

L'échantillonneur de gibbs à deux étapes crée une Chaîne de Markov à partir d'une loi jointe donnée de la manière suivante :

si deux variables aléatoires  $\theta_1$  et  $\theta_2$  ont pour densité jointe  $\pi(\theta_1, \theta_2)$ , avec les densités conditionnelles correspondantes  $\pi(\theta_2/\theta_1)$  et  $\pi(\theta_1/\theta_2)$  l'échantillonneur de gibbs à deux étapes génère une Chaîne de Markov  $(\theta_1^{(t)}, \theta_2^{(t)})$  selon les étapes suivantes :

prendre  $\Theta_0 = \theta_0$ , pour  $t = 1, 2, \dots, n$  générer :

$$1-\theta_2^{(t-1)} \sim \pi\left(\theta_2/\theta_1^{(t-1)}\right)$$

$$2-\theta_1^{(t-1)} \sim \pi\left(\theta_1/\theta_2^{(t-1)}\right)$$

L'implantation de l'algorithme est alors aisée, lorsqu'on peut simuler à la partir de deux densités conditionnelles. Lorsque on dispose d'une forme explicite de  $\pi(\theta_1, \theta_2)$  à une constante de normalisation -prés les densités conditionnelles  $\pi(\theta_2/\theta_1)$  et  $\pi(\theta_1/\theta_2)$  sont aussi disponibles sous forme explicite, par conséquent s'il n'est pas possible de simuler directement ces densités conditionnelles, des approximations MC ou MCMC peuvent être utilisées à la place.

Il est facilement que si  $(\theta_1^{(t)}, \theta_2^{(t)})$  suit la distribution de  $\pi$  alors il est de même pour  $(\theta_1^{(t+1)}, \theta_2^{(t+1)})$ , puis les deux étapes de l'itération  $t$  utilisent une simulation suivant la véritable densité conditionnelle, donc la convergence de l'algorithme de MC est assurée.

#### Exemple 2.4 (Génération de la loi beta-binomiales)

Considérons le couple de loi

$$X/\theta \sim \text{Bin}(n, \theta), \theta \sim \beta e(a, b)$$

qui donne la loi jointe

$$f(x, \theta) = \binom{n}{x} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{x+a-1} (1-\theta)^{n-x+b-1}$$

La loi conditionnelle correspondant de  $X/\theta$  est donnée ci-dessus on a par ailleurs

$$\theta/x \sim \beta e(x+a, n-x+b)$$

La loi beta, l'échantillonneur de gibbs associée peut être implanté comme suite :

*Nsim=500 #valeur initiale*

*n=15*

*a=3*

*b=7*

*X=T=array(0,dim=c(Nsim,1)) #initialisation des vecteurs*

*T[1]=rbeta(1,a,b) #initialisation de chaînes*

*X[1]=rbinom(1,n,T[1])*

*for(i in 2 :Nsim){*

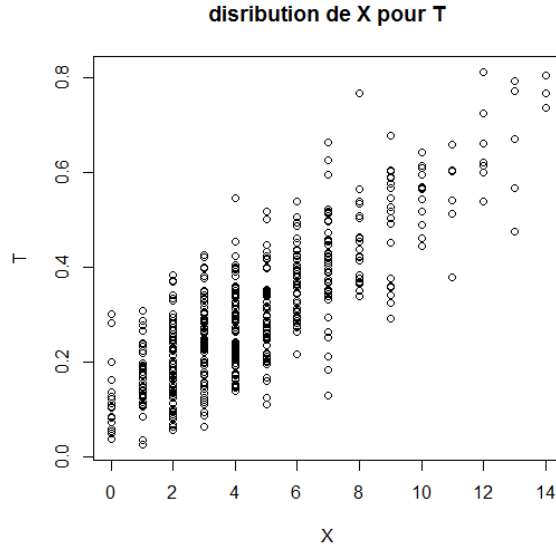
*X[i]=rbinom(1,n,T[i-1])*

*T[i]=rbeta(1,a+X[i],n-X[i]+b)*

*}#boucle d'échantillonnage*

*plot(X,T,main='distribution de X pour T')*





### L'échantillonneur de Gibbs à plusieurs étapes

Considérons le cas élémentaire  $\pi(\theta_1, \theta_2)$ , on suppose  $\pi(\theta_1/\theta_2)$  et  $\pi(\theta_2/\theta_1)$  disponibles, on peut alors générer ce qu'on appellera une séquence de gibbs de la manière suivante :

partant d'une valeur  $\theta_1^{(0)}$ , on génère  $\theta_2^{(0)}$  avec  $\pi(\cdot/\theta_1^{(0)})$  puis  $\theta_1^{(1)}$  avec  $\pi(\cdot/\theta_2^{(0)})$  puis  $\theta_2^{(1)}$  avec  $\pi(\cdot/\theta_1^{(1)})$  ainsi de suite.

Après  $t$  itération de ce schéma, il vient une séquence :  $(\theta_1^{(0)}, \theta_2^{(0)}, \theta_1^{(1)}, \theta_2^{(1)}, \dots, \theta_1^{(t)}, \theta_2^{(t)})$  pour  $t$  assez grand,  $\theta_1^{(t)}$  est une réalisation de  $\theta$ .

Etant donné :  $\theta^{(t)} = (\theta_1^{(t)}, \dots, \theta_p^{(t)})$

1-générer  $\theta_1^{(t+1)} \sim \pi(\theta_1/\theta_2^{(t)}, \dots, \theta_p^{(t)})$

2-générer  $\theta_2^{(t+1)} \sim \pi(\theta_2/\theta_1^{(t+1)}, \theta_3^{(t)}, \dots, \theta_p^{(t)})$

.

.

.

p-générer  $\theta_p^{(t+1)} \sim \pi(\theta_p/\theta_1^{(t+1)}, \theta_2^{(t+1)}, \dots, \theta_{p-1}^{(t+1)})$

Seules les lois conditionnelles de  $\pi$  sont utilisées pour la simulation. Donc, de même

pour un problème de grande dimension, toutes les simulations sont univariées.

**Exemple 2.5 (Casella et George [1990] )**

Soient les distributions conditionnelles suivantes :

$$\pi(\theta_1/\theta_2) \propto \theta_2 \exp(-\alpha\theta_1\theta_2), \quad (\theta_1 > 0)$$

$$\pi(\theta_2/\theta_1) \propto \theta_1 \exp(-\alpha\theta_2\theta_1), \quad (\theta_2 > 0)$$

On ne peut pas calculer la loi marginale de  $\theta_1$  (ni de  $\theta_2$ ) mais les lois-conditionnelles sont des lois exponentielles, donc sont faciles à simuler, comme  $E_\pi(\theta_1/\theta_2) = \theta_2$ , on pourra approcher  $E_\pi(\theta_1)$  par :  $\frac{1}{k} \sum_{i=1}^k \theta_1^{(i)}$  où  $\frac{1}{k} \sum_{i=1}^k \theta_2^{(i)}$ , les deux estimateurs sont totalement équivalents d'un point de vue probabiliste et convergent vers la même valeur.

### 2.4.4 Comparaison d'échantillonnage de Gibbs avec Metropolis-Hasting

-L'échantillonneur de gibbs peut être vu comme un analogue de l'algorithme de Metropolis-Hasting.

-La proposition concerne à tour de rôle chacune des dimensions de  $\Theta$ .

-La probabilité d'acceptation est systématiquement de 1.

-Il faut cependant être capable de déterminer ( ou simuler suivant) les lois  $P(\theta_j/X; \Theta_{-j})$ .

## 2.5 MCMC et convergence

Les résultats généraux que nous avons présenté sur la convergence des méthodes MCMC sont de nature asymptotique et, en pratique, le contrôle de convergence des approximations particulières produites par les méthodes MCMC est une question difficile. L'une d'entre elle consiste à tenter de prouver, rigoureusement des bonnes (probabilités) sur l'écart entre les approximations et leur limite. De telles bornes sont généralement difficiles à obtenir, et pour être exploitables en pratique, on distingue premièrement la convergence de MC puis la convergence de MCMC.

### 2.5.1 Convergence de Monte-Carlo

C'est la loi forte des grands nombres qui permet de justifier la convergence de la méthode, et le théorème centrale limite qui précise la vitesse de convergence.

#### Définition 2.1

L'erreur est définie par :  $\varepsilon_n = E(\theta) - \frac{1}{n}(\theta_1 + \theta_2 + \dots + \theta_n)$ .

Le théorème de la limite centrale donne un asymptotique l'erreur  $\varepsilon_n$ , mais de nature aléatoire. Il dit que la loi de l'erreur finie par ressembler à une loi gaussienne centrée.

Cela signifie que (d'après le théorème centrale limite). Si  $X$  est une variable aléatoire de loi  $\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$  et si  $f$  est une fonction continue bornée,  $E\left(f\left(\frac{\sqrt{n}}{\delta}\varepsilon_n\right)\right)$  converge vers

$$E(f(X)) = \int_{-\infty}^{+\infty} f(x) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx$$

#### Remarque 2.3

On peut aussi montrer que pour tout  $C_1$  et  $C_2$

$$\lim_{n \rightarrow +\infty} P\left(\frac{\delta}{\sqrt{n}}C_1 \leq \varepsilon_n \leq \frac{\delta}{\sqrt{n}}C_2\right) = \int_{C_1}^{C_2} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx$$

Dans les applications, on oublie le passage à la limite, et on remplace  $\varepsilon_n$  par une gaussienne centrée de variance  $\frac{\delta^2}{n}$

#### Exemple 2.6

On cherche à calculer  $I = E(\exp(\beta X))$ , avec  $X$  variable aléatoire gaussienne centrée réduite. Alors le calcul exacte montre que  $I = \exp\left(\frac{\beta^2}{2}\right)$  par la méthode de Monte-Carlo, on pose :  $X' = \exp(\beta X)$ ; donc  $I = E(X')$  la variance de la variable aléatoire  $X'$  vaut  $\delta^2 = \exp(2\beta^2) - \exp(\beta^2)$  l'erreur relative moyenne est :

$$\begin{aligned} e &= \frac{\delta}{I\sqrt{n}} \\ &= \sqrt{\frac{\exp(\beta^2) - 1}{n}} \end{aligned}$$

donc :  $n \simeq \frac{\exp(\beta^2) - 1}{\exp(2)} e^{-2}$  si on souhaite avoir  $e = 1$ , et si  $\beta = 5$ , on doit prendre  $n = 7.10^{10}$  une technique de réduction de variance sera utile dans ce cas.

En générale l'inégalité (2.3) convergs vers

$$\int_{\Theta} \frac{h(\theta) \pi(\theta/x) d\theta}{\int_{\Theta} \pi(\theta/x) d\theta}$$

## 2.5.2 Convergence des algorithmes de MH et Gibbs

Puisque l'algorithme de gibbs est un cas particulier de MH alors :

-Ergodicite : pour  $h$  tel que  $E_{\pi}(|h(\theta)|) < \infty$

$$\lim_{T \rightarrow +\infty} \frac{1}{T} \sum_{t=1}^T h(\theta^{(t)}) = \int h(\theta) f(\theta) d\theta$$

-Converge en variance totale

$$\lim_{n \rightarrow +\infty} \left\| \int K^n(\theta, \cdot) \mu(d\theta) - f \right\|_{TV} = 0$$

pour la loi initiale  $\mu$ ,  $K^n(\theta, \cdot)$  est le noyau de la chaîne après  $n$  transitions. En particulier

$$\lim_{t \rightarrow +\infty} P[\theta^{(t)} \in A] = \int_A f(\theta) d\theta$$

# Chapitre 3

## Applications

### 3.1 Introduction

Dans ce chapitre, nous utilisons des techniques de calcul basées sur la simulation, car il est un élément essentiel des outils et des statistiques, sous forme de programmes en R, nous appliquons ces techniques de deux façons. Au début, nous avons parlé à Monte-Carlo. Ensuite Metropolis-Hasting (MCMC).

### 3.2 Applications de Monte-Carlo (MC)

Dans cette section, nous avons simulés la méthode de Monte-Carlo à l'aide de deux distributions (Normale et Binomiale).

#### 3.2.1 Simulation de MC avec une distribution normale

► Si nous avons un stock des livres avec un taux de rendement gaussien (Normale). Le taux de rendement moyen est de 9% et l'écart type est de 12%, cela signifie essentiellement que notre stock aura une distribution normale du taux de rendement centrée à 9. La fonction "*rnorm*" générera des valeurs autour de 9 avec un écart-type de 12 qui se répartissent comme une distribution normale (voir annexe, programme 1 partie 1)

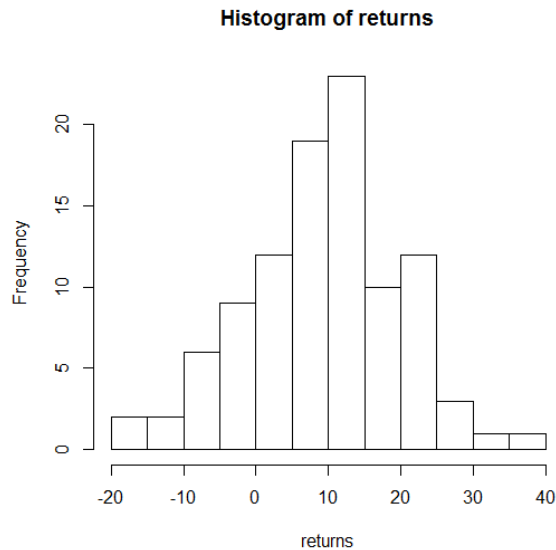


Figure 1 : Resultats de simulations de 100 tirages de la loi normale (9,12)

### Interprétation

La courbe montre la répartition des valeurs avec le schéma d'émergence de données colonnaire sur la forme de distribution normale. Autrement dire les retours vectoriels contiennent 100 nombres avec la moyenne 9 et l'écart-type 12, nous pouvons voir que les retours sont répartis comme une distribution normale.

►Maintenant, supposons que notre retour est annuel et nous aimerions voir quel sera notre retour en 4 ans. Nous pouvons commencer par choisir au hasard 4 échantillons à partir du vecteur de retour (voir annexe, programme 1 partie 2)

[1] 1.307647 10.492049 -4.440902 10.150590

Ci-dessus, on a des retours pour une année jusqu'à 4 années (voir annexe, programme 1 partie 3)

[1] 8.144335 7.739717 -9.631079 23.731298

Vous pouvez voir que les rendements sont différents cette fois, alors quels scénario nous devrions anticiper ? pourquoi ne nous considérons par un grand nombre de scénario ? Nous pouvons estimer le taux de rendement par échantillonnage à partir du vecteur de retour

pour un grand nombre de fois c-à-d 10000 fois, puis prendre la moyenne pour chaque année. Il s'agit d'une simulation et un échantillonnage aléatoire répété pour obtenir des résultats numériques (voir annexe, programme 1 partie 4)

	[,1]	[,2]	[,3]	[,4]
[1,]	2.325729	-17.949140	7.739717	21.788331
[2,]	2.141447	-10.704193	20.901110	20.311747
[3,]	10.492049	-9.631079	12.458658	6.549687
[4,]	15.688964	10.507150	-1.975553	-16.344384
[5,]	10.492049	3.305106	16.445481	1.147742
[6,]	9.344556	1.147742	26.296905	15.026501

Ci-dessus est un ensemble de données de 10000 livres de rendement pour chaque année, nous pouvons prendre la colonne pour voir ce que nous pouvons attendre pour chaque année (voir annexe, programme 1 partie 5)

[1] 9.036440 8.951381 8.845277 8.838054

Nous pourrions échantillonner un million de fois au lieu 10000, donc il faut régler la moyenne et le sd de la distribution, du taux de rendement.

### 3.2.2 Simulation de MC avec une distribution binomiale

On a la probabilité binomiale suivante :

$$P(X / N, r) = C_X^N r^X (1 - r)^{N-X}$$

►Il existe de nombreux outils facilement disponibles dans différents ensembles statistiques qui peuvent être utilisés pour représenter des échantillons des distributions standard (voir annexe, programme 2 partie 1)

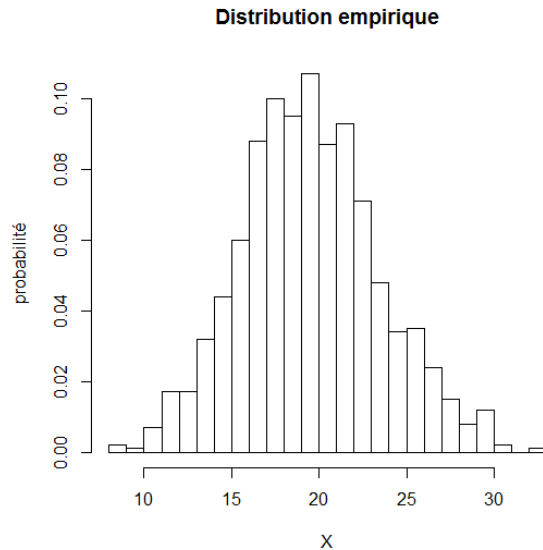


Figure 2 : Resultats de simulations de 1000 tirages du binomial (100,0.2)

### Interprétation

Générera 1000 valeurs aléatoires de  $X$  à partir d'une distribution Binomial  $(N,r)$  et tracer la répétition empirique le plus d'échantillonné, donc la distribution empirique représentera la vraie distribution.

►Si la distribution binomiale n'est pas disponible, il est nécessaire d'avoir le nombre aléatoire avec une distribution uniforme. Cette distribution uniforme fondamentale se trouve habituellement dans la plupart des logiciels. Dans les ordinateurs, les «valeurs aléatoires» ne sont pas vraiment aléatoires, mais pseudo-aléatoires, c'est-à-dire produit par un algorithme déterministe. Une fois que nous avons un générateur de variables  $U(0, 1)$ , alors en principe, nous pouvons générer toutes les autres distributions de manière plus ou moins rapide et plus ou moins précise. Et une fois que nous pouvons simuler une variable aléatoire  $X$  de sa distribution, il est très facile d'obtenir des distributions empiriques de toute transformation de  $X$ . En utilisant la simulation Monte-Carlo, nous avons pas besoin de résoudre la forme analytique de la densité de probabilité de la variable transformée  $g(X)$  (voir annexe, programme 2 partie 2)



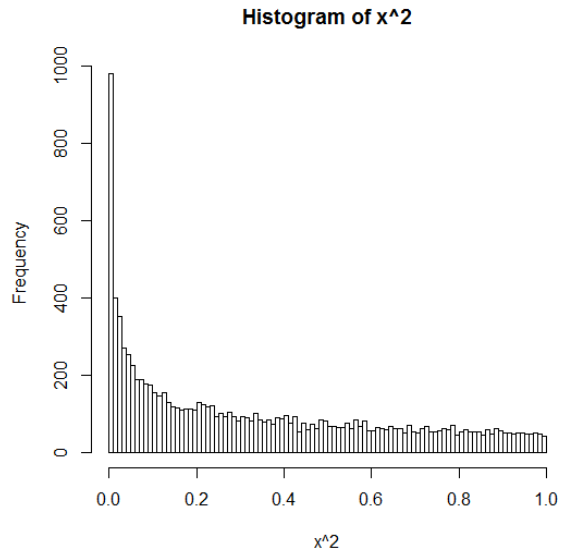


Figure 3 : Simulation de la densité  $g(x)=x^2$

### 3.3 Applications de MCMC

Bien qu'il existe certainement des bons logiciels pour faire la simulation, notamment BUGC ou JAGS. Dans ce travail nous donnons un exemple pour une régression linéaire, échantillonnée par MCMC.

Pour la création des données d'essai, Supposons une relation linéaire entre le prédicteur et la variable de réponse, nous prenons donc un modèle linéaire et ajoutons du bruit (voir annexe, programme 3 partie 1)

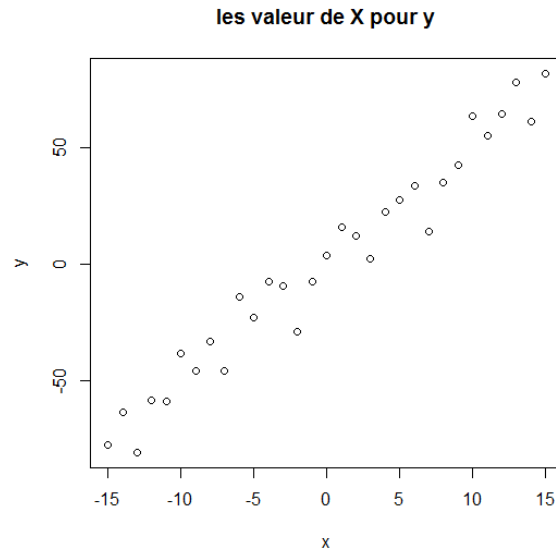


Figure 4 : Le nuage des points

### Interprétation

Cette courbe représente un nuage des points qui sont répartis sur l'axe  $x=y$ .

#### 3.3.1 La fonction de vraisemblance pour le modèle

Pour estimer les paramètres dans une analyse bayésienne, nous devons dériver la fonction vraisemblance pour le modèle que nous voulons adapter. La probabilité avec laquelle nous nous attendons à ce que les données observées se produisent conditionnellement aux paramètres du modèle que nous examinons. Donc, étant donnée que notre modèle linéaire  $y = ax + b + N(0, sd)$  prend les paramètres  $(a, b, sd)$ , en tant que contribution, nous devons retourner à la probabilité d'obtenir les données ci-dessus sous ce modèle. Cela semble plus compliqué que cela, comme vous le voyez dans le code, nous calculons simplement la différence entre les valeurs théoriques  $y = ax + b$ . Et la valeur observée  $y$ , et nous devons rechercher les densités de probabilité (en utilisant `dnorm`) pour que de telles déviations se produisent (voir annexe, programme 3 partie 2)

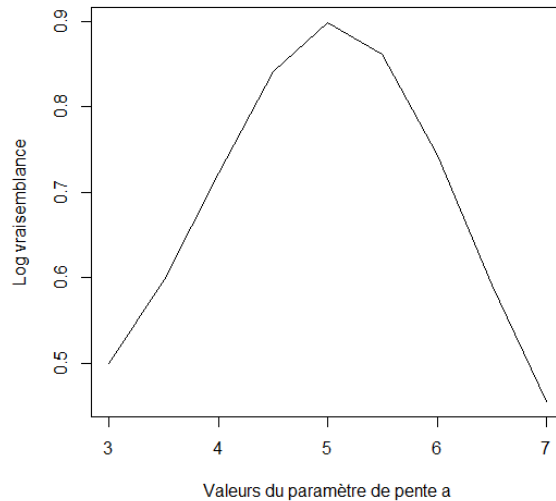


Figure 5 : valeurs du parametre de pente a

### Interprétation

Comme d'illustration, les valeurs du paramètre de pente a. Le résultat devrait une normale ressembler à l'intrigue à droite.

### 3.3.2 Comment définir l'a priori ?

La deuxième étape, c'est comme toujours dans les statistiques bayésiennes, nous devons spécifier la distribution a priori pour chaque paramètre. Pour simplifier, j'ai utilisé des distributions uniformes et des distributions normales pour les trois paramètres. Afin d'introduire la loi a priori dans logiciel R on utilise le (programme 3 partie 3)

### 3.3.3 Comment définir l'a posteriori ?

Le produit de l'a priori et de la vraisemblance est la quantité réelle sur laquelle le MCMC fonctionnera. Cette fonction s'appelle l'a posteriori, nous travaillons ici avec la somme parce que nous avons des logarithmes. Afin d'introduire la loi a posteriori dans logiciel R on utilise le (programme 3 partie 4)

### Remarque 3.1

*Lorsque nous utilisons la loi a priori (resp a posteriori) assez pour appeler la fonction priori (resp posteriori).*

### 3.3.4 Application de M-H

En principe, l'algorithme peut être utilisée pour échantillonner a partir de n'importe quelle fonction intégrable. Ceci est réalisé par :

1-À partir d'une valeur de paramètre aléatoire.

2-Choisir une nouvelle valeur de paramètre proche de l'ancienne valeur en fonction d'une certaine densité de probabilité appelée fonction de proposition.

3-Sauter à ce nouveau point avec une probabilité  $P(\text{nouveau}) / p(\text{ancien})$ , Où  $p$  est la fonction cible, et  $p > 1$  signifie sauter aussi

Il est logique de penser à ce qui fonctionne, mais pour le moment, je peux vous assurer que - lorsque nous exécutons cet algorithme, la distribution des paramètres qu'il visite converge vers la distribution cible  $p$ . Donc, avec le logiciel R (voir annexe, programme 3 partie 5) on a le résultat suivant :

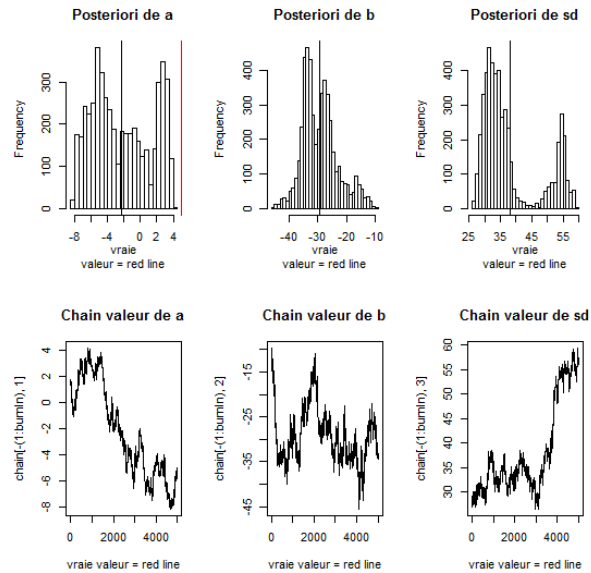


Figure 6 : La range superieur montre les estimations a posteriori pour la pente (a), l'interception (b) et l'ecart-type de l'erreur (sd). La range inferieur montre la chaine de Markov des valeurs des parametres

## Interprétation

Les premières étapes de l'algorithme peuvent être biaisées par la valeur initiale, et sont donc généralement mis au rebut pour une analyse plus approfondie. Une sortie intéressante à examiner est le taux d'acceptation qui peut être influencé par la fonction de proposition : généralement, plus les propositions sont proches, plus le taux d'acceptation est important. Les taux d'acceptation très élevés, cependant, ne sont généralement pas les deux : ce qui signifie que les algorithmes sont "rester" au même point, ce qui entraîne une sondage sous-optimal de l'espace des paramètres (mélange). On peut afficher les taux d'acceptations entre 20% et 30% sont optimales pour les applications typiques (plus ici). Enfin, nous pouvons tracer les résultats. Il existe des façons plus élégantes de faire des graphes. Les parcelles résultantes devraient ressembler à l'intrigue ci-dessous. Vous voyez

que nous récupérons plus ou moins les paramètres d'origine qui ont été utilisés pour créer nos données, et vous voyez également que nous obtenons une certaine zone autour des valeurs a posteriori les plus élevées qui ont également un certain soutien par les données, qui est l'équivalent des intervalles de confiances bayésiennes, et voilà les caractéristiques de notre modèle linéaire.

# en comparaison :

Call :

lm(formula = y ~ x)

Residuals :

Min	1Q	Median	3Q	Max
-20.3045	-8.6165	-0.1333	7.5156	20.5284

Coefficients

	Estimate	Std.Error	t value	Pr(> t )
(Intercept)	-1.8053	1.5442	-1.169	0.252
x	5.0975	0.1726	29.527	<2e-16 ***

Signif. codes 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error 11.1 on 29 degrees of freedom

Multiple R-squared 0.9456 Adjusted R-squared 0.9437

F-statistic 504.1 on 1 and 29 DF p-value < 2.2e-16

# Conclusion générale

Dans ce travail, nous avons étudié la théorie mathématique des méthodes de Monte Carlo par chaîne de Markov. Nous avons montré leurs intérêts et les conclusions significatives.

Les méthodes de Monte Carlo par chaînes de Markov permettent de modéliser beaucoup de problèmes. Leur nature probabiliste et hiérarchique convient parfaitement aux méthodes d'inférence bayésienne. Parmi celles-ci, l'échantillonneur de Metropolis-Hastings est souvent utilisé car il peut être appliqué dans la plupart des cas et notamment aux modèles linéaire de la forme  $y=ax+b$ .

Nous avons constaté sur des simulations peut converger beaucoup trop lentement pour être utilisé en pratique. Pour cela il est nécessaire d'étudier les théorèmes de limite centrale pour les procédures MCMC ce qui demande de profonds développements et sera abordé dans le futur.

# Résumé

Dans ce travail, nous avons essayé d'exposer quelques notions et avantages de l'approche bayésienne.

L'objectif de ce travail est d'étudier quelques méthodes de simulations exactement les méthodes de Monte Carlo par chaînes de Markov (MCMC).

Nous avons utilisé une modélisation bayésienne parce qu'elle offre plus de souplesse et d'objectivité dans les calculs statistiques.

Nous avons fait aussi quelques applications numériques et des méthodes de simulations (méthode de Monte Carlo, méthode de Monte Carlo par chaînes de Markov) illustrées à l'aide de langage R.

**Mots clés :**

Approche bayésienne, méthodes de simulations, méthodes de Monte Carlo par chaînes de Markov (MCMC), méthodes de Monte Carlo (MC).



## Abstract

In this work, we tried to expose some notions and advantages of the Bayesian approach.

The objective of this work is to study some methods of simulations exactly Monte Carlo methods by Markov chains (MCMC).

We used Bayesian modeling because it offers more flexibility and objectivity in statistical calculations.

We have also done some numerical applications and simulation methods (Monte Carlo method, Monte Carlo method using Markov chains) illustrated using R-language.

**Keywords :**

Bayesian approach, simulation methods, Monte Carlo methods using Markov chains (MCMC), Monte Carlo (MC) methods.

# Annexe

## Les lois de probabilités

►Loi Uniforme  $U(a, b)$  La densité est donnée par :

$$f(x) = \frac{1}{b-a} 1_{]a,b[}(x), \quad -\infty < a < b < +\infty$$

Les moments de  $x$  sont :

$$\begin{cases} E(x) = \frac{a+b}{2} \\ Var(x) = \frac{(b-a)^2}{12} \end{cases}$$

►Loi exponentielle  $\exp(\lambda)$

La densité est donnée par :

$$f(x) = \lambda \exp(-\lambda x) 1_{\mathbb{R}_+}(x), \quad x > 0, \lambda > 0$$

Les moments de  $x$  sont :

$$\begin{cases} E(x) = \frac{1}{\lambda} \\ Var(x) = \frac{1}{\lambda^2} \end{cases}$$

►Loi Gamma  $G(\alpha, \beta)$

La densité est donnée par :

$$f(x/\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x) 1_{\mathbb{R}_+^*}(x), \quad \alpha, \beta \in \mathbb{R}_+^*$$

Les moments de  $x$  sont :

$$\begin{cases} E(x) = \frac{\alpha}{\beta} \\ Var(x) = \frac{\alpha}{\beta^2} \end{cases}$$

►Loi Normale  $N(\mu, \sigma^2)$

La densité est donnée par :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \mu \in \mathbb{R}, \sigma \in \mathbb{R}_+^*$$

Les moments de  $x$  sont :

$$\begin{cases} E(x) = \mu \\ Var(x) = \sigma^2 \end{cases}$$

►Loi Bernoulli  $B(p)$

La densité est donnée par :

$$f(x/p) = p^x (1-p)^{1-x}, p \in ]0, 1[$$

Les moments de  $x$  sont :

$$\begin{cases} E(x) = p \\ Var(x) = p(1-p) \end{cases}$$

►Loi Binomiale  $B(n, p)$

La densité est donnée par :

$$P(X = x) = C_n^x p^x (1-p)^{n-x}, n \in \mathbb{N}^*, p \in ]0, 1[$$

Les moments de  $x$  sont :

$$\begin{cases} E(x) = np \\ Var(x) = np(1-p) \end{cases}$$

►Loi Beta  $B(a, b)$

La densité est donnée par :

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1-x)^{b-1} 1_{[0,1]}(x), a, b > 0$$

Les moments de  $x$  sont :

$$\begin{cases} E(x) = \frac{a}{a+b} \\ Var(x) = \frac{ab}{(a+b)^2(a+b+1)} \end{cases}$$

►Loi Beta-Binomiale *BetaBin* ( $n, a, b$ )

La densité est donnée par :

$$f(x) = \frac{\Gamma(x+b)\Gamma(n-x+b)\Gamma(n+1)\Gamma(a+b)}{\Gamma(x+1)\Gamma(n-x+1)\Gamma(a)\Gamma(b)\Gamma(n+a+b)}, \text{ où } n = 1, 2, \dots, a, b > 0$$

Les moments de  $x$  sont :

$$\begin{cases} E(x) = \frac{na}{a+b} \\ Var(x) = n(n+a+b) \frac{ab}{(a+b)^2(a+b+1)} \end{cases}$$

►Loi khi-deux  $\chi_n^2$

La khi-deux à  $n$  degrés de liberté est un cas particulier de la gamma :  $\chi_n^2 \sim \Gamma\left(\frac{1}{2}, \frac{n}{2}\right)$ .

La densité est donnée par :

$$f_{\chi^2}(x/n) = \begin{cases} \frac{\left(\frac{1}{2}\right)^{\frac{n}{2}} x^{\frac{n}{2}-1} \exp\left(-\frac{1}{2}x\right)}{\Gamma\left(\frac{n}{2}\right)} & \text{pour } x > 0 \\ 0 & \text{sinon} \end{cases}$$

Les moments de  $x$  sont :

$$\begin{cases} E(x) = n \\ Var(x) = 2n \end{cases}$$

►Loi khi-deux -inverse  $\chi_n^{-2}$

De même, une khi-deux-inverse est la loi de l'inverse d'une variable khi-deux. c'est donc un cas particulier de la gamma-inverse  $\chi_n^{-2} \sim \Gamma^{-1}\left(\frac{1}{2}, \frac{n}{2}\right)$ .

La densité est donnée par :

$$f_{\chi^{-2}}(x/n) = \begin{cases} \frac{\left(\frac{1}{2}\right)^{\frac{n}{2}} \left(\frac{1}{x}\right)^{\frac{n}{2}+1} \exp\left(-\frac{1}{2x}\right)}{\Gamma\left(\frac{n}{2}\right)} & \text{pour } x > 0 \\ 0 & \text{sinon} \end{cases}$$

Les moments de  $x$  sont :

$$\begin{cases} E(x) = \frac{1}{n-2}, & n > 2 \\ Var(x) = \frac{2}{(n-2)^2(n-4)}, & n > 4 \end{cases}$$

### ►Le modèle linéaire

$y = ax + b$  avec  $a \in \mathbb{R}^*$  et  $b \in \mathbb{R}$  est la relation linéaire Entre  $x$  et  $y$  et un modèle d'erreur normal  $N(0, sd)$  Avec l'écart type  $sd$ , utilisons donc le même modèle pour l'ajustement et voir si nous pouvons récupérer nos valeurs de paramètres d'origine.

## Les programmes

### Programme 1

#### Partie 1

```
set.seed(1729)
returns=rnorm(1 :100,mean=9,sd=12)#générer 100 valeur
hist(returns)#tracer un histogramme
```

#### Partie 2

```
sample(returns,size=4)
```

#### Partie 3

```
sample(returns,size=4)
```

#### Partie 4

```
X=matrix(ncol=4,nrow=10000)
for(année in 1 :4){#pour chaque année échantillonnez le retour de 10000 fois
for(i in 1 :10000){X[i,année]=sample(returns,1)}}
#afficher les 6 premiers lignes
head(X)
```

#### Partie 5

```
#taux de rendement attendu moyen
apply(X,2,mean)
```

### Programme 2

#### Partie 1

```
N=100
```

```

r=0.2
X<-rbinom(1000,N,r)
hist(X,freq=FALSE,min(X) :max(X),
xlab='X',ylab='probabilité',
main='Distribution empirique')

```

### Partie 2

```

for(i in 1 :1000){X[i]<-sum(runif(100)<0.2)}
x<- runif(10000)
hist(x^2,100)

```

### Programme 3

#### Partie 1

```

a<- 5
b<- 0
trueSd<- 10
n<-31#taille de l'échantillon
# Créer des valeurs X indépendantes
x <- (-(n-1)/2) :((n-1)/2)
# Créer des valeurs dépendantes selon ax + b + N(0,sd)
y <- a * x + b + rnorm(n=n,mean=0,sd=trueSd)
plot(x,y, main="les valeur de X pour y")

```

#### Partie 2

```

probabilité<- fonction(param){
a = param[1]
b = param[2]
sd = param[3]
pred = a*x + b
Possibilitéunique = dnorm(y, mean = pred, sd = sd, log = T)
Somme= sum(Possibilitéunique)
return(Somme)}
#tracer le profil de vraisemblance de la pente a
Valeursdepente<- fonction(x){return(probabilité(c(x,b, trueSd)))}
Unevraisemblance<- lapply(seq(3, 7, by=0.5), Valeursdepente)

```

```
plot (seq(3, 7, by=0.5), Unevraisemblance, type="l", xlab = "Valeurs du paramètre de
pente a", ylab = "Log vraisemblance")
```

### **Partie 3**

```
#La distribution a priori
priori<- function(param){
a = param[1]
b = param[2]
sd = param[3]
apriori = dunif(a, min=0, max=10, log = T)
bpriori = dnorm(b, sd = 5, log = T)
sdpriori = dunif(sd, min=0, max=30, log = T)
return(apriori+bpriori+sdpriori)}
```

### **Partie 4**

```
posteriori<- function(param){
return (likelihood(param) + priori(param))}
```

### **Partie 5**

```
valeur_initiale = c(4,0,10)
proposition_de_fonction<- function(param){
return(rnorm(3,mean = param, sd= c(0.1,0.5,0.3)))}
run_metropolis_MCMC<- function(valeur_initiale ,iterations){
chain = array(dim = c(iterations+1,3))
chain[1,] =valeur_initiale
for (i in 1 :iterations){
proposition= proposition_de_fonction(chain[i,])
probab = exp(posteriori(proposition) - posteriori(chain[i,]))
if (runif(1) < probab){
chain[i+1,] =proposition
}else{
chain[i+1,] = chain[i,]
}
}
}
chain = run_metropolis_MCMC(valeur_initiale,10000)
```

```

return(chain)
}
burnIn = 5000
acceptance = 1-mean(duplicated(chain[-(1 :burnIn),]))
par(mfrow = c(2,3))
hist(chain[-(1 :burnIn),1],nclass=30, , main="Posteriori de a", xlab="vraie
valeur = red line" )
abline(v = mean(chain[-(1 :burnIn),1]))
abline(v = a, col="red" )
hist(chain[-(1 :burnIn),2],nclass=30, main="Posteriori de b", xlab="vraie
valeur = red line")
abline(v = mean(chain[-(1 :burnIn),2]))
abline(v = b, col="red" )
hist(chain[-(1 :burnIn),3],nclass=30, main="Posteriori de sd", xlab="vraie
valeur = red line")
abline(v = mean(chain[-(1 :burnIn),3]) )
abline(v = trueSd, col="red" )
plot(chain[-(1 :burnIn),1], type = "l", xlab="vraie valeur = red line" , main= "Chain
valeur de a", )
abline(h = a, col="red" )
plot(chain[-(1 :burnIn),2], type = "l", xlab="vraie valeur = red line" , main= "Chain
valeur de b", )
abline(h = b, col="red" )
plot(chain[-(1 :burnIn),3], type = "l", xlab="vraie valeur = red line" , main= "Chain
valeur de sd", )
summary(lm(y~x))

```



# Bibliographie

- [1] **A. Guillin, J. M. Marin, et C. P. Robert.** Estimation bayésienne approximative par échantillonnage préférentiel. Université Paris IX Dauphine.
- [2] **Céline Baranger \_ Julien Mathiaud.** 2012/2013. Méthode de Monte-Carlo.
- [3] **Christian P. Robert et George Casella.** 2011. Méthodes de Monte-Carlo avec R. Springer-Verlag France. Université Paris \_ Dauphine, University of Florida.
- [4] **Djaweyda Ghouil et Préparé par : Nadia Bensadia, Widad Biad.** 2012 - 2013. Les méthodes de Monte Carlo par Chaînes de Markov et application. Université de Jijel, Faculte des Sciences Exactes et Informatique, Département de Mathématiques.
- [5] **Irisa -Inria Rennes.** 11/10/2006. Les méthodes de Monte-Carlo. Université Montpellier.
- [6] **Jean Bérard.** Introduction aux méthodes de simulation particulières.
- [7] **Jean \_ Jacques Boreux, Éric Parnet, Jacques Bernier.** 519/163. Pratique du calcul bayésien. Springer-Verlag France, Paris, 2010.
- [8] **Jean \_ Jacques Droesbeke, Jeanne Fine et Gilbert Sporta Éditeur.** (519/105). Méthodes bayésiennes en statistique. Éditions Technip, paris, 2002. Imprimé en France, ISBN 2-7108-0813-7
- [9] **Jean \_ Michel Marin, christain p. Robert.** les bases de la statistique bayésienne. I3M, Université Montpellier 2, Montpellier&Grest, INSEE, paris et Université paris Dauphine, paris&CRest, INSEE, paris
- [10] **Jerome Dupuis (LSP-UPS),** (Version septembre 2007). Statistique bayésienne et algorithmes MCMC
- [11] **Linda OTMANI et Abdlkader BENYETTOU** l'intégration des méthodes de Mont-Carlo par chaînes de Markov dans l'apprentissage des réseaux de neu-

rones. Université de sciences et de technologie d'ORAN. mohamed Bodiaf-Faculté des sciences-Département d'informatique-laboratoire SIMPA.

- [12] **Massé pierre\_yves, Meiniel William et Roussillon Pierre.** Paris, Cachan, Juin 2011. Synthèse de textures par les chaînes de Markov. Lionel Moisan : Université Paris Dextartes.
- [13] **Maxime HERVE.** version 4 2012-version 1 2010. Aide\_mémoire de statistique appliquée à la biologie (construire son étude et analyser les résultats à l'aide du logiciel R).
- [14] **Mme Halil Née Khorsi Rachida.** 10/03/2011. Inférence bayésienne en series chronologiques. Université Mammeri, Tiziouzou, Faculte des sciences département des mathématiques.
- [15] **Nicolas Baud et Vincent Porte.** Mai 2011. Méthdes de Monte-Carlo appliquée à la finance :Echantillonnage d'importance et stratification pour les processus de diffusion. Ecole nationale de la statistique et de l'Administration Economique (ENSAE).
- [16] **Nicolas Grenon\_Godbout.** 7/11/2015. Méthodes MCMC : améliorations d'un algorithme d'adaptation régionale et applications à la chimatologie. Département de Mathématiques et statistique\_ Université de Montréal\_ sous la direction de Mylène Bédard.
- [17] **Olivrer Cappé et Christian P. Robert.** Une approche Monte-Carlo adaptative pour l'approximation de loi a posteriori avec application à l'inférence de paramètres cosmologiques. LTCL, Télécome Paris, CNRS, rue Barraelt, 75634 Paris cedex 13, France Université Paris\_Dauphine, CREST-INSEE, place du Marichel de lattre du Tassigny, 7577 Paris cedex 16, France.
- [18] **Radu Stoica.** 22/10/2014. Modilisation probabiliste et inférence statistique pour l'analyse des données spatialisées. Université Lille1.
- [19] **Sylvain Rubenthaler.** 2016/2017. Méthde de Monte-Carlo. Université Nice Sophia Antipolis.
- [20] **Vincent Mazet.** 22/05/2003. Introduction aux méthodes de Monte-Carlo par chaînes de Markov. Université Herni Poincaré, 54506 Vandoeuvre-lés Nancy cedex.

# ملخص

في هذا العمل حاولنا شرح بعض المفاهيم و مزايا نهج النظرية الافتراضية  
الهدف من هذا العمل هو دراسة بعض طرق المحاكاة باستعمال طريقة المونتي كارلو  
بواسطة سلسلة ماركوف  
استخدمنا نماذج من النظرية الافتراضية لأنها توفر المزيد من المرونة والموضوعية في  
الحسابات الإحصائية  
ونحن أيضا قمنا بتطبيق طرق المحاكاة الرقمية (مونتي كارلو . مونتي كارلو بسلسلة  
ماركوف) والتي تتمثل في استخدام برنامج