



Faculté des Sciences Exactes et Informatique  
Département de Mathématiques

N° d'ordre :

N° de séries :

## Mémoire de fin d'études

Présenté pour l'obtention du diplôme de

### Master

Spécialité : Mathématiques.

Option : Probabilités et statistique.

### Thème

# Application Bayésienne de la méthode RDS : cas d'un modèle AR(1)

Présenté par :

- Saoudi Mohammed - Bouzaout Mohammed

Soutenu le 06 / 07 / 2019

Devant le jury :

Président :	Hassen Cheraitia	Maitre de conférences B	Université de Jijel
Encadreur :	Djoweyda GHouil	Maitre assistante A	Université de Jijel
Examineur :	Khawla Boudjerda	Maitre de conférences B	Université de Jijel

Promotion 2018/2019

## *Remerciements*

Tout travail réussi dans la vie nécessite en premier lieu la Bénédiction de **Dieu**, et ensuite l'aide et le support de plusieurs personnes. Nous tenons donc à remercier et à adresser notre reconnaissance à toute personne qui nous a aidé de loin ou de près afin de réaliser ce travail.

Nous voudrions tout d'abord adresser toute notre gratitude à la directrice de ce mémoire Madame **Ghouil Djoweyda** pour sa précieuse présence assistance, sa disponibilité et l'intérêt qu'elle a manifesté pour ce modeste travail, nous la remercions pour ses orientations et son enthousiasme envers notre travail. Les judicieux conseils et rigueur qu'elle nous a prodigué tout au long de ce travail. Nous avons pris un grand plaisir en travaillant avec elle.

Nous adressons également, nos remerciements chaleureux à Monsieur **Cheraitia Hassen** pour avoir accepté la présidence du jury de ce mémoire et pour son propre aide, nos remerciements vont également à Madame **Boudjerda Khawla** qui a bien voulu prendre la responsabilité d'évaluer ce travail. Merci aussi à tous nos collègues et amis, nous leurs exprimons notre amitié et nous leur souhaitons beaucoup de réussites.

Enfin, nous tenons à exprimer notre reconnaissance à Monsieur **Bensouilah Bachir**, et à tous les enseignants du département de mathématiques de l'université de Jijel, et en particulier Monsieur **Gherda Meabrouk** et Madame **Sellami. N.**

Un grand merci à nos familles, la famille **Saoudi** et **Bouzaout** qui nous ont toujours soutenues et à tous ceux qui ont participé à la réalisation de ce mémoire.

Mohammed & Mohammed.

---

# TABLE DES MATIÈRES

<b>Introduction générale</b>	<b>vi</b>
<b>1 Estimation Bayésienne et méthodes de Monte Carlo</b>	<b>1</b>
1.1 L'estimation paramétrique Bayésienne . . . . .	3
1.1.1 Introduction . . . . .	3
1.1.2 Principe de la statistique Bayésienne . . . . .	3
1.1.3 Notations et définitions . . . . .	4
1.1.4 Choix de la Loi a priori . . . . .	5
1.1.5 La Loi a posteriori . . . . .	7
1.1.6 Quelques notion de la Théorie de décision . . . . .	10
1.1.7 Estimateur Bayésien . . . . .	12
1.2 Les méthodes de Monte Carlo . . . . .	17
1.2.1 Introduction . . . . .	17
1.2.2 Description de la méthode . . . . .	17
1.2.3 Convergence de la méthode . . . . .	19
1.2.4 Simulation de Monte Carlo . . . . .	20

---

1.2.5	Échantillonnage Aléatoire . . . . .	24
1.2.6	Échantillonnage descriptif (ED) ou (DS) . . . . .	25
1.2.7	L'échantillonnage descriptif amélioré (EDA) ou (RDS) . . . . .	27
<b>2</b>	<b>Application de la méthode RDS sur un modèle AR(1)</b>	<b>30</b>
2.1	Notions de base en séries chronologiques . . . . .	32
2.1.1	Introduction . . . . .	32
2.1.2	Les composantes principales d'une série chronologique . . . . .	33
2.1.3	Stationnarité d'une série temporelle . . . . .	34
2.1.4	Processus aléatoires . . . . .	35
2.1.5	Le processus bruit blanc . . . . .	36
2.1.6	Le processus autorégressif d'ordre $p$ : AR( $p$ ) . . . . .	37
2.1.7	Le processus moyenne mobile d'ordre $q$ : MA( $q$ ) . . . . .	38
2.1.8	Le processus mixte ARMA( $p,q$ ) . . . . .	38

---

# LISTE DES TABLEAUX

1.1	Lois conjuguées usuelles. . . . .	6
1.2	Résultats de l'exemple de la méthode ED . . . . .	26
1.3	Résultats de l'exemple de la méthode RDS . . . . .	28
2.1	Les valeurs simulées de $\hat{\rho}$ et des variances pour $n = 10$ . . . . .	47
2.2	Les valeurs simulées de $\hat{\rho}$ et des variances pour $n = 30$ . . . . .	47
2.3	Les valeurs simulées de $\hat{\rho}$ et des variances pour $n = 50$ . . . . .	48
2.4	L'estimateur RDS général de $\hat{\rho}$ pour $n = 10$ . . . . .	48
2.5	L'estimateur RDS général de $\hat{\rho}$ pour $n = 30$ . . . . .	49
2.6	L'estimateur RDS général de $\hat{\rho}$ pour $n = 50$ . . . . .	49
2.7	Les valeurs simulées de $\hat{\theta}$ et des variances pour $n = 10$ . . . . .	50
2.8	Les valeurs simulées de $\hat{\theta}$ et des variances pour $n = 30$ . . . . .	50
2.9	Les valeurs simulées de $\hat{\theta}$ et des variances pour $n = 50$ . . . . .	51
2.10	L'estimateur RDS général de $\hat{\theta}$ . . . . .	51

---

## TABLE DES FIGURES

1.1	Principe de la statistique Bayésienne. . . . .	4
1.2	densité de la loi exponentielle de paramètre 1 pour $n=100$ . . . . .	21
1.3	densité de la loi de Cauchy de paramètre 1 pour $n=100$ . . . . .	22
2.1	Les données de la série . . . . .	40
2.2	La série chronologique . . . . .	40
2.3	Auto corrélation (ACF). . . . .	40
2.4	Auto corrélation partial (pacf). . . . .	41
2.5	Résultat d'estimation pour le modèle $MA(1)$ . . . . .	42
2.6	Résultat d'estimation pour le modèle $AR(1)$ . . . . .	42

---

# INTRODUCTION GÉNÉRALE

Tandis qu'avant les années 1990, l'inférence Bayésienne a été au mieux une entreprise difficile dans la pratique, réservée à un nombre restreint de chercheurs spécialisés et limitée à un nombre plutôt restreint des modèles ; Elle est devenue dans nos jours un procédé très accessible et performant qui peut assez facilement être appliqué à presque tout type de modèle. Et tout cela grâce aux algorithmes de calcul qui sont globalement numériques, comme les méthodes de Mont Carlo (*MC*) et les méthodes de Mont Carlo par chaînes de Markov (*MCMC*).

Pour un modèle paramétrique bien défini de densité où  $\theta$  est le paramètre inconnu indiquant la densité  $f(x/\theta)$ , la démarche Bayésienne consiste à traiter le paramètre  $\theta$  comme une variable aléatoire en lui associant une loi de probabilité sur l'espace  $\Theta$  dite loi a priori et noté  $\pi(\theta)$ . Cette loi représente pour un statisticien Bayésien l'ensemble des informations a priori disponible sur le paramètre  $\theta$  et dans un contexte pratique, elle regroupe aussi l'ensemble des opinions des experts.

L'approche Bayésienne est très flexible aussi pour l'analyse des séries temporelles. Les méthode *MC* et *MCMC* ont rendu même les méthodes les plus compliqués des séries temporelles favorables à l'analyse bayésienne. Beaucoup d'économistes ont trouvé les méthodes bayésiennes attrayantes non seulement pour des raison philosophiques, mais surtout pour leurs efficacité.

Pour plusieurs applications des modèles auto-régressifs et en raison du nombre important de paramètres impliqués les chercheurs préfèrent souvent d'utiliser des loi a priori non informatives malgré que l'examen de l'effet de cette utilisation sur les lois a posteriori

est relativement rare, voir Kadiyala et Karlsson (1997), Ni et Sun (2003).

Généralement ils travaillent avec une loi a priori constante pour les coefficients de régression et une loi de Jeffreys pour la matrice des covariance des erreurs, presque la combinaison de ces lois a priori conduit à des simulations simples de la loi a posteriori.

L'objectif de ce mémoire est d'une part de se familiariser un peu sur l'approche bayésienne et d'autre part de montrer l'efficacité des méthodes *MC* et plus particulièrement la méthode *RDS* pour l'estimation bayésienne des paramètres d'un modèles *AR(1)* comme type de série temporelle.

Ainsi ce mémoire est réalisé en deux chapitres, chaque chapitre contient deux parties essentielles. Le premier chapitre aborde dans la première partie quelques notions de base sur l'approche bayésienne, et dans la deuxième partie, une présentation du principe de fonctionnement d'une méthode *MC*, en se basant sur la méthode *RDS*.

Concernant le deuxième chapitre, nous avons donné une brève présentation sur les séries temporelles dans la première partie, et dans la deuxième nous avons effectué une estimation bayésienne des paramètres d'un modèle *AR(1)*, en appliquant la méthode *RDS*.

---

---

# CHAPITRE 1

---

## ESTIMATION BAYÉSIENNE ET MÉTHODES DE MONTE CARLO

**Partie 1 :**  
**L'estimation paramétrique**  
**Bayésienne**

## 1.1 L'estimation paramétrique Bayésienne

### 1.1.1 Introduction

Pour faire une analyse d'une modèle paramétrique l'estimation des paramètres inconnus est une étape essentielle et inévitable. Dans un sens Bayésien l'estimation d'un paramètre repose sur la minimisation du risque a posteriori dans l'espace des paramètres. Pour plusieurs applications des procédures Bayésiennes, la moyenne a posteriori est déclaré comme un estimateur Bayésien, mais il faut prendre acte que la moyenne a posteriori comme estimateur Bayésien est optimale que pour certaines fonctions de perte. Toutefois donc, la fonction de perte détermine la forme de l'estimateur Bayésien.

### 1.1.2 Principe de la statistique Bayésienne

La statistique Bayésienne est un ensemble de méthodes statistiques qui permettent de se plonger plus loin dans la définition des méthodes et dans l'inclusion des données a priori et de connaissance a priori (fusion des données avec des bases de connaissance).

Alors que la statistique classique repose sur la loi des observations, la statistique Bayésienne repose sur la loi a posteriori. La loi a posteriori peut s'interpréter comme un résumé (en un sens probabiliste) de l'information disponible sur  $\theta$ , une fois  $X$  observé. L'approche Bayésienne réalise en quelque sorte l'actualisation de l'information a priori par l'observation  $X$ , au travers de  $\pi(\theta/X)$ .

Le schéma ci-dessous résume la démarche Bayésienne dans le cadre de la statistique paramétrique inférentielle. Il faut également apparite, la modélisation stochastique des  $x_i$  comme étant des réalisations de variables aléatoires  $X_i$  (cette modélisation a priori disponible sur le paramètre  $\theta$ , au traverse de la loi a posteriori).[9]

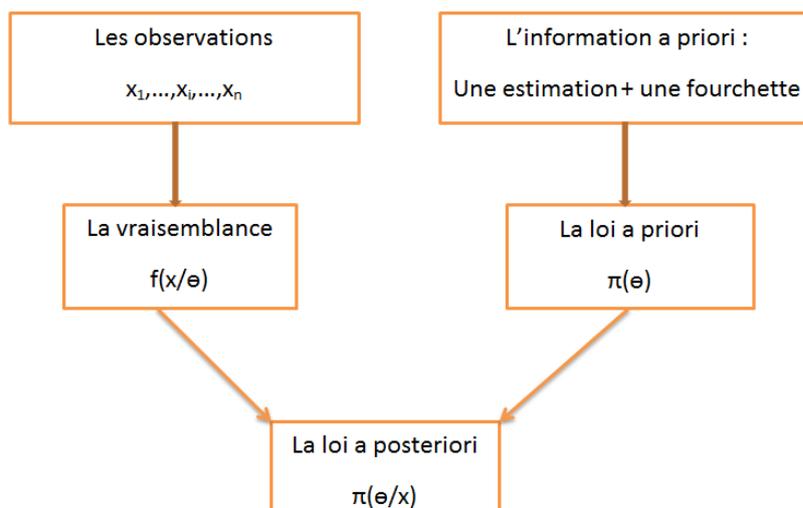


FIGURE 1.1 – Principe de la statistique Bayésienne.

### 1.1.3 Notations et définitions

Pour plus de clarté dans ce chapitre on prend les notations suivantes pour les différentes distributions rencontrées. Au long de ce mémoire, nous allons utiliser les notations suivantes :

- On dispose d'un échantillon  $X$ , tel que :  $X = (x_1, \dots, x_n)$  de taille  $n$ . Les observations  $x_i$  sont donc considérées comme des réalisations de variables aléatoires, notées  $X_i$ .
- $\Theta$  : l'espace des états de la nature (l'espace des paramètres dans le cas d'un problème statistique).
- $\pi(\theta)$  pour la distribution a priori de  $\theta$ .
- $f(x/\theta)$  pour la vraisemblance des données  $x$ .
- $\pi(\theta/x)$  pour la distribution a posteriori de  $\theta$ .
- $m(x)$  pour la distribution marginale de  $x$ .
- $L(\delta, \theta)$  pour une fonction de perte, tel que :  
 $\delta$  la règle de décision.

#### Définitions 1.1.

- On entend par information a priori sur le paramètre  $\theta$  toute information disponible sur  $\theta$  en dehors de celle apportée par les observations.

- *L'information a priori sur  $\theta$  est entaché d'incertitude (si ce n'était pas le cas, le paramètre  $\theta$  serait connu avec certitude et on n'aurait pas à l'estimer).*
- *Il est naturel de modéliser cette information a priori à travers d'une loi de probabilité, appelée loi a priori sa densité est notée  $\pi(\theta)$ .*
- *On fera dans ce chapitre systématiquement l'hypothèse que sachant  $\theta$ , les variables aléatoires  $X_i$  sont indépendantes ; Autrement dit on aura :  $f(x/\theta) = \prod_{i=1}^n f(x_i/\theta)$  est la vraisemblance de l'échantillon.[9]*

### 1.1.4 Choix de la Loi a priori

Le plus souvent on ne dispose pas de suffisamment d'information a priori sur le paramètre inconnu  $\theta$  pour construire la loi a priori. Dans la pratique on a recours à des lois usuelles (lois normales, lois gamma,...etc) ou à des lois dites conjuguées, l'information a priori étant alors utilisé pour déterminer les paramètres de la loi a priori, appelés hyper paramètres. En l'absence d'information a priori on introduira la notion de loi a priori non informative qui permet de rester dans un cadre bayésien, alors même que l'on ne dispose pas d'information a priori.

#### Définition 1.1 (Loi a priori conjuguées).

*La loi des observations étant supposée connue, on se donner une famille  $\mathcal{G}$  de loi des probabilité sur  $\Theta$ .*

*On suppose que la loi a priori appartient à  $\mathcal{G}$ . Si dans ces conditions, la loi a posteriori appartient encore à  $\mathcal{G}$ , on dit que la loi a priori est conjuguée.*

#### Exemples des lois conjuguées

Une loi conjuguée peut être déterminée en considérant la forme de la vraisemblance  $f(x/\theta)$  et en prenant une loi a priori de la même forme que cette dernière vue comme une fonction du paramètre. Les lois a priori conjuguées obtenues par ce procédé sont dites **naturelles**.

Le tableau suivant présent quelques familles de lois conjuguées.

$f(x \theta)$	$\pi(\theta)$	$\pi(\theta x)$
Normale $\mathcal{N}(\theta, \sigma^2)$	Normale $\mathcal{N}(\mu, \tau^2)$	$\mathcal{N}(\varrho(\sigma^2\mu + \tau^2x), \varrho\sigma^2\tau^2)$ $\varrho = 1/(\sigma^2 + \tau^2)$
Poisson $\mathcal{P}(\theta)$	Gamma $\mathcal{G}(\alpha, \beta)$	$\mathcal{G}(\alpha + x, \beta + 1)$
Gamma $\mathcal{G}(\nu, \theta)$	Gamma $\mathcal{G}(\alpha, \beta)$	$\mathcal{G}(\alpha + \nu, \beta + x)$
Binomiale $\mathcal{B}(n, \theta)$	Bêta $\mathcal{Be}(\alpha, \beta)$	$\mathcal{Be}(\alpha + x, \beta + n - x)$
Binomiale Négative $\mathcal{Neg}(m, \theta)$	Bêta $\mathcal{Be}(\alpha, \beta)$	$\mathcal{Be}(\alpha + m, \beta + x)$
Multinomiale $\mathcal{M}_k(\theta_1, \dots, \theta_k)$	Dirichlet $\mathcal{Be}(\alpha, \beta)$	$\mathcal{D}(\alpha_1 + x_1, \dots, \alpha_k + x_k)$
Normale $\mathcal{N}(\mu, 1/\theta)$	Gamma $\mathcal{G}(\alpha, \beta)$	$\mathcal{G}(\alpha + 0.5, \beta + (\mu - x)^2/2)$

TABLE 1.1 – Lois conjuguées usuelles.

**Définition 1.2 (Lois a priori non informatives).**

Partons d'un exemple pour introduire la notion de loi a priori non informative. On considère le modèle statistique suivant : les observations sont  $X_i/\theta$  i.i.d. suivent une loi de Bernoulli de paramètre  $\theta \in ]0, 1[$ . Il n'y a pas une unique loi a priori non informative pour le paramètre  $\theta$ . On peut en fait proposer différentes lois a priori non informatives.

- En l'absence d'information a priori sur  $\theta$ , il est naturel de proposer une loi uniforme sur  $\theta$  car elle donne probabilité égale aux intervalles de longueur  $l$  donnée, à savoir  $l$ .
- On peut également proposer la loi a priori (impropre) de Haldane

$$\pi(\theta) = [\theta(1 - \theta)]^{-1} \mathbf{1}_{]0,1[}$$

En arguant que  $E[\theta/x]$  est égal à l'estimateur du maximum de vraisemblance.

- Une alternative a été proposée par Jeffrey en 1960.

**Définition 1.3.**

Soit  $\theta$  un paramètre réel. On appelle loi a priori non informative de Jeffreys, la loi (éventuellement impropre) de densité  $\pi_j(\theta) = [I(\theta)]^{-\frac{1}{2}} \mathbf{1}_{\Theta}(\theta)$

$I(\theta)$  désigne l'information de Fisher apportée par  $x$  sur  $\theta$  tq :

$$I(\theta) = -E \left( \frac{\partial^2 \ln f(x/\theta)}{\partial \theta^2} \right)$$

**Définition 1.4 (Justification de la loi de Jeffrey).**

D'une part  $I(\theta)$  s'interprète comme la quantité d'information apportée par l'observation  $x$  sur  $\theta$ . D'autre part choisir  $\pi_j$  ne fait pas intervenir d'autre information que celle apportée par les observation (en fait, au travers de  $f(x/\theta)$ ). En ce sens, la loi a priori de Jeffrey est non informative.

**Remarque 1.1.** La loi a priori est une loi de probabilité qui résume toute l'information disponible sur le paramètre d'intérêt, avant le recueil des données. Donc, la loi a priori :

- Est une loi de probabilité.
- Elle s'applique au paramètre d'intérêt (une proportion, une moyenne, une variance, un coefficient de corrélation,...etc).
- Elle décrit ce que l'on sait sur ce paramètre avant que l'on ne recueille les données.
- Elle indique donc la probabilité que le paramètre ait telle ou telle valeur.

### 1.1.5 La Loi a posteriori

La formule de Bayes est un mécanisme d'apprentissage qui combine l'information a priori sur les paramètres et l'information apportée par les données en une information a posteriori sur les paramètres.

Précisément la formule de Bayes s'applique à la densité conjointe des paramètres et renvoie la densité conjointe des paramètres a posteriori c'est-à-dire conditionnellement aux observations.

La formule de Bayes appliquée à des densités de probabilité s'écrit [20] :

$$\pi(\theta/x) = \frac{f(x/\theta)\pi(\theta)}{m(x)} \quad (1.1)$$

où  $m(x)$  est la distribution marginale des données qui a ici le rôle d'une constante de normalisation pour que la densité a posteriori s'intègre bien à 1, puisque  $x$  et  $m(x)$  sont connus a posteriori de  $\theta$  évaluée en  $x$  est proportionnelle à  $f(x/\theta)\pi(\theta)$ . La formule (1.1) peut donc être réécrite sous la forme suivante, où  $\alpha$  est un symbole de proportionnalité :

$$\pi(\theta/x) \propto f(x/\theta)\pi(\theta) \quad (1.2)$$

cette expression montre que l'inférence bayésienne vérifie le principe de vraisemblance : a posteriori, l'information provenant des données provient exclusivement de la vraisemblance  $f(x/\theta)$ .

La formule de Bayes et le calcul des différentes lois marginales qui en découle font appel à de nombreux calculs d'intégrales multidimensionnelles pour les quelles la forme analytique peut être complexe.

Les méthodes de simulations de Monte Carlo permettent alors d'approcher ces intégrales.

**Définition 1.5 (la loi a posteriori).** [20]

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.** [20] On dispose d'un vecteur d'observations :  $x = (x_1, \dots, x_i, \dots, x_n)$  et on considère le modèle bayésien suivant :

$$X_i/\theta \rightsquigarrow \text{Bernoulli}(\theta) \text{ et } \theta \rightsquigarrow \text{Beta}(a, b) .$$

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

tell que  $s = \sum_{i=1}^n x_i$ , et comme  $\theta \rightsquigarrow \text{Beta}(a, b)$ , alors :

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

en appliquant la formule de Bayes on a :

$$\begin{aligned} \pi(\theta/x) &= \frac{f(x/\theta)\pi(\theta)}{\int_a^b f(x/\theta)\pi(\theta)d\theta} \\ &= \frac{\theta^s (1 - \theta)^{n-s} \frac{1}{\beta(a, b)} \theta^{a-1} (1 - \theta)^{b-1} \mathbb{1}_{[a,b]}(\theta)}{\frac{1}{\beta(a, b)} \int_a^b \theta^s (1 - \theta)^{n-s} \theta^{a-1} (1 - \theta)^{b-1} d\theta} \end{aligned}$$

Soit :

$$\begin{aligned} m(x) &= \frac{1}{\beta(a, b)} \int_a^b \theta^s (1 - \theta)^{n-s} \theta^{a-1} (1 - \theta)^{b-1} d\theta \\ &= \frac{1}{\beta(a, b)} \int_a^b \theta^{s+a-1} (1 - \theta)^{n+b-s-1} d\theta \\ &= \frac{\beta(s+a, n+b-s)}{\beta(a, b)} \end{aligned}$$

On pose :  $\alpha = s + a$ ,  $\beta = n + b - s$

Donc :

$$m(x) = \frac{\beta(\alpha, \beta)}{\beta(a, b)}$$

Alors :

$$\begin{aligned} \pi(\theta/x) &= \frac{\frac{1}{\beta(a, b)} \theta^{\alpha-1} (1-\theta)^{\beta-1} \mathbf{1}_{[0,1]}(\theta)}{\frac{\beta(\alpha, \beta)}{\beta(a, b)}} \\ &= \frac{1}{\beta(\alpha, \beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1} \mathbf{1}_{[0,1]}(\theta) \end{aligned}$$

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

**Définition 1.6 (Le raisonnement proportionnel).**

Il est par fois possible d'éviter le calcul de l'intégrale  $\int_{\Theta} f(x/\theta)\pi(\theta)d\theta$  en raisonnement proportionnellement.

Soient deux fonctions réelles  $f$  et  $g$  définies sur le même espace  $Y$ . On dit que  $f$  et  $g$  sont proportionnelles ce qui on note  $f \propto g$  s'il existe une constante  $a$  tel que  $f(y) = ag(y)$  pour tout  $y \in Y$ . Il est claire que la relation  $\propto$  est une relation d'équivalence c-à-dire : si  $f \propto g$  et  $g \propto h$  Alors  $f \propto h$ .

**Remarque 1.2.**

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  $f(x/\theta)\pi(\theta)$  sont effectivement proportionnelles la constante  $a$  qui apparait dans la définition est égale ici  $a$  :

$$\frac{1}{\int_{\Theta} f(x/\theta)\pi(\theta)d\theta}$$

à noter que cette quantité est bien une constante au sens où elle ne dépend pas de  $\theta$ .

**Exemple 1.2.** En reprenant le même modèle bayésien précédent

$$X_i/\theta \rightsquigarrow \text{Bernoulli}(\theta) \text{ et } \theta \rightsquigarrow \text{Beta}(a, b).$$

On a :

$$\begin{aligned}\pi(\theta/x) &= \frac{f(x/\theta)\pi(\theta)}{\int_{\Theta} f(x/\theta)\pi(\theta)d\theta} \\ &= \frac{\alpha f(x/\theta)\pi(\theta)}{\alpha \theta^s (1-\theta)^{n-s} \theta^{a-1} (1-\theta)^{b-1} \mathbf{1}_{[0,1]}(\theta)} \\ &= \frac{\alpha \theta^{s+a-1} (1-\theta)^{a+b-s-1} \mathbf{1}_{[0,1]}(\theta)}{\alpha \theta^{s+a-1} (1-\theta)^{a+b-s-1} \mathbf{1}_{[0,1]}(\theta)}\end{aligned}$$

Il s'agit d'une loi Beta( $s+a, n+b-s$ ) tel que  $s = \sum_{i=1}^n x_i$ .

**Définition 1.7 (Lois dérivées de la loi a posteriori).**

1 **La loi de couple**  $(x, \theta)$  sa densité est notée  $h(\theta, x)$ . On a donc :

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

2 **La loi marginale de  $x$**  : Sa densité es notée  $m(x)$ , on a donc :

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

### 1.1.6 Quelques notion de la Théorie de décision

Il est fréquent pour une personne d'avoir à prendre une décision en générale ou lui conseil de prendre la "meilleure" décision dans l'ensemble des tout les décision qui s'offrent à elle quel sues faut il donner à la notion de "meilleure" décision, il s'agira de formaliser le comportement d'un individu on d'un groupe confronter à une situation de choix en avenir insertion.

**Définition 1.8 (Les éléments de la théorie de décision).**

Un problème de décision est en générale fondé sur les quatre éléments suivantes :

- 1) Un ensemble de décision  $D$ .
- 2) un espace des paramètres  $\Theta$ .
- 3) un fonction de perte  $l(\delta, \theta)$  qui d'écrite la perte de prendre la décision  $\delta$  lorsque le paramètre est  $\theta$ .
- 4) un ensemble d'observation  $X$ .

**Définition 1.9 (Fonctions de pertes ).**

Soit  $\delta \in D$  une règle de décision. Une fonction de perte (coût) est une fonction mesurable de  $(\Theta \times D)$  à valeurs dans  $\mathbb{R}_+$ , notée  $l(\delta, \theta)$  et définie tel que :

- 1)  $\forall(\delta, \theta) \in (D \times \Theta), l(\delta, \theta) > 0$
- 2)  $\forall\theta \in \Theta, \exists\delta^* \in D, l(\delta^*(x), \theta) = 0$

Deux fonctions de pertes sont les plus utilisées dans l'estimation Bayésienne.

1) **La perte quadratique :**

C'est une fonction définie par

$$\begin{aligned} l : (\Theta \times D) &\longrightarrow \mathbb{R}_+ \\ (\delta, \theta) &\longrightarrow (\delta - \theta)^2 \end{aligned}$$

2) **La perte absolue :** C'est une fonction définie par

$$\begin{aligned} l : (\Theta \times D) &\longrightarrow \mathbb{R}_+ \\ (\delta, \theta) &\longrightarrow |\delta - \theta| \end{aligned}$$

**Définition 1.10 (le risque fréquentiste).**

Pour une fonction de perte donnée  $l(\delta, \theta)$ , la fonction de risque associée est :

$$\begin{aligned} R(\delta, \theta) &= E_\theta[l(\delta(x), \theta)] \\ &= \int_x l(\delta(x), \theta) f(x/\theta) d\mu(x) \end{aligned}$$

C'est une fonction de  $\theta$  et ne définit pas un ordre total sur  $D$ , elle ne permet donc pas de comparer toutes décisions et estimateurs.

**Définition 1.11 (Le risque a posteriori).**

Pour un Bayésien  $\theta$  est une variable aléatoire de distribution a priori  $\pi(\theta)$  et une fois que les données sont disponibles, la distribution estimée de  $\theta$  sera donnée par la loi a posteriori  $\pi(\theta/x)$  et le risque pertinent sera le risque a posteriori notée  $\rho(\pi, \delta/x)$  est défini comme suit :

$$\begin{aligned} \rho(\pi, \delta/x) &= E^\pi[l(\delta(x), \theta)/x] \\ &= \int_\Theta l(\delta/x, \theta) \pi(\theta/x) d\theta \end{aligned}$$

**Définition 1.12 (Le risque Bayésien).**

Pour une fonction de perte donnée le risque Bayésien noté  $r(\pi, \delta)$  est défini comme suit :

$$\begin{aligned} r(\pi, \delta) &= E[R(\delta, \theta)/x] \\ &= \int_\Theta R(\delta, \theta) \pi(\theta) d\theta \end{aligned}$$

**Définition 1.13 (La stratégie bayésienne).**

Soit  $\rho^n$  une structure statistique et soit  $\Theta$  l'espace des paramètres, soit  $\delta_B \in D$  une stratégie (règle de décision).  $\delta_B$  est dite une règle de Bayes ou stratégie Bayésienne ou estimateur Bayésien si :

$$r(\delta_B, \pi) = \inf_{\delta \in D} r(\delta, \pi)$$

**1.1.7 Estimateur Bayésien**

L'estimation Bayésienne est basée sur la minimisation d'une fonction de risque a posteriori. C'est la fonction de perte qui détermine la forme de l'estimateur Bayésien.

**Définition 1.14 (Définition de l'estimateur Bayésien).**

Un estimateur Bayésien est une règle de décision Bayésienne  $\delta_B$ , c-à-dire la règle de décision  $\delta_B$  qui minimise le risque Bayésien  $r(\delta, \pi)$

$$r(\delta_B, \pi) = \inf_{\delta \in D} r(\delta, \pi)$$

Pour obtenir la valeur de l'infimum il faut en théorie minimiser une intégrale double.

en effet ; On a :

$$\begin{aligned} r(\delta, \pi) &= \int_{\Theta} R(\delta, \theta) \pi(\theta) d\theta \\ &= \int_x l(\delta(x), \theta) f(x, \theta) d\mu(x) \\ &= \int_{\Theta} \int_x l(\delta, \theta) f(x/\theta) dx \pi(\theta) d\theta \\ &= \int_{\Theta} \int_x l(\delta, \theta) \frac{f(x/\theta) \pi(\theta)}{m(x)} m(x) dx d\theta \\ &= \int_x \left( \int_{\Theta} l(\delta, \theta) \pi(\theta/x) d\theta \right) m(x) dx \\ &= \int_x \rho(\pi, \delta/x) m(x) dx \end{aligned}$$

Donc pour minimiser  $r(\delta, \pi)$  pour toute valeur de  $x$ , il suffit de minimiser la fonction de risque a posteriori. Alors un estimateur Bayésien est la règle de décision qui minimise le risque a posteriori  $\rho(\delta, \pi/x)$ .

### 1. L'estimateur Bayésien Sous la perte quadratique

Comme nous l'avons vu, une fonction de perte quadratique est une fonction :  $l : (D \times \theta) \rightarrow \mathbb{R}^+$ . donnée par :

$$l(\delta, \theta) = (\delta - \theta)^2$$

Ainsi : Soit

$$\begin{aligned} f(\delta, x) &= \rho(\pi, \delta/x) \\ &= E_{\pi}(l(\delta, \theta)) \\ &= \int_{\Theta} l(\delta, \theta) \pi(\theta/x) d\theta \\ &= \int_{\Theta} (\delta - \theta)^2 \pi(\theta/x) d\theta \\ &= \int_{\Theta} (\delta^2 - 2\delta\theta + \theta^2) \pi(\theta/x) d\theta \\ &= \delta^2 \int_{\Theta} \pi(\theta/x) d\theta - 2\delta \int_{\Theta} \theta \pi(\theta/x) d\theta + \int_{\Theta} \theta^2 \pi(\theta/x) d\theta \\ &= E[\theta^2/x] - 2\delta E[\theta/x] + \delta^2 \end{aligned}$$

On cherche à minimiser  $f(\delta, x)$ , donc il faut trouver  $\delta$  que vérifie :

$$\frac{\partial}{\partial \delta} f(\delta, x) = 0$$

- $\frac{\partial}{\partial \delta} f(\delta, x) = \frac{\partial}{\partial \delta} [E[\theta^2/x] - 2\delta E[\theta/x] + \delta^2] = 2\delta - 2E[\theta/x]$
- $\frac{\partial}{\partial \delta} f(\delta, x) = 0$   
 $\implies 2\delta - 2E[\theta/x] = 0$   
 $\implies \delta_B = E[\theta/x]$

#### Définition 1.15.

*Sous la perte quadratique, l'estimateur Bayésienne est la moyenne de la loi a posteriori.*

**Exemple 1.3 (d'après l'exemple précédent).** *On considère le modèle Bayésien suivant :*

$$X_i/\theta \rightsquigarrow \text{Bernoulli}(\theta) \text{ et } \theta \rightsquigarrow \text{Beta}(a, b).$$

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

D'où :

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

## 2. l'estimateur bayésienne Sous la perte absolue

On a :

$$l(\delta, \theta) = |\delta - \theta|$$

$$\begin{aligned} f(\delta, x) &= \rho(\pi, \delta/x) \\ &= E_{\pi}(l(\delta, \theta)) \\ &= \int_{\theta} l(\delta, \theta) \pi(\theta/x) d\theta \\ &= \int_{\theta} |\delta - \theta| \pi(\theta/x) d\theta \\ &= \int_{\theta_1}^{\delta} (\theta - \delta) \pi(\theta/x) d\theta + \int_{\delta}^{\theta_2} (\delta - \theta) \pi(\theta/x) d\theta \end{aligned}$$

On cherche  $\delta$  qui vérifié :

$$\frac{\partial}{\partial \delta} f(\delta, x) = 0$$

- $\frac{\partial}{\partial \delta} f(\delta, x) = - \int_{\theta_1}^{\delta} \pi(\theta/x) d\theta + \int_{\delta}^{\theta_2} \pi(\theta/x) d\theta$
- $\frac{\partial}{\partial \delta} f(\delta, x) = 0$   
 $\implies \int_{\theta_1}^{\delta} \pi(\theta/x) d\theta = \int_{\delta}^{\theta_2} \pi(\theta/x) d\theta$

$\sigma$  est bien la médian de la loi a posteriori.

### Définition 1.16.

*Sous la perte absolue, l'estimateur Bayésien est la médiane de la loi a posteriori.*

## 3. L'estimateur du maximum a posteriori (MAP)

On appelle estimateur *MAP* tout estimateur  $\delta^{\pi} \in \arg \max_{\theta} \pi(\theta/x)$ . Cette notion est la contre partie bayésien du maximum de vraisemblance fréquentiste. Il a le grand avantage de ne pas dépendre d'une fonction de perte et est utile pour les approches théoriques. Ses inconvénients sont les mêmes que l'estimateur du maximum de vraisemblance : non unicité, instabilité (dus aux calculs d'optimisation) et dépendance vis-à-vis de la mesure de référence (dominant  $\Theta$ ). En outre, il ne vérifie pas la non invariance par paramétrisation qui peut apparaître importante intuitivement.

**Remarques 1.1.**

- Il faut signaler que le mode de la loi a posteriori est parfois retenus comme estimation Bayésien de  $\theta$  si l'on suspect que la vraisemblance est multimodale. Dans ce cas, on a :

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

- L'estimateur Bayésien des  $h(\theta) \in \mathbb{R}$  est par définition  $E_\pi[h(\theta)/x]$ .

**Proposition 1.1 (Propriétés de l'estimateur de Bayes).**

P1. L'estimateur de Bayes est admissible.

P2. L'estimateur de Bayes est biaisé.

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

P3. L'estimateur de Bayes est convergent en probabilité (quand la taille de l'échantillon  $n \rightarrow +\infty$ ).

P4. La loi a posteriori peut être asymptotiquement (c.à.d. pour de grandes valeurs de  $n$ ) approximée par une loi normale  $\mathcal{N}(E[\theta/x], \text{Var}[\theta/x])$ .

**Remarque 1.3.** Cette dernière propriété est particulièrement utile pour construire des intervalles de confiance a posteriori.

**Partie 2 :**

**Les méthodes de Monte Carlo**

## 1.2 Les méthodes de Monte Carlo

### 1.2.1 Introduction

La méthode de Monte-Carlo a vu son essor à partir de la fin de la seconde guerre mondiale, essentiellement dans le cadre du projet américain "Manhattan" concernant le développement de l'arme nucléaire, cette époque correspond également à la construction des premiers "ordinateurs". Ce projet étant été classé "secret défense", il est difficile de savoir exactement qui parmi ses pionniers : Von Neumann, Ulam et Metropolis a proposé le nom de "Monte-Carlo". Quoi qu'il en soit, ce terme fait référence aux jeux de hasard : la capitale de la principauté de Monaco. Le terme méthode de Monte-Carlo désigne une famille de méthodes algorithmiques visant à calculer une valeur numérique approchée en utilisant des procédés aléatoires, c'est-à-dire des techniques probabilistes. D'un point de vue purement mathématique, une méthode de Monte-Carlo peut servir au calcul d'intégrales (simples ou multiples) et à la résolution d'équations aux dérivées partielles, de systèmes linéaires et de problèmes d'optimisation. D'une manière plus globale, elle permet d'avancer toutes les possibilités liées à une observation, en quantifiant le risque associé (ou l'incertitude statistique). Pratiquement, ces techniques sont couramment utilisées dans divers domaines (physique, ingénierie, finance...). Dans ce chapitre, nous avons présenté la méthode de Monte-Carlo du point de vue de son application pour le calcul d'intégrale. En particulier, pour appliquer la méthode en question, il faut savoir simuler les variables aléatoires, par la méthode d'inversion, la méthode *DS* et la méthode *RDS*. Le problème traité est donc le calcul de  $\int_x \varphi(x)d\pi(x)$ , où  $\pi$  est une loi de probabilité et  $\varphi$  est une fonction donnée.

### 1.2.2 Description de la méthode

Pour utiliser une méthode de Monte-Carlo, on doit tout d'abord mettre sous la forme d'une espérance la quantité que l'on cherche à calculer. C'est souvent simple (calcul d'intégrale par exemple) mais peut-être plus compliqué (équations aux dérivées partielles par exemple).

A l'issue de cette étape, il reste à calculer une quantité de la forme  $E[X]$ ; l'espérance mathématique de  $X$ .

Une application classique des méthodes Monte-Carlo est le calcul de quantités du

type :

$$I = E[\varphi(X)] = \int_x \varphi(x)f(x)dx$$

où  $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$  est une fonction donnée et  $X$  un vecteur aléatoire de densité  $f$  suivant laquelle on sait simuler. Dans ce contexte, l'estimateur Monte-Carlo de base est défini par

$$\hat{I}_n = \frac{1}{n} \sum_{i=1}^n \varphi(X_i),$$

où les  $X_i$  sont générées de façon i.i.d. selon  $f$ .

**Exemple 1.4.**

Soit  $I = \int_0^2 e^{-x^2} dx$ .

On utilise la méthode de MC pour calculer l'intégrale  $I$  sous forme d'une espérance.

On pose :

1.  $h(x) = e^{-x^2}$
2.  $f(x)$  suit une loi uniforme  $\mathcal{U}_{[0,2]}$

alors :

$$\varphi(x) = \frac{h(x)}{f(x)} = 2e^{-x^2}$$

l'intégrale  $I$  devient :  $I = \int_0^2 2e^{-x^2} \frac{1}{2} dx = E[2 \exp^{-x^2}]$  où  $X \rightsquigarrow \mathcal{U}_{[0,2]}$

d'après la loi forte des grands nombres, l'intégrale  $I$  est estimée par :

$$\hat{I} = \frac{\sum_{i=0}^2 2e^{-x^2}}{n}$$

**Remarques 1.2.**

1. la méthode de monte-carlo ne peut donc s'utiliser que pour des variables aléatoires intégrables.
2. Pour l'approche Bayésienne la quantité d'intérêt prend souvent la forme d'une espérance. Par exemple sous la perte quadratique, l'estimateur Bayésien est la moyenne de la loi a posteriori  $E_\pi[\theta/x]$ .

Il est clair que l'estimateur  $\hat{I}_n$  est sans biais, c'est-à-dire que  $E[\hat{I}_n] = I$ . Mieux, la loi forte des grands nombres assure qu'il est convergent.

### 1.2.3 Convergence de la méthode

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

**Théorème 1.1 (Loi forte des grands nombres).** [16]

Soit  $X_1, \dots, X_n$  une suite de variables aléatoire indépendante et identiquement distribuées. Si  $E|\varphi(X)| < +\infty$ , alors

$$\hat{I}_n = \frac{1}{n} \sum_{i=1}^n \varphi(X_i) = \frac{\varphi(X_1) + \varphi(X_2) + \dots + \varphi(X_n)}{n} \xrightarrow[n \rightarrow \infty]{p.s.} E[\varphi(X)] = I$$

**Théorème 1.2 (Central limit).** Soit  $(I_n)_{n \in \mathbb{N}}$  une suite de variable aléatoire i.i.d de carrée intégrable :  $E[(\varphi(X^2))] < \infty$ , alors

$$\sqrt{n}(\hat{I}_n - I) \xrightarrow[n \rightarrow \infty]{l} \mathcal{N}(0, \sigma^2)$$

avec

$$\sigma^2 = \text{var}(\varphi(x)) = E[\varphi(X^2)] - E[\varphi(X)]^2 = \int \varphi(X^2)f(x)dx - I^2$$

Ainsi, lorsque  $n$  est grand, notre estimateur suit à peu près une loi normale : avec un abus de notation flagrant, on a  $\hat{I}_n \approx \mathcal{N}(I, \frac{\sigma^2}{n})$ , c'est-à-dire que  $\hat{I}_n$  tend vers  $I$  avec une vitesse en  $O(\frac{1}{\sqrt{n}})$ . Plus précisément, le TCL doit permettre de construire des intervalles de confiance. Cependant, en général, l'écart-type  $\sigma$  est lui aussi inconnu, il va donc falloir l'estimer. Qu'à cela ne tienne, la méthode Monte-Carlo en fournit justement un estimateur à peu de frais puisque basé sur le même échantillon  $(X_1, \dots, X_n)$ , à savoir

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \varphi^2(X_i) - \hat{I}_n^2 \xrightarrow[n \rightarrow \infty]{p.s.} \sigma^2$$

par la loi des grands nombres pour le premier terme et la proposition de la loi des grands nombres pour le second. Le lemme de Slutsky implique donc que

$$\sqrt{n} \frac{\hat{I}_n - I}{\hat{\sigma}_n} \xrightarrow[n \rightarrow \infty]{\ell} \mathcal{N}(0, 1)$$

## 1.2.4 Simulation de Monte Carlo

La simulation de M.C est une expérience de l'échantillon c-à-d, que le nombre aléatoire sont utilisés pour générer seulement quelques réponse parmi toutes les réponses possibles qui correspondent à tout les permutations des nombres aléatoires. Le fait d'utiliser d'échantillonnage en simulation cause toujours des erreurs d'échantillonnage car les estimateurs obtenues sont des fonctions des valeurs de la variable d'entrée. Parmi les techniques qui réduisent les erreurs d'échantillonnage, les méthodes de réduction de la variance telles que : la méthode d'échantillonnage préférentiels, la méthode de variable de contrôle, la méthode d'échantillonnage stratifié ... et plusieurs d'autre.

### Méthode d'inversion

La méthode d'inversion est la plus simple des méthodes générales de simulation. Elle consiste à composer l'inverse de la fonction de répartition  $F$  de la distribution à simuler avec un générateur de la loi uniforme sur  $[0, 1]$ .  $F$  n'étant pas toujours inversible au sens classique, on défini son inverse de la façon suivante :

$$\forall u \in [0, 1], F^{-1}(u) = \inf\{x; F(x) \geq u\}.$$

De point de vue théorique, le principe de la méthode est justifié dans la proposition suivante :

#### Proposition 1.2.

*Soient  $F$  une fonction de répartition sur  $\mathbb{R}$  et  $U$  une variable aléatoire de loi uniforme sur  $[0, 1]$ . Alors la variable aléatoire*

$$X = F^{-1}(U)$$

*a pour fonction de répartition  $F$ .*

Pour démontrer cette proposition nous avons besoin du lemme suivant :

#### Lemme 1.1.

*Pour tout  $u \in [0, 1], x \in \mathbb{R}; F^{-1}(u) \leq x \Leftrightarrow u \leq F(x)$ .*

#### Démonstration du lemme.

Pour l'implication directe, la croissance de  $F$  donne

$$F(F^{-1}(u)) \leq F(x),$$

puisque la fonction de répartition est continue à droite

$$F(F^{-1}(u)) \geq u,$$

nous avons donc  $F(x) \geq u$ , quant à l'implication réciproque elle est triviale.

### Démonstration de la proposition.

$\forall x \in \mathbb{R}, P(X \leq x) = P(F^{-1}(u) \leq x) = P(u \leq F(x)) = F(x)$  Cette proposition permet d'obtenir les formules analytiques utiles pour simuler un grand nombre de lois, nous introduisons quelques exemples illustratifs :

#### Exemples 1.1. 1. La loi exponentielle :[16]

soit  $X$  une variable aléatoire réelle qui suit une loi exponentielle de paramètre  $\lambda$ . La fonction inverse de sa fonction de répartition vaut pour  $u \in [0, 1]$ ,

$$F^{-1}(u) = \frac{-\ln(1-u)}{\lambda}$$

ou encore

$$\frac{-\ln u}{\lambda}$$

Le code R suivant donne 100 observations de la loi exponentielle de paramètre 1

```
N=100
```

```
U=runif(N)
```

```
X=-log(1-U)
```

```
plot(density(X))
```

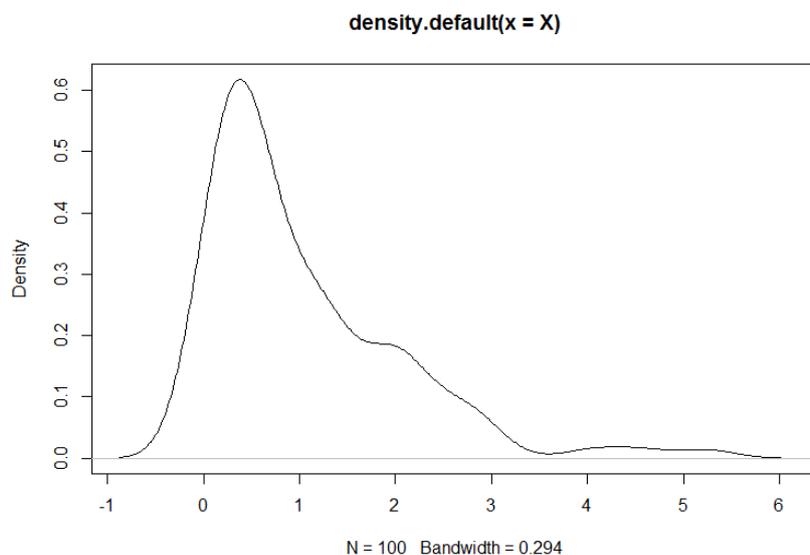


FIGURE 1.2 – densité de la loi exponentielle de paramètre 1 pour n=100.

2. La Loi de Cauchy : soit  $X$  une variable aléatoire réelle de loi de Cauchy de paramètre  $a$ , alors la variable

$$F^{-1}(U) = a \tan\left(\pi\left(U - \frac{1}{2}\right)\right),$$

suit une loi de Cauchy de paramètre  $a$ . Nous avons proposé le programme suivant qui permet de simuler 100 observations de la loi de Cauchy de paramètre 1 :

```
N = 100
U = runif(100)
X = numeric(N)
for(i in 1 : N)
  X[i] = tan(pi * (U[i] - 1/2))
plot(density(X))
```

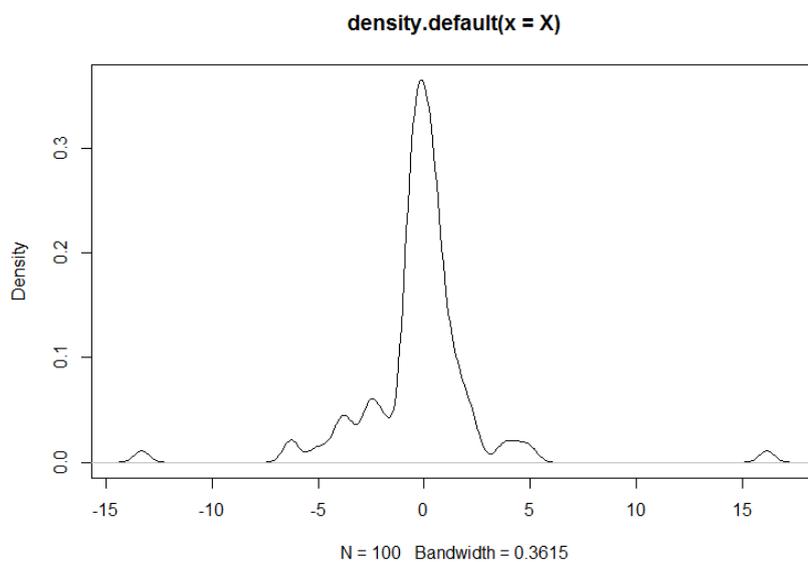


FIGURE 1.3 – densité de la loi de Cauchy de paramètre 1 pour  $n=100$ .

*On peut aussi simuler les variables discrètes par la méthode d'inversion.*

**Lois discrètes**

Soit  $X$  une variable aléatoire réelle discrète de loi  $p_i = P(X = x_i)$ , nous supposons que les éventualités sont rangé par ordre croissant

$$x_i < x_{i+1}, i \in 1 \dots n, n \in \mathcal{N}$$

considérons la fonction de répartition de  $X$  est définie par :

$$f(x) = \begin{cases} 0 & \text{si } x < x_1 \\ p_1 + \dots + p_i & \text{si } x_i \leq x < x_{i+1} \end{cases}$$

-Nous pouvons définir l'algorithme de simulation par la méthode d'inversion comme suit :

$i \leftarrow 1$

choix  $\leftarrow$  Random

Tant que (choix  $> F_i$ ) faire

$i \leftarrow i+1$

fin Tante que

$X \leftarrow x_i$

**Exemple 1.5.** *2- la loi de poisson*

*Si  $X$  suit une loi de poisson de paramétriser  $\lambda$  il n'y a pas d'expression simple pour la fonction de répartition et l'ensemble des valeurs possibles est infini.*

*Il faut donc calculer les valeurs  $F_i$  au fur et à mesure.*

*L'algorithme relatif à une telle simulation est donc :*

$X \leftarrow 0$

$p \leftarrow \exp(-\lambda)$

$F \leftarrow p$

choix  $\leftarrow$  Random

Tant que (choix  $> F$ ) faire

$X \leftarrow X+1$

$$p \leftarrow p\lambda/X$$

$$F \leftarrow F+p$$

*fin Tant que*

$$X \leftarrow X-1$$

*Remarquons que la méthode d'inversion n'est exacte qu' à condition de connaître l'expression explicite de  $F^{-1}$ .*

*C'est rarement la plus efficace pour les variables à densité.*

## 1.2.5 Échantillonnage Aléatoire

Un échantillonnage aléatoire est basé sur le principe que tous les éléments de la population ont une probabilité égale (non nulle) de faire partie de l'échantillon : c'est une méthode permette de choisir  $n$  unités parmi les  $N$  de la population de façon aléatoire.

### Définition 1.17.

*Un échantillon aléatoire de taille  $n$  de la variable aléatoire  $X$  est un vecteur aléatoire  $(X_1, X_2, \dots, X_n)$  où*

*$X_i, i = 1..n$  sont des variables aléatoires indépendantes et de même loi que la variable aléatoire  $X$ .*

*En utilisant l'échantillonnage aléatoire, considérons le cas d'un problème de simulation avec une variable aléatoire d'entrée et définissons les estimateurs de la simulation dans les deux cas suivants :*

**Premier cas :** *Une variable de sortie possédant  $k$  paramètres à estimer,  $\theta_j, j = 1, 2, \dots, k$ .*

*Dans ce cas, les estimateurs de la simulation obtenus sont définis par :*

$$Y_j = F_j(u_1, u_2, \dots, u_n), \quad j = 1, 2, \dots, k$$

*où*

1)  $Y_j$  est l'estimateur du paramètre inconnu  $\theta_j$ ,  $j = 1, 2, \dots, k$

2)  $F_j$   $j = 1, 2, \dots, k$  est la fonction de simulation, généralement définie par un programme informatique, reliant les valeurs de la variable d'entrée et chaque estimateur,

3)  $u_1, u_2, \dots, u_n$  sont des nombres aléatoires indépendants uniformément distribués entre 0 et 1.

En simulation, une histoire crée un estimateur pour chaque paramètre étudié. Chaque histoire crée donc  $k$  estimateurs car nous avons  $k$  paramètres à estimer.

**Deuxième cas :**  $m$  variables de sortie dont chacune possède  $k$  paramètres à estimer suivants :

$$\theta_{j1}, \theta_{j2}, \dots, \theta_{jk} \quad j = 1, 2, \dots, m$$

Dans ce cas, les estimateurs de la simulation obtenus sont définis par :

$$\begin{aligned} Y_{1l} &= F_{1l}(u_1, u_2, \dots, u_n) \\ &\vdots \\ Y_{jl} &= F_{jl}(u_1, u_2, \dots, u_n) \\ &\vdots \\ Y_{ml} &= F_{ml}(u_1, u_2, \dots, u_n), \quad l = 1, 2, k \end{aligned}$$

où

$Y_{jl}$ ,  $j = 1, 2, \dots, m$  sont les estimateurs des paramètres inconnus  $\theta_{jl}$ .

Notons que dans ce cas, chaque histoire crée  $km$  estimateurs.

### 1.2.6 Échantillonnage descriptif (ED) ou (DS)

Proposé comme une approche alternative à la simulation de Monte Carlo, l'échantillonnage descriptif, (Descriptif Sampling DS)(Saliby, 1980 et 1990) est basé sur une sélection totalement déterministe des valeurs d'échantillonnage d'entrée et de leurs permutations aléatoire. En tant que tel, l'échantillonnage descriptif évite l'effet ensembliste du à la méthode de MC et conduit à des estimations plus précises de simulation.[26]

#### Procédure d'échantillonnage

Supposons que le problème de simulation contient une variable aléatoire d'entrée  $X$  de fonction de répartition  $F$ , et soit  $F^{-1}$  son inverse telle que  $X = F^{-1}(R)$  où  $R \sim \mathcal{U}[0, 1]$ . La génération d'un échantillon descriptif, se fait la manière suivant :

1. Calculer les nombres  $r_i = \frac{i - 0.5}{n}$ ,  $i = \overline{1, n}$ .  
ça permet de subdiviser l'intervalle  $[0, 1]$  en  $n$  sous intervalles équiprobables, tel que  $r_i$  représente le milieu de  $i^{\text{ème}}$  sous intervalle.
2. Calculer les valeurs de l'échantillon  $x_i$  telles que

$$x_i = F^{-1}(r_i), \quad i = 1, 2, \dots, n.$$

3. permuter aléatoirement l'ensemble des valeurs  $x_i$ ,  $i = 1, 2, \dots, n$ .
4. Stocker la suite obtenue dans un nouveau vecteur  $y$  et l'utiliser lorsque c'est nécessaire.

À la fin de l'histoire, l'estimateur  $ED$  du paramètre  $\theta$  est donné par :

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^n y_i$$

Cette méthode a été critiquée pour son biais de l'échantillonnage qu'elle peut produire. les estimateurs prochains par l'algorithme  $ED$  sont des estimateurs biaisés .  
La taille de l'échantillon doit être connue au début de la simulation.

**Exemple 1.6.** Estimer la moyenne de la loi exponentielle de paramètre ( $\lambda = 1$ ),

$$r_i = \frac{i - 0.5}{n} \quad i = 1, 2, \dots, n .$$

Soit  $n = 4$ .

i	$r_i$	$x_i = -\ln(1 - r_i)$
1	0.125	2.079
2	0.375	0.981
3	0.675	0.471
4	0.875	0.133
Total		3.664

TABLE 1.2 – Résultats de l'exemple de la méthode ED

L'estimateur de la moyenne est :

$$\hat{\lambda} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{3.664}{4} = 1.22.$$

### 1.2.7 L'échantillonnage descriptif amélioré (EDA) ou (RDS)

Pour réduire le risque du biais de la méthode précédente, une approche a été proposée, qui consiste en un bloc qui doit être situé à l'intérieur d'un générateur visant à distribuer aléatoirement des sous-ensembles réguliers de tailles de nombre premiers  $p_q, p_{q+1}; \dots$  pour tout  $q$  dans un ordre aléatoire à la demande de la simulation. Nous nous arrêtons de produire ces sous suites lorsque la simulation s'arrête (le critère d'arrêt est la durée de la simulation dans une exécution). Dans cette approche, chaque exécution est déterminée par un bloc de différents nombres premiers. Si on a besoin de  $M$  exécutions répétées, il est nécessaire de considérer  $M$  blocs de sous-ensembles réguliers  $m_1; m_2; \dots; m_M$ . Les nombres premiers et les valeurs de la variable aléatoire des sous-ensembles ne sont pas les mêmes pour toutes les exécutions répliquées.

#### Principe

Cette méthode considère un bloc d'ensemble de nombres régulier  $r_i$  de cardinal des nombres premier  $p_i$  et distribue les observations  $x_i$  à la demande de la simulation, on arrête le processus de génération lorsque  $m$  nombre premier  $p_i$  tels que  $\sum_{i=1}^m p_i = n$  sont générés qui conduisent à  $m$  sous histoire. À la fin de chaque histoire on calcul l'estimateur résultant. On obtient  $m$  estimateurs *RDS* :  $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_m$ .

l'estimateur final est

$$\hat{\theta} = \frac{\hat{\theta}_1 + \hat{\theta}_2 + \dots + \hat{\theta}_m}{m}$$

**Exemple 1.7.** *Simuler la loi exponentielle avec  $\mathbb{E}(X) = 1$  tel que  $r_i = \frac{i - 0.5}{p} \rightsquigarrow \mathcal{U}_{[0,1]}$ ,  $i = \overline{1, p}$ .*

*Avec la méthode d'inversion on obtient :*

$$x_i = -\ln(1 - r_i) \quad i = \overline{1, p} \text{ et on estime sa moyenne.}$$

*On génère  $p_1 = 7$ ,  $p_2 = 11$  et  $p_3 = 13$ .*

i	$r_i$	$x_i$	$r_i$	$x_i$	$r_i$	$x_i$
1	0.071	0.071	0.045	0.047	0.038	0.039
2	0.214	0.241	0.136	0.147	0.115	0.123
3	0.357	0.442	0.227	0.248	0.192	0.214
4	0.5	0.693	0.318	0.383	0.269	0.314
5	0.643	1.030	0.409	0.526	0.346	0.425
6	0.786	1.540	0.5	0.693	0.423	0.550
7	0.927	2.639	0.591	0.894	0.5	0.693
8		T=6.659	0.682	1.145	0.577	0.860
9			0.773	1.482	0.654	1.061
10			0.864	1.992	0.731	1.312
11			0.955	3.091	0.808	1.649
12				T=10.647	0.885	2.159
13					0.962	3.258
						T=12.662

TABLE 1.3 – Résultats de l'exemple de la méthode RDS

Nous calculons les trois moyennes  $\bar{Y}_1$ ,  $\bar{Y}_2$  et  $\bar{Y}_3$

$$\bar{Y}_1 = \frac{6.659}{7} = 0.951.$$

$$\bar{Y}_2 = \frac{10.648}{11} = 0.968.$$

$$\bar{Y}_3 = \frac{12.662}{13} = 0.974.$$

L'estimation de la moyenne en utilisant cette méthode est obtenue par :

$$\hat{y} = \frac{1}{3} \sum_{i=1}^3 \bar{y}_i = 0.965. [28]$$

### Algorithme RDS

La procédure de génération d'échantillonnage descriptif amélioré est la suivante :

1. On génère  $m$  nombre premier  $p_1, p_2, \dots, p_m$
2. On choisit un nombre premier aléatoirement  $p_i$  et on effectue la sous histoire suivante :

a– On calcule les nombre  $r_i = \frac{i - 0.5}{p_i}$ ,  $i = 1, 2, \dots, m$ .

b– On mélange les  $r_i$ , et on les met dans un vecteur  $R$ .

c- On calcule les valeurs  $x_i = F^{-1}(R_i)$ .

d- On calcule  $\hat{\theta}_i = \frac{\sum_{i=1}^m x_i}{m}$ .

**Les avantages :**

- Elle garde les avantages de la méthode précédente de plus cette approche élimine la nécessité de déterminer à l'avance la taille de l'échantillon.
- Elle élimine le problème du biais dans la méthode précédente
- Les estimateurs produits par EDA sont sans biais.

---

---

## CHAPITRE 2

---

### APPLICATION DE LA MÉTHODE RDS SUR UN MODÈLE AR(1)

**Partie 1 :**

**Quelque notions sur les séries  
chronologiques**

## 2.1 Notions de base en séries chronologiques

### 2.1.1 Introduction

Ce présent chapitre montre comment l'analyse Bayésienne peut faire face à des structures de dépendance en étudiant des modèles standards de séries chronologiques. Nous n'allons pas proposer une théorie unifiée mais plutôt, nous allons décrire via des exemples les éléments essentiels dans l'étude Bayésienne des séries temporelles en se basant sur un modèle auto-régressif.

La théorie des séries chronologiques (ou temporelles) abordée dans ce cours est appliquée de nos jours dans des domaines aussi variés que l'économétrie, la médecine ou la démographie, pour n'en citer qu'une petite partie. On s'intéresse à l'évolution au cours du temps d'un phénomène, dans le but de décrire, expliquer puis prévoir ce phénomène dans le futur. On dispose ainsi d'observations à des dates différentes, c'est à dire d'une suite de valeurs numériques indicées par le temps. On peut songer par exemple : au volume des exportations mensuelle au niveau d'un port (économie), au relevé des concentrations d'un produit chimique chaque 5 heures (chimie), à la série trimestrielle de la consommation d'électricité par un foyer algérien (physique).

Une fois que la série temporelle est observée, divers problèmes liés à cette série doivent être résolus. Deux problèmes intéressants forment souvent l'objectif principal d'une série temporelle : le premier concerne la prévision qui consiste à prévoir les valeurs futures de la série à partir de ses valeurs passées. Et le deuxième concerne la détection des outliers résultants. La résolution de ces problèmes passe par la modélisation de la série, l'estimation des paramètres inconnus dans le modèle et l'ajustement.

Nous allons nous intéresser dans notre travail à l'estimation Bayésienne des paramètres d'un modèle  $AR(1)$ .

Ainsi, ce chapitre contient deux parties : dans la première partie nous donnons quelques notions de base sur les séries temporelles en se basant sur les modèles  $AR$ . Dans la deuxième partie nous allons appliquer la méthode  $RDS$  abordée dans le premier chapitre pour estimer les paramètres d'un modèle  $AR(1)$  dans un sens Bayésien.

Une série chronologique est un processus stochastique  $(X_t)_{t \in \mathbb{Z}}$ , où  $t$  représente le temps. Dans ce chapitre, plusieurs concepts importants liés à l'analyse de séries chronologiques seront abordés. Parmi ceux-ci, on retrouve les notions d'autocorrélation, de stationnarité

et de bruit blanc. On énoncera également un résultat de Wold (1948) concernant les séries chronologiques stationnaires.

### Définition 2.1.

Une série chronologique est une suite des variables aléatoires observées séquentiellement dans le temps (une suite  $(X_t)$  d'observation chiffrées d'un même phénomène ordonnées dans le temps) :

$$\begin{aligned} X : T \subseteq \mathbb{N} \subseteq \mathbb{Z} &\longrightarrow \mathbb{R} \\ t &\longmapsto X(t) \end{aligned}$$

On écrit  $\{X_t, t \in T\}$ .

## 2.1.2 Les composantes principales d'une série chronologique

Le but de la décomposition d'une série chronologique est de distinguer dans l'évolution de la série : une tendance des variations saisonnières qui se répètent chaque année et des variations accidentelles.

L'intérêt de ceci est d'une part de mieux comprendre, de mieux décrire l'évolution de la série et d'autre part de prévoir son évolution.

Ainsi donc, on distingue trois composantes principales d'une série chronologique :

### 1. La tendance $T(t)$ (the trend) :

Cette composante correspond à l'évolution fondamentale de la série, peut être croissante ou non croissante, linéaire ou exponentiel.

### 2. La saisonnalité (variation saisonnière) $S(t)$ :

Les variations saisonnières sont des variations périodiques à l'intérieur d'une année et qui se reproduisent de façon plus ou moins permanente d'une année sur l'autre donc c'est une fonction périodique de période  $m$ , elle est notée :  $S(t)$ .

d'où :  $S(t + km) = S(t), \forall k \in \mathbb{Z}$ .

### 3. Les variations résiduelles ou accidentelles $\varepsilon(t)$ :

les variations résiduelles sont des fluctuations irrégulières et imprévisibles, elles sont supposées en général de faible amplitude. On note  $\varepsilon(t), \forall t \in \mathbb{Z}$  et  $(\varepsilon(t))_{t \in \mathbb{Z}}$  et est appelée

aussi bruit-blanc (white noise) .

Alors le modèle de décomposition classique d'une série chronologique s'écrit comme suit :

$$X(t) = T(t) + S(t) + \varepsilon(t) \text{ (s'appelle le modèle additif).}$$

ou

$$X(t) = T(t)S(t)\varepsilon(t) \text{ (s'appelle le modèle multiplicatif).}$$

### Remarques 2.1.

1.  $T(t)$  et  $S(t)$  sont appelées les composantes déterministes (ou la partie déterministe de la série).
2.  $\varepsilon(t)$  est la partie aléatoire ou la partie non déterministe.

### 2.1.3 Stationnarité d'une série temporelle

Dans ce qui suit nous donnons sous forme de notes, les divers types de stationnarité d'une série temporelle

1. La série temporelle  $\{X_t, t \in T\}$  est dite **stationnaire en moyenne** lorsque la moyenne de chacune des variables de la suite est identique.i.e

$$E(X_t) = E(X_0), \forall t \in T$$

2. De même, on dit que cette série est **stationnaire en variance** lorsque :

$$Var(X_t) = Var(X_0), \forall t \in T$$

3. La série  $(X_t)_t$  est dite **stationnaire au sens strict** ou **strictement (fortement) stationnaire** si le vecteur  $X_{t_1}, X_{t_2}, \dots, X_{t_n}$  a la même loi que le vecteur  $X_{t_1+k}, X_{t_2+k}, \dots, X_{t_n+k}$  pour tout  $t_1, t_2, \dots, t_n \in \mathbb{Z}$  et  $n \in \mathbb{N}$ .
4. Et enfin, elle est dite **stationnaire de second ordre ou au sens faible** si les trois conditions suivantes sont vérifiées :

- $E(X_t) = m, \forall t \in \mathbb{Z}$
- $E((X_t)^2) < \infty, \forall t \in \mathbb{Z}$
- $cov(X_t, X_{t+h}) = cov(X_{t-1}, X_{t-1+h}) = \dots = cov(X_0, X_h) = \gamma(h)$

La stationnarité au sens strict est peu réalisable en pratique sauf un peu dans le cas gaussien, et est plus exigeant que le concept de la stationnarité faible comme l'indique le lemme suivant.

**Lemme 2.1.** [4]

*Si la série  $X_t$  est strictement stationnaire et  $E((X_t)^2) < \infty$ , alors  $X_t$  est faiblement stationnaire. La réciproque est en général.*

*Après avoir donnée la définition d'un processus bruit blanc, nous introduisons un théorème fondamental de l'analyse des séries temporelles stationnaires.*

### 2.1.4 Processus aléatoires

On appelle processus stochastique un ensemble de variables aléatoires  $X(t), t \in T$  où  $T$  est l'ensemble des instants en lesquels le processus est défini ( $T = \mathcal{Z}$ )

Un processus aléatoire est caractérisé par ses moments et particulièrement le premier et le second moment. À savoir :

**la moyenne :** la fonction moyenne d'un processus aléatoire  $(X_t)_{t=1, T}$  notée  $\mu(t)$  est définie telle que :

$$\mu(t) = E(X_t) = \frac{1}{T} \sum_{t=1}^T X_t$$

**la variance :** la fonction variance, notée  $\sigma^2(t)$ , est définie telle que :

$$\sigma^2(t) = \text{var}(X_t) = \frac{1}{T} \sum_{t=1}^T (X_t - E(X_t))^2$$

**la fonction d'auto-covariance :** on appelle la fonction d'auto-covariance, notée  $\gamma(h)$  ou acvf, est définie telle que :

$$\gamma(h) = \text{cov}(X_t - X_{t+h}) = E[(X_t - E(X_t))(X_{t+h} - E(X_{t+h}))]$$

**la fonction d'auto corrélation :** on appelle la fonction d'auto corrélation d'ordre  $h \in \mathcal{N}$  de  $X_t$  est on notée  $\rho(h)$  la fonction :

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}$$

**La fonction d'autocorrélation partielle : "pacf" :**

Le coefficient de corrélation partiel  $\phi_{kk}$  entre les deux variables  $X_1$  et  $X_k$  d'un processus stochastique  $X_t$  est le coefficient de corrélation entre les deux variables  $X_1$  et  $X_k$  auquel

on retranche les observations  $(X_2, \dots, X_{k-1})$ , on utilisera l'abréviation *acf* pour la fonction d'autocorrélation et *pacf* pour la fonction d'autocorrélation partielle, elle s'écrit comme suit

$$\phi_{kk} = \frac{|R^*(k)|}{|R(k)|}$$

où  $R(k)$  est la matrice d'autocorrélation et  $R^*(k)$  est la matrice obtenue après le remplacement de la dernière colonne de la matrice  $R(k)$  par  $(\rho(1), \rho(2), \dots, \rho(k))$ .

**Théorème 2.1 (Théorème de Wold).** [4]

*Tout processus stationnaire du second ordre peut être représenté sous la forme :*

$$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} + \mu_t$$

où

- .  $\varepsilon_t$  est le bruit blanc
- .  $\mu_t$  est une composante linéaire déterministe avec  $\text{cov}(\mu_t, \varepsilon_{t-j}) = 0$
- .  $\psi_j$  des paramètres satisfaisant :  $\psi_0 = 1, \psi_j \in \mathbb{R}, \sum_{j=0}^{\infty} (\psi_j)^2 < \infty$

*La condition sur les  $\psi_j$  assure l'existence des moments d'ordre deux du processus.*

### 2.1.5 Le processus bruit blanc

Soit le processus  $\{X_t, t \in \mathbb{Z}\}$ . Si pour tout n-uple du temps  $t_1 < t_2 < \dots < t_n$ , les variables aléatoires réelles  $X_{t_2} - X_{t_1}, \dots, X_{t_n} - X_{t_{n-1}}$  sont indépendantes, il s'agit d'un processus à accroissements indépendants.

Un bruit blanc est un processus stochastique à accroissements non corrélés. Il est dit bruit blanc "fort" si les accroissements sont indépendants. On l'appelle aussi processus purement aléatoire.

Un bruit blanc est donc tel que :

- .  $E[X_t] = m, \forall t \in \mathbb{Z}$
- .  $V[X_t] = \sigma^2, \forall t \in \mathbb{Z}$
- .  $\text{cov}[X_t, X_{t+k}] = \gamma_x(k) = 0, \forall t \in \mathbb{Z}, \forall k \in \mathbb{Z}$

Si  $E[X_t] = 0$ , le bruit blanc est centré.

Dans le reste de ce travail, en écrivant bruit blanc, nous sous entendons bruit blanc stationnaire et on note souvent  $\varepsilon_t$ .

### 2.1.6 Le processus autorégressif d'ordre $p$ : $AR(p)$

L'étude des modèles auto-régressifs ( $AR(p)$ ) constitue l'un des problèmes fondamentaux posés par l'analyse des séries chronologiques en économétrie et en statistique. De façon générale, l'analyse des modèles auto régressifs permet d'établir les contrôles, de faciliter l'élaboration des prévisions et de conduire à la réduction des fluctuations indésirables.

Dans la vie courante, on rencontre de nombreux phénomènes observés qui peuvent être modélisés par ces processus, ils sont définis par la relation :

$$X(t) = \varepsilon(t) + \sum_{k=1}^p \varphi_k X(t-k)$$

Où  $\varepsilon(t)$  est le bruit blanc.

**Théorème 2.2.** *On dit que le processus  $AR(p)$  est stationnaire si :*

i)  $E(X(t)) = c \quad \forall t \in \mathcal{Z}$  et  $c$  une constante .

ii)  $E(X(t)^2) < \infty \quad \forall t \in \mathcal{Z}$ .

iii)  $cov(X_t, X_{t+h}) = \gamma(h)$ .

**Théorème 2.3.** *Le processus  $AR(p)$  est causal et stationnaire ssi le polynôme :*

$\varphi(z) = 1 - \sum_{k=1}^p \varphi_k z^k$  est tel que :

$\varphi(z) \neq 0, \forall z \in \mathbb{C},$  avec  $|z| \leq 1$ .

ou bien

$\varphi(z) = 0, \forall z \in \mathbb{C},$  avec  $|z| > 1$ .

**Remarque 2.1.** *par définition tout  $AR(p)$  est un processus inversible.*

### 2.1.7 Le processus moyenne mobile d'ordre $q$ : $MA(q)$

Un processus moyenne mobile d'ordre  $q$  est un processus  $\{X_t, t \in \mathbb{Z}\}$  dont les variables vérifient

$$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} \quad (2.1)$$

où  $\theta_1, \theta_2, \dots, \theta_q$  sont des paramètres réel, et  $\varepsilon_t$  est un bruit blanc centré. En utilisant l'opérateur retarde  $\beta$  ( $\beta X_t = X_{t-1}$ ) l'équation (2.1) est équivalente à

$$X(t) = \theta(\beta)\varepsilon(t)$$

avec

$$\theta(\beta) = 1 + \theta_1 \beta + \dots + \theta_q \beta^q = \sum_{j=0}^q \theta_j \beta^j$$

Un processus satisfait une représentation  $MA(q)$  est toujours stationnaire sans tenir compte des valeurs de  $\theta_1, \theta_2, \dots, \theta_q$ .

En effet

$$(i) E(X_t) = E(\varepsilon_t) \sum_{j=0}^q \theta_j = 0$$

$$(ii) V(X_t) = V(\varepsilon_t) \sum_{j=0}^q \theta_j^2 = \sigma_\varepsilon^2 \sum_{j=0}^q \theta_j^2$$

$$(iii) cov(X_t, X_{t+h}) = \sigma_\varepsilon^2 \sum_{j=0}^q \theta_j \theta_{j+1}$$

Les conditions de la stationnarité de second ordre sont toujours vérifiées, d'où la stationnarité. Mais contrairement au processus autoregressifs, le processus moyenne mobile n'est pas toujours inversible et pour qu'il le soit, il faut que son polynôme caractéristique ait toutes ces racines à l'intérieur du cercle unité.

### 2.1.8 Le processus mixte $ARMA(p, q)$

On va introduire une classe de processus linéaire très importante dans la modélisation des séries chronologique, les processus  $ARMA$ . Ces processus sont linéaires et sous certaines conditions stationnaires.

#### Définition 2.2.

Un processus auto-régressif en moyenne mobile d'ordres  $p, q$  (tels que  $p \geq 0, q \geq 0$  et  $p + q \geq 0$ ) est un processus qui peut s'écrire :

$$X_t = \sum_{k=1}^p \varphi_k X_{t-k} + \sum_{j=1}^q \theta_j \varepsilon_{t-j}, \quad \text{pour tout } t > 0 \quad (2.2)$$

telle que  $\varepsilon(t)$  est un bruit blanc, et  $\varphi_1, \varphi_2, \dots, \varphi_p (\varphi_p \neq 0)$  et  $\theta_1, \theta_2, \dots, \theta_q (\theta_q \neq 0)$  sont des constants.

En utilisant l'opérateur de retard  $\beta$ , l'équation (2.1) revient :

$$\varphi(\beta)X(t) = \theta(\beta)\varepsilon(t)$$

telle que :

$\varphi(\beta) = \sum_{k=0}^p \varphi_k \beta^k$ , ( $\varphi_0 = 1$ ) est un polynôme en  $\beta$  de degré  $p$ .

et  $\theta(\beta) = \sum_{j=1}^q \theta_j \beta^j$ , ( $\theta_0 = 0$ ) est un polynôme en  $\beta$  de degré  $q$ .

### Remarque 2.2.

1. Si  $p = 0$  et  $q > 0$ , les processus  $ARMA(0, q)$  sont notés  $MA(q)$  de la forme :

$$X(t) = \varepsilon(t) + \sum_{j=1}^q \theta_j \varepsilon(t - j)$$

est appelés processus à moyenne mobile d'ordre  $q$ .

2. Si  $q = 0$  et  $p > 0$ , les processus  $ARMA(p, 0)$  sont notés simplement  $AR(p)$  de la forme :

$$X(t) = \varepsilon(t) + \sum_{k=1}^p \varphi_k X(t - k)$$

et appelés processus auto-régressifs d'ordre  $p$ .

### Exemple 2.1. Modélisation d'une série temporelle[8]

L'exemple suivant concerne la donnée de l'hormone lutéinisante dans 48 échantillons de sang prélevés toutes les 10 minutes chez une femme. Ces données sont retirés de la base de donnée **lh** qui se trouve dans le package **datasets**.

Ces données peuvent être modélisés par une série chronologique, sous R on suit les étapes suivantes :

`>help(lh)` : affiche les informations sur la série chronologique.

`>data(lh)` : fait appel aux données de la série chronologique.

`>try(lh)` : affiche la série chronologique.

```
Time Series:
Start = 1
End = 48
Frequency = 1
 [1] 2.4 2.4 2.4 2.2 2.1 1.5 2.3 2.3 2.5 2.0 1.9 1.7 2.2 1.8 3.2 3.2 2.7 2.2 2.2
 [20] 1.9 1.9 1.8 2.7 3.0 2.3 2.0 2.0 2.9 2.9 2.7 2.7 2.3 2.6 2.4 1.8 1.7 1.5 1.4
 [39] 2.1 3.3 3.5 3.5 3.1 2.6 2.1 3.4 3.0 2.9
> |
```

FIGURE 2.1 – Les données de la série

`>plot(lh)` : Pour tracer la série chronologique.

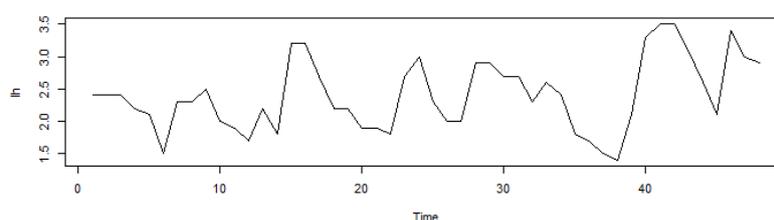


FIGURE 2.2 – La série chronologique

Cette série chronologique peut être modélisée par un  $MA(q)$  ou bien  $AR(p)$ , et pour choisir le meilleur modèle il faut prendre le plus petit AIC, pour cela on calcule l'ordre  $q$  de modèle  $MA(q)$  et l'ordre  $p$  du modèle  $AR(p)$ , en utilisant les codes R suivants :

`>acf(lh)` : Calcule l'ordre du processus  $MA(q)$ .

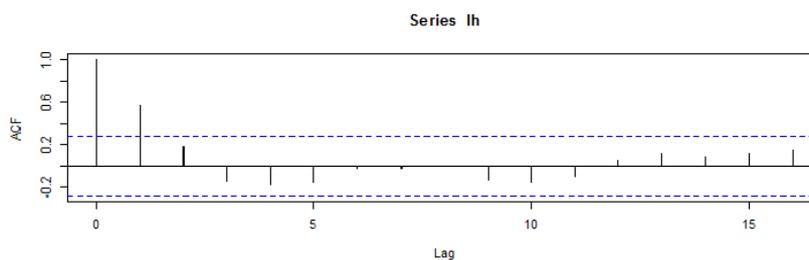


FIGURE 2.3 – Auto corrélation (ACF).

D'après la figure on constate que  $q = 1$ .

`>pacf(lh)` : Calcule l'ordre du processus  $AR(p)$ .

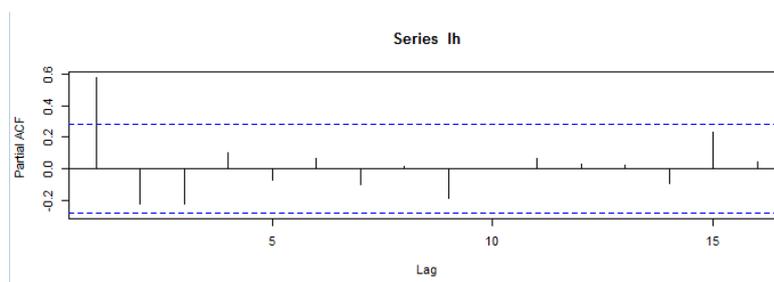


FIGURE 2.4 – Auto corrélation partial (pacf).

D'après la figure on constate que  $p = 1$ .

Donc la série chronologique peut être écrite sous forme d'un modèle  $MA(1)$  :

$$X_t = \mu + \hat{\theta}\varepsilon_{t-1} + \varepsilon_t$$

tell que :  $\varepsilon_t \sim W.N(0, \sigma_\varepsilon^2)$ .

Ou bien sous forme d'un modèle  $AR(1)$  :

$$X_t = \mu + \hat{\phi}(X_{t-1} - \mu) + \varepsilon_t$$

tell que :  $\varepsilon_t \sim W.N(0, \sigma_\varepsilon^2)$ .

Une estimation des paramètres se fait la fonction **arima**, en utilisant le package **States**.

Pour le modèle  $MA(1)$  :

`>arima(lh, order=c(0,0,1))`

```

Call:
arima(x = lh, order = c(0, 0, 1))

Coefficients:
      ma1  intercept
      0.4810   2.4051
s.e.   0.0944   0.0979

sigma^2 estimated as 0.2123:  log likelihood = -31.05,  aic = 68.1
> |

```

FIGURE 2.5 – Résultat d'estimation pour le modèle  $MA(1)$ .

D'où le modèle  $MA(1)$  s'écrit comme suit :

$$X_t = 2.4051 + 0.4810\varepsilon_{t-1} + \varepsilon_t$$

tell que :  $\varepsilon_t \sim W.N(0, \sigma_\varepsilon^2)$ .

Pour le modèle  $AR(1)$  :

>arima(lh,order=c(1,0,0))

```

Call:
arima(x = lh, order = c(1, 0, 0))

Coefficients:
      ar1  intercept
      0.5739   2.4133
s.e.   0.1161   0.1466

sigma^2 estimated as 0.1975:  log likelihood = -29.38,  aic = 64.76
> |

```

FIGURE 2.6 – Résultat d'estimation pour le modèle  $AR(1)$ .

D'où le modèle  $AR(1)$  s'écrit comme suit :

$$\begin{aligned} X_t &= 2.4133 + 0.5739(X_{t-1} - 2.4133) + \varepsilon_t \\ &= 1.0283 + 0.5739X_{t-1} + \varepsilon_t \end{aligned}$$

tell que :  $\varepsilon_t \sim W.N(0, \sigma_\varepsilon^2)$ .

D'après les résultats on remarque que pour le modèle  $AR(1)$  ;  $AIC = 64.76$  ; qui est inférieure à celui du modèle  $MA(1)$  ( $AIC = 68.1$ ).

Alors le meilleur modèle est le modèle  $AR(1)$ .

**Partie 2 :**

**Exemple d'application : Modèle**

**AR(1)**

L'objectif de cette partie est d'appliquer la méthode *RDS* abordée dans le premier chapitre de ce mémoire pour estimer les paramètres d'un modèle *AR(1)* dans un sens bayésien. Notre étude est une généralisation des résultats obtenus par **Fellag et Ibazizen** (2003).

On considère le modèle auto-régressif d'ordre 1 (*AR(1)*), défini comme suit :

$$X_t = \rho X_{t-1} + Y_t, \quad t = \dots, -1, 0, 1, \dots \quad (2.3)$$

où  $0 \leq \rho < 1$  et les  $Y_t$  sont i.i.d et  $Y_t \sim \exp(\theta)$ ,  $\theta > 0$ , et de densité

$$f(y) = \theta \exp(-\theta y) I_{[0, \infty[}(y), \quad \theta > 0$$

On suppose  $X_1 \sim \exp((1 - \rho)\theta)$  sachant que le processus (2.3) est stationnaire.

La vraisemblance pour  $x = (x_1, x_2, \dots, x_n)$  est donnée par :

$$f(x/\theta, \rho) = (1 - \rho)\theta^n e^{-\theta(n\bar{x} - \rho s)} I_A(x)$$

où  $A = \{x, x_1 > 0, x_t - \rho x_{t-1} \geq 0, t = 1, 2, \dots, n\}$  et

$$n\bar{x} = \sum_{i=1}^n x_i, \quad s = n\bar{x} - (x_n - x_1)$$

Les estimateurs du maximum de vraisemblance  $\rho_0$  et  $\theta_0$  pour  $\rho$  et  $\theta$  respectivement sont introduits par Andél (1988) comme suit :

$$\rho_0 = \min\left(1, \frac{x_2}{x_1}, \dots, \frac{x_n}{x_1}\right), \theta_0 = \frac{n}{n\bar{x} - \rho_0 s}$$

Turkman(1990) a considéré une analyse bayésienne pour le modèle (2.3) basée sur une loi a priori non informative définie comme suit :

$$\pi(\theta, \rho) \propto \frac{1}{\theta(1 - \rho)} I_{[0, +\infty[[0, 1[}(\theta, \rho) \quad (2.4)$$

En 2003, Fellag et Ibazizen ont proposé une analyse bayésienne pour le même modèle en utilisant une loi a priori pour  $\rho$  et  $\theta$  plus générale que celle définie par : [11]

$$\pi(\theta, \rho) \propto \frac{\rho^{\beta-1}}{\theta(1 - \rho)} I_{[0, +\infty[[0, 1[}(\theta, \rho), \quad \beta > 0 \quad (2.5)$$

qui contient comme cas special ( $\beta = 1$ ) la loi a priori proposé par Turkman(1990).

En 2016, Benhamada, Boudadi et Ghouil ont proposé une loi a priori plus générale que celle dans (2.5)

$$\pi(\theta, \rho) \propto \frac{\rho^{\beta-1}}{\theta(1-\mu\rho)^\lambda} I_{[0,+\infty[[[0,1[}(\theta, \rho), \quad \beta > 0, \lambda > 0, \mu > 0 \quad (2.6)$$

Elles ont fait une estimation bayésienne des paramètres du modèle précédent, en utilisant les méthodes *MCMC* [3].

Dans le but de faire une comparaison entre les méthodes *MCMC* et la méthode *RDS*, nous visons dans ce mémoire à faire une estimation bayésienne des paramètres du modèle précédent, en utilisant la même loi a priori(2.6) par la méthode *RDS* proposée par Tari et Dehmani (2006)[2].

Comme nous l'avons vu précédemment, la méthode *RDS* consiste à générer un échantillon de taille  $n$  selon la loi de la variable d'entrée  $X$  suivant la procédure de l'échantillonnage *RDS*.

D'abord, nous calculons la loi a posteriori  $\pi(\theta, \rho/x)$  . En appliquant la formule de Bayes, cette loi a posteriori est donnée par :

$$\pi(\theta, \rho/x) \propto f(x/\theta, \rho)\pi(\theta, \rho)$$

Alors :

$$\pi(\theta, \rho/x) \propto \rho^{\beta-1}(1-\rho)(1-\mu\rho)^{-\lambda}\theta^{n-1}e^{-\theta(n\bar{x}-\rho s)} I_{[0,+\infty[[[0,1[}(\theta, \rho) \quad (2.7)$$

Sous l'hypothèses que  $\rho$  et  $\theta$  sont indépendants, les distributions marginales complètes de  $\theta$  et  $\rho$  sont respectivement :

$$F = \pi(\theta/x) \propto \theta^{n-1}e^{-\theta(n\bar{x}-\rho s)} \quad (2.8)$$

$$H = \pi(\rho/x) \propto \rho^{\beta-1}(1-\rho)(1-\mu\rho)^{-\lambda}e^{\theta\rho s} \quad (2.9)$$

Par la méthode d'inversion on trouve :

$$\theta = F^{-1}(u) = \frac{(n-1)\log(\theta) - \log(u)}{n\bar{x} - \rho s}$$

$$\rho = H^{-1}(u) = \frac{\log(u) - (\beta-1)\log(\rho) - \log(1-\rho) + \lambda\log(1-\mu\rho)}{\theta s}$$

Pour chaque paramètre, la procédure d'échantillonnage *RDS* est la suivante :

1. On génère aléatoirement un vecteur de nombre premier  $p = (5, 2, 11, 19, 17, 13, 7)$ .
2. Pour chaque nombre premier  $p_i$ , on calcule les nombres  $r_i = \frac{i - 0.5}{p_i}$   $i = \overline{1, 7}$ .
3. On mélange les  $r_i$ , et on les sauvegarde dans un nouveau vecteurs  $R$ .
4. On calcule les nombre :

$$\theta_i = F^{-1}(R_i) \quad \text{pour } \theta$$

$$\rho_i = H^{-1}(R_i) \quad \text{pour } \rho$$

Pour chaque sous histoire, on calcule les estimateurs :

$$\hat{\theta}_i = \frac{\sum_{i=1}^{P_i} \theta_i}{P_i}$$

$$\hat{\rho}_i = \frac{\sum_{i=1}^{P_i} \rho_i}{P_i}$$

Les estimateurs finaux pour  $\theta$  et  $\rho$  sont :

$$\hat{\theta}_i = \frac{\sum_{i=1}^k \hat{\theta}_i}{k}$$

$$\hat{\rho}_i = \frac{\sum_{i=1}^k \hat{\rho}_i}{k}$$

$k$  est le nombre des sous histoires.

Nous avons fait des simulations pour différentes valeurs de  $n$ ,  $\rho$ ,  $\theta$ ,  $\beta$ ,  $\lambda$  et  $\mu$  dans le but d'étudier l'effet de changement de ces paramètres sur les estimateurs.

Les résultats de la simulations sont donnés dans les tableaux suivants :

TABLE 2.1 – Les valeurs simulées de  $\hat{\rho}$  et des variances pour  $n = 10$ .

$\rho$	Nombre Premier	$\mu = 0.5; \lambda = 4.3$ $\theta = 1.5; \beta = 2.1$	$\mu = 0.8; \lambda = 4$ $\theta = 1.5; \beta = 2.1$	$\mu = 0.9; \lambda = 4.5$ $\theta = 3.5; \beta = 2.1$
0.1	5	0.1011323(0.0003507)	0.1009333(0.0040893)	0.1045495(0.0017033)
	2	0.1004633(0.0024378)	0.1011459(0.0028660)	0.1055691(0.0012737)
	11	0.100166(0.00406200)	0.1003265(0.0047937)	0.0942315(0.0016085)
$\rho$	Nombre premier	$\mu = 0.7; \lambda = 0.8$ $\theta = 2.3; \beta = 3.8$	$\mu = 0.2; \lambda = 0.4$ $\theta = 2.3; \beta = 3.9$	$\mu = 0.1; \lambda = 0.2$ $\theta = 2; \beta = 4$
0.3	5	0.2982337(0.009894)	0.3167504(0.009074)	0.3028101(0.0075490)
	2	0.306878(0.0077829)	0.3095141(0.006251)	0.3061694(0.0057908)
	11	0.322971(0.0134178)	0.2918598(0.008883)	0.2967908(0.0083530)

TABLE 2.2 – Les valeurs simulées de  $\hat{\rho}$  et des variances pour  $n = 30$ .

$\rho$	Nombre Premier	$\mu = 0.5; \lambda = 4.3$ $\theta = 2.3; \beta = 3$	$\mu = 0.5; \lambda = 4.3$ $\theta = 2.3; \beta = 4$	$\mu = 0.5; \lambda = 4.8$ $\theta = 2.7; \beta = 3$
0.1	5	0.1005185(0.000603)	0.1179965(0.000306)	0.1003532(0.00061027)
	2	0.1005914(0.000457)	0.1282757(0.000279)	0.1001984(0.00046047)
	11	0.1000946(0.000686)	0.1207308(0.000306)	0.1002897(0.00069959)
	19	0.1003459(0.000733)	0.122045(0.0003952)	0.1006633(0.00074862)
$\rho$	Nombre premier	$\mu = 0.2; \lambda = 4.5$ $\theta = 2.3; \beta = 13$	$\mu = 0.1; \lambda = 0.1$ $\theta = 0.8; \beta = 15$	$\mu = 0.2; \lambda = 3$ $\theta = 1.4; \beta = 13$
0.3	5	0.3009284(0.000370)	0.3070051(0.000268)	0.3004573(0.00035938)
	2	0.3040358(0.000297)	0.3040256(0.000208)	0.3021126(0.00028999)
	11	0.3001257(0.000416)	0.3093967(0.000308)	0.3006806(0.00041238)
	19	0.2983184(0.000434)	0.2920554(0.000290)	0.2988642(0.00043034)

TABLE 2.3 – Les valeurs simulées de  $\hat{\rho}$  et des variances pour  $n = 50$ .

$\rho$	Nombre Premier	$\mu = 0.5; \lambda = 4.3$ $\theta = 2.3; \beta = 4$	$\mu = 0.1; \lambda = 4.5$ $\theta = 2.8; \beta = 4.7$	$\mu = 0.8; \lambda = 4$ $\theta = 2.8; \beta = 4.5$
0.1	5	0.1000103(0.0002205)	0.1006318(0.00013090)	0.1000322(0.00015895)
	2	0.1011612(0.0001741)	0.1022559(0.00010525)	0.1016677(0.00012752)
	11	0.1002079(0.0002510)	0.1012845(0.00015051)	0.0999781(0.00018040)
	19	0.1000143(0.0002654)	0.1023661(0.00016263)	0.0987409(0.00018622)
	17	0.1018579(0.0002725)	0.1014512(0.00015818)	0.1001002(0.00018950)
$\rho$	Nombre premier	$\mu = 0.2; \lambda = 10$ $\theta = 2.5; \beta = 16$	$\mu = 0.1; \lambda = 0.1$ $\theta = 0.8; \beta = 15$	$\mu = 0.8; \lambda = 4.5$ $\theta = 3.5; \beta = 18$
0.3	5	0.3012184(0.000241)	0.3074518(0.000269)	0.3001725(0.00019568)
	2	0.3026317(0.000192)	0.3055394(0.000210)	0.3000124(0.00015446)
	11	0.3001816(0.000292)	0.308457(0.0003065)	0.3002983(0.00022104)
	19	0.3037953(0.000270)	0.3056884(0.000317)	0.3023521(0.00023656)
	17	0.2928461(0.000269)	0.3147651(0.000333)	0.2983877(0.00022822)

TABLE 2.4 – L'estimateur RDS général de  $\hat{\rho}$  pour  $n = 10$ .

$\rho = 0.1$		
$\mu = 0.5; \lambda = 4.3$ $\theta = 1.5; \beta = 2.1$	$\mu = 0.8; \lambda = 4$ $\theta = 1.5; \beta = 2.1$	$\mu = 0.9; \lambda = 4.5$ $\theta = 3.5; \beta = 2.1$
0.1005872 (0.0033349)	0.1008019 (0.0039163)	0.10145 (0.0015285)
$\rho = 0.3$		
$\mu = 0.7; \lambda = 0.8$ $\theta = 2.3; \beta = 3.8$	$\mu = 0.2; \lambda = 0.4$ $\theta = 2.3; \beta = 3.9$	$\mu = 0.1; \lambda = 0.2$ $\theta = 2; \beta = 4$
0.309361 (0.0103649)	0.3041872 (0.0080697)	0.3019235 (0.0072310)

TABLE 2.5 – L'estimateur RDS général de  $\hat{\rho}$  pour  $n = 30$ .

$\rho = 0.1$		
$\mu = 0.5; \lambda = 4.3$ $\theta = 2.3; \beta = 3$	$\mu = 0.5; \lambda = 4.3$ $\theta = 2.3; \beta = 4$	$\mu = 0.5; \lambda = 4.8$ $\theta = 2.7; \beta = 3$
0.1003876 (0.0006202)	0.122262 (0.0003367)	0.1003762 (0.0006297)
$\rho = 0.3$		
$\mu = 0.2; \lambda = 4.5$ $\theta = 2.3; \beta = 13$	$\mu = 0.1; \lambda = 0.1$ $\theta = 0.8; \beta = 15$	$\mu = 0.2; \lambda = 3$ $\theta = 1.4; \beta = 13$
0.3008512 (0.0003798)	0.3013207 (0.0002689)	0.3005287 (0.0003742)

TABLE 2.6 – L'estimateur RDS général de  $\hat{\rho}$  pour  $n = 50$ .

$\rho = 0.1$		
$\mu = 0.5; \lambda = 4.3$ $\theta = 2.3; \beta = 4$	$\mu = 0.1; \lambda = 4.5$ $\theta = 2.8; \beta = 4.7$	$\mu = 0.8; \lambda = 4$ $\theta = 2.8; \beta = 4.5$
0.1006689 (0.0002368)	0.1015979 (0.0001414)	0.1001038 (0.0001685)
$\rho = 0.3$		
$\mu = 0.2; \lambda = 10$ $\theta = 2.5; \beta = 16$	$\mu = 0.1; \lambda = 0.1$ $\theta = 0.8; \beta = 15$	$\mu = 0.8; \lambda = 4.5$ $\theta = 3.5; \beta = 18$
0.3001346 (0.0002530)	0.308303 (0.0002876)	0.3002446 (0.0002071)

### Interprétation des résultats de $\hat{\rho}$

D'après les tables (2.1), (2.2), (2.3), (2.4), (2.5), (2.6), on a les valeurs calculées de  $\hat{\rho}$  pour différentes valeurs de  $\rho$ ,  $\theta$ ,  $\beta$ ,  $\lambda$  et  $\mu$ . Pour chaque cas, on remarque que  $\hat{\rho}$  est très proche de la vraie valeur prise. Pour chaque  $n$ , on a des valeurs différentes de  $\lambda$ ,  $\beta$  et  $\mu$ , lorsque  $n$  augmente, il faut augmenter les valeurs de  $\lambda$ ,  $\beta$  et  $\mu$ .

TABLE 2.7 – Les valeurs simulées de  $\hat{\theta}$  et des variances pour  $n = 10$ .

$\theta$	nombre premier	$\rho=0.6$	$\rho=0.65$	$\rho=0.7$
2.1	5	2.190057(0.0834410)	2.143859(0.0799580)	1.9704310(0.067540)
	2	2.145595(0.1062592)	2.227328(0.1145080)	2.338050(0.1261760)
	11	2.010547(0.0567230)	1.962777(0.0540590)	2.130500(0.0636930)
2.2	5	2.172562(0.0730470)	2.222674(0.0764560)	2.189015(0.0741580)
	2	2.192042(0.0470750)	2.104249(0.0893700)	2.262842(0.1033490)
	11	2.222215(0.0623645)	2.336118(0.0689210)	1.918439(0.0464790)
2.3	5	2.315801(0.0746690)	2.312336(0.0744450)	2.276885(0.0721800)
	2	2.33044(0.09719048)	2.290425(0.0938810)	2.295743(0.0943170)
	11	2.377486(0.0648650)	2.350731(0.0634138)	2.322545(0.0619020)

TABLE 2.8 – Les valeurs simulées de  $\hat{\theta}$  et des variances pour  $n = 30$ .

$\theta$	nombre premier	$\rho=0.6$	$\rho=0.65$	$\rho=0.7$
2.1	5	2.187401(0.0103120)	2.114191(0.0096330)	2.110186(0.009596)
	2	2.071908(0.0101280)	2.162945(0.0110370)	1.985015(0.009265)
	11	2.218229(0.0098410)	2.157942(0.0093130)	2.130417(0.009077)
	19	1.992867(0.0075540)	2.019517(0.0077570)	2.158707(0.008863)
2.2	5	1.982224(0.0074640)	2.25880(0.00969200)	2.220766(0.009368)
	2	2.13081(0.00938901)	2.222746(0.0102160)	2.205951(0.010062)
	11	2.200569(0.0085754)	2.206832(0.0086240)	2.204308(0.008604)
	19	2.605910(0.0114702)	2.176763(0.0080030)	2.245941(0.008520)
2.3	5	2.312565(0.0090680)	2.311989(0.0090640)	2.293646(0.008920)
	2	2.274034(0.0095000)	2.318269(0.0098730)	2.313443(0.009832)
	11	2.333632(0.0086410)	2.285038(0.0082850)	2.325634(0.008582)
	19	2.331926(0.0082500)	2.287645(0.0079400)	2.281955(0.007901)

TABLE 2.9 – Les valeurs simulées de  $\hat{\theta}$  et des variances pour  $n = 50$ .

$\theta$	nombre premier	$\rho=0.6$	$\rho=0.65$	$\rho=0.7$
2.1	5	2.244864(0.00389000)	2.26111(0.0039470)	1.970276(0.002997)
	2	2.084109(0.00353500)	2.004798(0.003271)	2.031274(0.003358)
	11	2.029093(0.00304000)	1.999286(0.002951)	2.19380(0.0035540)
	19	2.085000(0.00314500)	2.084464(0.003112)	2.161303(0.003346)
	17	2.046461(0.00301800)	2.146171(0.003320)	2.100408(0.003180)
2.2	5	2.114136(0.00304200)	2.19656(0.0032830)	2.209159(0.003321)
	2	2.243678(0.00360103)	2.186931(0.003421)	2.225046(0.003541)
	11	2.148288(0.00301262)	2.246935(0.003295)	2.166633(0.003064)
	19	2.155487(0.00294770)	2.221313(0.003130)	2.251183(0.003215)
	17	2.251087(0.00289600)	2.166648(0.002995)	2.198909(0.003085)
2.3	5	2.272162(0.00313700)	2.318015(0.003265)	2.307287(0.003234)
	2	2.324093(0.00343900)	2.291944(0.003345)	2.331569(0.003462)
	11	2.29026(0.003063000)	2.312944(0.003124)	2.310627(0.003118)
	19	2.325605(0.00307500)	2.315892(0.003049)	2.301942(0.003012)
	17	2.344828(0.00350846)	2.286845(0.002980)	2.266314(0.002936)

TABLE 2.10 – L'estimateur RDS général de  $\hat{\theta}$ .

n	$\theta$	$\rho=0.6$	$\rho=0.65$	$\rho=0.7$
10	2.1	2.115400(0.082141)	2.111321(0.082842)	2.146327(0.085805)
	2.2	2.195606(0.077465)	2.221014(0.078249)	2.217151(0.074662)
	2.3	2.342142(0.078908)	2.317831(0.077247)	2.298391(0.076133)
30	2.1	2.117601(0.009459)	2.113649(0.009435)	2.096081(0.009200)
	2.2	2.222987(0.009224)	2.216285(0.009134)	2.219242(0.009139)
	2.3	2.313039(0.008865)	2.300735(0.008790)	2.303670(0.008890)
50	2.1	2.097905(0.003320)	2.099166(0.003320)	2.091412(0.003287)
	2.2	2.201283(0.003222)	2.203677(0.003225)	2.210186(0.003245)
	2.3	2.292641(0.003122)	2.305128(0.003154)	2.303548(0.003152)

### Interprétation des résultats de $\hat{\theta}$

Contrairement à  $\rho$ , on a la distribution marginale de  $\theta$  ne dépend pas des paramètres  $\lambda$ ,  $\beta$  et  $\mu$ . Elle dépend seulement de  $n$  et  $\rho$ , pour cela nous avons pris des différentes valeurs de  $n$  et  $\rho$  pour déterminer l'effet de ce changement sur la valeur de l'estimateur. En général et pour chaque cas et d'après les tables (2.7), (2.8), (2.9) et (2.10) les valeurs de  $\hat{\theta}$  sont très proches des vraies valeurs prises.

---

## CONCLUSION ET PERSPECTIVE

Au long de ce travail, nous avons aboutit à estimer les paramètres d'un modèle  $AR(1)$  dans un sens bayésien en utilisant la méthode  $RDS$ . Les résultats de cette estimation, confirment encore une fois la flexibilité de l'approche bayésienne pour les modèle auto-régressifs.

Ces résultats montrent aussi comment la méthode  $RDS$  est plus pratique pour l'approche bayésienne, parce que la quantité que l'on cherche à calculer prend généralement la forme d'une espérance mathématique (calcul de  $E[\theta/x]$ ). Donc on peut gagner du temps par rapport à l'approche classique.

En comparant nos résultats avec ceux obtenus par En Benhamada, Boudadi et Ghouil en 2016, qui ont fait une estimation Bayésienne des paramètres du même modèle en utilisant les méthodes  $MCMC$  (2.5). On constate qu'il n'ya pas une grande différence entre les estimateurs calculés par  $RDS$  et les estimateurs calculés par  $MCMC$ . On peut dire donc que les deux méthodes sont flexibles pour l'estimation bayésienne des paramètres d'un modèle  $AR(1)$ .

En perspective, il serait très intéressant de généraliser ce travail en choisissant une loi a priori qui permet de meilleures performances d'estimation bayésienne, ou bien d'appliquer la méthode  $RDS$  sur un autre type de séries temporelles ; on peut choisir l'exemple (2.1) de la première partie du chapitre 2. Ou plus simplement aborder la question de prévision de ce même modèle après l'estimation des paramètres.

---

# BIBLIOGRAPHIE

- [1] **J. Andèl**, *On AR(1) processus with exponential white noise*. Comm. In Stat. theory and Methods, 17, 1481-1495.
- [2] **Amaral Turkman, M. A. (1990)**. *bayésienne analysis of an autoregressive process with exponential white noise*. statistics, 4, 601-608.
- [3] **S. Benhamada, Z. Boudadi**. *Aspects de la Robustesse Bayésienne* Mémoire pour l'obtention du diplôme de Mastre (2015,2016), Université Mohamed Seddik Ben Yahia - Jijel
- [4] **P.J. Brockwell et R. Davis** , *An introduction to time series forecasting*. Second edition. Springer-Verlag,(2000) New York, USA.
- [5] **F. Campilb et V. Rossi (2006)**. *Les méthodes de Monte Carlo : exposé introductif*. Université Montpellier.
- [6] **A. Charpentier**, *Cours de Séries Temporelles*. Fédération Française des Sociétés d'Assurance (FFSA)-Centre de Recherche en Economie et en Statistique (CREST),1998.
- [7] **Christian, P. Robert**. *Le Choix Bayésien : Principes et pratique*. (2006) Springer-Verlag France, Paris.
- [8] **P.J. Diggle**. *Time Series : A Biostatistical Introduction*. Oxford 1990, table A.1, series 3
- [9] **J. DUPUIS**, *Statistique bayésienne et algorithmes MCMC* . (LSP-UPS)(Version Septembre 2007)IMAT (Master 1).

- [10] **L. Elie, B. Lapeyre.** *Introduction aux Méthodes de Monte-Carlo.* Lien : [cermics.enpc.fr/bl/PS/SIMULATION-X/polymonte-carlo-x.pdf](http://cermics.enpc.fr/bl/PS/SIMULATION-X/polymonte-carlo-x.pdf). Septembre 2001.
- [11] **H. Fellag, M. Ibazizen.** *Bayesian estimation of an AR(1) process with exponential white noise,* Statistics, October 2003, Vol. 37(5), pp. 365-372.
- [12] **X. Gourdon ,** *Les maths en tête : Analyse.* second édition, 2008.
- [13] **J.M. Hammersley and D. C. Handscomb.** *Monte Carlo methods.* Oxford University, 1967.
- [14] **C. Gouieroux et A. Montfort,** *Séries temporelles et modèles dynamiques.* second édition, 1997.
- [15] **W. K. Hastings.** *Monte carlo sampling methods using markov chains and their applications.* Biometrika, 57(1) :97109, 1970.
- [16] **D.Hettak et Gadoum,** *L. Etude de la gestion des stocks de la matière première à l'entreprise alcost,* mémoire de fin d'études recherche opérationnelle, Université A.Mira Béjaia, 2007.
- [17] **B. Lapeyre, E. Pardoux, R. Sentis.** *Méthodes de Monte-Carlo pour les équations de transport et de diffusion.* Series : Mathématiques et Applications, Vol. 29.
- [18] **M. Lavine, L. Wasserman, R.L. Wolpert , (1991).** *Bayesian inference with specified marginals.* J. Amer. Statist. Assoc., 86, 400-403.
- [19] **N. Metropolis,** *The Begining of the Monte Carlo Method.* Los Alamos Science, 15(584) : 125130, 1987.
- [20] **R. Meyer et J. Yu,** *Bugs for a bayesian analysis of stochastic volatility models.* The Econometrics Journal, 3(2) :198-215, 2000.
- [21] **P.D. Moral et N. Bartoli.** *Simulation et algorithmes stochastiques : une introduction avec applications.* Cepaduès, Toulouse, 2001.
- [22] **M.Ourbih,** *Cours de 3<sup>ème</sup> année STID "Simulation a événement discret",* Département Mathématique, Université de béjaia, 2012.
- [23] **E. Parent, J. Bernier,.** *Le raisonnement bayésien : Modélisation et Inférence.* Springer-Verlag (2007) France, Paris.
- [24] **H. Raiffa et R. Schlaifer,** *Applied statistical decision theory.* Boston, Massachusetts : Harvard Business School, 1961.
- [25] **E. Saliby,** *Descriptive sampling : an improvement over latin hypercube sampling.* In Winter simulation Conference (1997), pp. 230–233.

- [26] **E. Saliby**, *Descriptive sampling : an improvement over latin hypercube sampling*. In Winter simulation Conference (1997), pp. 230–233.
- [27] **N.Sellami**, *Cours de 2<sup>ème</sup> année Master "séries chronologiques"*, Département Mathématique, Université de jijel , 2018/2019.
- [28] **M.Tari et A. Dahmani**, *Refined descriptive sampling : A better approach to monte carlo simulation*. Simulation Modelling Practice and Theory 14, 2 (2006), 143–160.

## **Résumé**

Dans ce mémoire de master nous avons essayé de donner au lecteur le langage nécessaire pour établir une estimation d'un paramètre inconnu dans un sens bayésienne, ainsi que la procédure d'échantillonner par la méthode *RDS*.

Nous sommes intéressés à un modèle  $AR(1)$   $x_t = \rho x_{t-1} + y_t$  où  $0 \leq \rho < 1$  et  $y_t$  sont des variables aléatoires indépendantes suivant la loi exponentielle de paramètre  $\theta$  (inconnu). Nous avons estimé les paramètres  $\rho$  et  $\theta$  par la méthode *RDS* proposée par Tari et Dehmani (2006). Notre étude a été basée sur les résultats obtenus par Fellag et Ibazizn (2003).

## **Mots clés :**

Estimation Bayésienne, Monte Carlo, *RDS*, Processus autorégressif.

## **Abstract**

In this master thesis we have tried to give the reader the necessary language to establish a Bayesian estimation of an unknown parameter, as well as, the procedure of sampling by the *RDS* method. We are interested in an  $AR(1)$  model  $x_t = \rho x_{t-1} + y_t$  where  $0 \leq \rho < 1$  and  $y_t$  are independent random variables according to the exponential law of parameter  $\theta$  (unknown). We estimate the parameters  $\rho$  and  $\theta$  by the *RDS* method proposed by Tari and Dehmani (2006). Our study was based on the results obtained by Fellag and Ibazizn (2003).

## **Keywords :**

Bayesian estimation, Monte Carlo, *RDS*, Autoregressive process.