

*République Algérienne Démocratique et Populaire*  
*Ministère de l'Enseignement Supérieur et de la Recherche*  
*Scientifique*  
*Université Mouloud Mammeri de Tizi Ouzou*  
*Faculté des Sciences*  
*Département des Mathématiques*  
*Laboratoire LMPA*



Filière : Mathématiques  
Spécialité : Probabilités & Statistique

---

*Inférence dans les modèles autorégressifs à  
coefficients aléatoires RCAR(1).*

---

*Présenté par*  
Melle MENOUCHE Ourdia

*Sous la direction de*  
Mme ATIL Lynda

Soutenue le 27 septembre 2023 devant le jury :

Mr. FELLAG Hocine	Professeur	UMMTO	Président
Mme. ATIL Lynda	MCA	UMMTO	Rapporteur
Mme. BELKACEM Cherifa	MCB	UMMTO	Examinatrice

Année Universitaire : 2022/2023

*Je dédie ce travail à la mémoire de ma mère, Berkane Sadia, qui nous a quittée avant de pouvoir témoigner de l'achèvement de cette étape de ma vie. Ma mère a toujours été une source constante d'inspiration et de soutien, son amour inconditionnel et son dévouement ont été des moteurs essentiels de mes succès académiques.*

*Malgré la peine que sa perte a engendré, je suis profondément honorée de lui consacrer ce travail, qui symbolise pour moi non seulement une réalisation personnelle, mais aussi un hommage à la femme qui m'a élevée et a toujours cru en moi. Je suis convaincue que si elle était encore parmi nous, elle serait fière de mon parcours académique et de cette réalisation. Son influence perdurera dans mon cœur et continuera de guider mes projets futurs. Maman, Je te suis redevable de ce que je suis aujourd'hui, et je m'efforcerai toujours de maintenir ta fierté envers moi. Que le bon Dieu t'accueille dans son éternel paradis.*

# Remerciements

*Je tiens à exprimer mes sincères remerciements à tous ceux qui ont joué un rôle essentiel dans la réalisation de ce mémoire. En premier lieu, je souhaite adresser ma profonde gratitude à ma promotrice, Madame Atil Lynda, pour son accompagnement précieux tout au long de la réalisation de ce travail. Son expertise, sa disponibilité et son soutien constant ont été les piliers de ma réussite. Elle m'a encouragé à repousser mes limites, à explorer de nouvelles perspectives et à atteindre des sommets académiques que je n'aurais jamais crus possibles.*

*De plus, je souhaite adresser mes sincères remerciements aux membres du jury, Mme Belkacem et Mr Fellag, pour l'honneur qu'ils m'ont fait en acceptant de consacrer leur temps et leur expertise à l'évaluation de ce travail.*

*Je tiens également à exprimer ma reconnaissance envers mon père et mes frères, Farid et Yacine. Je tiens à accorder une mention spéciale à mon frère Makhlouf, qui occupe une place unique et irremplaçable dans ma vie. Son soutien inestimable a été le moteur qui m'a permis d'atteindre cette étape cruciale. Je souhaite également remercier chaleureusement ma meilleure copine Sarah et sa famille, qui ont toujours été là pour moi.*

*Enfin, l'équipe pédagogique du master Probabilités et Statistiques mérite ma gratitude pour leur enseignement précieux. En résumé, ma reconnaissance est profonde envers toutes les personnes qui ont joué un rôle essentiel dans la concrétisation de ce travail de recherche.*

# Résumé

Ce mémoire se penche sur l'inférence statistique dans les modèles autorégressifs à coefficient aléatoire d'ordre 1, RCAR(1). L'objectif principal de cette étude est d'estimer les paramètres clés de ces modèles, en utilisant trois méthodes différentes : les moindres carrés, la méthode du maximum de vraisemblance et l'approche Bayésienne hiérarchique basée sur la méthode de Monte-Carlo par chaînes de Markov (MCMC). La méthode des moindres carrés consiste à minimiser la somme des carrés des résidus et à dériver par rapport aux paramètres inconnus. Ensuite, la méthode du maximum de vraisemblance nous permet d'estimer les paramètres via la fonction de vraisemblance. Enfin, dans le cas de la méthode Bayésienne hiérarchique l'échantillonneur de Gibbs permet de générer des échantillons à partir d'une distribution a posteriori, qui, une fois moyennés, donne la valeur estimée du paramètre inconnu. L'efficacité des trois méthodes est comparée en fonction de la moyenne de l'erreur quadratique moyenne (AMSE) pour des données de simulation.

**Mots-clés :** Modèle autorégressif à coefficients aléatoires d'ordre 1 (RCAR(1)), estimation, stationnarité, méthode des moindres carrés, estimation du maximum de vraisemblance, approche Bayésienne hiérarchique, modèle autorégressif dynamique à coefficients aléatoires d'ordre un (RCDAR(1,1)), échantillonnage de Gibbs, la moyenne de l'erreur quadratique moyenne (AMSE).

# Abstract

This thesis focuses on statistical inference in first-order random coefficient autoregressive models RCAR(1). The main objective of this study is to estimate the key parameters of these models using three different methods : least-squares, maximum likelihood estimation, and a hierarchical Bayesian approach based on Markov Chain Monte Carlo (MCMC) sampling. The least squares method consists in minimizing the sum of the squared residuals and computing derivatives with respect to the unknown parameters. Next, maximum likelihood estimation enables parameter estimation through the likelihood function. Finally, in the case of the hierarchical Bayesian method, the Gibbs sampler is employed to generate samples from a posterior distribution, which, when averaged, provides the estimated value of the unknown parameter. The efficiency of these three methods is compared based on the average mean squared error (AMSE) for simulated data.

**Keywords :** First order of random coefficient autoregressive model (RCAR(1)), estimation, stationarity, least-squares method, maximum likelihood estimation, hierarchical Bayesian approach, first order of random coefficient dynamic autoregressive model (RCDAR(1,1)), Gibbs sampling, average mean square error (AMSE).

# Table des matières

<b>Introduction générale</b>	<b>9</b>
<b>1 Généralités et paradigme Bayésien</b>	<b>12</b>
1.1 Introduction . . . . .	12
1.2 Théorème de Bayes . . . . .	13
1.3 Présentation du modèle Bayésien . . . . .	14
1.4 Inférence statistique Bayésienne . . . . .	16
1.4.1 Approche décisionnelle, Estimation ponctuelle . . . . .	16
1.4.2 Choix de la loi a priori . . . . .	21
1.4.3 Estimation par intervalles . . . . .	26
1.5 Modèle Bayésien hiérarchique et méthodes Monte-Carlo par chaîne de Markov . . . . .	28
1.5.1 L'analyse Bayésienne hiérarchique . . . . .	28
1.5.2 Méthode de Monte-Carlo par chaîne de Markov . . . . .	29
1.5.2.1 Algorithme de Metropolis-Hastings . . . . .	30
1.5.2.2 L'échantillonnage de Gibbs . . . . .	31
1.6 Notions de base en séries chronologiques . . . . .	32
1.6.1 Processus stochastiques . . . . .	32
1.6.2 Une série chronologique . . . . .	32
1.6.3 Théorème de Wold . . . . .	33
1.6.4 Opérateur retard et opérateur différence . . . . .	34
1.6.5 Stationnarité . . . . .	34
1.6.5.1 Série stationnaire au sens strict ou stationnarité forte	34
1.6.5.2 Stationnarité d'ordre deux ou la stationnarité faible	35
1.6.6 Ergodicité . . . . .	36
1.6.7 Processus bruit blanc (White noise) . . . . .	38

1.6.8	Marche aléatoire (Random Walk) . . . . .	38
1.6.9	Fonction d'autocovariance . . . . .	39
1.6.10	Fonction d'autocorrélation (ACF) . . . . .	39
1.6.11	Fonction d'autocorrélation partielle (PACF) . . . . .	40
1.6.12	Quelques modèles de séries temporelles . . . . .	41
<b>2</b>	<b>Modèle autorégressif à coefficients aléatoires</b>	<b>43</b>
2.1	Introduction . . . . .	43
2.2	Étude des modèles autorégressifs à coefficients aléatoires . . . . .	44
2.2.1	Modèle autorégressif à coefficients aléatoires d'ordre $p$ . . . . .	44
2.2.2	Modèle autorégressif à coefficient aléatoire d'ordre un . . . . .	44
2.3	Conditions de stationnarité . . . . .	45
2.3.1	L'existence d'une solution faiblement stationnaire . . . . .	45
2.3.2	L'existence d'une solution strictement stationnaire . . . . .	46
2.4	L'approche fréquentiste dans l'estimation d'un modèle RCAR(1) . . . . .	50
2.4.1	Estimation par la méthode des moindres carrés . . . . .	50
2.4.1.1	Calcul de l'estimateur par la méthode des moindres carrés . . . . .	51
2.4.1.2	La consistance et la normalité asymptotique de l'estimateur des moindres carrés . . . . .	55
2.4.2	Estimation des paramètres d'un modèle RCAR(1) par la méthode du maximum de vraisemblance . . . . .	56
2.4.2.1	Calcul de l'estimateur du maximum de vraisemblance . . . . .	57
2.4.2.2	La consistance forte de l'estimateur du maximum de vraisemblance . . . . .	60
2.4.2.3	Le théorème central limite . . . . .	65
<b>3</b>	<b>Application numérique</b>	<b>77</b>
3.1	Introduction . . . . .	77
3.2	Application sur le modèle RCAR(1) . . . . .	78
3.2.1	Présentation du modèle RCAR(1) . . . . .	78
3.2.2	Méthodes d'estimation des paramètres . . . . .	79
3.2.2.1	La méthode des moindres carrés . . . . .	79
3.2.2.2	La méthode du maximum de vraisemblance . . . . .	80
3.2.2.3	La méthode Bayésienne hiérarchique . . . . .	81
3.2.3	Étude de simulation . . . . .	85
3.3	Application sur le modèle RCDAR(1,1) . . . . .	92
3.3.1	Présentation du modèle RCDAR( $p,q$ ) et RCDAR(1,1) . . . . .	92

## TABLE DES MATIÈRES

---

3.3.2	Estimation des paramètres d'un modèle RCDAR(1,1) par la méthode Bayésienne hiérarchique . . . . .	93
3.3.2.1	Choix des densités a priori . . . . .	94
3.3.2.2	Inférence a posteriori par échantillonnage de Gibbs	95
3.3.3	Étude de Simulation . . . . .	96
	<b>Conclusion générale</b>	<b>101</b>

# Table des figures

3.1	Le graphique de la série temporelle pour les données générées pour $n = 100$ . . . . .	86
3.2	Le graphique de la série temporelle pour les données générées pour $n = 300$ . . . . .	87
3.3	Le graphique de la série temporelle pour les données générées pour $n = 500$ . . . . .	88
3.4	Le graphique de la série temporelle des données générées à partir d'un modèle AR(1) pour $y_t$ ( $\eta = 0.1, 0.5$ , et $0.9$ ). . . . .	97
3.5	Le graphique de la série temporelle des données générées à partir d'un modèle RCDAR(1,1) pour $X_t$ ( $\eta = 0.1$ ). . . . .	98
3.6	Le graphique de la série temporelle des données générées à partir d'un modèle RCDAR(1,1) pour $X_t$ ( $\eta = 0.5$ ). . . . .	99
3.7	Le graphique de la série temporelle des données générées à partir d'un modèle RCDAR(1,1) pour $X_t$ ( $\eta = 0.9$ ). . . . .	99

# Liste des tableaux

1.1	Quelques estimateurs de Bayes usuels. . . . .	19
1.2	Lois a priori conjuguées usuelles . . . . .	22
3.1	La moyenne de l'erreur quadratique moyenne (AMSE) de l'estimation des paramètres par la méthode des moindres carrés . . . . .	89
3.2	La moyenne de l'erreur quadratique moyenne (AMSE) de l'estimation des paramètres par la méthode du maximum de vraisemblance	90
3.3	La moyenne de l'erreur quadratique moyenne (AMSE) de l'estimation des paramètres par la méthode Bayésienne hiérarchique . . . .	91

# Introduction générale

L'évolution des séries temporelles est étroitement liée au développement de la statistique, de l'informatique et à l'évolution des besoins en analyse des données au fil du temps. Les séries temporelles, bien que présentes depuis l'Antiquité sous la forme d'enregistrements chronologiques tels que les données météorologiques, les prix des marchandises et les événements historiques, n'ont véritablement pris forme que grâce aux progrès dans ces domaines clés.

Au XIXe siècle, des statisticiens éminents tels que Francis Galton et George Udny Yule ont posé les bases des méthodes formelles d'analyse des séries temporelles. Cependant, ce n'est que plus tard, au XXe siècle, avec le développement de la théorie des probabilités, que l'analyse des séries temporelles s'est vraiment formalisée. Des modèles probabilistes, comme les modèles ARIMA, ont vu le jour et sont devenus des outils fondamentaux pour la modélisation et la prévision des séries temporelles. Ces derniers ont été initialement conçus pour modéliser des séries temporelles stationnaires, où les statistiques clés, telles que la moyenne et la variance, restent stables au fil du temps. Ces modèles combinent des composantes autoregressives (AR) et des composantes de moyenne mobile (MA) pour saisir les dépendances temporelles et les motifs saisonniers des données.

Cependant, les modèles ARIMA présentent des limites importantes. Ils ont du mal à traiter efficacement les données non stationnaires, les tendances à long terme et les changements structurels dans les données.

Face à ces limites, les chercheurs ont ressenti le besoin de disposer de modèles plus flexibles pour modéliser des séries temporelles complexes. Ils ont cherché des solutions permettant aux coefficients des équations de régression de varier au fil du temps, afin de mieux capturer les changements potentiels dans les relations entre

---

les variables.

C'est ainsi que les modèles RCAR (Random Coefficient Autoregressive) ont vu le jour. Les modèles RCAR constituent une extension des modèles ARIMA, mais avec une innovation cruciale : les coefficients des équations de régression sont traités comme des variables aléatoires suivant certaines distributions. Depuis leur apparition, les propriétés de base ont été dérivées par de nombreux auteurs. Par exemple, Consilka(1974,1976)[12][11] et Andel (1976)[1] ont examiné la stabilité et la stationnarité au second ordre du modèle RCAR. Nicholls and Quinn (1982)[26] ont, en plus, étudié l'inférence statistique. Ces modèles sont intensivement utilisés dans divers domaines tels que l'analyse des séries temporelles en économie, en finance et en biologie, en médecine, en météorologie, en traitement du signal, et dans d'autres domaines.

Les modèles RCAR sont particulièrement utiles lorsque les données présentent des tendances, des cycles, des ruptures structurelles ou des fluctuations non linéaires. Ils peuvent capturer ces variations de manière plus précise que les modèles ARIMA traditionnels, qui supposent des coefficients constants.

Les modèles RCAR ont été construits pour répondre à la nécessité de modéliser des séries temporelles complexes et variées de manière plus flexible, tout en permettant la prise en compte des changements structurels, de la non-stationnarité et en offrant des approches d'inférence avancées. Leur genèse découle des défis de modélisation des données temporelles dans un large éventail d'applications.

Notre travail se base principalement sur la thèse de Dazhe Wang intitulée "Frequentist and Bayesian analysis of Random Coefficient Autoregressive Models" (2003)[32]. Nous utiliserons également des références telles que Hwang and Basawa (1993) [20] et Quinn and Nicholls (1981) [28]. Notre objectif principal est de développer une compréhension approfondie des méthodes d'inférence statistique pour le modèle autoregressif à coefficient aléatoire d'ordre un, noté par RCAR(1). Nous chercherons à explorer les outils et les techniques qui permettent d'estimer ces paramètres. Notre évaluation repose principalement sur la moyenne de l'erreur quadratique moyenne (AMSE) comme critère de comparaison. Nous cherchons à déterminer quelle méthode offre les estimations les plus précises et les plus fiables des paramètres des modèles RCAR(1). Cette application nous permet de mieux comprendre comment ces méthodes fonctionnent.

Notre mémoire est composé de trois chapitres, dans le premier nous présentons quelques généralités sur l'inférence statistique Bayésienne et Bayésienne hiérarchique.

---

Nous introduisons également des concepts fondamentaux liés aux séries chronologiques, jetant ainsi les fondations nécessaires pour la compréhension des modèles RCAR.

Le deuxième chapitre se penche sur le modèle RCAR(1) en détail. Nous introduisons le modèle autorégressif à coefficient aléatoire RCAR d'ordre  $p$  et nous considérons en particulier un RCAR d'ordre 1. Nous donnons ensuite les conditions de stationnarité d'un modèle RCAR(1) et nous nous concentrons sur les méthodes d'estimation fréquentistes telles que la méthode des moindres carrés et la méthode du maximum de vraisemblance.

Dans le troisième chapitre nous comparons les méthodes d'inférence dans les modèles RCAR(1). Nous mettons en lumière les avantages et les limites de chaque méthode, en nous appuyant sur des données de simulation et des calculs basés sur la méthode de Monte-Carlo par chaînes de Markov (MCMC). L'utilisation de l'algorithme d'échantillonnage de Gibbs dans notre calcul MCMC constitue un aspect important de notre approche.

# Généralités et paradigme Bayésien

## 1.1 Introduction

La statistique Bayésienne a été nommée en l'honneur du révérend Thomas Bayes, un mathématicien britannique du 18<sup>ème</sup> siècle. Son article posthume de 1763 [5], intitulé "An Essay to Solving a Problem in the Doctrine of Chances", présente sa méthode pour mettre à jour les probabilités en fonction de nouvelles données. Malgré cela, sa méthode est restée largement ignorée pendant de nombreuses décennies. Au fil des siècles, de nombreux mathématiciens et statisticiens, dont Pierre-Simon de Laplace, ont réussi au développement des statistiques Bayésiennes. Cependant, elles ont connu une renaissance grâce aux avancées de l'informatique et aux nouvelles méthodes de simulation.

Au début du 20<sup>ème</sup> siècle, des statisticiens tels que Harold Jeffreys et Bruno de Finetti ont commencé à développer et à promouvoir la statistique Bayésienne. Cependant, la statistique fréquentiste était largement prédominante à l'époque, et la statistique Bayésienne n'a pas été largement acceptée avant les années 1960 et 1970. L'utilisation de méthodes de simulation telles que l'échantillonnage de Monte-Carlo a permis la manipulation et l'analyse de distributions de probabilité complexes.

La statistique Bayésienne est une approche de la statistique qui se base sur l'inférence Bayésienne pour traiter les données et effectuer des dépenses. L'approche Bayésienne repose sur le théorème de Bayes, qui permet de mettre à jour

les croyances à propos d'un événement en fonction de l'obtention de nouvelles données. Cette approche est particulièrement utile lorsque les données sont rares, ou lorsqu'il est important de prendre en compte des connaissances antérieures ou des hypothèses subjectives. Aujourd'hui, la statistique Bayésienne est largement utilisée dans de nombreux domaines, notamment en science, en ingénierie, en médecine, en finance et en économie pour traiter des problèmes tels que l'estimation des paramètres, la prévision des séries chronologiques, la classification, la prise de décision et l'analyse des réseaux.

Ce chapitre donnera un aperçu général de l'inférence statistique Bayésienne et Bayésienne hiérarchique ainsi que certaines notions sur les séries chronologiques.

## 1.2 Théorème de Bayes

Considérons les événements aléatoires  $A$  et  $B$  tels que  $P(B) \neq 0$ .  $P(A|B)$  et  $P(B|A)$  sont définis respectivement comme étant la probabilité de  $A$ , conditionnellement à la réalisation de  $B$ , et la probabilité de  $B$ , conditionnellement à la réalisation de  $A$ , de la manière suivante :

$$P(A|B) = \frac{P(A, B)}{P(B)} \quad \text{et} \quad P(B|A) = \frac{P(B, A)}{P(A)}$$

En utilisant la relation  $P(A, B) = P(B, A)$ , qui représente la probabilité que les deux événements  $A$  et  $B$  se produisent simultanément, on peut établir la relation entre les deux probabilités conditionnelles  $P(A|B)$  et  $P(B|A)$ .

En remplaçant  $P(A|B)$  par  $P(B|A)$  qui est égale à  $P(B|A)P(A)$ , on peut déduire la relation entre les deux probabilités conditionnelles  $P(A|B)$  et  $P(B|A)$  :

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Cette relation, connue sous le nom de théorème de Bayes, représente en réalité un principe d'actualisation. Elle décrit comment la vraisemblance de  $P(A)$  est mise à jour vers  $P(A|B)$  une fois que l'événement  $B$  a été observé.

Bayes (1763)[5] a formulé une version continue de ce résultat : pour deux variables aléatoires  $X$  et  $Y$  avec une distribution conditionnelle  $f(x|y)$  et une distribution marginale  $g(y)$ , la distribution conditionnelle de  $y$  sachant  $x$  est donnée par :

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

Ce théorème d'inversion est naturellement justifié d'un point de vue probabiliste. Cependant, Bayes et Laplace sont allés plus loin en considérant que l'incertitude concernant le paramètre  $\theta$  d'un modèle peut être décrit par une distribution de probabilité  $\pi$  sur  $\Theta$ , appelée distribution a priori. L'inférence est alors basée sur la distribution conditionnelle de  $\theta$  étant donné  $x$ ,  $\pi(\theta|x)$ , qui est appelée distribution a posteriori et est définie par :

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

La formule de Bayes, également connue sous le nom d'équation de Bayes, peut être présentée de la manière suivante, étant donné que le dénominateur est indépendant de  $\theta$  :

$$\pi(\theta|x) \propto f(x|\theta)\pi(\theta)$$

La transition de la distribution a priori à la distribution a posteriori des paramètres du modèle statistique, exprimée par la formule de Bayes, peut être interprétée comme une mise à jour ou une actualisation des connaissances en fonction des observations.

## 1.3 Présentation du modèle Bayésien

**Définition 1.1.** *On appelle modèle statistique Bayésien la donnée d'un modèle statistique paramétré  $(\mathcal{X}, \mathcal{A}, \mathcal{P}_{\theta})$ , où  $\theta \in \Theta$ .*

**Définissons les espaces intervenant dans un modèle Bayésien**

- **Espace des observations**

Noté  $\mathcal{X}$ , il représente l'ensemble des résultats suite à une étude d'un phénomène.

- **Espace des actions ou décisions**

Noté  $\mathcal{A}$ , il représente l'ensemble des actions ou décisions à prendre après l'obtention de l'information.

- **Espace des états de la nature**

Noté  $\Theta$ , il représente l'espace des paramètres inconnus  $\theta$ .

- **Espace des règles de décisions**

Noté  $\mathcal{D}$ , il représente l'ensemble des règles de décisions qu'on définit comme une application de  $\mathcal{X}$  dans  $\mathcal{A}$  notée par  $\delta$ .

$$\begin{aligned} \delta &: \mathcal{X} \longrightarrow \mathcal{A} \\ x_i &\longrightarrow \delta(x_i) = a_i \end{aligned}$$

$a_i$  étant la  $i^{\text{ème}}$  action.

### Définissons les lois de probabilités qui interviennent dans la statistique Bayésienne

La statistique est utilisée pour inférer des phénomènes passés ou futurs en utilisant une approche probabiliste. Dans l'analyse Bayésienne, le paramètre inconnu  $\theta$  est considéré comme aléatoire et l'espace des paramètres  $\Theta$  est muni d'une probabilité  $\pi$ , tel que  $(\Theta, \mathcal{A}, \pi)$  forme un espace probabilisé.

- **Loi a priori**

Notée par  $\pi(\theta)$ , désigne la distribution du paramètre inconnu  $\theta$ , elle porte l'information sur ce dernier. L'appellation *a priori* exprime le fait qu'elle a été établie préalablement à l'observation des données  $x$ .

- **Loi des observations**

C'est la loi conditionnelle de  $x$  sachant  $\theta$  ou fonction de vraisemblance, sa densité est notée par  $f(x|\theta)$ . Elle est donnée par :

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

- **Loi du couple  $(\theta, x)$**

Cette loi est appelée aussi loi jointe, sa densité est notée  $f(\theta, x)$ .

$$f(\theta, x) = f(x|\theta)\pi(\theta)$$

- **Loi marginale de  $x$**

Elle est notée  $f(x)$  et elle est calculée de la manière suivante :

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

- **Loi a posteriori**

C'est la loi conditionnelle de  $\theta$  sachant  $x$ , sa densité est noté  $\pi(\theta|x)$ , cette loi est une actualisation de la loi a priori  $\pi(\theta)$  aux vu de l'observation  $x$ .

En vertu du théorème de Bayes, on peut exprimer cette loi comme suit :

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

**Remarque 1.1.** *Nous constatons que la loi a posteriori reste la pièce maîtresse de l'estimation Bayésienne.*

## 1.4 Inférence statistique Bayésienne

### 1.4.1 Approche décisionnelle, Estimation ponctuelle

**Définition 1.2.** *(Fonction coût)*

On appelle fonction coût (ou fonction perte) toute fonction  $L$  :

$$\begin{aligned} L : \Theta \times \mathcal{A} &\longrightarrow \mathbb{R}^+ \\ (\theta, \delta(x)) &\longrightarrow L(\theta, \delta(x)) \end{aligned}$$

$L(\theta, \delta(x))$  évalue le coût ou la perte associé à la décision  $a = \delta(x)$  quand le paramètre vaut  $\theta$ . Elle permet de calculer la perte pour une mauvaise décision.

**Définition 1.3.** *Soit  $L(\theta, \delta(x))$  une fonction coût et soit  $\pi$  une loi a priori pour  $\theta$ . On appelle estimateur de Bayes, la règle de décision  $\delta^\pi(x)$  qui vérifie :*

$$\delta^\pi(x) = \arg \min_{\delta(x) \in \mathcal{D}} \mathbb{E}^\pi[L(\theta, \delta(x))|x]$$

L'intérêt de l'approche décisionnelle est de trouver la meilleure évaluation possible pour la fonction de  $\theta$ , ce qui peut entraîner au mieux un coût nul lorsque ce paramètre est connu. Dans le cas contraire, la fonction coût perdrait sa continuité, ce qui pourrait empêcher le choix d'une procédure de décision.

Pour obtenir un critère de comparaison à partir d'une fonction coût, nous considérons la moyenne de celui-ci, c'est ce que on appelle le risque fréquentiste.

**Définition 1.4.** (*Risque fréquentiste*)

Pour une fonction coût  $L(\theta, \delta(x))$  donnée, le risque fréquentiste est défini comme la moyenne de cette fonction, représentant le coût moyen associé à une règle de décision. Le risque fréquentiste est noté  $R(\theta, \delta(x))$ .

$$\begin{aligned} R(\theta, \delta(x)) &= \mathbb{E}_\theta[L(\theta, \delta(x))] \\ &= \int_{\mathcal{X}} L(\theta, \delta(x)) f(x|\theta) dx \end{aligned}$$

Le risque fréquentiste permet d'établir un critère de comparaison entre les décisions. On dira que la stratégie  $\delta_1$  est préférable à  $\delta_2$  et on notera  $\delta_1 < \delta_2$  si

$$R(\theta, \delta_1) \leq R(\theta, \delta_2), \quad \forall \theta \in \Theta$$

Cette définition permet d'établir un pré-ordre sur l'ensemble  $\mathcal{D}$  des décisions. Cependant, ce pré-ordre est partiel car il ne permet pas de comparer deux règles de décision telles que :

$$R(\theta_1, \delta_1) < R(\theta_1, \delta_2)$$

et

$$R(\theta_2, \delta_1) > R(\theta_2, \delta_2)$$

On définit alors le coût a posteriori.

**Définition 1.5.** (*Coût a posteriori*)

Le coût a posteriori  $\rho(\pi, \delta|x)$  est défini comme la moyenne du coût par rapport à la loi a posteriori. Il est également appelé risque a posteriori.

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

On considère maintenant la moyenne du risque fréquentiste sous la loi a priori, qui est appelée le risque de Bayes (ou risque intégré).

**Définition 1.6.** (*Risque de Bayes*)

Le risque Bayésien noté  $r(\pi, \delta(x))$ , est défini pour une fonction coût donnée par :

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

Ainsi, le risque de Bayes  $r(\pi, \delta(x))$  peut également être exprimé comme la moyenne du risque a posteriori  $\rho(\theta, \delta(x))$  suivant la loi marginale  $f(x)$ .

**Définition 1.7.** (*Estimateur de Bayes*)

On appelle estimateur de Bayes associé à une fonction coût  $L$  et à une distribution a priori  $\pi$ , toute décision  $\delta^\pi$  qui minimise le risque intégré  $r(\pi, \delta(x))$ .

On a :

$$\delta^\pi(x) = \arg \min_{\delta(x) \in \mathcal{D}} r(\pi, \delta(x))$$

### Propriétés de l'estimateur de Bayes

- L'estimateur de Bayes est admissible.
- L'estimateur de Bayes est biaisé.
- L'estimateur de Bayes est convergent en probabilité lorsque la taille de l'échantillon  $n$  tend vers l'infini.
- La loi a posteriori peut être asymptotiquement (c'est-à-dire pour de grandes valeurs de  $n$ ) approximée par une loi normale  $\mathcal{N}(\mathbb{E}[\theta|x], \mathbb{V}[\theta|x])$ .

Le tableau (1.1) ci-dessous représente quelques estimateurs de Bayes du paramètre  $\theta$  sous la fonction coût quadratique pour les lois a priori conjuguées des familles exponentielles usuelles.

<i>Loi de <math>x</math></i>	<i>Loi conjuguée</i>	<i>Moyenne a posteriori</i>
Normale $\mathcal{N}(\theta, \sigma^2)$	Normale $\mathcal{N}(\mu, \tau^2)$	$\frac{(\sigma^2\mu + \tau^2x)}{(\sigma^2 + \tau^2)}$
Poisson $\mathcal{P}(\theta)$	Gamma $\Gamma(\alpha, \beta)$	$\frac{\alpha + x}{\beta + 1}$
Gamma $\Gamma(v, \theta)$	Gamma $\Gamma(\alpha, \beta)$	$\frac{\alpha + v}{\beta + x}$
Binomiale $\mathcal{B}(n, \theta)$	Bêta $\mathcal{B}e(\alpha, \beta)$	$\frac{\alpha + x}{\alpha + \beta + n}$
Binomiale Négative Neg $(n, \theta)$	Bêta $\mathcal{B}e(\alpha, \beta)$	$\frac{\alpha + n}{\alpha + \beta + x + n}$
Multinomiale $\mathcal{M}_k(n; \theta_1, \dots, \theta_k)$	Dirichlet $\mathcal{D}(\alpha_1, \dots, \alpha_k)$	$\frac{\alpha_i + x_i}{(\sum_j \alpha_j + n)}$
Normale $\mathcal{N}(\mu, \frac{1}{\theta})$	Gamma $\Gamma(\alpha/2, \beta/2)$	$\frac{\alpha + 1}{\beta + (\mu - x)^2}$

TABLE 1.1 – Quelques estimateurs de Bayes usuels.

L'estimateur  $\delta^\pi$  peut être obtenu de manière analytique ou numérique, en fonction de la complexité du coût  $L$  et de la loi a posteriori  $\pi(\theta|x)$ . Dans le cas des fonctions coût classiques, les estimateurs Bayésiens correspondants sont souvent des caractéristiques usuelles de la distribution a posteriori.

Nous présentons ci-dessous les trois fonctions coût usuelles, qui sont appréciées pour leur utilités et leur simplicités. Elles nous permettent d'obtenir facilement les estimateurs Bayésiens les plus couramment utilisés.

### La fonction coût quadratique

La fonction coût quadratique, introduite par Legendre (1805) and Gauss (1810), est sans conteste le critère d'évaluation le plus commun. Elle est particulièrement intéressante lorsque l'espace des paramètres est borné et permet un choix plus subjectif du coût.

**Définition 1.8.** *La fonction coût quadratique est définie par :*

$$L(\theta, \delta(x)) = (\theta - \delta(x))^2$$

*une variante de cette fonction coût est le coût quadratique pondérée définie par :*

$$L(\theta, \delta(x)) = g(\theta)(\theta - \delta(x))^2$$

**Remarque 1.2.** *Sous la fonction coût quadratique, l'estimateur de Bayes est la moyenne a posteriori, également connu sous le nom d'estimateur MMSE (Minimum Mean Square Error).*

### La fonction coût absolue

On peut également obtenir l'estimateur en utilisant la fonction coût absolue dont l'expression s'écrit comme suit :

**Définition 1.9.** *Soit la fonction coût absolue définie par :*

$$L(\theta, \delta(x)) = \begin{cases} \theta - \delta(x) & \text{si } \theta > \delta(x) \\ \delta(x) - \theta & \text{si } \theta \leq \delta(x) \end{cases} \Rightarrow L(\theta, \delta(x)) = |\theta - \delta(x)|.$$

**Remarque 1.3.** *Sous la fonction coût absolue, l'estimateur de Bayes est la médiane a posteriori.*

### La fonction coût 0-1

Ce coût est surtout utilisé dans l'approche classique des tests d'hypothèse, il a été proposé par Neyman and Pearson.

**Définition 1.10.** *La fonction coût 0 – 1 est définie par :*

$$L(\theta, \delta(x)) = \begin{cases} 0 & \text{si } |\theta - \delta(x)| < \varepsilon \\ 1 & \text{si } |\theta - \delta(x)| \geq \varepsilon \end{cases}, \forall \varepsilon > 0.$$

**Remarque 1.4.** *Sous la fonction coût 0 – 1, l'estimateur de Bayes est appelé le maximum a posteriori (MAP), il représente le mode de la distribution a posteriori.*

### 1.4.2 Choix de la loi a priori

Le choix des lois a priori est une étape fondamentale dans l'analyse Bayésienne et constitue une différence notable avec la statistique fréquentiste. Les stratégies de choix sont diverses et peuvent être motivées par différents points de vue. Ce choix peut être fondé sur des expériences passées, les connaissances subjectives d'un statisticien ou la facilité du calcul. Enfin, ces stratégies peuvent également tenir compte du fait qu'on ne sait rien par des lois non informatives.

Dans ce qui suit, nous allons introduire différents types de lois a priori.

#### Loi a priori conjuguée naturelle

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

L'approche a priori conjuguée, introduite par Raiffa and Schlaifer en (1961), peut être considérée comme un point de départ pour l'élaboration de distributions a priori fondées sur des informations a priori limitées.

**Définition 1.11.** *Une famille  $\mathcal{F}$  de distribution sur  $\Theta$  est dite conjuguée pour la loi  $f(x|\theta)$  si  $\forall \pi \in \mathcal{F}$ , la distribution a posteriori  $\pi(\theta|x)$  appartient également à  $\mathcal{F}$ . L'avantage des familles conjuguées est avant tout la facilité des calculs.*

**Remarque 1.5.** *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. Les lois a priori obtenues par ce procédé sont dites conjuguées naturelles.*

Le tableau (1.2) ci-dessus contient quelques lois a priori conjuguées usuelles.

$f(x \theta)$	$\pi(\theta)$	$\pi(\theta x)$
Normale $\mathcal{N}(\theta, \sigma^2)$	Normale $\mathcal{N}(\mu, \tau^2)$	$\mathcal{N}\left(\frac{\sigma^2\mu + \tau^2x}{\sigma^2 + \tau^2}, \frac{\sigma^2\tau^2}{\sigma^2 + \tau^2}\right)$
Poisson $\mathcal{P}(\theta)$	Gamma $\Gamma(\alpha, \beta)$	$\Gamma(\alpha + x, \beta + 1)$
Gamma $\Gamma(v, \theta)$	Gamma $\Gamma(\alpha, \beta)$	$\Gamma(\alpha + v, \beta + x)$
Binomiale $\mathcal{B}(n, \theta)$	Beta $\mathcal{B}e(\alpha, \beta)$	$\mathcal{B}e(\alpha + x, \beta + n - x)$
Binomiale Négative Neg $(m, \theta)$	Beta $\mathcal{B}e(\alpha, \beta)$	$\mathcal{B}e(\alpha + m, \beta + x)$
Multinomiale $\mathcal{M}_k(\theta_1, \dots, \theta_k)$	Dirichlet $\mathcal{D}(\alpha_1, \dots, \alpha_k)$	$\mathcal{D}(\alpha_1 + x_1, \dots, \alpha_k + x_k)$
Normale $\mathcal{N}(\mu, \frac{1}{\theta})$	Gamma $\Gamma(\alpha, \beta)$	$\Gamma(\alpha + 0.5, \beta + \frac{(\mu-x)^2}{2})$

TABLE 1.2 – Lois a priori conjuguées usuelles

En règle générale, les lois a priori conjuguées sont associées à un type spécifique de lois d'échantillonnage qui permettent de les obtenir systématiquement. Ces lois sont souvent regroupées en familles appelées familles exponentielles.

### Famille exponentielle

**Définition 1.12.** On appelle famille exponentielle à  $s$  paramètres toute famille de loi de distribution  $\{P_\theta\}$  dont la densité a la forme suivante :

$$f(x|\theta) = h(x) \cdot \exp\left(\sum_{i=1}^s \eta_i(\theta) T_i(x) - B(\theta)\right)$$

avec

- $\eta_i(\cdot)$ ,  $B(\cdot)$  sont des fonctions de paramètre  $\theta$ .
- $T_i(\cdot)$  sont des statistiques (fonction de  $x$ ).
- $h(\cdot)$  est une fonction de  $x$ .

**Définition 1.13.** *Le modèle exponentielle s'écrit sous la forme dite canonique en reparamétrisant  $\eta_i(\theta)$  par  $\theta_i$ , on obtient dès lors l'équation suivante :*

$$f(x|\theta) = h(x)\exp\left(\sum_{i=1}^s \theta_i T_i(x) - B(\theta)\right)$$

Le résultat suivant donne la forme des lois naturelles conjuguées dans le cas du modèle exponentiel.

**Proposition 1.1.** *Une famille exponentielle conjuguée pour  $f(x|\theta)$  est donnée par*

$$\pi(\theta|\mu, \lambda) = K(\mu, \lambda)\exp(\theta\mu - \lambda B(\theta))$$

où  $K(\mu, \lambda)$  est la constante de normalisation et la loi a posteriori correspondante est donnée comme suit :

$$\pi(\theta|x) \propto \exp\left(\theta(\mu + x) - (\lambda + 1)B(\theta)\right)$$

### Lois a priori non-informatives

Une loi a priori non-informative représente une ignorance sur le problème considéré, mais ne signifie pas que l'on ne sache absolument rien sur la distribution statistique du paramètre. Les lois a priori non-informatives sont des lois qui fournissent une information sur le paramètre à estimer dont le poids dans l'inférence est réduit. Ce sont des distributions vagues qui, a priori ne favorisent aucune valeur, ni aucune hypothèse particulière. En conséquence, elles permettent aux données de "parler pour elles-mêmes". Ces lois a priori ont des avantages : elles sont faciles à formuler, elles donnent l'apparence d'objectivité dans le choix de la loi a priori et elles évitent l'utilisation de lois a priori subjectives mal appliquées. Les différentes méthodes proposées pour obtenir ce type de lois a priori ont pour point commun de n'utiliser comme source d'information que la forme de la fonction de vraisemblance  $f(x|\theta)$  définie par le modèle.

Les lois non-informatives peuvent être considérées comme des lois de référence auxquelles chacun pourrait avoir recours quand toute information a priori est absente ou minime. Certaines de ces lois sont plus utiles ou plus efficaces que d'autres mais ne peuvent être perçues comme moins informatives que d'autres. Il est désormais largement admis qu'il n'existe pas d'a priori absolument non-informative (Kass and Wasserman, 1996)[21], comme le fait aussi remarquer Lindley (1990)[22] : "l'erreur est de les interpréter (les lois a priori non-informative) comme des représentations

d'une complète ignorance".

Dans ce qui suit, nous allons présenter deux techniques de construction des lois non informatives.

### Lois a priori uniformes (ou lois a priori de Laplace )

La loi a priori non informative la plus simple et la plus utilisée est la loi uniforme. En effet, ce choix repose sur l'équiprobabilité des valeurs possible du paramètre  $\theta$  sur  $\Theta$  : si  $\Theta$  est un ensemble de taille  $k$ , alors

$$\pi(\theta_i) = \frac{1}{k}, \quad \forall i$$

Laplace a été le premier à utiliser des techniques non-informatives puisque, bien que ne disposant pas d'information a priori pour les paramètres qu'il étudiait, il munit ces paramètres d'une loi qui prend en compte son ignorance en donnant la même vraisemblance à chaque valeur possible, soit donc en utilisant une loi uniforme. Son raisonnement, appelé plus tard principe de la raison insuffisante, se fondait sur l'équiprobabilité des événements élémentaires.

### Lois a priori de Jeffreys

Les lois a priori de Jeffreys, établies en 1946 et 1961, sont basées sur l'information de Fisher qui représente une mesure de la quantité d'informations sur  $\theta$  contenue dans l'observation  $x$ .

$$I(\theta) = \mathbb{E}_\theta \left[ \left( \frac{\partial \log f(x|\theta)}{\partial \theta} \right)^2 \right]$$

Dans le cas où  $\theta$  est un paramètre unidimensionnel et que le domaine de  $x$  ne dépend pas de  $\theta$ , alors cette information est égale à :

$$I(\theta) = -\mathbb{E}_\theta \left[ \left( \frac{\partial^2 \log f(x|\theta)}{\partial^2 \theta} \right) \right] \tag{1.1}$$

**Définition 1.14.** Soit  $\theta \in \mathbb{R}$ , on appelle loi a priori non-informative de Jeffreys, la loi a priori de la forme :

$$\pi(\theta) = C\sqrt{I(\theta)}$$

où

- $C$  une constante de normalisation.

- $I(\theta)$  l'information de Fisher apportée par  $x$  sur  $\theta$ .

Dans le cas où  $\theta$  est un paramètre multidimensionnel, on définit la matrice d'information de Fisher  $I(\theta)$  par la généralisation de l'équation (1.1).

Pour  $\theta$ , les éléments de  $I(\theta)$  sont données par :

$$I_{ij}(\theta) = -\mathbb{E}_{\theta} \left[ \left( \frac{\partial^2 \log f(x|\theta)}{\partial \theta_i \partial \theta_j} \right) \right], \quad (i, j = 1, \dots, k)$$

Et la loi non-informative de Jeffreys est alors définie par :

$$\pi(\theta) \propto [\det(I(\theta))]^{1/2}$$

**Remarque 1.6.**

- La loi a priori de Jeffreys est invariante par transformation bijective.
- La loi a priori de Jeffreys offre une méthode automatique pour obtenir une loi a priori non-informative pour n'importe quel modèle paramétrique, et permet de retrouver des estimateurs classiques.
- Ce type de construction de lois non-informatives mène très souvent à des lois a priori impropres.

**Lois a priori impropres**

Une loi impropre (ou généralisée) est une mesure  $\sigma$ -finie sur l'espace des paramètres  $\Theta$ , c'est-à-dire une mesure  $\pi$  telle que

$$\int_{\Theta} \pi(\theta) d\theta = +\infty$$

Ces lois sont obtenues lorsqu'on dispose des critères subjectifs ou théoriques sur la distribution a priori du paramètre, qui conduisent à une mesure  $\sigma$ -finie sur  $\Theta$  plutôt qu'à une mesure de probabilité. Les lois a priori impropres sont utiles dans les modèles non-informatifs cependant, elles ne peuvent être utilisées que si la condition suivante est vérifiée :

$$m_{\pi}(x) = \int_{\Theta} f(x|\theta)\pi(\theta)d\theta < \infty$$

En conclusion, l'usage de lois a priori impropres est justifié si la loi a posteriori est propre car elle ne dépend pas de la constante multiplicative de la loi a priori. Dans ce cas ces lois sont utiles du moins tant que la loi a posteriori existe car l'inférence Bayésienne se fonde sur la loi a posteriori  $\pi(\theta|x)$ .

Une difficulté pratique dans l'utilisation des lois impropres est de vérifier la condition d'intégrabilité

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

surtout dans des situations complexes, comme les modèles hiérarchiques. Cependant cette difficulté ne doit pas être considérée comme un inconvénient car les nouvelles techniques de calcul Bayésien comme les algorithmes MCMC ne nécessitent pas dans la pratique de vérifier cette condition.

### 1.4.3 Estimation par intervalles

L'inférence statistique est une démarche d'inversion puisqu'elle vise à remonter des effets aux causes, des observations aux paramètres, seule la théorie Bayésienne réalise cette inversion de façon cohérente. Dans la conception classique de l'intervalle de confiance les bornes observées sont des grandeurs aléatoires qui varient d'un échantillon à un autre.

#### Intervalle de crédibilité Bayésien

L'intervalle de crédibilité est l'équivalent Bayésien de l'intervalle de confiance dans la statistique classique. L'approche Bayésienne présente l'avantage de construire directement une région de confiance, tandis que l'approche classique nécessite une stratégie particulière telle que l'utilisation d'une fonction pivotale.

En revanche, l'interprétation de ces deux derniers est légèrement différente : les intervalles de crédibilité Bayésiens traitent leurs limites comme fixes et le paramètre estimé comme variable aléatoire, tandis que les intervalles de confiance fréquentistes traitent leurs limites comme variables aléatoires et le paramètre estimé comme fixe.

De plus, les intervalles de crédibilité Bayésiens utilisent la connaissance a priori, alors que les intervalles de confiance fréquentistes ne le font pas.

**Définition 1.15.** *On appelle intervalle de confiance de niveau  $1 - \alpha$  de paramètre  $\theta$  tout intervalle  $I$ , tel que*

$$P(\theta \in I) = 1 - \alpha$$

pour  $\alpha \in [0, 1]$  fixé.

Si  $\alpha$  augmente l'intervalle de confiance est plus petit.

**Définition 1.16.** (Fonction pivotale)

Une fonction pivotale de paramètre  $\theta$  est une fonction des observations  $(x_1, \dots, x_n)$  et dont la loi ne dépend pas de  $\theta$ .

**Définition 1.17.** Soit le modèle Bayésien  $(\mathcal{X}, \mathcal{A}, P_0, \pi(\theta))$ . On suppose que le paramètre à estimer est  $\theta$ . On dit que  $I$  l'intervalle de crédibilité de niveau  $\alpha$  si

$$P^{\pi(\cdot|x)}(\theta \in I) \geq 1 - \alpha$$

avec  $P^{\pi(\cdot|x)}$  la loi de probabilité dont la densité a posteriori est  $\pi(\theta|x)$ .

On pose généralement  $I = C_x$ , on l'appelle aussi région  $\alpha$ -crédible.

**Définition 1.18.** (Région  $\alpha$ -crédible)

On dira qu'une région est  $\alpha$ -crédible de plus forte densité a posteriori (PFDP, ou HPD pour Highest Posterior Density), si elle est définie de la manière suivante :

$$C_x^\pi = \{\theta, \pi(\theta|x) \geq k\}$$

où  $k$  est la plus grande valeur telle que :

$$P(\theta \in C_x^\pi) \geq 1 - \alpha$$

Par conséquent, si  $C_x^\pi$  est une région PFDP, alors elle est également  $\alpha$ -crédible.

**Définition 1.19.** (Intervalle de confiance a priori )

On appelle intervalle de confiance a priori  $J$  de niveau  $1 - \alpha$ , l'intervalle pour lequel on a :

$$P(\theta \in J) = 1 - \alpha$$

où  $P$  est la distribution de densité a priori  $\pi(\cdot)$ .

$$P(\theta \in J) = \int_J \pi(\theta) d\theta$$

**Définition 1.20.** (Intervalle de confiance a posteriori)

On appelle intervalle de confiance a posteriori  $I$  de niveau  $1 - \alpha$ , l'intervalle pour lequel on a :

$$P(\theta \in I|x) = \int_I \pi(\theta|x) d\theta = 1 - \alpha$$

## 1.5 Modèle Bayésien hiérarchique et méthodes Monte-Carlo par chaîne de Markov

Dans cette section, nous présentons les concepts de base de l'approche Bayésienne hiérarchique puis, des méthodes d'inférence de type Monte-Carlo par chaîne de Markov (MCMC).

### 1.5.1 L'analyse Bayésienne hiérarchique

Un des problèmes d'une approche Bayésienne classique noté par les fréquentistes est l'incertitude concernant la loi a priori. Une approche qui essaie de rectifier ce problème est l'analyse Bayésienne hiérarchique qui met des mesures a priori sur les paramètres de la loi a priori  $\pi(\theta)$ .

**Définition 1.21.** *On appelle modèle Bayésien hiérarchique de niveau  $n$ , un modèle statistique Bayésien avec densité conditionnelle de  $x$  sachant  $\theta$ ,  $f(x|\theta)$ , la densité a priori  $\pi(\theta)$  décomposée en  $n$  densités conditionnelles et une densité marginale, c'est-à-dire :*

$$\begin{aligned} x|\theta &\sim f(x|\theta), \\ \theta|\theta_1 &\sim \pi_1(\theta, \theta_1), \\ \theta_1|\theta_2 &\sim \pi_2(\theta_1, \theta_2), \\ &\vdots \\ \theta_{n-1}|\theta_n &\sim \pi_n(\theta_{n-1}|\theta_n) \\ \theta_n &\sim \pi_{n+1}(\theta_n) \end{aligned}$$

Les paramètres  $\theta_i$  sont appelés hyperparamètres de niveau  $i$  ( $1 \leq i \leq n$ ) et  $\pi_{n+1}(\theta_n)$  est une loi marginale. Par conséquent

$$\pi(\theta) = \int_{\Theta_1 \times \dots \times \Theta_n} \pi_1(\theta|\theta_1) \dots \pi_n(\theta_{n-1}|\theta_n) \pi_{n+1}(\theta_n) d\theta_1 \dots d\theta_n$$

Habituellement, nous nous limitons au cas où  $n = 1$  (Bayésien classique).

Dans ce cas nous avons, la loi a posteriori de  $\theta$  est donnée par :

$$\begin{aligned}\pi(\theta|x) &= \int \pi(\theta|\theta_1, x)\pi(\theta_1|x)d\theta_1 \\ \pi(\theta|\theta_1, x) &= \frac{f(x|\theta)\pi_1(\theta|\theta_1)}{f_1(x|\theta_1)} \\ f_1(x|\theta_1) &= \int f(x|\theta)\pi_1(\theta|\theta_1)d\theta \\ \pi(\theta_1|x) &= \frac{f_1(x|\theta_1)\pi_2(\theta_1)}{f(x)} \\ f(x) &= \int f_1(x|\theta_1)\pi_2(\theta_1)d\theta_1\end{aligned}$$

Un aspect positif de l'analyse Bayésienne hiérarchique est qu'elle augmente la robustesse de l'analyse Bayésienne classique, puisqu'elle réduit l'arbitraire sur le choix de l'hyperparamètre (parfois reporté à un niveau plus élevé) et établit une moyenne des réponses Bayésiennes conjuguées.

La décomposition d'une loi a priori  $\pi$  en plusieurs composantes  $\pi_1, \pi_2, \dots, \pi_n$  permet parfois d'obtenir des approximations plus aisées de certaines quantités a posteriori.

Un inconvénient des modèles hiérarchiques est qu'ils ne permettent généralement pas un calcul explicite des estimateurs de Bayes, même lorsque les niveaux successifs sont conjugués, et il faut donc avoir recours à des techniques numériques d'approximation.

Nous citons, par exemple, deux méthodes générales pour construire des estimateurs Bayésiens hiérarchiques :

- La première est la méthode de Monte-Carlo (Robert, 2001)[30].
- La deuxième méthode est la méthode de Monte-Carlo avec fonction d'importance (Robert, 2001)[30].

### 1.5.2 Méthode de Monte-Carlo par chaîne de Markov

Les méthodes de Monte-Carlo par chaîne de Markov (Markov Chain Monte Carlo en anglais et abrégé en MCMC) génèrent une suite de variables aléatoires  $(\theta_1, \dots, \theta_n, \dots)$  et, hormis la première à laquelle on attribue une valeur arbitraire, chacune d'entre elles dépend uniquement de celle qui la précède. Les calculs sont ensuite poursuivis en appliquant à cette séquence une loi des grands nombres pour

les chaînes markoviennes ergodiques de forme identique.

Dans ce qui suit, nous allons voir deux des algorithmes MCMC les plus connus : tout d'abord, l'algorithme de Metropolis-Hastings, puis l'algorithme de Gibbs.

### 1.5.2.1 Algorithme de Metropolis-Hastings

L'algorithme de Metropolis-Hastings développé initialement par Metropolis et al. (1953)[23] puis généralisé par Hastings (1970)[18], repose sur l'utilisation d'une loi dite instrumentale selon laquelle on peut facilement générer des variables aléatoires. Le principe est de générer des échantillons aléatoires à partir de cette loi instrumentale. Ces échantillons sont ensuite corrigés afin qu'ils se comportent asymptotiquement comme des observations aléatoires de la distribution cible ou stationnaire.

**Algorithme Metropolis-Hastings, sous sa formulation générale, s'écrit :**

Pour  $\theta^{(0)}$  est une valeur initiale, on définit par récurrence les valeurs de  $\theta^{(t)}$ . A l'étape  $t$ , à partir de  $\theta^{(t-1)}$ ,  $\theta^{(t)}$  est construit en tirant un  $\theta'$  à l'aide d'une distribution de probabilité instrumentale :  $\theta' \sim q(\cdot|\theta^{(t-1)})$ .  $\theta^{(t)}$  est alors donné par :

$$\theta^{(t)} = \begin{cases} \theta' & \text{avec une probabilité } \alpha(\theta', \theta^{(t-1)}) \\ \theta^{(t-1)} & \text{avec une probabilité } 1 - \alpha(\theta', \theta^{(t-1)}) \end{cases}$$

où

$$\alpha(\theta', \theta^{(t-1)}) = \min \left( \frac{\pi(\theta')q(\theta^{(t-1)}|\theta')}{\pi(\theta^{(t-1)})q(\theta'|\theta^{(t-1)})}, 1 \right)$$

La loi de densité  $\pi(\theta)$  est souvent appelée loi cible ou loi objet, tandis que la loi de densité  $q(\cdot|\theta)$  est dite loi de proposition. Une propriété stupéfiante de cet algorithme est d'autoriser un nombre infini de lois de proposition produisant toute une chaîne de Markov convergeant vers la loi d'intérêt.

**Remarque 1.7.** *Notons qu'il est possible suivant cette construction de rester au même endroit après une itération. On peut alors montrer en écrivant la condition de balance, que pour ce choix de  $\alpha$ , on obtient une chaîne de Markov de loi stationnaire  $\pi$ . Cette chaîne de Markov est ergodique si et seulement si  $(\theta^{(t)})_t$  est irréductible et apériodique.*

### 1.5.2.2 L'échantillonnage de Gibbs

L'échantillonnage de Gibbs développé par Geman and Geman (1984)[16] aussi connu sous le nom de l'algorithme de Gibbs. Cet algorithme est très populaire parmi les méthodes MCMC grâce à la simplicité de ses calculs. Considérons le vecteur de paramètres  $\theta = (\theta_1, \dots, \theta_K)$  de dimension  $K$  de densité a posteriori  $\pi(\theta|X)$ . Supposons qu'il est possible de simuler selon toutes les distributions conditionnelles complètes a posteriori  $\pi(\theta_k|X, \theta_{-k})$ ,  $k = 1, \dots, K$  et où  $\theta_{-k}$  désigne le vecteur  $\theta$  privé de la composante  $k$ . Pour simuler suivant la distribution a posteriori jointe  $\pi(\theta|X)$ , l'échantillonneur de Gibbs génère les paramètres  $\theta_k$ ,  $k = 1, \dots, K$  itérativement suivant leurs lois conditionnelles complètes.

#### Algorithme d'échantillonnage de Gibbs :

Soit le vecteur de valeurs initiales  $\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_K^{(0)})$ .  $\theta^{(t)}$  sachant  $\theta^{(t-1)}$  est obtenu par :

1. Générer  $\theta_1^{(t)}$  suivant la distribution conditionnelle complète  $\pi(\theta_1|X, \theta_2^{(t-1)}, \dots, \theta_K^{(t-1)})$ ,
2. Générer  $\theta_2^{(t)}$  suivant la distribution conditionnelle complète  $\pi(\theta_2|X, \theta_1^{(t)}, \theta_3^{(t-1)}, \dots, \theta_K^{(t-1)})$ ,
3. ...
4. Générer  $\theta_K^{(t)}$  suivant la distribution conditionnelle complète  $\pi(\theta_K|X, \theta_1^{(t)}, \dots, \theta_{K-1}^{(t)})$ ,
5.  $\theta^{(t)} = (\theta_1^{(t)}, \dots, \theta_K^{(t)})$ .

L'échantillonneur de Gibbs est un cas particulier de l'algorithme de Metropolis-Hastings dans lequel nous acceptons toujours ( $\rho = 1$ ), il présente donc l'avantage de bouger dans l'espace à chaque itération. Il a également l'avantage de ne pas avoir à choisir une distribution de proposition. L'inconvénient de cette méthode est qu'elle nécessite de savoir générer suivant chacune des distributions conditionnelles complètes. Pour pallier ce problème, lorsque nous avons une distribution conditionnelle complète qui ne découle pas d'une loi convenable, il est alors possible d'utiliser une étape de Metropolis-Hastings dans l'échantillonneur de Gibbs, nous parlons alors d'algorithme de Metropolis-Hastings dans Gibbs.

Sous les conditions de régularité et après un nombre suffisamment important d'itérations, c'est-à-dire  $t > t_0$ , il a été montré que ces deux méthodes permettent

d'obtenir un échantillon  $\theta^{(t)}$  qui converge en distribution vers la distribution cible  $\pi(\theta|X)$  (Robert and Casella, 2004) [31]. Ainsi, les ensembles  $\{\theta^{(t)}, t > t_0\}$  et  $\{\theta_i^{(t)}, t > t_0, i = 1, \dots, K\}$  dans le cas de l'échantillonneur de Gibbs peuvent être considérés comme des échantillons suivant la distribution jointe  $\pi(\theta|X)$  et suivant les distributions marginales  $\pi(\theta_i|X)$  respectivement. Les premières  $t_0$  itérations sont appelées la période de “burn-in”<sup>1</sup>.

## 1.6 Notions de base en séries chronologiques

Dans ce qui suit, nous allons présenter certaines notions et définitions importantes en séries chronologiques.

### 1.6.1 Processus stochastiques

Un processus stochastique est une famille de variables aléatoires indexées par le temps dont l'objectif principal est la représentation des phénomènes aléatoires qui évoluent dans le temps. Pour chaque instant du temps “ $t$ ”, la valeur de la quantité étudiée  $X_t$  est une variable aléatoire. L'ensemble des valeurs  $X_t$  quand  $t$  varie est appelé processus aléatoire  $\{X_t; t \in \mathbb{Z}\}$ .

Le concept des processus stochastiques joue un rôle très important dans la modélisation des séries chronologiques.

Soit  $(\Omega, \mathcal{F}, \mathcal{P})$  un espace probabilisé.

**Définition 1.22.** *Un processus stochastique ou processus aléatoire noté  $\{X_t; t \in \mathbb{T}\}$  est une famille de variables aléatoires, définies sur un même espace probabilisé  $(\Omega, \mathcal{F}, \mathcal{P})$ .*

- Si  $T$  est dénombrable, nous disons que le processus est discret.
- Si  $T$  est un intervalle, nous disons que le processus est continu.

### 1.6.2 Une série chronologique

**Définition 1.23.** *Une série chronologique est aussi appelée série temporelle ou chronique est une suite finie d'observations  $(X_t)_{t=1, \dots, T}$ . Mathématiquement, une série chronologique est une réalisation (trajectoire) d'un processus aléatoire.*

---

1. période de chauffe

Ainsi une série temporelle  $(X_t)_{t=1,\dots,T}$  se décompose classiquement en :

- Une tendance ou trend  $m_t$ .
- Un effet saisonnier  $s_t$ .
- Une erreur  $\epsilon_t$ .

Deux modèles sont généralement proposés :

- Modèle additif

$$X_t = m_t + s_t + \epsilon_t$$

avec

$$\mathbb{E}(\epsilon_t) = 0$$

- Modèle multiplicatif

$$X_t = m_t \cdot s_t \cdot \epsilon_t$$

avec

$$\mathbb{E}(\epsilon_t) = 1$$

### 1.6.3 Théorème de Wold

Le théorème de Wold (1938) est le théorème fondamental de l'analyse des séries temporelles stationnaires.

**Théorème 1.1.** (*Décomposition de Wold*)

Tout processus stationnaire d'ordre deux  $\{X_t; t \in \mathbb{Z}\}$  peut être représenté sous la forme :

$$X_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j} + \kappa_t$$

où les paramètres  $\psi_j$  satisfont  $\psi_0 = 1$ ,  $\psi_j \in \mathbb{R}$ ,  $\forall j \in \mathbb{N}^*$ ,  $\sum_{j=0}^{\infty} \psi_j^2 < \infty$ , et  $\epsilon_t$  est un bruit blanc i.i.d  $(0, \sigma_\epsilon^2)$ .

Le terme  $\kappa_t$  désigne la composante linéaire déterministe telle que  $\text{cov}(\kappa_t, \epsilon_{t-j}) = 0$ ,  $\forall j \in \mathbb{Z}$ .

**Remarque 1.8.**

- Si on omet la composante déterministe  $\kappa_t$ , tout processus stationnaire peut s'écrire comme une somme pondérée infinie de chocs passés.
- Généralement, on prend  $\kappa_t = m = \mathbb{E}(X_t)$ .

$$X_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j} + m$$

avec  $m$  terme constant (intercept).

- La condition  $\psi_0 = 1$  est une condition de normalisation.
- La condition de sommabilité  $\sum_{j=0}^{\infty} \psi_j^2 < \infty$  assure l'existence des moments d'ordre deux. Sous cette condition, nous disons alors que  $X_t$  converge en moyenne quadratique.

### 1.6.4 Opérateur retard et opérateur différence

#### Opérateur retard

**Définition 1.24.** L'opérateur retard, noté  $B$  (backward) ou  $L$  (lag), décale le processus d'une unité de temps vers le passé. Il est défini comme suit :

$$BX_t = X_{t-1}$$

Si on appelle cet opérateur  $k$  fois, on décalera le processus de  $k$  unités de temps.

$$\underbrace{BBB \dots B}_{k \text{ fois}} X_t \dots = B^k X_t = X_{t-k}$$

#### Opérateur différence

**Définition 1.25.** L'opérateur de différence, noté  $\Delta$  ou  $\nabla$ , est défini comme la différence entre un processus et sa version retardée d'une unité de temps. Cet opérateur se construit à partir de l'opérateur retard :

$$\Delta X_t = \nabla X_t = (1 - B)X_t = X_t - X_{t-1}$$

### 1.6.5 Stationnarité

Nous commençons par définir une série stationnaire au sens strict (ou stationnarité forte) pour ensuite étudier les propriétés de la stationnarité du second ordre (ou stationnarité faible).

#### 1.6.5.1 Série stationnaire au sens strict ou stationnarité forte

Soit  $\{X_t; t \in \mathbb{Z}\}$  une série temporelle.

**Définition 1.26.** La série  $\{X_t; t \in \mathbb{Z}\}$  est dite strictement ou fortement stationnaire si  $\forall$  le  $n$ -uplet du temps  $t_1 < t_2 < \dots < t_n$ , tel que  $t_i \in \mathbb{Z}$  et pour tout  $h \in \mathbb{Z}$  avec  $t_i + h \in \mathbb{Z}$ ,  $\forall i$ , ( $i = 1, \dots, n$ ), la suite  $(X_{t_1+h}, \dots, X_{t_n+h})$  a la même loi de probabilité que la suite  $(X_{t_1}, \dots, X_{t_n})$ .

Une autre façon équivalente de définir la stationnarité forte (ou stationnarité stricte) est la suivante :

**Définition 1.27.** *La série  $\{X_t; t \in \mathbb{Z}\}$  est dite stationnaire au sens strict si, pour toutes les valeurs  $j_1, \dots, j_n$  la distribution jointe de la suite  $(X_t, X_{t+j_1}, \dots, X_{t+j_n})$  dépend uniquement des intervalles de temps  $j_1, \dots, j_n$  et est indépendante de la période  $t$ .*

Nous notons que toutes les caractéristiques (moments) d'un processus strictement stationnaire sont invariantes dans le temps. Cependant, cette définition de la stationnarité est considérée comme trop forte et très exigeante, car elle nécessite la connaissance de la loi conjointe du processus, ce qui est rarement réalisable en pratique, sauf dans des cas très spécifiques.

Toutefois, plusieurs propriétés essentielles des processus aléatoires peuvent être obtenues uniquement à partir des moments du premier et du second ordre.

La stationnarité de ces deux moments peut donc être suffisante pour expliquer la stationnarité du processus.

Pour cette raison, nous avons besoin d'un concept de stationnarité moins fort et qui peut être rencontré dans la pratique.

### 1.6.5.2 Stationnarité d'ordre deux ou la stationnarité faible

Dans l'analyse des séries chronologiques, la plupart des résultats et des méthodes sont basés sur l'hypothèse de stationnarité faible.

**Définition 1.28.** *Un processus  $(X_t)_{t \in \mathbb{Z}}$  est dit stationnaire au second ordre, ou stationnaire au sens faible, ou stationnaire d'ordre deux si les trois conditions suivantes sont satisfaites :*

1.  $\forall t \in \mathbb{Z}, \mathbb{E}(X_t^2) < \infty$ .
2.  $\forall t \in \mathbb{Z}, \mathbb{E}(X_t) = m$ , indépendante de  $t$ .
3.  $\forall (t, h) \in \mathbb{Z}^2, cov(X_t, X_{t+h}) = \mathbb{E}[(X_{t+h} - m)(X_t - m)] = \gamma(h)$ , indépendante de  $t$ .

La première condition  $\mathbb{E}(X_t^2) < \infty$ , signifie l'existence (ou la convergence) des moments d'ordre deux.

La seconde condition  $\mathbb{E}(X_t) = m$  porte sur les moments d'ordre un et signifie que l'espérance du processus doit être indépendante du temps.

Enfin, la troisième condition porte sur les moments d'ordre deux, résumés par la fonction d'autocovariance qui dépend seulement de la différence des instants.

En résumé, un processus est stationnaire de second ordre si l'ensemble de ses moments sont indépendants du temps.

**Remarque 1.9.** *Dans les processus du second ordre, la stationnarité stricte implique la stationnarité faible (la réciproque n'est pas vraie sauf pour les processus Gaussiens).*

### 1.6.6 Ergodicité

L'ergodicité est une propriété d'un processus stochastique qui signifie que les propriétés statistiques du processus calculées à partir d'une seule trajectoire (ou réalisation) de ce processus sont représentatives de l'ensemble des trajectoires possibles du processus sur une période suffisamment longue.

**Définition 1.29.** *Une suite stationnaire  $\{X_t, t \in \mathbb{Z}\}$  est dite ergodique si elle vérifie la loi forte des grands nombres.*

**Définition 1.30.** [13]

*Une transformation  $T : (\Omega, \mathcal{A}) \rightarrow (\Omega, \mathcal{A})$  définie sur l'espace probabilisé  $(\Omega, \mathcal{A}, \mathbb{P})$  est dite bijective bimesurable et  $\mathbb{P}$ -invariante si elle est bijective, mesurable, d'inverse mesurable et si  $\mathbb{P}(T(A)) = \mathbb{P}(A)$  pour tout  $A \in \mathcal{A}$ .*

*On note  $\mathcal{I} = \{A \in \mathcal{A} / T(A) = A\}$  la sous-tribu de  $\mathcal{A}$  formée des éléments invariants par  $T$ . Une transformation est ergodique si, de plus,  $A \in \mathcal{I}$  implique  $\mathbb{P}(A) = 0$  ou 1.*

*Remarquons que  $\mathcal{I} = \{A \in \mathcal{A} / T(A) = A\}$  est une sous-tribu de  $\mathcal{A}$ .*

#### Lien avec les processus stationnaires

Soit  $X = (X_n)_{n \in \mathbb{Z}}$  un processus réel stationnaire défini sur l'espace probabilisé  $(\Omega, \mathcal{A}, \mathbb{P})$ . Alors la loi image  $\mathbb{P}_X$  est une probabilité sur  $(\mathbb{R}^{\mathbb{Z}}, \mathcal{B}(\mathbb{R}^{\mathbb{Z}}))$ . La tribu  $\mathcal{B}(\mathbb{R}^{\mathbb{Z}})$  est celle engendrée par les pavés élémentaires  $A = \prod_{k \in \mathbb{Z}} A_k$  où  $A_k = \mathbb{R}$  sauf pour un ensemble fini d'indices  $k$ .

La transformation  $T$  définie par  $T(x)_i = (x_{i+1})$  pour  $x = (x_i)_{i \in \mathbb{Z}}$  vérifie

$T(A) = \prod_{k \in \mathbb{Z}} A_{k+1}$ , elle est bijective, bimesurable et  $\mathbb{P}$ -invariante; on l'appelle opérateur de décalage.

On note  $\mathcal{J} = X^{-1}(\mathcal{I})$ , la tribu image de  $\mathcal{I}$  par  $X$ .

Lorsque  $T$  est ergodique, c'est-à-dire  $\mathbb{P}(A) = 0$  ou  $1$  quand  $A \in \mathcal{J}$ , on dit que le processus  $X = (X_n)_{n \in \mathbb{Z}}$  est ergodique.

### Importance de l'ergodicité

- L'ergodicité est cruciale dans l'analyse des processus stochastiques, car elle permet de caractériser les propriétés statistiques d'un processus à partir d'une seule réalisation suffisamment longue.
- Elle est essentielle pour obtenir la convergence presque sûre d'une moyenne empirique spatiale, ce qui est fondamental dans de nombreux domaines.

### Différence ou lien avec la stationnarité

Bien que les concepts d'ergodicité et de stationnarité soient liés à l'analyse des séries temporelles, ils sont distincts :

- Stationnarité : Un processus stochastique est dit stationnaire lorsque ses propriétés statistiques restent constantes au fil du temps. Il peut s'agir de la stationnarité en sens strict (toutes les propriétés statistiques restent constantes) ou de la stationnarité au sens large (seule la moyenne et la variance restent constantes, la corrélation peut varier). La stationnarité est une hypothèse sur les propriétés statistiques du processus.
- Ergodicité : Un processus stochastique est ergodique lorsque les propriétés statistiques calculées à partir d'une seule réalisation sont représentatives de l'ensemble du processus sur une période suffisamment longue. L'ergodicité est une propriété sur le comportement du processus au fil du temps et sur la représentativité statistique des trajectoires individuelles.

**Remarque 1.10.** *Le théorème de l'ergodicité (1.2) est un résultat fondamental en statistiques et en théorie des probabilités.*

**Théorème 1.2.** *(Théorème de l'ergodicité) [32]*

*Soit  $\xi = (\xi_1, \xi_2, \dots)$  une suite aléatoire strictement stationnaire et ergodique avec*

$\mathbb{E}|\xi_1| < \infty$ . Alors :

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \xi_k = \mathbb{E}(\xi_1) \quad \text{presque sûrement}$$

### 1.6.7 Processus bruit blanc (White noise)

L'exemple le plus connu de processus stationnaire est le processus bruit blanc (ou White noise) noté  $\mathcal{BB}$ , il existe deux types de bruit blanc.

**Définition 1.31.** *Un processus  $\{X_t, t \in \mathbb{Z}\}$  est un bruit blanc faible, si et seulement si :*

- $\mathbb{E}(X_t) = 0$  cela signifie que l'espérance du processus est indépendante du temps, et de plus qu'elle est nulle.
- $\gamma(h) = \mathbb{E}(X_t X_{t-h}) = \begin{cases} \sigma^2 & \text{si } h = 0. \\ 0 & \text{si } h \neq 0. \end{cases}$

*Cela implique bien entendu l'indépendance de la fonction d'autocovariance par rapport au temps. En d'autres termes, cela signifie que les bruits blancs sont des processus stationnaires particuliers sans "mémoire".*

$\{X_t, t \in \mathbb{Z}\}$  est un bruit blanc fort si et seulement si les  $X_t$  sont i.i.d,

$$\mathbb{E}(X_t) = 0$$

et

$$\mathbb{V}(X_t) = \sigma^2$$

Nous noterons  $X_t \sim BB(0, \sigma^2)$ .

### 1.6.8 Marche aléatoire (Random Walk)

**Définition 1.32.** *Un processus  $(X_t)$  indicé par  $\mathbb{N}$  suit une « marche aléatoire » si, pour tout  $t \in \mathbb{N}$ , il vérifie*

$$X_t = X_{t-1} + \epsilon_t$$

où  $X_0$  est une valeur arbitraire .

$$\epsilon_t \sim BB(0, \sigma^2)$$

Si  $X_0$  est choisie dans  $L^2(\Omega, \mathcal{A}, \mathbb{P})$  et indépendante de  $\epsilon_t$ , alors on peut exprimer explicitement le processus comme suit :

$$X_t = X_0 + \sum_{k=1}^t \epsilon_k$$

Cette expression elle ne garantit pas une stationnarité en espérance, car  $\mathbb{E}(X_t) = \mathbb{E}(X_0)$ . Cependant, un calcul direct montre que  $\mathbb{V}(X_t) = \mathbb{V}(X_0) + \sigma^2 t$ , ce qui entraîne une non-stationnarité en variance.

### 1.6.9 Fonction d'autocovariance

**Définition 1.33.** Soit  $\{X_t; t \in \mathbb{Z}\}$  une série stationnaire.

La fonction  $k \rightarrow \gamma_k = \mathbb{E}(X_t X_{t-k})$  est appelée fonction d'autocovariance de la série. La fonction d'autocovariance d'un processus  $\{X_t; t \in \mathbb{Z}\}$  centré et stationnaire de second ordre est donc définie par :

$$\gamma(t, s) = \mathbb{E}(X_t X_s) - \mathbb{E}(X_t)\mathbb{E}(X_s) = \gamma(t - s)$$

**Remarque 1.11.**

- La fonction d'autocovariance est paire  $\gamma_k = \gamma_{-k}$ .
- La matrice des autocovariances d'ordre  $k$ , est définie par :

$$\Gamma_k = (\gamma(i - j))_{i,j=1,\dots,k} = \begin{pmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(k-1) \\ \gamma(1) & \gamma(0) & \cdots & \gamma(k-2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(k-1) & \gamma(k-2) & \cdots & \gamma(0) \end{pmatrix}$$

- La fonction d'autocovariance fournit des informations sur les liaisons linéaires temporelles qui existent entre les différentes composantes de la série  $X_t$ .

### 1.6.10 Fonction d'autocorrélation (ACF)

**Définition 1.34.** Soit  $\{X_t; t \in \mathbb{Z}\}$  une série temporelle stationnaire de moyenne  $\mathbb{E}(X_t) = m$ .

La fonction d'autocorrélation de  $\{X_t\}$  (ACF), notée  $\rho_k$ , est définie comme suit :

$$k \rightarrow \rho(k) = \frac{\text{cov}(X_t, X_{t-k})}{\sqrt{\mathbb{V}(X_t) \cdot \mathbb{V}(X_{t-k})}} = \frac{\gamma_k}{\gamma_0}$$

avec  $\gamma_0 = \mathbb{V}(X_t) > 0$

**Remarque 1.12.**

- $\rho(k)$  représente l'expression du lien linéaire entre le présent et le passé d'ordre  $k$  de la série temporelle.
- La fonction  $\rho(k)$  est paire,  $\rho(k) = \rho(-k)$ .
- La représentation graphique de la fonction d'autocorrélation en fonction de  $k$  est appelée *corrélogramme*.
- L'ACF d'une série stationnaire est une caractéristique très utile pour en comprendre la structure. Sous des conditions générales, Brockwell et Davis (2009)[10] montrent que, le coefficient d'autocorrélation empirique  $\hat{\rho}(k)$  est un estimateur convergent de  $\rho(k)$ .
- D'après le Théorème Central Limite, la variable centrée  $t_{\rho_k}$  ci-dessous suit une loi normale centrée réduite :

$$t_{\rho_k} = \frac{\hat{\rho}_k - \rho_k}{\mathbb{V}(\hat{\rho}_k)} \sim N(0, 1), \forall k \in \mathbb{Z}$$

où  $\mathbb{V}(\hat{\rho}_k)$  désigne l'estimateur de la variance empirique des estimateurs  $\hat{\rho}_k$  :

$$\mathbb{V}(\hat{\rho}_k) = \frac{1}{N} \sum_{j=-N}^N \hat{\rho}_k^2, \quad N < k$$

En utilisant la symétrie des  $\rho_k$ , nous montrons que :

$$\mathbb{V}(\hat{\rho}_k) = \frac{1}{N} \left( 1 + 2 \sum_{j=1}^n \hat{\rho}_k^2 \right)$$

### 1.6.11 Fonction d'autocorrélation partielle (PACF)

La PACF, ou fonction d'autocorrélation partielle, c'est l'ACF obtenue en supprimant les corrélations dues aux termes courts.

#### Rôle de la PACF

On peut montrer que la PACF d'un modèle AR(p) est nulle à partir de l'ordre  $p + 1$ , car la PACF d'un modèle AR s'exprime comme suit :

$$\phi_{kk} = \text{Corr}(X_t, X_{t+k} | X_{t+1}, X_{t+2}, \dots, X_{t+k-1})$$

et pour un modèle AR(p) :

$$\phi_{kk} = 0 \quad \text{si} \quad k \geq p + 1$$

Finalement, l'ACF décroît lentement vers zéro lorsque  $k$  tend vers l'infini et la PACF s'annule à partir de  $k = p + 1$ .

Dans le cas d'un modèle MA( $q$ ), c'est l'inverse, car c'est l'ACF qui s'annule à partir de  $k = q + 1$ .

### 1.6.12 Quelques modèles de séries temporelles

Il existe plusieurs modèles de séries temporelles utilisés pour modéliser et analyser des données temporelles. Voici quelques-uns des modèles les plus courants :

#### Processus autorégressif d'ordre $p$ , AR( $p$ )

Un processus autorégressif est un modèle de régression pour les séries temporelles dans lequel la série est expliquée par ses propres valeurs passées plutôt que par d'autres variables.

**Définition 1.35.** (Brockwell, P. J. Davis, R. A 2002)[9]

On dit qu'un processus stationnaire  $\{X_t; t \in \mathbb{Z}\}$  admet une représentation autorégressive d'ordre  $p$ , notée AR( $p$ ) s'il vérifie l'équation suivante :

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \phi_3 X_{t-3} + \dots + \phi_p X_{t-p} + \epsilon_t$$

où

$$\{\epsilon_t; t \in \mathbb{Z}\} \sim \mathcal{BB}(0, \sigma^2)$$

et  $\phi_1, \dots, \phi_p$  sont des coefficients réels tels que  $\phi_p \neq 0$ .

$$X_t = \sum_{k=1}^p \phi_k X_{t-k} + \epsilon_t \Leftrightarrow \Phi(B)X_t = \epsilon_t$$

avec  $\Phi$  étant le polynôme de degré  $p$  dont les coefficients sont  $(1, -\phi_1, \dots, -\phi_p)$ .

#### Processus autorégressif d'ordre un, AR(1)

**Définition 1.36.** Un processus autorégressif d'ordre un AR(1) est défini par l'équation suivante :

$$X_t = \alpha + \phi X_{t-1} + \epsilon_t, t = 0, \pm 1, \pm 2, \dots$$

avec  $\phi \in \mathbb{R}$  et  $\epsilon_t \sim \mathcal{BB}(0, \sigma_\epsilon^2)$

**Propriétés 1.1.** Quelques propriétés des processus AR(1) :

- i) Si  $\phi = \pm 1$ , le processus  $(X_t)$  n'est pas stationnaire.
- ii) Si  $|\phi| < 1$ , le processus  $(X_t)$  est stationnaire.
- iii) Si  $|\phi| > 1$ , cas explosif.

## Modèle moyenne mobile d'ordre $q$ , MA( $q$ )

**Définition 1.37.** (Brockwell, P. J. Davis, R. A 2002)[9]

Un processus stationnaire  $\{X_t; t \in \mathbb{Z}\}$  admet une représentation moyenne mobile d'ordre  $q$  noté MA( $q$ ) s'il vérifie l'équation suivante :

$$X_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q}$$

où  $\{\epsilon_t; t \in \mathbb{Z}\} \sim \mathcal{BB}(0, \sigma^2)$  et  $\theta_1, \dots, \theta_q$  sont des coefficients réels tels que  $\theta_q \neq 0$ .

$$X_t = \sum_{k=1}^q \theta_k \epsilon_{t-k} \Leftrightarrow X_t = \theta(B) \epsilon_t$$

avec  $\theta_0 = 1$  et  $\theta(B)$  étant le polynôme de degré  $q$  dont les coefficients sont  $\{1, \theta_1, \dots, \theta_q\}$ .

## Modèle autorégressif moyenne mobile d'ordre $p$ et $q$ , ARMA( $p, q$ )

Un modèle ARMA( $p, q$ ), également connu sous le nom de modèle autorégressif moyenne mobile d'ordre  $p$  et  $q$ , est un type de modèle de série chronologique qui combine à la fois des termes autorégressifs (AR) et des termes moyennes mobiles (MA).

**Définition 1.38.** (Brockwell, P. J. Davis, R. A 2002)[9]

Un modèle ARMA( $p, q$ ) est défini par l'équation suivante :

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \phi_3 X_{t-3} - \dots - \phi_p X_{t-p} = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q}$$

$$\sum_{k=0}^p \phi_k X_{t-k} = \sum_{k=0}^q \phi_k \epsilon_{t-k}$$

avec  $\phi_0 = 1$  et  $\theta_0 = 1$ , nous pourrions retrouver une écriture polynomiale  $\Theta(B)$  et  $\Phi(B)$  :

$$\Phi(B)X_t = \Theta(B)\epsilon_t$$

avec :

$\Phi(B)$  est le polynôme de degré  $p$  dont les coefficients sont  $\{1, -\phi_1, \dots, -\phi_p\}$ .

$\Theta(B)$  est le polynôme de degré  $q$  dont les coefficients sont  $\{1, \theta_1, \dots, \theta_q\}$ .

# Modèle autorégressif à coefficients aléatoires

## 2.1 Introduction

Au cours des 40 dernières années, un intérêt croissant a été porté aux modèles de séries chronologiques. L'un des modèles majeurs qui a suscité l'attention est l'étude des modèles autorégressifs à coefficients aléatoires, souvent abrégés en RCAR. Il s'agit d'une extension du modèle AR. Ce modèle a été introduit et étudié par plusieurs auteurs, tels que Andel(1976)[1], Quinn and Nicholls(1981)[28]. Dans ce chapitre, nous allons aborder le modèle autorégressif à coefficients aléatoires d'ordre  $p$ , noté RCAR( $p$ ), de manière générale. Ensuite, nous nous concentrerons sur le cas particulier d'ordre un, appelé RCAR(1), et étudierons ses conditions de stationnarité. Nous présenterons ensuite plusieurs méthodes d'estimation dans le cadre de l'approche fréquentiste, telles que la méthode des moindres carrés, la méthode du maximum de vraisemblance dans l'estimation des paramètres.

## 2.2 Étude des modèles autorégressifs à coefficients aléatoires

### 2.2.1 Modèle autorégressif à coefficients aléatoires d'ordre $p$

**Définition 2.1.** *Un modèle autorégressif à coefficients aléatoires d'ordre  $p$ , noté  $RCAR(p)$  est représenté sous la forme suivante :*

$$X_t = \sum_{i=1}^p (\alpha_i + \beta_t) X_{t-i} + \epsilon_t, \quad t \in \mathbb{Z}$$

Où

- $\{\epsilon_t, t \in \mathbb{Z}\}$  est une suite de variables aléatoires indépendantes identiquement distribuées de moyenne nulle et une variance  $\sigma^2 > 0$ .
- $\{\beta_t, t \in \mathbb{Z}\}$  est une suite de variables aléatoires indépendantes identiquement distribuées.
- $\alpha_i$  sont des constantes,  $\forall i = 1 \dots n$ .

Nous nous intéressons au cas particulier du modèle  $RCAR(1)$ , modèle autorégressif à coefficient aléatoire d'ordre un donné par l'expression suivante :

### 2.2.2 Modèle autorégressif à coefficient aléatoire d'ordre un

Considérons deux suites de variables aléatoires  $\{\epsilon_t, t \in \mathbb{Z}\}$  et  $\{\beta_t, t \in \mathbb{Z}\}$  vérifiant les hypothèses suivantes :

- A1 :  $\{\epsilon_t\}$  est une suite de variables aléatoires indépendantes avec une moyenne nulle et une variance  $\sigma^2 > 0$ .
- A2 :  $\{\beta_t\}$  est une suite de variables aléatoires indépendantes avec une moyenne  $\mu_\beta$  et une variance  $\sigma_\beta^2 > 0$ .
- A3 :  $\{\beta_t\}$  et  $\{\epsilon_t\}$  sont indépendantes.

**Définition 2.2.** *Un processus  $\{X_t, t \in \mathbb{Z}\}$  est dit autorégressif à coefficient aléatoire d'ordre 1, noté  $RCAR(1)$ , s'il vérifie l'équation suivante :*

$$X_t = \alpha + \beta_t X_{t-1} + \epsilon_t, \quad t \in \mathbb{Z} \tag{2.1}$$

Si  $\alpha = 0$

$$X_t = \beta_t X_{t-1} + \epsilon_t, \quad t \in \mathbb{Z} \quad (2.2)$$

## 2.3 Conditions de stationnarité

### 2.3.1 L'existence d'une solution faiblement stationnaire

**Théorème 2.1.** [20] (*S.Y.Hwang and I.V.Basawa 1993*)

Si

$$\mathbb{E}(\beta_t^2) < 1$$

c'est à dire

$$\mu_\beta^2 + \sigma_\beta^2 < 1$$

alors le processus (2.2) admet une unique solution stationnaire ergodique dans  $\mathbb{L}^2$  donnée par :

$$X_t = \epsilon_t + \sum_{j=1}^{\infty} \left( \prod_{i=0}^{j-1} \beta_{t-i} \right) \epsilon_{t-j} \quad (2.3)$$

Hwang and Basawa (1993)[20], démontrent ce théorème en se référant à Nicholls and Quinn (1982) et Breiman (1968). Ils notent d'abord que (2.3) satisfait l'équation (2.2) et que  $X_t$  dans (2.2) est stationnaire et ergodique.

**Démonstration.** En réécrivant le modèle (2.3) sous la forme suivante :

$$X_t = \beta_t X_{t-1} + \epsilon_t$$

$$X_t = \beta_t(\beta_{t-1} X_{t-2} + \epsilon_{t-1}) + \epsilon_t$$

$$X_t = \beta_t \beta_{t-1} X_{t-2} + \beta_t \epsilon_{t-1} + \epsilon_t$$

$$X_t = \beta_t \beta_{t-1}(\beta_{t-2} X_{t-3} + \epsilon_{t-2}) + \beta_t \epsilon_{t-1} + \epsilon_t$$

$$X_t = \beta_t \beta_{t-1} \beta_{t-2} X_{t-3} + \beta_t \beta_{t-1} \epsilon_{t-2} + \beta_t \epsilon_{t-1} + \epsilon_t$$

.

.

.

$$X_t = \prod_{j=0}^{p-1} \beta_{t-j} X_{t-p} + \sum_{j=1}^{p-1} \left( \prod_{i=0}^{j-1} \beta_{t-i} \epsilon_{t-j} \right) + \epsilon_t$$

et donc d'après l'inégalité de B. Tchebychev, lorsque  $p \rightarrow \infty$  nous obtenons :

$$\begin{aligned} \mathbb{P} \left( \left| \prod_{j=0}^{p-1} \beta_{t-j} X_{t-p} \right| > \epsilon \right) &\leq \frac{\mathbb{E} \left( \prod_{j=0}^{p-1} \beta_{t-j} X_{t-p} \right)^2}{\epsilon^2} \\ \mathbb{P} \left( \left| \prod_{j=0}^{p-1} \beta_{t-j} X_{t-p} \right| > \epsilon \right) &\leq \frac{\left( \prod_{j=0}^{p-1} \mathbb{E}(\beta_{t-j} X_{t-p}) \right)^2}{\epsilon^2} \\ \mathbb{P} \left( \left| \prod_{j=0}^{p-1} \beta_{t-j} X_{t-p} \right| > \epsilon \right) &\leq \frac{\prod_{j=0}^{p-1} \mathbb{E}(\beta_{t-j}^2) \mathbb{E}(X_{t-p}^2)}{\epsilon^2} \\ \mathbb{P} \left( \left| \prod_{j=0}^{p-1} \beta_{t-j} X_{t-p} \right| > \epsilon \right) &\leq \frac{(\sigma_\beta^2 + \mu_\beta^2)^p + \mathbb{E}(X_{t-p}^2)}{\epsilon^2} \rightarrow 0 \end{aligned}$$

Étant donné que

$$\mathbb{E}(X_{t-p}^2) < \infty$$

et

$$\sigma_\beta^2 + \mu_\beta^2 < 1$$

Il en résulte que  $\{X_t, t \in \mathbb{Z}\}$  est stationnaire et ergodique.

### 2.3.2 L'existence d'une solution strictement stationnaire

Nous commençons par énoncer le résultat de Billingsley (1995) [7], qui sera utilisé dans la démonstration du théorème (2.8).

**Lemme 2.1.** [6]

*Soit  $\{X_t, t \in \mathbb{Z}\}$  un processus stochastique. Considérons un processus stochastique  $\{Y_t, t \in \mathbb{Z}\}$  avec  $Y_t = \Phi(\dots, X_{t-1}, X_t, X_{t+1}, \dots)$ , où  $\Phi$  est une fonction mesurable. Si  $\{X_t, t \in \mathbb{Z}\}$  est strictement stationnaire et ergodique, en particulier, si les  $X_t$  sont indépendants et identiquement distribués, alors  $\{Y_t, t \in \mathbb{Z}\}$  est strictement stationnaire et ergodique.*

Quinn (1982)[29] a étudié le problème de l'existence d'une solution strictement stationnaire et ergodique pour les modèles bilinéaires, définis par l'équation

$$X_n = aX_{n-1} + be_nX_{n-1} + e_n$$

où  $(e_n)$  est une suite de variables aléatoires strictement stationnaire et ergodique.

Il a montré que la condition  $\mathbb{E}(\ln |a + be_n|) \leq 0$  est nécessaire et suffisante pour l'existence de cette solution. Les mêmes arguments appliqués à un modèle RCAR(1) donne

**Théorème 2.2.** [29] *Considérons l'équation suivante :*

$$Y_n = \beta_n Y_{n-1} + e_n \quad (2.4)$$

où  $\{(\beta_n, e_n), n \in \mathbb{Z}\}$ , est une suite de vecteurs aléatoires strictement stationnaire et ergodique avec :

$$\mathbb{E}|\ln |e_n|| < \infty \quad \text{et} \quad \mathbb{E}|\ln |\beta_n|| < \infty$$

Si  $\mathbb{E}(\ln |\beta_n|) < 0$ , alors il existe une solution  $\{Y_n\}$  strictement stationnaire et ergodique pour l'équation (2.4), et telle que pour tout  $t \in \mathbb{Z}$ ,  $Y_n$  est mesurable par rapport à  $\mathcal{F}_n = \sigma(\{(\beta_t, e_t), t \leq n\})$  la tribu engendrée par  $\{(\beta_t, e_t), t \leq n\}$ .

**Démonstration.** Nous supposons que

$$\mathbb{E}(\ln |\beta_n|) = a < 0$$

Nous définissons la suite  $\{T_j\}$  avec  $j \in \mathbb{N}$  par :

$$T_0 = e_n \quad \text{et} \quad T_j = \prod_{i=0}^{j-1} \beta_{n-i} e_{n-j} \quad \text{pour} \quad j \geq 1$$

Soit  $S_r = \sum_{j=1}^r T_j$ , nous avons :

$$\ln(|T_j|) = \ln \prod_{i=0}^{j-1} |\beta_{n-i}| |e_{n-j}|$$

D'où

$$\ln(|T_j|) = \ln |e_{n-j}| + \sum_{i=0}^{j-1} \ln |\beta_{n-i}|$$

En multipliant les deux membres de l'équation par  $\frac{1}{j}$ , nous obtenons :

$$\frac{1}{j} \ln(|T_j|) = \frac{1}{j} \ln |e_{n-j}| + \frac{1}{j} \sum_{i=0}^{j-1} \ln |\beta_{n-i}|$$

Par le théorème de l'ergodicité (1.2),  $j \rightarrow \infty$  nous allons obtenir :

$$\frac{1}{j} \ln(|T_j|) \xrightarrow{ps} \mathbb{E}(\ln(|\beta_n|))$$

et

$$|T_j|^{\frac{1}{j}} \xrightarrow{ps} \exp(\mathbb{E}(\ln(|\beta_n|))) = \exp(a) < 1$$

Posons maintenant toutes les réalisations pour lesquelles la convergence ci-dessus est vérifiée on les notes  $(T_j^*)$ .

Puis nous donnons  $\delta > 0$ , tel que  $\delta + a < \rho < 1$  pour certain  $\rho$ , nous pouvons trouver  $R_\delta$  tel que :

$$|T_j^*| < \rho^j < 1 \quad \text{et} \quad \forall j \in R_\delta$$

D'où,  $\sum_{j=0}^r |T_j^*|$  converge et par conséquent  $S_r = \sum_{j=0}^r T_j$  converge presque sûrement quand  $r \rightarrow \infty$ .

Maintenant nous posons :

$$\begin{aligned} Y_t^* &= \lim_{r \rightarrow \infty} S_r \\ &= \lim_{r \rightarrow \infty} (e_n + \sum_{j=1}^r T_j) \\ &= e_n + \sum_{j=1}^{\infty} \prod_{i=0}^{j-1} \beta_{n-i} e_{n-j} \end{aligned}$$

Il est évident que  $Y_t^*$  est une solution  $\mathcal{F}_n$  mesurable pour l'équation (2.4). De plus par hypothèse  $\{(e_n, \beta_n)\}$  est une suite de vecteurs strictement stationnaire et ergodique et par le lemme (2.1) on conclut que  $\{Y_t^*\}$  est strictement stationnaire et ergodique.

**Théorème 2.3.** (Dazhe Wang 2003)[32]

Soit  $\{X_t; t \in \mathbb{Z}\}$  un modèle autorégressif à coefficients aléatoires d'ordre 1, et supposons que les hypothèses A1 et A3 (énoncées dans la présentation du modèle RCAR(1)(2.2.2)) soient satisfaites. Si les  $\{\beta_t, \epsilon_t\}$  ont des distributions normales, alors une condition suffisante pour l'existence d'une solution strictement stationnaire et ergodique est :

$$\ln(\sigma_\beta^2) < \zeta + \ln(2) - 2 \int_0^1 \frac{1 - \exp[-\lambda(1 - \omega^2)]}{1 - \omega^2} d\omega \quad (2.5)$$

où  $\zeta \approx 0,57721$ , désigne la constante d'Euler et  $\lambda = \frac{\mu_\beta^2}{2\sigma_\beta^2}$ .

**Démonstration.** Soit  $e_t = \alpha + \epsilon_t$ . Il est clair que  $\{(\beta_t, e_t), t \in \mathbb{Z}\}$  est une suite strictement stationnaire et ergodique.

Il est facile de voir que

$$\mathbb{E}|\ln |e_t|| < \infty \quad \text{et} \quad \mathbb{E}|\ln |\beta_t|| < \infty$$

Afin de faire appel au théorème (2.2), nous devons montrer que

$$\mathbb{E}(\ln |\beta_t|) = \mathbb{E}(\ln |\mu_\beta + \sigma_\beta v_t|) < 0$$

où

$$v_t \sim \mathcal{N}(0, 1)$$

Nous sommes confrontés à deux cas :

- Si  $\mu_\beta = 0$ , nous aurons alors :

$$\begin{aligned} \mathbb{E}(\ln |\beta_t|) &= \mathbb{E}(\ln |\sigma_\beta v_t|) \\ &= \mathbb{E}(\ln |\sigma_\beta|) + \mathbb{E}(\ln |v_t|) \\ &= \frac{1}{2} \ln(\sigma_\beta^2) - \frac{1}{2}(\zeta + \ln(2)) \end{aligned}$$

Par conséquent

$$\mathbb{E}(\ln |\beta_t|) < 0$$

- Si  $\mu_\beta \neq 0$ , nous aurons alors :

$$\begin{aligned} \mathbb{E}(\ln |\beta_t|) &= \mathbb{E}(\ln |\mu_\beta + \sigma_\beta v_t|) \\ &= \mathbb{E}(\ln(|\mu_\beta| |1 + \frac{\sigma_\beta}{\mu_\beta} v_t|)) \\ &= \mathbb{E}(\ln(|\mu_\beta|) + \mathbb{E}(\ln(|1 + \frac{\sigma_\beta}{\mu_\beta} v_t|))) \\ &= \ln(|\mu_\beta|) + \mathbb{E}(\ln(|1 + \frac{\sigma_\beta}{\mu_\beta} v_t|)) \\ &= \frac{1}{2} \ln(|\mu_\beta|^2) + \mathbb{E}(\ln(|1 + \frac{\sigma_\beta}{\mu_\beta} v_t|)) \\ &= \frac{1}{2} \ln(|\mu_\beta|^2) + \mathbb{E}(\ln(|1 + Z|)) \end{aligned}$$

Avec

$$Z \sim \mathcal{N}(0, \frac{\sigma_\beta^2}{\mu_\beta^2})$$

En utilisant l'évaluation de  $\mathbb{E}(\ln |1 + Z|)$  dans [29] nous allons obtenir :

$$\mathbb{E}(\ln |\beta_t|) < 0 \iff \ln(\sigma_\beta^2) < \zeta + \ln(2) - 2 \int_0^1 \frac{1 - \exp[-\lambda(1 - \omega^2)]}{1 - \omega^2} d\omega$$

**Remarque 2.1.** *La méthode de Box-Jenkins est largement utilisée pour la modélisation et la prévision des séries chronologiques, notamment pour les modèles ARIMA (Autoregressive Integrated Moving Average). Elle se base sur l'identification, l'estimation et la validation des modèles pour analyser et prédire les données temporelles.*

*Cependant, la méthode de Box-Jenkins n'est pas directement applicable aux modèles RCAR, notamment des RCAR(1) ou des modèles d'ordre supérieur sans adaptation. Les modèles RCAR nécessitent une approche statistique différente pour estimer les paramètres, en tenant compte de la variabilité des coefficients. Pour cela, des méthodes spécifiques d'inférence statistique, telles que les méthodes de Monte-Carlo par chaînes de Markov (MCMC) ou les méthodes Bayésiennes, sont souvent utilisées, prenant en considération la nature aléatoire des coefficients.*

*En résumé, bien que la méthode de Box-Jenkins soit une approche courante pour les modèles ARIMA, elle ne peut être directement appliquée aux modèles RCAR en raison de la nature aléatoire des coefficients. Des méthodes alternatives doivent être utilisées pour modéliser et estimer les modèles RCAR.*

## **2.4 L'approche fréquentiste dans l'estimation d'un modèle RCAR(1)**

Il existe différentes méthodes d'estimation des paramètres d'un modèle autorégressif à coefficient aléatoire d'ordre un. Dans cette section, nous allons faire l'inférence statistique de ces modèles par l'application de certaines méthodes usuelles d'estimation largement utilisées : la méthode des moindres carrés et la méthode du maximum de vraisemblance. Ces deux méthodes ont été étudiées par de nombreux auteurs, et nous montrerons ensuite la consistance et la normalité des estimateurs obtenus.

### **2.4.1 Estimation par la méthode des moindres carrés**

La première méthode proposée est basée sur le critère des moindres carrés et se déroule en deux étapes afin d'effectuer l'estimation.

### 2.4.1.1 Calcul de l'estimateur par la méthode des moindres carrés

Nicholls and Quinn (1981)[28] considèrent l'échantillon  $\{X_0, X_1, \dots, X_n\}$  du modèle (2.2). L'estimateur des moindres carrés noté  $\hat{\mu}_{\beta, MC}$  du paramètre  $\mu_\beta$  tel que  $\mu_\beta = \mathbb{E}(\beta_t)$  est obtenu en minimisant la somme des carrés des erreurs.

Soit  $\mathcal{F}_t = \sigma(X_s, s \leq t)$  la tribu engendrée par le processus jusqu'à l'instant  $t$ , nous avons :

$$\begin{aligned}\mathbb{E}(X_t | \mathcal{F}_{t-1}) &= \mathbb{E}(\beta_t X_{t-1} + \epsilon_t | \mathcal{F}_{t-1}) \\ \mathbb{E}(X_t | \mathcal{F}_{t-1}) &= \mathbb{E}(\beta_t X_{t-1} | \mathcal{F}_{t-1}) + \mathbb{E}(\epsilon_t | \mathcal{F}_{t-1})\end{aligned}\tag{2.6}$$

Comme  $\beta_t$  et  $\epsilon_t$  sont indépendantes de  $\mathcal{F}_{t-1}$ , et que  $X_{t-1}$  est  $\mathcal{F}_{t-1}$  mesurable.

En outre, nous savons que :

- $\mathbb{E}(X_{t-1} | \mathcal{F}_{t-1}) = X_{t-1}$
- $\mathbb{E}(\beta_t | \mathcal{F}_{t-1}) = \mu_\beta$
- $\mathbb{E}(\epsilon_t | \mathcal{F}_{t-1}) = 0$

En substituant dans l'équation (2.6), nous pouvons conclure que :

$$\mathbb{E}(X_t | \mathcal{F}_{t-1}) = \mu_\beta X_{t-1}$$

Soit  $u_t = X_t - \mu_\beta X_{t-1}$ , nous pouvons donc écrire :

$$\begin{aligned}\mathbb{E}(u_t | \mathcal{F}_{t-1}) &= \mathbb{E}(X_t - \mu_\beta X_{t-1} | \mathcal{F}_{t-1}) \\ \mathbb{E}(u_t | \mathcal{F}_{t-1}) &= \mathbb{E}(X_t | \mathcal{F}_{t-1}) - \mathbb{E}(\mu_\beta X_{t-1} | \mathcal{F}_{t-1}) \\ \mathbb{E}(u_t | \mathcal{F}_{t-1}) &= 0\end{aligned}$$

De la même manière, nous pouvons également écrire que :

$$\begin{aligned}\mathbb{E}(u_t^2 | \mathcal{F}_{t-1}) &= \mathbb{E}((X_t - \mu_\beta X_{t-1})^2 | \mathcal{F}_{t-1}) \\ \mathbb{E}(u_t^2 | \mathcal{F}_{t-1}) &= \mathbb{E}(X_t^2 - 2\mu_\beta X_t X_{t-1} + \mu_\beta^2 X_{t-1}^2 | \mathcal{F}_{t-1}) \\ \mathbb{E}(u_t^2 | \mathcal{F}_{t-1}) &= \mathbb{E}(X_t^2 | \mathcal{F}_{t-1}) - 2\mu_\beta \mathbb{E}(X_t X_{t-1} | \mathcal{F}_{t-1}) + \mathbb{E}(\mu_\beta^2 X_{t-1}^2 | \mathcal{F}_{t-1})\end{aligned}$$

Puisque :

$$\mathbb{E}(X_t^2 | \mathcal{F}_{t-1}) = \mathbb{E}(\beta_t^2 X_{t-1}^2 | \mathcal{F}_{t-1}) + \mathbb{E}(\epsilon_t^2 | \mathcal{F}_{t-1}) + 2\mathbb{E}((\beta_t \epsilon_t) X_{t-1} | \mathcal{F}_{t-1})$$

Comme nous avons :

$$\begin{aligned}\mathbb{E}(X_t^2|\mathcal{F}_{t-1}) &= X_{t-1}^2\mathbb{E}(\beta_t^2|\mathcal{F}_{t-1}) + \mathbb{E}(\epsilon_t^2|\mathcal{F}_{t-1}) \\ \mathbb{E}(X_t^2|\mathcal{F}_{t-1}) &= X_{t-1}^2(\sigma_\beta^2 + \mu_\beta^2) + \sigma^2\end{aligned}$$

Par conséquent, nous pouvons conclure que :

$$\begin{aligned}\mathbb{E}(u_t^2|\mathcal{F}_{t-1}) &= X_{t-1}^2(\sigma_\beta^2 + \mu_\beta^2) + \sigma^2 + \mu_\beta^2 X_{t-1}^2 - 2\mu_\beta^2 X_{t-1}^2 \\ \mathbb{E}(u_t^2|\mathcal{F}_{t-1}) &= \sigma^2 + \sigma_\beta^2 X_{t-1}^2\end{aligned}$$

Nous avons

$$\sum_{t=1}^n u_t^2 = \sum_{t=1}^n (X_t - \mu_\beta X_{t-1})^2$$

La première étape consiste à estimer  $\mu_\beta$  en minimisant  $\sum_{t=1}^n u_t^2$  par rapport à  $\mu_\beta$ .

Puisque

$$\frac{\partial}{\partial \mu_\beta} \sum_{t=1}^n u_t^2 = \frac{\partial}{\partial \mu_\beta} \sum_{t=1}^n (X_t - \mu_\beta X_{t-1})^2$$

Donc

$$\frac{\partial}{\partial \mu_\beta} \sum_{t=1}^n u_t^2 = \frac{\partial}{\partial \mu_\beta} \sum_{t=1}^n (X_t^2 - 2\mu_\beta X_{t-1} X_t + \mu_\beta^2 X_{t-1}^2)$$

En dérivant par rapport à  $\mu_\beta$ , nous obtenons :

$$\frac{\partial}{\partial \mu_\beta} \sum_{t=1}^n u_t^2 = \sum_{t=1}^n (-2X_{t-1} X_t + 2\mu_\beta X_{t-1}^2)$$

En dérivant une seconde fois, nous observons que :

$$\frac{\partial^2}{\partial^2 \mu_\beta} \sum_{t=1}^n u_t^2 = 2 \sum_{t=1}^n (X_{t-1}^2) > 0$$

Étant donné que les conditions nécessaires sont satisfaites et que la deuxième dérivée est strictement positive, nous pouvons donc conclure :

$$\frac{\partial}{\partial \mu_\beta} \sum_{t=1}^n u_t^2 = 0$$

Cela équivaut à affirmer que :

$$\sum_{t=1}^n (-2X_t X_{t-1} + 2\mu_\beta X_{t-1}^2) = 0$$

De la même manière, nous pouvons conclure :

$$\hat{\mu}_{\beta,MC} = \frac{\sum_{t=1}^n X_t X_{t-1}}{\sum_{t=1}^n X_{t-1}^2} \quad (2.7)$$

La deuxième étape de l'estimation des moindres carrés implique la formation du résidu  $\hat{u}_t$  en utilisant l'estimateur  $\hat{\mu}_{\beta,MC}$ , permettant ainsi de minimiser par rapport à  $\sigma^2$  et  $\sigma_\beta^2$ .

$$\hat{u}_t = X_t - \hat{\mu}_{\beta,MC} X_{t-1}$$

Ceci équivaut à minimiser

$$l(\sigma^2, \sigma_\beta^2) = \sum_{t=1}^n (\hat{u}_t^2 - \sigma^2 - \sigma_\beta^2 X_{t-1}^2)^2$$

En dérivant par rapport à ces deux paramètres, nous obtenons :

$$\frac{\partial l(\sigma^2, \sigma_\beta^2)}{\partial \sigma_\beta^2} = \frac{\partial \sum_{t=1}^n (\hat{u}_t^2 - \sigma^2 - \sigma_\beta^2 X_{t-1}^2)^2}{\partial \sigma_\beta^2}$$

Donc

$$\frac{\partial l(\sigma^2, \sigma_\beta^2)}{\partial \sigma_\beta^2} = -2 \sum_{t=1}^n X_{t-1}^2 (\hat{u}_t^2 - \sigma^2 - \sigma_\beta^2 X_{t-1}^2)$$

Nous constatons que la deuxième dérivée est strictement positive :

$$\frac{\partial l(\sigma^2, \sigma_\beta^2)}{\partial \sigma_\beta^2} = \frac{\partial (-2 \sum_{t=1}^n X_{t-1}^2 (\hat{u}_t^2 - \sigma^2 - \sigma_\beta^2 X_{t-1}^2))}{\partial \sigma_\beta^2}$$

$$\frac{\partial l(\sigma^2, \sigma_\beta^2)}{\partial \sigma_\beta^2} = 2 \sum_{t=1}^n X_{t-1}^4 > 0$$

Ensuite, Il est nécessaire de résoudre :

$$\begin{aligned} \frac{\partial l(\sigma^2, \sigma_\beta^2)}{\partial \sigma_\beta^2} &= 0 \\ -2 \sum_{t=1}^n X_{t-1}^2 (\hat{u}_t^2 - \sigma^2 - \sigma_\beta^2 X_{t-1}^2) &= 0 \end{aligned}$$

Nous pouvons conclure que :

$$\hat{\sigma}_{\beta, MC}^2 = \frac{\sum_{t=1}^n X_{t-1}^2 \hat{u}_t^2 - \sigma^2 \sum_{t=1}^n X_{t-1}^2}{\sum_{t=1}^n X_{t-1}^4}$$

Nous estimons  $\hat{\sigma}_{MC}^2$  de la même manière :

$$\begin{aligned} \frac{\partial l(\sigma^2, \sigma_\beta^2)}{\partial \sigma^2} &= \frac{\partial}{\partial \sigma^2} \sum_{t=1}^n (\hat{u}_t^2 - \sigma^2 - \sigma_\beta^2 X_{t-1}^2) \\ \frac{\partial l(\sigma^2, \sigma_\beta^2)}{\partial \sigma^2} &= -2 \sum_{t=1}^n (\hat{u}_t^2 - \sigma^2 - \sigma_\beta^2 X_{t-1}^2) = 0 \end{aligned}$$

Donc

$$n \hat{\sigma}_{MC}^2 = \sum_{t=1}^n (\hat{u}_t^2 - \sigma_\beta^2 X_{t-1}^2)$$

Par conséquent, nous pouvons conclure que l'estimateur de  $\hat{\sigma}_{MC}^2$  est donné par :

$$\hat{\sigma}_{MC}^2 = n^{-1} \left( \sum_{t=1}^n (\hat{u}_t^2 - \hat{\sigma}_{\beta, MC}^2 X_{t-1}^2) \right)$$

Soit

$$\hat{\sigma}_{MC}^2 = n^{-1} \sum_{t=1}^n \hat{u}_t^2 - \hat{\sigma}_{\beta, MC}^2 \bar{z}$$

avec

$$\bar{z} = n^{-1} \sum_{t=1}^n X_{t-1}^2$$

### 2.4.1.2 La consistance et la normalité asymptotique de l'estimateur des moindres carrés

L'estimateur obtenu par la méthode des moindres carrés est à la fois consistant et présente une normalité asymptotique.

Dans l'étude menée par Nicholls and Quinn (1981)[28], l'objectif est de déterminer la loi asymptotique de  $\hat{\mu}_{\beta,MC}$ .

La première étape consiste à voir la consistance de l'estimateur  $\hat{\mu}_{\beta,MC}$  et à établir que :

$$\hat{\mu}_{\beta,MC} \xrightarrow{\text{ps}} \frac{\gamma(1)}{\gamma(0)} = \mu_{\beta}$$

lorsque

$$n \rightarrow \infty$$

où :

$$\gamma(0) = \frac{\sigma^2}{(1 - \mu_{\beta}^2 - \sigma_{\beta}^2)} \quad \text{et} \quad \gamma(k) = \mu_{\beta}^k \gamma(0), \quad \forall k \in \{0, 1, \dots, k\}$$

Pour déterminer la loi asymptotique de  $\hat{\mu}_{\beta,MC}$ , les auteurs posent les conditions suivantes :

- $(\epsilon_t, \beta_t)$  sont des suites de variables aléatoires indépendantes et identiquement distribuées (iid).
- $\mathbb{E}(X_t^4) < \infty$

Ce qui implique que le processus  $X_t$  est markovien. Par conséquent, il suffit que le moment d'ordre quatre de  $\beta_t$  soit inférieur à un, c'est-à-dire  $\mathbb{E}(\beta_t^4) < 1$ .

**Théorème 2.4.** [28](Nicholls, D.F. and Quinn, B.G 1981)

Lorsque  $n \rightarrow \infty$  et sous les conditions  $\sigma_{\beta}^2 + \mu_{\beta}^2 < 1$  et  $\mathbb{E}(X_t^4) < \infty$  nous obtenons que :

$$\sqrt{n}(\hat{\mu}_{\beta,MC} - \mu_{\beta}) \xrightarrow{\mathcal{L}} \mathcal{N}(0, v^2)$$

avec

$$v^2 = \frac{\sigma^2 \mathbb{E}(X_0^2) + 2\sigma_{\beta\epsilon} \mathbb{E}(X_0^3) + \sigma_{\beta}^2 \mathbb{E}(X_0^4)}{\gamma(0)^2}$$

où

$$\text{cov}(\epsilon_t, \beta_t) = \sigma_{\beta\epsilon}$$

**Démonstration.** Pour démontrer ce résultat, les auteurs procèdent de la manière suivante :

$$\begin{aligned} \hat{\mu}_{\beta,MC} - \mu_\beta &= \frac{\sum_{t=1}^n X_t X_{t-1}}{\sum_{t=1}^n X_{t-1}^2} - \mu_\beta \\ &= \frac{\sum_{t=1}^n X_t X_{t-1} - \mu_\beta \sum_{t=1}^n X_{t-1}^2}{\sum_{t=1}^n X_{t-1}^2} \\ &= \frac{\sum_{t=1}^n (\beta_t X_{t-1} + \epsilon_t) X_{t-1} - \mu_\beta \sum_{t=1}^n X_{t-1}^2}{\sum_{t=1}^n X_{t-1}^2} \\ &= \frac{\sum_{t=1}^n X_{t-1} (\tilde{\beta}_t X_{t-1} + \epsilon_t)}{\sum_{t=1}^n X_{t-1}^2} \end{aligned}$$

$$\tilde{\beta}_t = \beta_t - \mu_\beta, \quad t \in \mathbb{Z}.$$

Nous notons que la loi de  $\tilde{\beta}_t$  dépend de  $\mu_\beta$ .

Si  $(X_{t-1}(\tilde{\beta}_t X_{t-1} + \epsilon_t))$  est stationnaire et ergodique, d'après le théorème central limite (TCL) des martingales stationnaires.

$$\frac{1}{n} \sum_{t=1}^n X_{t-1} (\tilde{\beta}_t X_{t-1} + \epsilon_t) \xrightarrow{\mathcal{L}} \mathcal{N}(0, v^2 \gamma(0)^2) \quad (2.8)$$

et d'après le théorème de l'ergodicité (1.2)

$$\frac{1}{n} \sum_{t=1}^n X_{t-1}^2 \xrightarrow{\text{ps}} \gamma(0)^2 \quad (2.9)$$

de (2.8) et (2.9) le résultat s'ensuit.

### 2.4.2 Estimation des paramètres d'un modèle RCAR(1) par la méthode du maximum de vraisemblance

Dans cette section nous considérons les hypothèses suivantes :

- A4 :  $\mathbb{E}(\epsilon_t^4) < \infty$  et  $\mathbb{E}(\beta_t^4) < \infty$ .

- A5 :  $\sigma^2 \geq \delta_1 > 0$  et  $\sigma_\beta^2 \geq \delta_2 > 0$  tels que  $\delta_1, \delta_2 > 0$ .
- A6 :  $(\mu_\beta, \sigma_\beta^2) \in D$  où  $D$  est un ensemble fermé contenu dans la région de stationnarité stricte et d'ergodicité donnée par l'inégalité (2.5).

L'hypothèse A4 sera utilisée pour démontrer le théorème central limite pour l'estimateur du maximum de vraisemblance, et les hypothèses A5 et A6 sont imposées afin que l'ensemble des paramètres du modèle RCAR(1).

$$\Theta = \{(\alpha, \mu_\beta, \sigma_\beta^2, \sigma^2) : \alpha \in \mathbb{R}, (\mu_\beta, \sigma_\beta^2) \in \mathcal{D}, \sigma^2 \geq \delta_1, \sigma_\beta^2 \geq \delta_2\} \text{ soit compact.}$$

### 2.4.2.1 Calcul de l'estimateur du maximum de vraisemblance

Considérons l'échantillon  $(X_0, X_1, \dots, X_n)$  à partir d'un modèle RCAR(1),  $\{X_t, t \in \mathbb{Z}\}$  est strictement stationnaire, ergodique et  $\mathcal{F}_t$  mesurable et satisfait l'équation (2.1) sous les conditions A1 – A6, nous fixons un  $X_0$  et nous dérivons la fonction de vraisemblance en supposant la normalité de  $\beta_t$  et  $\epsilon_t$ .

Les principales propriétés de l'espérance conditionnelle nous permettent d'écrire :

$$\begin{aligned} \mathbb{E}(X_t|X_{t-1}) &= \mathbb{E}(\alpha + \beta_t X_{t-1} + \epsilon_t|X_{t-1}) \\ &= \mathbb{E}(\alpha + \beta_t X_{t-1}|X_{t-1}) + \mathbb{E}(\epsilon_t|X_{t-1}) \\ &= \alpha + \mu_\beta X_{t-1} \end{aligned}$$

L'application de ces propriétés à la variance nous permet d'écrire :

$$\begin{aligned} \mathbb{V}(X_t|X_{t-1}) &= \mathbb{V}(\alpha + \beta_t X_{t-1} + \epsilon_t|X_{t-1}) \\ &= \mathbb{V}(\alpha + \beta_t X_{t-1}|X_{t-1}) + \mathbb{V}(\epsilon_t|X_{t-1}) \\ &= \sigma^2 + \sigma_\beta^2 X_{t-1}^2 \end{aligned}$$

Par conséquent, la fonction de vraisemblance (conditionnelle à  $X_0$ ) a la forme suivante :

$$\begin{aligned} f_n(X_1, X_2, \dots, X_n|X_0) &= \prod_{i=1}^n f(X_i|X_{i-1}) \\ &= (2\pi)^{-\frac{n}{2}} \prod_{t=1}^n \left\{ (\sigma^2 + \sigma_\beta^2 X_{t-1}^2)^{-\frac{1}{2}} \exp \left[ -\frac{(X_t - \alpha - \mu_\beta X_{t-1})^2}{2(\sigma^2 + \sigma_\beta^2 X_{t-1}^2)} \right] \right\} \end{aligned}$$

Nous considérons une nouvelle fonction  $L_n$  avec les paramètres suivants  $(\alpha, \mu_\beta, \sigma_\beta^2, \sigma^2)$

pour faciliter la notation.

$$f_n(X_1, X_2, \dots, X_n | X_0) = L_n(\alpha, \mu_\beta, \sigma_\beta^2, \sigma^2)$$

Nous posons :

$$\tilde{l}_n(\alpha, \mu_\beta, \sigma_\beta^2, \sigma^2) = -\frac{2}{n} \ln[L_n(\alpha, \mu_\beta, \sigma_\beta^2, \sigma^2)] - \ln(2\pi)$$

Nous aurons alors :

$$\begin{aligned} \tilde{l}_n(\alpha, \mu_\beta, \sigma_\beta^2, \sigma^2) &= -\frac{2}{n} \ln \left[ (2\pi)^{-\frac{n}{2}} \prod_{t=1}^n \left\{ (\sigma^2 + \sigma_\beta^2 X_{t-1}^2)^{-\frac{1}{2}} \exp \left( -\frac{(X_t - \alpha - \mu_\beta X_{t-1})^2}{2(\sigma^2 + \sigma_\beta^2 X_{t-1}^2)} \right) \right\} \right] \\ &\quad - \ln(2\pi) \\ &= \frac{1}{n} \sum_{t=1}^n \ln(\sigma^2 + \sigma_\beta^2 X_{t-1}^2) + \frac{1}{n} \sum_{t=1}^n \frac{(X_t - \alpha - \mu_\beta X_{t-1})^2}{\sigma^2 + \sigma_\beta^2 X_{t-1}^2} \end{aligned}$$

Il est pratique de minimiser  $\tilde{l}_n(\alpha, \mu_\beta, \sigma_\beta^2, \sigma^2)$  au lieu de maximiser  $L_n(\alpha, \mu_\beta, \sigma_\beta^2, \sigma^2)$ .

Pour cela nous posons  $\tau = \frac{\sigma_\beta^2}{\sigma^2}$ , et nous donnons la fonction  $\bar{l}_n$  qui dépend de  $\tau$ .

$$\begin{aligned} \bar{l}_n(\alpha, \mu_\beta, \tau, \sigma^2) &= \tilde{l}_n(\alpha, \mu_\beta, \sigma_\beta^2, \sigma^2) \\ &= \ln(\sigma^2) + n^{-1} \sum_{t=1}^n \ln(1 + \tau X_{t-1}^2) + \sigma^{-2} n^{-1} \sum_{t=1}^n \frac{(X_t - \alpha - \mu_\beta X_{t-1})^2}{(1 + \tau X_{t-1}^2)} \end{aligned}$$

En dérivant par rapport aux paramètres  $\alpha$  et  $\mu_\beta$ , nous obtenons :

$$\begin{aligned} \frac{\partial}{\partial \alpha} \bar{l}_n(\alpha, \mu_\beta, \tau, \sigma^2) &= -2\sigma^{-2} n^{-1} \sum_{t=1}^n \frac{(X_t - \alpha - \mu_\beta X_{t-1})}{1 + \tau X_{t-1}^2} \\ \frac{\partial}{\partial \mu_\beta} \bar{l}_n(\alpha, \mu_\beta, \tau, \sigma^2) &= -2\sigma^{-2} n^{-1} \sum_{t=1}^n \frac{(X_t - \alpha - \mu_\beta X_{t-1}) X_{t-1}}{1 + \tau X_{t-1}^2} \end{aligned}$$

Maintenant, nous avons :

$$\begin{aligned} \frac{\partial}{\partial \alpha} \bar{l}_n(\alpha, \mu_\beta, \tau, \sigma^2) &= 0 \\ \frac{\partial}{\partial \mu_\beta} \bar{l}_n(\alpha, \mu_\beta, \tau, \sigma^2) &= 0 \end{aligned}$$

nous trouvons

$$\begin{aligned}\hat{\mu}_\beta &= \hat{\mu}_{\beta,n}(\tau) = \frac{c_1 c_2 - c_3 c_4}{c_1 c_5 - c_4^2} \\ \hat{\alpha} &= \hat{\alpha}_n(\tau) = \frac{c_3 - \mu_{\beta,n}(\tau) c_4}{c_1}\end{aligned}$$

et

$$\hat{\sigma}^2 = \sigma_n^2(\tau) = n^{-1} \sum_{t=1}^n \frac{(X_t - \alpha_n(\tau) - \mu_{\beta,n}(\tau) X_{t-1})^2}{1 + \tau X_{t-1}^2}$$

où

$$\begin{aligned}c_1 &= \sum_{t=1}^n \frac{1}{1 + \tau X_{t-1}^2} \\ c_2 &= \sum_{t=1}^n \frac{X_t X_{t-1}}{1 + \tau X_{t-1}^2} \\ c_3 &= \sum_{t=1}^n \frac{X_t}{1 + \tau X_{t-1}^2} \\ c_4 &= \sum_{t=1}^n \frac{X_{t-1}}{1 + \tau X_{t-1}^2} \\ c_5 &= \sum_{t=1}^n \frac{X_{t-1}^2}{1 + \tau X_{t-1}^2}\end{aligned}$$

L'estimateur du maximum de vraisemblance de  $\hat{\alpha}_n, \hat{\mu}_{\beta,n}, \hat{\sigma}_n^2, \hat{\sigma}_{\beta,n}^2$ , peut être obtenu en calculant  $\hat{\tau}_n$ .

Nous pouvons alors considérer le minimum de la fonction suivante :

$$l_n(\alpha, \mu_\beta, \tau) = \inf_{\sigma^2} \bar{l}_n(\alpha, \mu_\beta, \tau, \sigma^2) - 1$$

Où

$$\bar{l}_n(\alpha, \mu_\beta, \tau, \sigma^2) = \ln \sigma^2 + n^{-1} \sum_{t=1}^n \ln(1 + \tau X_{t-1}^2) + \sigma^{-2} n^{-1} \sum_{t=1}^n \frac{(X_t - \alpha - \mu_\beta X_{t-1})^2}{1 + \tau X_{t-1}^2}$$

Calculons  $\inf_{\sigma^2} \bar{l}_n(\alpha, \mu_\beta, \tau, \sigma^2)$ , nous obtenons alors :

$$\frac{\partial \bar{l}_n(\alpha, \mu_\beta, \tau, \sigma^2)}{\partial \sigma^2} = \sigma^{-2} - \sigma^{-4} n^{-1} \sum_{t=1}^n \frac{(X_t - \alpha - \mu_\beta X_{t-1})^2}{1 + \tau X_{t-1}^2}$$

D'où

$$\sigma^2 = n^{-1} \sum_{t=1}^n \frac{(X_t - \alpha - \mu_\beta X_{t-1})^2}{1 + \tau X_{t-1}^2}$$

Par suite

$$\begin{aligned} \inf_{\sigma^2} \bar{l}_n(\alpha, \mu_\beta, \tau, \sigma^2) &= \ln \left( n^{-1} \sum_{t=1}^n \frac{(X_t - \alpha - \mu_\beta X_{t-1})^2}{1 + \tau X_{t-1}^2} \right) \\ &\quad + n^{-1} \sum_{t=1}^n \ln(1 + \tau X_{t-1}^2) + 1 \end{aligned}$$

D'où

$$l_n(\alpha, \mu_\beta, \tau) = \ln \left( n^{-1} \sum_{t=1}^n \frac{(X_t - \alpha - \mu_\beta X_{t-1})^2}{1 + \tau X_{t-1}^2} \right) + n^{-1} \sum_{t=1}^n \ln(1 + \tau X_{t-1}^2)$$

$\hat{\tau}_n$  est obtenu en minimisant la fonction  $l_n(\hat{\alpha}, \hat{\mu}_\beta, \tau_n)$  par rapport à  $\tau$ . Comme  $\tau = \frac{\sigma_\beta^2}{\sigma^2}$ , alors  $\hat{\sigma}_{\beta,n}^2 = \hat{\tau}_n \hat{\sigma}_n^2$ .

#### 2.4.2.2 La consistance forte de l'estimateur du maximum de vraisemblance

La consistance forte de  $\hat{\alpha}_n$ ,  $\hat{\mu}_{\beta,n}$ ,  $\hat{\sigma}_n^2$  et  $\hat{\sigma}_{\beta,n}^2$  sera démontrée en examinant la fonction  $l_n(\alpha, \mu_\beta, \tau)$ .

Il convient de noter que le minimum de  $l_n(\alpha, \mu_\beta, \tau)$  est le même que celui de  $l_n^*(\alpha, \mu_\beta, \tau)$  défini par :

$$\begin{aligned}
 l_n^*(\alpha, \mu_\beta, \tau) &= l_n(\alpha, \mu_\beta, \tau) - l_n(\alpha_0, \mu_{\beta,0}, \tau_0) \\
 &= n^{-1} \sum_{t=1}^n \ln(1 + \tau X_{t-1}^2) + \ln \left( n^{-1} \sum_{t=1}^n \frac{(X_t - \alpha - \mu_\beta X_{t-1})^2}{1 + \tau X_{t-1}^2} \right) \\
 &\quad - n^{-1} \sum_{t=1}^n \ln(1 + \tau_0 X_{t-1}^2) - \ln \left( n^{-1} \sum_{t=1}^n \frac{(X_t - \alpha_0 - \mu_{\beta,0} X_{t-1})^2}{1 + \tau_0 X_{t-1}^2} \right) \\
 &= n^{-1} \sum_{t=1}^n \ln \left( \frac{1 + \tau X_{t-1}^2}{1 + \tau_0 X_{t-1}^2} \right) + \ln \left( n^{-1} \sum_{t=1}^n \frac{(X_t - \alpha - \mu_\beta X_{t-1})^2}{1 + \tau X_{t-1}^2} \right) \\
 &\quad - \ln \left( n^{-1} \sum_{t=1}^n \frac{(X_t - \alpha_0 - \mu_{\beta,0} X_{t-1})^2}{1 + \tau_0 X_{t-1}^2} \right)
 \end{aligned}$$

Où  $\alpha_0, \mu_{\beta,0}$  et  $\tau_0$  désignent certaines valeurs fixes de  $\alpha, \mu_\beta$  et  $\tau$ .

**Lemme 2.1.** *Sous les hypothèses A1 – A3 (énoncées dans la section (2.2.2)), la variable aléatoire  $Y_t$  définie par  $Y_t = X_t^2$  est presque sûrement non constante pour tout  $t$ , où  $\{X_t, t \in \mathbb{Z}\}$  est une solution strictement stationnaire et ergodique de l'équation (2.1).*

**Lemme 2.2.** *Considérons la fonction de la forme suivante :*

$$g(x) = \frac{ax^2 + bx + c}{dx^2 + e}$$

où  $d > 0$  et  $e > 0$ . Alors il existe une constante  $M > 0$  telle que  $|g(x)| \leq M$  pour tout  $x$ .

**Démonstration.** Prenons  $M = \left| \frac{a}{d} \right| + \left| \frac{b}{2\sqrt{de}} \right| + \left| \frac{c}{e} \right| + 1$ .

**Théorème 2.5.**  $\{X_t, t \in \mathbb{Z}\}$  une suite strictement stationnaire, ergodique et  $\mathcal{F}_t$  une solution mesurable qui satisfait l'équation (2.1) avec

$\theta = (\alpha, \mu_\beta, \sigma_\beta^2, \sigma^2)' = (\alpha_0, \mu_{\beta,0}, \sigma_{\beta,0}^2, \sigma_0^2)' = \theta_0$  sous les hypothèses A1, A3, A5 et A6 (énoncées dans les sections (2.2.2) et (2.4.2)). Définissons  $\pi = (\alpha, \mu_\beta, \tau)'$  où

$\tau = \frac{\sigma_\beta^2}{\sigma^2}$ . Soit  $\Theta^*$  l'image de  $\Theta$  par l'application continue  $T : \theta \rightarrow \pi$ , où  $\Theta$  est l'espace

des paramètres du modèle RCAR(1). Soit  $\pi_0 = (\alpha_0, \mu_{\beta,0}, \tau_0)'$  où  $\tau_0 = \frac{\sigma_{\beta,0}^2}{\sigma_0^2}$ . Alors

la  $\lim_{n \rightarrow \infty} l_n^*(\alpha, \mu_\beta, \tau)$  existe presque sûrement pour tout  $(\alpha, \mu_\beta, \tau) \in \Theta^*$  et sa limite  $l^*(\alpha, \mu_\beta, \tau)$  est minimisée uniquement sur  $\Theta^*$  en  $\pi_0$  à condition que  $\pi_0 \in \text{int}(\Theta^*)$ .

**Démonstration.** Nicholls and Quinn [26] montrent que sous les hypothèses A1 – A3 et A5 – A6 (énoncées dans les sections (2.2.2) et (2.4.2)),  $\delta_1$  et  $\delta_2$  peuvent être choisies de telle sorte que  $(\alpha_0, \mu_{\beta,0}, \sigma_{\beta,0}^2, \sigma_0^2)' \in \text{int}(\Theta)$  et que l'ensemble  $\Theta$  soit compact.

Puisque  $T$  est une application continue et  $\Theta$  est compact, nous pouvons conclure que  $\Theta^*$  est également compact et que  $\pi_0 \in \text{int}(\Theta^*)$  car  $\pi_0 = T(\theta_0)$ .

Maintenant, selon le théorème de l'ergodicité (1.2), les trois termes dans  $l_n^*(\alpha, \mu_\beta, \tau)$  convergent respectivement vers

$$\mathbb{E} \left[ \ln \left( \frac{1 + \tau X_{t-1}^2}{1 + \tau_0 X_{t-1}^2} \right) \right], \ln \left[ \mathbb{E} \left( \frac{(X_t - \alpha - \mu_\beta X_{t-1})^2}{1 + \tau X_{t-1}^2} \right) \right] \text{ et } \ln \left[ \mathbb{E} \left( \frac{(X_t - \alpha_0 - \mu_{\beta,0} X_{t-1})^2}{1 + \tau_0 X_{t-1}^2} \right) \right]$$

à condition que ces espérances sont finies.

Pour le premier terme, il est facile de voir que

$$\max \left( 1, \frac{\tau}{\tau_0} \right) \geq \frac{1 + \tau X_{t-1}^2}{1 + \tau_0 X_{t-1}^2} \geq \min \left( 1, \frac{\tau}{\tau_0} \right)$$

Par conséquent

$$\mathbb{E} \left[ \ln \left( \frac{1 + \tau X_{t-1}^2}{1 + \tau_0 X_{t-1}^2} \right) \right] < \infty$$

Pour le deuxième terme, d'après le lemme (2.2), nous avons :

$$\begin{aligned} \mathbb{E} \left( \frac{(X_t - \alpha - \mu_\beta X_{t-1})^2}{1 + \tau X_{t-1}^2} \right) &= \left| \mathbb{E} \frac{(X_t - \alpha - \mu_\beta X_{t-1})^2}{1 + \tau X_{t-1}^2} \right| \\ &\leq \mathbb{E} \left| \frac{(X_t - \alpha - \mu_\beta X_{t-1})^2}{1 + \tau X_{t-1}^2} \right| \\ &< \infty \end{aligned}$$

Pour le troisième terme, nous avons :

$$0 < \mathbb{E} \left( \frac{(X_t - \alpha_0 - \mu_{\beta,0} X_{t-1})^2}{1 + \tau_0 X_{t-1}^2} \right) = \sigma_0^2 < \infty$$

De plus,  $\mathbb{E} \left[ \frac{(X_t - \alpha - \mu_\beta X_{t-1})^2}{1 + \tau X_{t-1}^2} \right] > 0$ , car sinon, nous aurions  $X_t = \alpha + \mu_\beta X_{t-1} = 0$  presque sûrement, ce qui est exclu par les hypothèses A5 et A6 (énoncées dans la section (2.4.2)).

Ensuite, en appliquant le théorème de l'ergodicité (1.2), nous avons que  $l_n^*(\alpha, \mu_\beta, \tau)$  converge presque sûrement vers

$$\begin{aligned}
 l^*(\alpha, \mu_\beta, \tau) &= \mathbb{E} \left[ \ln \left( \frac{1 + \tau X_{t-1}^2}{1 + \tau_0 X_{t-1}^2} \right) \right] + \ln \left[ \mathbb{E} \left( \frac{(X_t - \alpha - \mu_\beta X_{t-1})^2}{1 + \tau X_{t-1}^2} \right) \right] - \ln(\sigma_0^2) \\
 \\
 \mathbb{E} \left[ \frac{(X_t - \alpha - \mu_\beta X_{t-1})^2}{1 + \tau X_{t-1}^2} \right] &= \mathbb{E} \left[ \frac{(X_t - \alpha_0 - \mu_{\beta,0} X_{t-1}) + (\alpha_0 - \alpha) + (\mu_{\beta,0} - \mu_\beta) X_{t-1}}{1 + \tau X_{t-1}^2} \right]^2 \\
 &= \mathbb{E} \left[ \frac{(X_t - \alpha_0 - \mu_{\beta,0} X_{t-1})^2}{1 + \tau X_{t-1}^2} \right] + \mathbb{E} \left[ \frac{[(\alpha_0 - \alpha) + (\mu_{\beta,0} - \mu_\beta) X_{t-1}]^2}{1 + \tau X_{t-1}^2} \right] \\
 &\geq \mathbb{E} \left[ \frac{(X_t - \alpha_0 - \mu_{\beta,0} X_{t-1})^2}{1 + \tau X_{t-1}^2} \right] \\
 &\geq \mathbb{E} \left[ \frac{\mathbb{E}((X_t - \alpha_0 - \mu_{\beta,0} X_{t-1})^2) | \mathcal{F}_{t-1}}{1 + \tau X_{t-1}^2} \right] \\
 &\geq \mathbb{E} \left[ \frac{\sigma_0^2 + \sigma_{\beta,0}^2 X_{t-1}^2}{1 + \tau X_{t-1}^2} \right] \\
 &= \sigma_0^2 \mathbb{E} \left[ \frac{1 + \tau_0 X_{t-1}^2}{1 + \tau X_{t-1}^2} \right]
 \end{aligned}$$

Pour avoir l'égalité, il suffit que  $(\alpha_0 - \alpha) + (\mu_{\beta,0} - \mu_\beta) X_{t-1} = 0$  presque sûrement, c'est-à-dire lorsque  $\alpha = \alpha_0$  et  $\mu_\beta = \mu_{\beta,0}$ , car  $X_{t-1}$  est une variable aléatoire non constante d'après le lemme (2.1).

Par conséquent,

$$\begin{aligned}
 \inf_{(\alpha, \mu_\beta)} l^*(\alpha, \mu_\beta, \tau) &= l^*(\alpha_0, \mu_{\beta,0}, \tau) \\
 &= \mathbb{E} \left[ \ln \left( \frac{1 + \tau X_{t-1}^2}{1 + \tau_0 X_{t-1}^2} \right) \right] + \ln \left[ \mathbb{E} \left( \frac{(X_t - \alpha_0 - \mu_{\beta,0} X_{t-1})^2}{1 + \tau X_{t-1}^2} \right) \right] - \ln(\sigma_0^2) \\
 &= \mathbb{E} \left[ \ln \left( \frac{1 + \tau X_{t-1}^2}{1 + \tau_0 X_{t-1}^2} \right) \right] + \ln \left[ \mathbb{E} \left( \sigma^2 \frac{1 + \tau X_{t-1}^2}{1 + \tau_0 X_{t-1}^2} \right) \right] - \ln(\sigma_0^2) \\
 &= \mathbb{E} \left[ \ln \left( \frac{1 + \tau X_{t-1}^2}{1 + \tau_0 X_{t-1}^2} \right) \right] + \ln \left[ \mathbb{E} \left( \frac{1 + \tau_0 X_{t-1}^2}{1 + \tau X_{t-1}^2} \right) \right]
 \end{aligned}$$

et  $l^*(\alpha, \mu_\beta, \tau) = \inf_{(\alpha, \mu_\beta)} l^*(\alpha, \mu_\beta, \tau)$  seulement lorsque  $\alpha = \alpha_0$  et  $\mu_\beta = \mu_{\beta,0}$ .

Maintenant, pour toute variable aléatoire non négative  $Y$  d'espérance 1, nous avons par l'inégalité de Jensen :

$$\mathbb{E}[\ln(Y)] \leq \ln[\mathbb{E}(Y)] = 0$$

l'égalité étant vraie lorsque  $Y = 1$  presque sûrement.

Soit  $W = c^{-1} \frac{1+\tau_0 X_{t-1}^2}{1+\tau X_{t-1}^2}$ , où  $c = \mathbb{E} \left[ \frac{1+\tau_0 X_{t-1}^2}{1+\tau X_{t-1}^2} \right]$ . Il est clair que  $\mathbb{E}(W) = 1$  et

$$\begin{aligned} \inf_{(\alpha, \mu_\beta)} l^*(\alpha, \mu_\beta, \tau) &= \mathbb{E} \left[ \ln \left( \frac{1 + \tau X_{t-1}^2}{1 + \tau_0 X_{t-1}^2} \right) \right] + \ln \left[ \mathbb{E} \left\{ \frac{1 + \tau_0 X_{t-1}^2}{1 + \tau X_{t-1}^2} \right\} \right] \\ &= \mathbb{E} [\ln(c^{-1} W^{-1})] + \ln [\mathbb{E}(Wc)] \\ &= \mathbb{E} [-\ln(cW)] + \ln(c) \\ &= -\mathbb{E} [\ln(W)] \\ &\geq 0 \end{aligned}$$

L'égalité n'est vérifiée que lorsque  $W = c^{-1} \frac{1+\tau_0 X_{t-1}^2}{1+\tau X_{t-1}^2} = 1$  presque sûrement. D'où  $(\tau_0 - c\tau)X_{t-1}^2 + (1 + c) = 0$  presque sûrement.

Par le lemme (2.1), cela se produit uniquement lorsque  $\tau_0 - c\tau = 0$  et  $1 - c = 0$ , c'est-à-dire lorsque  $\tau = \tau_0$ . Par conséquent,  $l^*(\alpha, \mu_\beta, \tau)$  est minimisé uniquement lorsque  $\alpha = \alpha_0$ ,  $\mu_\beta = \mu_{\beta,0}$  et  $\tau = \tau_0$ .

**Corollaire 2.1.**  $\lim_{n \rightarrow \infty} \tilde{l}_n(\alpha, \mu_\beta, \sigma_\beta^2, \sigma^2)$  existe presque sûrement et elle est minimisée uniquement lorsque  $\alpha = \alpha_0$ ,  $\mu_\beta = \mu_{\beta,0}$ ,  $\sigma_\beta^2 = \sigma_{\beta,0}^2$  et  $\sigma^2 = \sigma_0^2$ .

**Démonstration.** Tout d'abord, rappelons que :

$$\begin{aligned} l_n^*(\alpha, \mu_\beta, \tau) &= l_n(\alpha, \mu_\beta, \tau) - l_n(\alpha_0, \mu_{\beta,0}, \tau_0) \\ &= \inf_{\sigma^2} \bar{l}_n(\alpha, \mu_\beta, \tau, \sigma^2) - 1 - l_n(\alpha_0, \mu_{\beta,0}, \tau_0) \\ &= \inf_{\sigma^2} \tilde{l}_n(\alpha, \mu_\beta, \sigma_\beta^2, \sigma^2) - 1 - l_n(\alpha_0, \mu_{\beta,0}, \tau_0) \end{aligned}$$

En utilisant le théorème (2.5) et la définition de  $l_n(\alpha, \mu_\beta, \tau)$ , nous constatons que  $\lim_{n \rightarrow \infty} \tilde{l}_n(\alpha, \mu_\beta, \sigma_\beta^2, \sigma^2)$  existe presque sûrement et elle est uniquement minimisée lorsque  $\alpha = \alpha_0$ ,  $\mu_\beta = \mu_{\beta,0}$  et  $\sigma^2 = \sigma^{2*} = \mathbb{E} \left[ \frac{(X_t - \alpha_0 - \mu_{\beta,0} X_{t-1})^2}{1 + \tau_0 X_{t-1}^2} \right]$  et  $\sigma_\beta^2 = \tau_0 \sigma^{2*}$ .

Mais

$$\begin{aligned}\sigma^{2*} &= \mathbb{E} \left\{ \mathbb{E} \left[ (X_t - \alpha_0 - \mu_{\beta,0} X_{t-1})^2 \mid \mathcal{F}_{t-1} \right] / (1 + \tau_0 X_{t-1}^2) \right\} \\ &= \mathbb{E} \left[ (\sigma_0^2 + \sigma_{\beta,0}^2 X_{t-1}^2) / (1 + \tau_0 X_{t-1}^2) \right] \\ &= \sigma_0^2\end{aligned}$$

Alors  $\lim_{n \rightarrow \infty} \tilde{l}_n(\alpha, \mu_\beta, \sigma_\beta^2, \sigma^2)$  est minimisée uniquement lorsque  $\alpha = \alpha_0$ ,  $\mu_\beta = \mu_{\beta,0}$ ,  $\sigma_\beta^2 = \sigma_{\beta,0}^2$  et  $\sigma^2 = \sigma_0^2$ .

Le théorème (2.5) est le résultat principal nécessaire dans la démonstration du théorème suivant, qui énonce la consistance forte des estimateurs du maximum de vraisemblance.

**Théorème 2.6.** [32] *Soit  $l_n(\alpha, \mu_\beta, \tau)$  minimisé sur  $\Theta^*$  en  $\alpha = \hat{\alpha}_n$ ,  $\mu_\beta = \hat{\mu}_{\beta,n}$  et  $\tau = \hat{\tau}_n$  où  $\Theta^*$  est défini dans le théorème (2.5). Soit  $\hat{\pi}_n = (\hat{\alpha}_n, \hat{\mu}_{\beta,n}, \hat{\tau}_n)'$ . Alors  $\hat{\pi}_n$  converge presque sûrement vers  $\pi_0 = (\alpha_0, \mu_{\beta,0}, \tau_0)$  à condition que  $\pi_0 \in \text{int}(\Theta^*)$ . De plus  $\hat{\sigma}_{\beta,n}^2$  et  $\hat{\sigma}_n^2$  convergent presque sûrement vers  $\sigma_\beta^2$  et  $\sigma^2$  respectivement.*

### 2.4.2.3 Le théorème central limite

Tout d'abord, nous présentons une version du théorème central limite pour les martingales, comme indiqué dans Billingsley(1961)[6], qui sera utile pour dériver les distributions asymptotiques des estimateurs du maximum de vraisemblance pour un modèle RCAR(1).

**Théorème 2.7.** [6]

*Soit  $\{\xi_t\}$  une suite de variables aléatoires ayant la propriété que  $\xi_t$  peut être exprimée comme une fonction qui ne dépend pas de  $t$ , qui est mesurable par rapport à  $\mathcal{F}_t$  générée par une suite  $\{\alpha_t, \alpha_{t-1}, \dots\}$  de variables aléatoires strictement stationnaires et ergodiques. De plus, supposons que  $\mathbb{E}(\xi_t \mid \mathcal{F}_{t-1}) = 0$  et  $\mathbb{E}(\xi_t^2) = c^2 < \infty$  alors  $(c_n^2)^{-\frac{1}{2}} \sum_{t=1}^n \xi_t$  converge en distribution vers une variable aléatoire normale standard.*

**Lemme 2.3.** *Soit  $f_n(\theta)$  une suite de fonctions continues et dérivables sur un compact  $\Theta$ . Si  $\left| \limsup_{n \rightarrow \infty} \sup_{\theta \in \Theta} \frac{\partial f_n(\theta)}{\partial \theta} \right| < \infty$ , alors  $f_n(\theta)$  est équicontinue sur  $\Theta$ .*

**Démonstration.** Nous devons montrer que pour tout  $\epsilon > 0$ , il existe un entier  $N$  et un  $\delta > 0$ , qui dépend de  $\epsilon$ , tels que  $|f_n(\theta_1) - f_n(\theta_2)| < \epsilon$  pour  $n > N$  lorsque  $\|\theta_1 - \theta_2\| < \delta$ .

Puisque  $f_n(\theta)$  est continue et dérivable sur  $\Theta$ , par le théorème des accroissements finis, nous avons pour  $\theta_1, \theta_2 \in \Theta$  :

$$f_n(\theta_1) - f_n(\theta_2) = (\theta_1 - \theta_2)' \frac{\partial f_n(\theta^*)}{\partial \theta}$$

où

$$\theta^* = \lambda \theta_1 + (1 - \lambda) \theta_2 \quad \text{pour un certain } \lambda \in (0, 1)$$

Maintenant, étant donné que

$$\left| (\theta_1 - \theta_2)' \frac{\partial f_n(\theta^*)}{\partial \theta} \right|^2 \leq \|\theta_1 - \theta_2\|^2 \left\| \frac{\partial f_n(\theta^*)}{\partial \theta} \right\|^2$$

Il s'ensuit que  $f_n(\theta)$  est équicontinue si

$$\limsup_{n \rightarrow \infty} \sup_{\theta \in \Theta} \left\| \frac{\partial f_n(\theta)}{\partial \theta} \right\| < \infty$$

Le lemme mentionné ci-dessus peut être utilisé pour montrer l'équicontinuité de la deuxième dérivée de  $\tilde{l}_n(\theta)$ ,  $\left\{ \frac{\partial^2 \tilde{l}_n(\theta)}{\partial \theta \partial \theta'} \right\}$ , sur un voisinage compact  $N(\theta_0)$  de  $\theta_0$ .

En suivant ce lemme, il suffit de montrer que chaque élément de la troisième dérivée de  $\tilde{l}_n(\theta)$  est uniformément borné sur un voisinage de  $\theta_0$ .

Par exemple, pour

$$\frac{\partial^2 \tilde{l}_n(\theta)}{\partial \mu_\beta^2} = 2n^{-1} \sum_{t=1}^n \lambda_t^{-1} X_{t-1}^2$$

où  $\lambda_t = \sigma^2 + \sigma_\beta^2 X_{t-1}^2$ , nous avons :

$$\begin{aligned} \left| \frac{\partial^3 \tilde{l}_n(\theta)}{\partial \mu_\beta^2 \partial \sigma^2} \right| &= 2n^{-1} \left| \sum_{t=1}^n \lambda_t^{-2} X_{t-1}^2 \right| \\ &\leq 2n^{-1} \sum_{t=1}^n |\lambda_t^{-1} X_{t-1}^2| |\lambda_t^{-1}| \\ &\leq \frac{2}{\sigma_\beta^2 \sigma^2} \\ &< \infty \end{aligned}$$

On procède de la même manière pour les autres éléments de la dérivée de  $\left\{ \frac{\partial^2 \tilde{l}_n(\theta)}{\partial \theta \partial \theta'} \right\}$ .

**Lemme 2.4.** Soit  $\{\theta_n\}$  une suite de variables aléatoires telle que  $\theta_n$  converge presque sûrement vers une constante  $\theta_0$ . Soit  $\{f_n(\cdot)\}$  une suite de fonctions telles que  $f_n(\theta)$  converge uniformément presque sûrement vers une fonction continue  $f(\theta)$  sur un voisinage  $N(\theta_0)$  de  $\theta_0$ . Alors  $f_n(\theta_n)$  converge presque sûrement vers  $f(\theta_0)$ .

**Démonstration.** Soit  $\omega$  tel que  $\theta_n(\omega)$  converge vers  $\theta_0(\omega)$ . Alors, pour  $n$  suffisamment grand, nous avons  $\theta_n(\omega) \in N(\theta_0)$ . Maintenant, fixons  $\omega$ . Nous avons alors :

$$\begin{aligned} |f_n(\theta_n) - f(\theta_0)| &= |f_n(\theta_n) - f(\theta_n) + f(\theta_n) - f(\theta_0)| \\ &\leq |f_n(\theta_n) - f(\theta_n)| + |f(\theta_n) - f(\theta_0)| \end{aligned}$$

Puisque  $f_n(\cdot)$  converge uniformément vers  $f(\cdot)$ , alors pour  $n \geq N$ , nous avons

$$|f_n(\theta_n(\omega)) - f(\theta_n(\omega))| < \frac{\epsilon}{2}$$

$f(\cdot)$  est continue sur  $N(\theta_0)$ , donc il existe un entier  $N \in \mathbb{N}$  tel que si  $n \geq N$ , alors

$$|f_n(\theta_n) - f(\theta_0)| < \epsilon$$

Donc  $f_n(\theta_n)$  converge presque sûrement vers  $f(\theta_0)$ .

**Lemme 2.5.** La suite  $\left\{ \frac{\partial^2 \tilde{l}_n(\theta)}{\partial \theta \partial \theta'} \right\}$  converge uniformément et presque sûrement sur un compact au voisinage  $N(\theta_0)$  de  $\theta_0$  vers  $\left\{ \frac{\partial^2 \tilde{l}(\theta)}{\partial \theta \partial \theta'} \right\}$ , où  $\tilde{l}(\theta) = \lim_{n \rightarrow \infty} \tilde{l}_n(\theta)$ .

**Démonstration.** Tout d'abord, définissons

$$\lambda_t = \sigma^2 + \sigma_\beta^2 X_{t-1}^2 \quad \text{et} \quad u_t = X_t - \alpha - \mu_\beta X_{t-1}$$

Nous avons

$$\begin{aligned} \mathbb{E}(u_t | \mathcal{F}_{t-1}) &= \mathbb{E}[(X_t - \alpha_0 - \mu_{\beta,0} X_{t-1}) + (\alpha_0 - \alpha) + (\mu_{\beta,0} - \mu_\beta) X_{t-1} | \mathcal{F}_{t-1}] \\ &= (\alpha_0 - \alpha) + (\mu_{\beta,0} - \mu_\beta) X_{t-1} \end{aligned}$$

et

$$\begin{aligned} \mathbb{E}(u_t^2 | \mathcal{F}_{t-1}) &= \mathbb{E} \left\{ [(X_t - \alpha_0 - \mu_{\beta,0} X_{t-1}) + (\alpha_0 - \alpha) + (\mu_{\beta,0} - \mu_\beta) X_{t-1}]^2 | \mathcal{F}_{t-1} \right\} \\ &= \sigma_0^2 + \sigma_{\beta,0}^2 X_{t-1}^2 + [(\alpha_0 - \alpha) + (\mu_{\beta,0} - \mu_\beta) X_{t-1}]^2 \end{aligned}$$

Les dérivées secondes de  $\tilde{l}_n(\theta)$  sont données élément par élément par :

$$\begin{aligned} \frac{\partial^2 \tilde{l}_n(\theta)}{\partial \alpha^2} &= 2n^{-1} \sum_{t=1}^n \lambda_t^{-1} \\ \frac{\partial^2 \tilde{l}_n(\theta)}{\partial \alpha \partial \mu_\beta} &= 2n^{-1} \sum_{t=1}^n \lambda_t^{-1} X_{t-1} \\ \frac{\partial^2 \tilde{l}_n(\theta)}{\partial \alpha \partial \sigma_\beta^2} &= 2n^{-1} \sum_{t=1}^n \lambda_t^{-2} u_t X_{t-1}^2 \\ \frac{\partial^2 \tilde{l}_n(\theta)}{\partial \alpha \partial \sigma^2} &= 2n^{-1} \sum_{t=1}^n \lambda_t^{-2} u_t \\ \frac{\partial^2 \tilde{l}_n(\theta)}{\partial \mu_\beta^2} &= 2n^{-1} \sum_{t=1}^n \lambda_t^{-1} X_{t-1}^2 \\ \frac{\partial^2 \tilde{l}_n(\theta)}{\partial \mu_\beta \partial \sigma_\beta^2} &= 2n^{-1} \sum_{t=1}^n \lambda_t^{-2} u_t X_{t-1}^3 \\ \frac{\partial^2 \tilde{l}_n(\theta)}{\partial \mu_\beta \partial \sigma^2} &= 2n^{-1} \sum_{t=1}^n \lambda_t^{-2} u_t X_{t-1} \\ \frac{\partial^2 \tilde{l}_n(\theta)}{\partial (\sigma_\beta^2)^2} &= 2n^{-1} \sum_{t=1}^n \lambda_t^{-3} u_t^2 X_{t-1}^4 - n^{-1} \sum_{t=1}^n \lambda_t^{-2} X_{t-1}^4 \\ \frac{\partial^2 \tilde{l}_n(\theta)}{\partial \sigma_\beta^2 \partial \sigma^2} &= 2n^{-1} \sum_{t=1}^n \lambda_t^{-3} u_t^2 X_{t-1}^2 - n^{-1} \sum_{t=1}^n \lambda_t^{-2} X_{t-1}^2 \\ \frac{\partial^2 \tilde{l}_n(\theta)}{\partial (\sigma^2)^2} &= 2n^{-1} \sum_{t=1}^n \lambda_t^{-3} u_t^2 - n^{-1} \sum_{t=1}^n \lambda_t^{-2} \end{aligned}$$

Sous l'hypothèse A5 (énoncée dans la section (2.4.2)), nous pouvons borner l'espérance de chacun des termes ci-dessus en utilisant le lemme (2.2). Par conséquent, en utilisant le théorème de l'ergodicité (1.2),  $\left\{ \frac{\partial^2 \tilde{l}_n(\theta)}{\partial \theta \partial \theta'} \right\}$  converge presque sûrement vers

une matrice qui est égale à  $\frac{\partial^2 \tilde{l}(\theta)}{\partial \theta \partial \theta'}$ , et est donnée élément par élément par :

$$\begin{aligned} \frac{\partial^2 \tilde{l}(\theta)}{\partial \alpha^2} &= 2\mathbb{E}(\lambda_t^{-1}) \\ \frac{\partial^2 \tilde{l}(\theta)}{\partial \alpha \partial \mu_\beta} &= 2\mathbb{E}(\lambda_t^{-1} X_{t-1}) \\ \frac{\partial^2 \tilde{l}(\theta)}{\partial \alpha \partial \sigma_\beta^2} &= 2\mathbb{E}(\lambda_t^{-2} u_t X_{t-1}^2) \\ \frac{\partial^2 \tilde{l}(\theta)}{\partial \alpha \partial \sigma^2} &= 2\mathbb{E}(\lambda_t^{-2} u_t) \\ \frac{\partial^2 \tilde{l}(\theta)}{\partial \mu_\beta^2} &= 2\mathbb{E}(\lambda_t^{-1} X_{t-1}^2) \\ \frac{\partial^2 \tilde{l}(\theta)}{\partial \mu_\beta \partial \sigma_\beta^2} &= 2\mathbb{E}(\lambda_t^{-2} u_t X_{t-1}^3) \\ \frac{\partial^2 \tilde{l}(\theta)}{\partial \mu_\beta \partial \sigma^2} &= 2\mathbb{E}(\lambda_t^{-2} u_t X_{t-1}) \\ \frac{\partial^2 \tilde{l}(\theta)}{\partial (\sigma_\beta^2)^2} &= 2\mathbb{E}(\lambda_t^{-3} u_t^2 X_{t-1}^4) - \mathbb{E}(\lambda_t^{-2} X_{t-1}^4) \\ \frac{\partial^2 \tilde{l}(\theta)}{\partial \sigma_\beta^2 \partial \sigma^2} &= 2\mathbb{E}(\lambda_t^{-3} u_t^2 X_{t-1}^2) - \mathbb{E}(\lambda_t^{-2} X_{t-1}^2) \\ \frac{\partial^2 \tilde{l}(\theta)}{\partial (\sigma^2)^2} &= 2\mathbb{E}(\lambda_t^{-3} u_t^2) - \mathbb{E}(\lambda_t^{-2}) \end{aligned}$$

De plus, par le lemme (2.3),  $\frac{\partial^2 \tilde{l}_n(\theta)}{\partial \theta \partial \theta'}$  est équicontinue sur un voisinage compact  $N(\theta_0)$  de  $\theta_0$ , d'où  $\left\{ \frac{\partial^2 \tilde{l}_n(\theta)}{\partial \theta \partial \theta'} \right\}$  converge uniformément et presque sûrement sur un voisinage de  $\theta_0$  vers  $\left\{ \frac{\partial^2 \tilde{l}(\theta)}{\partial \theta \partial \theta'} \right\}$ .

**Lemme 2.6.** *Pour toute suite  $\{\theta_n\}$  qui converge presque sûrement vers  $\theta_0$ , nous avons  $\left\{ \frac{\partial^2 \tilde{l}_n(\theta_n)}{\partial \theta \partial \theta'} \right\}$  qui converge presque sûrement vers  $\frac{\partial^2 \tilde{l}(\theta_0)}{\partial \theta \partial \theta'} = V$ , où  $V$  est une*

matrice symétrique et donnée par :

$$\begin{pmatrix} V_{11} & V_{12} & V_{13} & V_{14} \\ & V_{22} & V_{23} & V_{24} \\ & & V_{33} & V_{34} \\ & & & V_{44} \end{pmatrix} = \begin{pmatrix} 2\mathbb{E}(\lambda_{t,0}^{-1}) & 2\mathbb{E}(\lambda_{t,0}^{-1}X_{t-1}) & 0 & 0 \\ & 2\mathbb{E}(\lambda_{t,0}^{-1}X_{t-1}^2) & 0 & 0 \\ & & \mathbb{E}(\lambda_{t,0}^{-2}X_{t-1}^4) & \mathbb{E}(\lambda_{t,0}^{-2}X_{t-1}^2) \\ & & & \mathbb{E}(\lambda_{t,0}^{-2}) \end{pmatrix} \quad (2.10)$$

où

$$\lambda_{t,0} = \sigma_0^2 + \sigma_{\beta,0}^2 X_{t-1}^2$$

**Démonstration.** En conséquence du lemme (2.4) et du lemme (2.5), nous avons que  $\left\{ \frac{\partial^2 \tilde{l}_n(\theta_n)}{\partial \theta \partial \theta'} \right\}$  converge presque sûrement vers  $\frac{\partial^2 \tilde{l}(\theta_0)}{\partial \theta \partial \theta'} = V$ , tel que chaque élément  $V_{ij}$  de  $V$  est donné par :

$$V_{11} = \frac{\partial^2 \tilde{l}(\theta_0)}{\partial \alpha^2} = 2\mathbb{E}(\lambda_{t,0}^{-1})$$

$$V_{12} = V_{21} = \frac{\partial^2 \tilde{l}(\theta_0)}{\partial \alpha \partial \mu_\beta} = 2\mathbb{E}(\lambda_{t,0}^{-1} X_{t-1})$$

$$V_{13} = V_{31} = \frac{\partial^2 \tilde{l}(\theta_0)}{\partial \alpha \partial \sigma_\beta^2} = 2\mathbb{E}(\lambda_{t,0}^{-2} u_{t,0} X_{t-1}^2) = 2\mathbb{E}[\lambda_{t,0}^{-2} X_{t-1}^2 \mathbb{E}(u_{t,0} | \mathcal{F}_{t-1})] = 0$$

$$V_{14} = V_{41} = \frac{\partial^2 \tilde{l}(\theta_0)}{\partial \alpha \partial \sigma^2} = 2\mathbb{E}(\lambda_{t,0}^{-2} u_{t,0}) = 0$$

$$V_{22} = \frac{\partial^2 \tilde{l}(\theta_0)}{\partial \mu_\beta^2} = 2\mathbb{E}(\lambda_{t,0}^{-1} X_{t-1}^2)$$

$$V_{23} = V_{32} = \frac{\partial^2 \tilde{l}(\theta_0)}{\partial \mu_\beta \partial \sigma_\beta^2} = 2\mathbb{E}(\lambda_{t,0}^{-2} u_{t,0} X_{t-1}^3) = 0$$

$$V_{24} = V_{42} = \frac{\partial^2 \tilde{l}(\theta_0)}{\partial \mu_\beta \partial \sigma^2} = 2\mathbb{E}(\lambda_{t,0}^{-2} u_{t,0} x_{t-1}) = 0$$

$$\begin{aligned} V_{33} &= \frac{\partial^2 \tilde{l}(\theta_0)}{\partial(\sigma_\beta^2)^2} = 2\mathbb{E}(\lambda_{t,0}^{-3} u_{t,0}^2 X_{t-1}^4) - \mathbb{E}(\lambda_{t,0}^{-2} X_{t-1}^4) \\ &= 2\mathbb{E}[\lambda_{t,0}^{-3} X_{t-1}^4 \mathbb{E}(u_{t,0}^2 | \mathcal{F}_{t-1})] - \mathbb{E}(\lambda_{t,0}^{-2} X_{t-1}^4) = \mathbb{E}(\lambda_{t,0}^{-2} X_{t-1}^4) \end{aligned}$$

$$V_{34} = V_{43} = \frac{\partial^2 \tilde{l}(\theta_0)}{\partial\sigma_\beta^2 \partial\sigma^2} = 2\mathbb{E}(\lambda_{t,0}^{-3} u_{t,0}^2 X_{t-1}^2) - \mathbb{E}(\lambda_{t,0}^{-2} X_{t-1}^2) = \mathbb{E}(\lambda_{t,0}^{-2} X_{t-1}^2)$$

$$V_{44} = \frac{\partial^2 \tilde{l}(\theta_0)}{\partial(\sigma^2)^2} = 2\mathbb{E}(\lambda_{t,0}^{-3} t u_{t,0}^2) - \mathbb{E}(\lambda_{t,0}^{-2}) = \mathbb{E}(\lambda_{t,0}^{-2})$$

Nous utilisons le fait que :

$$\mathbb{E}(u_{t,0} | \mathcal{F}_{t-1}) = 0 \quad \text{et} \quad \mathbb{E}(u_{t,0}^2 | \mathcal{F}_{t-1}) = \lambda_{t,0}$$

où  $u_{t,0} = X_t - \alpha_0 - \mu_{\beta,0} X_{t-1}$

**Théorème 2.8.** *Soit  $\{X_t, t \in \mathbb{Z}\}$  une suite strictement stationnaire, ergodique et une solution  $\mathcal{F}_t$  mesurable pour l'équation (2.1) sous les hypothèses A1 – A6 (énoncées dans la section (2.2.2) et (2.4.2)). Alors  $\sqrt{n}(\hat{\theta}_n - \theta_0)$  converge vers une loi normale de moyenne nulle et de matrice de covariance  $V^{-1} W V^{-1}$ , où  $V$  est donnée dans (2.10) et  $W$  est une matrice symétrique ses éléments sont donnés par :*

$$\begin{aligned} W_{11} &= 4\mathbb{E}(\lambda_{t,0}^{-1}) \quad , \quad W_{12} = 4\mathbb{E}(\lambda_{t,0}^{-1} X_{t-1}) \\ W_{13} &= 2\mathbb{E}(u_{t,0} \lambda_{t,0}^{-3} X_{t-1}^2 \eta_{t,0}) \quad , \quad W_{14} = 2\mathbb{E}(u_{t,0} \lambda_{t,0}^{-3} \eta_{t,0}) \\ W_{22} &= 4\mathbb{E}(\lambda_{t,0}^{-1} X_{t-1}^2) \quad , \quad W_{23} = 2\mathbb{E}(u_{t,0} \lambda_{t,0}^{-3} X_{t-1}^3 \eta_{t,0}) \\ W_{24} &= 2\mathbb{E}(u_{t,0} \lambda_{t,0}^{-3} X_{t-1} \eta_{t,0}) \quad , \quad W_{33} = \mathbb{E}(\lambda_{t,0}^{-4} X_{t-1}^4 \eta_{t,0}^2) \\ W_{34} &= \mathbb{E}(\lambda_{t,0}^{-4} X_{t-1}^2 \eta_{t,0}^2) \quad , \quad W_{44} = \mathbb{E}(\lambda_{t,0}^{-2} \eta_{t,0}^2) \end{aligned}$$

où

$$\lambda_{t,0} = \sigma_0^2 + \sigma_{\beta,0}^2 X_{t-1}^2 \quad , \quad u_{t,0} = X_t - \alpha_0 - \mu_{\beta,0} X_{t-1} \quad \text{et} \quad \eta_{t,0} = u_{t,0}^2 - \lambda_{t,0}$$

De plus, si  $\{\beta_t\}$  et  $\{\epsilon_t\}$  sont conjointement normales, alors la matrice de covariance se réduit à  $2V^{-1}$ .

**Démonstration.** Nous montrons d'abord que la matrice  $V$  est définie positive. Puisque  $V$  est diagonale par bloc, il suffit de montrer que chaque bloc est défini

positif.

Considérons la sous-matrice :

$$\begin{aligned} \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix} &= \begin{pmatrix} 2\mathbb{E}(\lambda_{t,0}^{-1}) & 2\mathbb{E}(\lambda_{t,0}^{-1}X_{t-1}) \\ 2\mathbb{E}(\lambda_{t,0}^{-1}X_{t-1}) & 2\mathbb{E}(\lambda_{t,0}^{-1}X_{t-1}^2) \end{pmatrix} \\ &= \mathbb{E} \left[ 2\lambda_{t,0}^{-1} \begin{pmatrix} 1 \\ X_{t-1} \end{pmatrix} \begin{pmatrix} 1 & X_{t-1} \end{pmatrix} \right] = \mathbb{E}(zz') \end{aligned}$$

où

$$z = (\sqrt{2}\lambda_{t,0}^{-\frac{1}{2}}, \sqrt{2}\lambda_{t,0}^{-\frac{1}{2}}X_{t-1})'$$

Pour tout vecteur  $a = (a_1, a_2)'$ , nous avons

$$\mathbb{E}(a'zz'a) = \mathbb{E}[2\lambda_{t,0}^{-1}(a_1 + a_2X_{t-1})^2] \geq 0$$

Et l'égalité est vérifiée uniquement lorsque  $a_1 = a_2 = 0$  selon le lemme (2.1).

En suivant la même méthode, il est facile de montrer que la sous-matrice.

$$\begin{pmatrix} V_{33} & V_{34} \\ V_{43} & V_{44} \end{pmatrix}$$

est définie positive. Ainsi,  $V$  est définie positive.

Ensuite, rappelons que dans le lemme (2.5) nous avons montré que la dérivée seconde de  $\{\frac{\partial^2 \tilde{l}_n(\theta)}{\partial \theta \partial \theta'}\}$  converge uniformément et presque sûrement sur un voisinage compact  $N(\theta_0)$  de  $\theta_0$  vers  $\frac{\partial^2 \tilde{l}(\theta)}{\partial \theta \partial \theta'}$ .

Le développement de Taylor de  $\frac{\partial^2 \tilde{l}_n(\hat{\theta}_n)}{\partial \theta}$  autour de  $\theta_0$  donne que

$$\frac{\partial^2 \tilde{l}_n(\hat{\theta}_n)}{\partial \theta} = \frac{\partial^2 \tilde{l}_n(\theta_0)}{\partial \theta} + \frac{\partial^2 \tilde{l}_n(\tilde{\theta}_n)}{\partial \theta \partial \theta'}(\hat{\theta}_n - \theta_0)$$

où  $\tilde{\theta}_n$  est sur le segment de droite entre  $\hat{\theta}_n$  et  $\theta_0$ .

Puisque  $\hat{\theta}_n$  converge presque sûrement vers  $\theta_0$ , alors  $\tilde{\theta}_n$  converge presque sûrement vers  $\theta_0$ . Par conséquent, puisque  $\{\frac{\partial^2 \tilde{l}(\theta)}{\partial \theta \partial \theta'}\}$  converge uniformément sur  $N(\theta_0)$ ,  $\frac{\partial^2 \tilde{l}(\tilde{\theta}_n)}{\partial \theta \partial \theta'}$

converge presque sûrement vers la matrice  $V = \frac{\partial^2 \tilde{l}(\theta_0)}{\partial \theta \partial \theta'}$  d'après le lemme (2.6).

Puisque  $\hat{\theta}_n$  minimise  $\tilde{l}_n(\theta)$ , nous avons  $\frac{\partial \tilde{l}_n(\hat{\theta}_n)}{\partial \theta} = 0$ .

Si nous supposons que  $\sqrt{n} \frac{\partial \tilde{l}(\theta_0)}{\partial \theta}$  converge vers une distribution limite de moyenne nulle et de matrice de covariance  $W$ , alors  $\sqrt{n}(\hat{\theta}_n - \theta_0)$  aura la même distribution asymptotique que  $-V^{-1} \sqrt{n} \frac{\partial \tilde{l}(\theta_0)}{\partial \theta}$ . Par conséquent, d'après le théorème de Slutsky,  $\sqrt{n}(\hat{\theta}_n - \theta_0)$  converge vers une distribution normale limite avec une moyenne nulle et une matrice de covariance  $V^{-1} W V^{-1}$ .

Il reste à montrer que  $\sqrt{n} \frac{\partial \tilde{l}_n(\theta_0)}{\partial \theta}$  est asymptotiquement normal avec une moyenne nulle et une matrice de covariance  $W$ .

Rappelons que :

$$\tilde{l}_n(\theta_0) = n^{-1} \sum_{t=1}^n \ln(\sigma_0^2 + \sigma_{\beta,0}^2 X_{t-1}^2) + n^{-1} \sum_{t=1}^n \frac{(X_t - \alpha_0 - \mu_{\beta,0} X_{t-1})^2}{\sigma_0^2 + \sigma_{\beta,0}^2 X_{t-1}^2}$$

Il s'ensuit que :

$$\begin{aligned} \frac{\partial \tilde{l}_n(\theta_0)}{\partial \alpha} &= -2n^{-1} \sum_{t=1}^n \lambda_{t,0}^{-1} u_{t,0} \\ \frac{\partial \tilde{l}_n(\theta_0)}{\partial \mu_{\beta}} &= -2n^{-1} \sum_{t=1}^n \lambda_{t,0}^{-1} u_{t,0} X_{t-1} \\ \frac{\partial \tilde{l}_n(\theta_0)}{\partial \sigma_{\beta}^2} &= n^{-1} \sum_{t=1}^n \lambda_{t,0}^{-1} X_{t-1}^2 - n^{-1} \sum_{t=1}^n \lambda_{t,0}^{-2} u_{t,0}^2 X_{t-1}^2 \\ \frac{\partial \tilde{l}_n(\theta_0)}{\partial \sigma^2} &= n^{-1} \sum_{t=1}^n \lambda_{t,0}^{-1} - n^{-1} \sum_{t=1}^n \lambda_{t,0}^{-2} u_{t,0}^2 \end{aligned}$$

Soit

$$\xi_t(a) = \lambda_{t,0}^{-2} [2u_{t,0} \lambda_{t,0} (a_1 + a_2 X_{t-1}) + \eta_{t,0} (a_4 + a_3 X_{t-1}^2)]$$

où  $a = (a_1, a_2, a_3, a_4)'$ , alors il est facile de voir que

$$n^{-1} \sum_{t=1}^n \xi_t(a) = -a' \frac{\partial \tilde{l}_n(\theta_0)}{\partial \theta}$$

Maintenant, étant donné que

$$\begin{aligned} \mathbb{E}(u_{t,0} | \mathcal{F}_{t-1}) &= \mathbb{E}(X_t - \alpha_0 - \mu_{\beta,0} X_{t-1} | \mathcal{F}_{t-1}) = 0 \\ \mathbb{E}(u_{t,0}^2 | \mathcal{F}_{t-1}) &= \mathbb{E}[(X_t - \alpha_0 - \mu_{\beta,0} X_{t-1})^2 | \mathcal{F}_{t-1}] = \lambda_{t,0} \end{aligned}$$

Alors nous avons

$$\mathbb{E}(\eta_{t,0}|\mathcal{F}_{t-1}) = \mathbb{E}(u_{t,0}^2 - \lambda_{t,0}|\mathcal{F}_{t-1}) = 0$$

Donc

$$\mathbb{E}[\xi_t(a)|\mathcal{F}_{t-1}] = 0$$

Par définition,  $\xi_t(a)$  est strictement stationnaire et ergodique, et nous verrons toute suite  $\mathbb{E}[\xi_t^2(a)]$  est finie.

Par conséquent, par le théorème (2.7),  $n^{-1/2} \sum_{t=1}^n \xi_t(a)$  converge vers une distribution limite gaussienne avec une moyenne nulle et une variance de  $\mathbb{E}[\xi_t^2(a)]$ .

Cette variance peut être exprimée sous la forme  $a'Wa$  où  $W$  est symétrique et ne dépend pas de  $a$ . Ainsi,  $\sqrt{n} \frac{\partial \bar{l}_n(\theta_0)}{\partial \theta}$  converge vers une distribution normale avec une moyenne nulle et une matrice de covariance  $W$ .

Il reste à montrer que  $\mathbb{E}[\xi_t^2(a)]$  est finie pour chaque  $a$ .

Soit

$$\begin{aligned} \xi_t^2(a) &= 4\lambda_{t,0}^{-2}u_{t,0}^2(a_1 + a_2X_{t-1})^2 + \lambda_{t,0}^{-4}(a_4 + a_3X_{t-1}^2)^2\eta_{t,0}^2 \\ &\quad + 4\lambda_{t,0}^{-3}u_{t,0}(a_1 + a_2X_{t-1})(a_4 + a_3X_{t-1}^2)\eta_{t,0} \end{aligned} \quad (2.11)$$

Il est facile de voir que les trois termes dans (2.11) ont des espérances bornées .

Par exemple, pour le premier terme, selon le lemme (2.2)

$$\begin{aligned} \mathbb{E} \left| 4\lambda_{t,0}^{-2}u_{t,0}^2(a_1 + a_2X_{t-1})^2 \right| &= \mathbb{E} \left| 4\lambda_{t,0}^{-2}(a_1 + a_2X_{t-1})^2 \mathbb{E}(u_{t,0}^2|\mathcal{F}_{t-1}) \right| \\ &= 4\mathbb{E} \left| \lambda_{t,0}^{-1}(a_1 + a_2X_{t-1})^2 \right| \\ &< \infty \end{aligned}$$

Montrons que le deuxième terme est fini :

Comme  $(u_{t,0}|\mathcal{F}_{t-1}) \sim \mathcal{N}(0, \lambda_{t,0})$ , alors  $\mathbb{E}(u_{t,0}^4|\mathcal{F}_{t-1}) = 3\lambda_{t,0}^2$ .

$$\begin{aligned} \mathbb{E}(\eta_{t,0}^2|\mathcal{F}_{t-1}) &= \mathbb{E}[(u_{t,0}^2 - \lambda_{t,0})^2] \\ &= \mathbb{E}(u_{t,0}^4 + \lambda_{t,0}^2 - 2\lambda_{t,0}u_{t,0}^2|\mathcal{F}_{t-1}) \\ &= \mathbb{E}(u_{t,0}^4|\mathcal{F}_{t-1}) + \lambda_{t,0}^2 - 2\lambda_{t,0}\mathbb{E}(u_{t,0}^2|\mathcal{F}_{t-1}) \\ &= \mathbb{E}(u_{t,0}^4|\mathcal{F}_{t-1}) - \lambda_{t,0}^2 \\ &= 2\lambda_{t,0}^2 \end{aligned}$$

Nous avons

$$\begin{aligned} \mathbb{E} \left| \lambda_{t,0}^{-4} (a_4 + a_3 X_{t-1}^2)^2 \eta_{t,0}^2 \right| &= \mathbb{E} \left| \frac{(a_4 + a_3 X_{t-1}^2)^2}{(\sigma_0^2 + \sigma_{\beta,0}^2 X_{t-1}^2)^2} \right| \\ &\leq \left( \frac{|a_3|}{\sigma_{\beta,0}^2} + \frac{|a_4|}{\sigma_0^2} \right)^2 \\ &< \infty \end{aligned}$$

Par la même méthode, nous montrons que le troisième terme est fini.

Par conséquent,  $\mathbb{E}[\xi_t^2(a)] < \infty$

$$\begin{aligned} \mathbb{E}(\xi_t^2(a)) &= 4\mathbb{E}(\lambda_{t,0}^{-2} u_{t,0}^2 (a_1 + a_2 X_{t-1})^2) + \mathbb{E}(\lambda_{t,0}^{-4} (a_4 + a_3 X_{t-1}^2)^2 \eta_{t,0}^2) \\ &\quad + 4\mathbb{E}(\lambda_{t,0}^{-3} u_{t,0} (a_1 + a_2 X_{t-1}) (a_4 + a_3 X_{t-1}^2) \eta_{t,0}) \end{aligned}$$

La matrice  $W$  peut être obtenue en exprimant  $\mathbb{E}[\xi_t^2(a)]$  dans l'équation (2.11), sous la forme d'une matrice  $a'Wa$ , c'est-à-dire, chaque élément  $W_{ij}$ ,  $i, j = 1, 2, 3$  et 4, peut être trouvé en évaluant la composante de  $\mathbb{E}[\xi_t^2(a)]$  de la forme  $a_i W_{ij} a_j$ .

En notant que  $\mathbb{E}(u_{t,0} | \mathcal{F}_{t-1}) = 0$  et  $\mathbb{E}(u_{t,0}^2 | \mathcal{F}_{t-1}) = \lambda_{t,0}$ , nous avons :

$$\begin{aligned} W_{11} &= 4\mathbb{E}(u_{t,0}^2 \lambda_{t,0}^{-2}) = \mathbb{E}(\lambda_{t,0}^{-1}) \\ W_{12} &= W_{21} = 4\mathbb{E}(u_{t,0}^2 \lambda_{t,0}^{-2} X_{t-1}) = 4\mathbb{E}(\lambda_{t,0}^{-1} X_{t-1}) \\ W_{13} &= W_{31} = 2\mathbb{E}(u_{t,0} \lambda_{t,0}^{-3} X_{t-1}^2 \eta_{t,0}) \\ W_{14} &= W_{41} = 2\mathbb{E}(u_{t,0} \lambda_{t,0}^{-3} \eta_{t,0}) \\ W_{22} &= 4\mathbb{E}(u_{t,0}^2 \lambda_{t,0}^{-2} X_{t-1}^2) = 4\mathbb{E}(\lambda_{t,0}^{-1} X_{t-1}^2) \\ W_{23} &= W_{32} = 2\mathbb{E}(u_{t,0} \lambda_{t,0}^{-3} X_{t-1}^3 \eta_{t,0}) \\ W_{24} &= W_{42} = 2\mathbb{E}(u_{t,0} \lambda_{t,0}^{-3} X_{t-1}^2 \eta_{t,0}) \\ W_{33} &= \mathbb{E}(\lambda_{t,0}^{-4} X_{t-1}^4 \eta_{t,0}^2) \\ W_{34} &= W_{43} = \mathbb{E}(\lambda_{t,0}^{-4} X_{t-1}^2 \eta_{t,0}^2) \\ W_{44} &= \mathbb{E}(\lambda_{t,0}^{-4} \eta_{t,0}^2) \end{aligned} \tag{2.12}$$

Si  $\beta_t$  et  $\epsilon_t$  sont conjointement normales, alors  $u_{t,0}|\mathcal{F}_{t-1} \sim \mathcal{N}(0, \lambda_{t,0})$ , donc

$$\mathbb{E}(u_{t,0}\eta_{t,0}|\mathcal{F}_{t-1}) = \mathbb{E}(u_{t,0}^3 - u_{t,0}\lambda_{t,0}|\mathcal{F}_{t-1}) = 0$$

$$\mathbb{E}(u_{t,0}^4|\mathcal{F}_{t-1}) = 3\lambda_{t,0}^2$$

et

$$\mathbb{E}(\eta_{t,0}^2|\mathcal{F}_{t-1}) = \mathbb{E}(u_{t,0}^4 - 2u_{t,0}^2\lambda_{t,0} + \lambda_{t,0}^2|\mathcal{F}_{t-1}) = 2\lambda_{t,0}^2$$

Par conséquent, l'expression de  $W$  dans (2.12) devient :

$$W_{11} = 4\mathbb{E}(\lambda_{t,0}^{-1})$$

$$W_{12} = W_{21} = 4\mathbb{E}(\lambda_{t,0}^{-1}X_{t-1})$$

$$W_{13} = W_{14} = W_{23} = W_{24} = 0$$

$$W_{22} = 4\mathbb{E}(\lambda_{t,0}^{-1}X_{t-1}^2)$$

$$W_{33} = \mathbb{E}(\lambda_{t,0}^{-4}X_{t-1}^4\eta_{t,0}^2) = 2\mathbb{E}(\lambda_{t,0}^{-2}X_{t-1}^4)$$

$$W_{34} = \mathbb{E}(\lambda_{t,0}^{-4}X_{t-1}^2\eta_{t,0}^2) = 2\mathbb{E}(\lambda_{t,0}^2X_{t-1}^2)$$

$$W_{44} = \mathbb{E}(\lambda_{t,0}^{-2}\eta_{t,0}^2) = 2\mathbb{E}(\lambda_{t,0}^2)$$

Nous remarquons que  $W = 2V$  sous la condition que  $\epsilon_t$  et  $\beta_t$  soient conjointement normales. Donc

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \sim \mathcal{N}(0, 2V^{-1})$$

## Application numérique

### 3.1 Introduction

Après avoir posé les bases théoriques et exploré les méthodes d'inférence statistique dans les chapitres précédents, nous entamons maintenant une phase essentielle de notre étude, qui est l'application pratique de ces concepts sur des données simulées.

Dans ce troisième chapitre, nous nous inspirerons des deux articles suivants : le premier, intitulé 'A Comparison of Parameter Estimation Methods for the First-Order of Random Coefficient Autoregressive Model', a été rédigé par Araveeporn Autcha et publié dans le 'Thailand Statistician' en 2022 [2]. Le second article, intitulé 'Bayesian Analysis of Random Coefficient Dynamic AutoRegressive Model', a été écrit par Araveeporn Autcha et publié dans le 'Thailand Statistician' en 2012 [3]. Nous allons appliquer les méthodes des moindres carrés (MC), maximum de vraisemblance (MV) et l'approche Bayésienne hiérarchique pour estimer deux paramètres inconnus dans un modèle autorégressif à coefficients aléatoires d'ordre 1, RCAR(1). Nous allons utiliser l'algorithme d'échantillonnage de Gibbs dans le cadre de l'approche Bayésienne hiérarchique pour générer des échantillons à partir de la distribution a posteriori. L'efficacité de ces méthodes a été évaluée en utilisant la moyenne de l'erreur quadratique moyenne (AMSE) comme critère de comparaison, révélant que la méthode des moindres carrés se comporte généralement mieux, à l'exception de certaines situations de données particulières. Nous allons également réaliser des simulations numériques en utilisant le logiciel R. Ces simulations ont permis de valider nos méthodes d'estimation dans un contexte pratique. Enfin, nous allons explorer le modèle autorégressif dynamique à coefficients aléatoires

noté RCDAR(p,q), qui est un cas particulier du RCAR(p). Ceci se fait en ajoutant des variables exogènes au modèle RCAR, conduisant à deux variables. Des simulations de Monte-Carlo ont été effectuées dans divers scénarios, notamment pour des données stationnaires, faiblement stationnaires, tendant vers la non-stationnarité et enfin la non-stationnarité. Les résultats ont montré que nos méthodes d'estimation étaient particulièrement adaptées aux échantillons de grande taille surtout dans le cas classique et aux données faiblement stationnaires. Ce chapitre constitue le passage crucial de la théorie à l'application pratique, nous permettant de mieux comprendre comment nos méthodes fonctionnent et de les comparer dans différents contextes.

## 3.2 Application sur le modèle RCAR(1)

Dans notre application, notre modèle s'inspire de l'article suivant : "A Comparison of Parameter Estimation Methods for the First-Order of Random Coefficient Autoregressive Model" écrit par Araveeporn, Autcha et publié dans le 'Thailand Statistician' en 2022[2]. Notre intérêt se porte sur l'estimation des paramètres  $\theta = (\alpha, \mu_\beta)$ , en supposant que  $\sigma_\epsilon^2$  et  $\sigma_\beta^2$  sont des valeurs constantes.

### 3.2.1 Présentation du modèle RCAR(1)

Un modèle autorégressif à coefficient aléatoire d'ordre 1, noté RCAR(1) est représenté sous la forme suivant :

$$\begin{aligned} X_t &= \alpha + \beta_t X_{t-1} + \epsilon_t, \quad t=1,2,3,\dots,n, \\ \text{et } \beta_t &= \mu_\beta + \sigma_\beta \nu_t \end{aligned}$$

où

- $\beta_t$  sont des variables aléatoires i.i.d. avec une moyenne  $\mu_\beta$  et une variance  $\sigma_\beta^2$ .
- $\epsilon_t$  sont des variables aléatoires i.i.d. avec une moyenne nulle et une variance  $\sigma_\epsilon^2$ .

Le modèle RCAR(1) peut être réécrit comme suit :

$$\begin{aligned} X_t &= \alpha + \beta_t X_{t-1} + \epsilon_t \\ &= \alpha + \mu_\beta X_{t-1} + u_t \end{aligned}$$

où  $u_t = \sigma_\beta \nu_t X_{t-1} + \epsilon_t$

alors  $\nu_t$  est une variable aléatoire ayant une moyenne nulle et une variance de un et étant indépendante de  $\epsilon_t$ .

### 3.2.2 Méthodes d'estimation des paramètres

Pour l'estimation des paramètres  $\theta = (\alpha, \mu_\beta)$  du modèle RCAR(1), l'auteur Autcha Araveeporn (2022) [2] propose l'utilisation du critère des moindres carrés, de la méthode du maximum de vraisemblance et de la méthode Bayésienne hiérarchique basée sur la méthode MCMC.

#### 3.2.2.1 La méthode des moindres carrés

La première méthode d'estimation que nous proposons est le critère des moindres carrés pour estimer les paramètres  $\theta = (\alpha, \mu_\beta)$  en minimisant la somme des carrés des résidus.

Soit  $\mathcal{F}_t$  l'ensemble d'informations jusqu'au temps  $t$ , et  $u_t = X_t - \mu_\beta X_{t-1} - \alpha$ , alors nous pouvons voir que  $\mathbb{E}(u_t | \mathcal{F}_{t-1}) = 0$ ,  $\mathbb{E}(u_t^2 | \mathcal{F}_{t-1}) = \sigma_\epsilon^2 + \sigma_\beta^2 X_{t-1}^2$  et  $\mathbb{V}(u_t | \mathcal{F}_{t-1}) = \sigma_\epsilon^2 + \sigma_\beta^2 X_{t-1}^2$ .

Étant donné un échantillon  $X_1, X_2, \dots, X_n$ , le paramètre  $\theta = (\alpha, \mu_\beta)$  est estimé en minimisant  $\sum_{t=1}^n u_t^2$ . Ainsi, l'estimateur des moindres carrés  $\hat{\theta}_{MC} = (\hat{\alpha}_{MC}, \hat{\mu}_{\beta, MC})$  est donné par

$$\sum_{t=2}^n (u_t)^2 = \sum_{t=2}^n (X_t - \alpha - \mu_\beta X_{t-1})^2$$

Nous calculons les dérivées par rapport aux paramètres  $\hat{\alpha}_{MC}$  et  $\hat{\mu}_{\beta, MC}$ , ce qui nous conduit ensuite à obtenir :

$$\hat{\alpha}_{MC} = \frac{\sum_{t=2}^n X_{t-1}^2 \sum_{t=1}^n X_t - \sum_{t=2}^n X_t X_{t-1} \sum_{t=2}^n X_{t-1}}{n \sum_{t=2}^n X_{t-1}^2 - (\sum_{t=2}^n X_{t-1})^2}$$

et

$$\hat{\mu}_{\beta, MC} = \frac{n \sum_{t=2}^n X_t X_{t-1} - \sum_{t=1}^n X_t \sum_{t=2}^n X_{t-1}}{n \sum_{t=2}^n X_{t-1}^2 - (\sum_{t=2}^n X_{t-1})^2}$$

Pour le modèle RCAR(1), il peut être ajusté comme suit :

$$\hat{X}_t = \hat{\alpha}_{MC} + \hat{\mu}_{\beta, MC} X_{t-1}, \quad t = 2, 3, \dots, n$$

### 3.2.2.2 La méthode du maximum de vraisemblance

La méthode du maximum de vraisemblance tient compte des observations  $\{X_1, X_2, \dots, X_n\}$  et de la densité des probabilités (comme la méthode des moindres carrés) et de la densité des probabilités de l'échantillon.

Ensuite, Autcha Araveeporn (2022)[2] considère les données observées suivent une distribution normale, et nous obtenons

$$\mathbb{E}(X_t|X_{t-1}) = \alpha + \mu_\beta X_{t-1} \quad \text{et} \quad \mathbb{V}(X_t|X_{t-1}) = \sigma_\epsilon^2 + \sigma_\beta^2 X_{t-1}$$

Maintenant, la fonction de vraisemblance est définie comme suit :

$$\begin{aligned} L(\theta) &= L(\theta|X_t, X_{t-1}) = \prod_{t=2}^n f(X_t|X_{t-1}) \\ &= \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \prod_{t=2}^n (\sigma_\epsilon^2 + \sigma_\beta^2 X_{t-1}^2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} \sum_{t=2}^n \frac{(X_t - \alpha - \mu_\beta X_{t-1})^2}{\sigma_\epsilon^2 + \sigma_\beta^2 X_{t-1}^2}\right\} \end{aligned}$$

Il est plus pratique de travailler en définissant le paramètre  $v = \frac{\sigma_\beta^2}{\sigma_\epsilon^2}$ , de sorte qu'il peut être écrit comme suit :

$$L(\theta) = \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \sigma^2 \prod_{t=2}^n (1 + v X_{t-1}^2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma_\epsilon^2} \sum_{t=2}^n \frac{(X_t - \alpha - \mu_\beta X_{t-1})^2}{1 + v X_{t-1}^2}\right\}$$

Ensuite, nous prenons le logarithme népérien ( $\ln$ ) de la fonction de vraisemblance, puis nous procédons au calcul des dérivées partielles de  $\ln(L(\theta))$  par rapport aux paramètres  $\alpha$  et  $\mu_\beta$ .

$$\frac{\partial}{\partial \alpha} \ln(L(\theta)) = -\frac{1}{2\sigma_\epsilon^2} \sum_{t=2}^n \frac{(X_t - \alpha - \mu_\beta X_{t-1})}{1 + v X_{t-1}^2}$$

$$\frac{\partial}{\partial \mu_\beta} \ln(L(\theta)) = -\frac{1}{2\sigma_\epsilon^2} \sum_{t=2}^n \frac{(X_t - \alpha - \mu_\beta X_{t-1})X_{t-1}}{1 + v X_{t-1}^2}$$

Soit  $\frac{\partial}{\partial(\alpha, \mu_\beta)} \ln(L(\theta)) = 0$ . Nous obtenons alors

$$\hat{\alpha} = \frac{\sum_{t=2}^n \frac{X_t}{1+v X_{t-1}^2} - \mu_\beta \sum_{t=2}^n \frac{X_{t-1}}{1+v X_{t-1}^2}}{\sum_{t=2}^n \frac{1}{1+v X_{t-1}^2}} \quad (3.1)$$

$$\hat{\mu}_\beta = \frac{\sum_{t=2}^n \frac{X_t X_{t-1}}{1+v X_{t-1}^2} - \alpha \sum_{t=2}^n \frac{X_{t-1}}{1+v X_{t-1}^2}}{\sum_{t=2}^n \frac{X_{t-1}^2}{1+v X_{t-1}^2}} \quad (3.2)$$

À partir de (3.1) et (3.2), nous pouvons le réécrire comme suit :

$$\hat{\alpha} = \frac{c_1 - \hat{\mu}_\beta c_2}{c_3} \quad \text{et} \quad \hat{\mu}_\beta = \frac{c_4 - \hat{\alpha} c_2}{c_5}$$

où

$$c_1 = \sum_{t=2}^n \frac{X_t}{1 + v X_{t-1}^2}, \quad c_2 = \sum_{t=2}^n \frac{X_{t-1}}{1 + v X_{t-1}^2}, \quad c_3 = \sum_{t=2}^n \frac{1}{1 + v X_{t-1}^2}$$

$$c_4 = \sum_{t=2}^n \frac{X_{t-1}^2}{1 + v X_{t-1}^2}, \quad c_5 = \sum_{t=2}^n \frac{X_t X_{t-1}}{1 + v X_{t-1}^2}$$

Enfin, nous obtenons les estimateurs suivants pour  $\alpha$  et  $\mu_\beta$  :

$$\hat{\alpha}_{MV} = \frac{c_1 c_4 - c_5 c_2}{c_3 c_4 - c_2^2}$$

et

$$\hat{\mu}_{\beta, MV} = \frac{c_3 c_5 - c_1 c_2}{c_3 c_4 - c_2^2}$$

Pour les valeurs ajustées observées du modèle RCAR(1), elles peuvent être exprimées comme suit :

$$\hat{X}_t = \hat{\alpha}_{MV} + \hat{\mu}_{\beta, MV} X_{t-1}, \quad t = 2, 3, \dots, n$$

### 3.2.2.3 La méthode Bayésienne hiérarchique

Dans le cadre de la méthode Bayésienne, la méthode Monte-Carlo par chaînes de Markov (MCMC) permet d'estimer la forme de la distribution a posteriori. Morton et Finkenstadt (2005)[25] ont utilisé cette méthode pour modéliser le modèle des individus susceptible-infecté-récupéré<sup>1</sup> pour les maladies infectieuses. L'estimateur Bayésien considéré est la moyenne de la distribution a posteriori obtenu grâce à la méthode MCMC.

---

1. Susceptible-Infecté-Récupéré" (SIR) est un modèle épidémiologique utilisé pour décrire la propagation d'une maladie infectieuse au sein d'une population. Dans ce modèle, les individus sont divisés en trois catégories :

Susceptibles (S) : Ce sont les individus qui sont susceptibles d'être infectés par la maladie s'ils entrent en contact avec des individus infectés.

Infectés (I) : Ce sont les individus qui sont actuellement infectés par la maladie et qui peuvent la transmettre aux individus susceptibles.

Récupérés (R) : Ce sont les individus qui ont été infectés par la maladie, mais qui se sont rétablis et sont maintenant immunisés, ce qui signifie qu'ils ne peuvent pas être infectés à nouveau et ne peuvent pas transmettre la maladie.

## Estimation des paramètres pour le modèle RCAR(1)

Pour réaliser une estimation Bayésienne pour le modèle RCAR(1), nous proposons le modèle hiérarchique à trois niveaux.

- Au premier niveau se trouve la distribution conditionnelle des données  $X_t$  étant donné les variables aléatoires observées  $X_{t-1}$ , les coefficients  $\alpha$ ,  $\beta_t$  et  $\sigma_\epsilon^2$ .
- Le deuxième niveau consiste en la distribution conditionnelle de  $\beta_t$  étant donné les paramètres  $\mu_\beta$  et  $\sigma_\beta^2$ .
- Enfin, le dernier niveau représente la distribution a priori de  $\theta$  définie comme  $\theta = (\alpha, \mu_\beta)^T$ .

En conséquence, étant donné les variables de l'échantillon  $X_1, X_2, \dots, X_n$ , nous sommes en mesure d'exprimer le modèle RCAR(1) dans la structure hiérarchique suivante :

$$\begin{aligned} X_t | X_{t-1}, \alpha, \beta_t, \sigma_\epsilon^2 &\sim \mathcal{N}(\alpha + \beta_t X_{t-1}, \sigma_\epsilon^2), \\ \beta_t | \mu_\beta, \sigma_\beta^2 &\sim \mathcal{N}(\mu_\beta, \sigma_\beta^2), \\ (\alpha, \mu_\beta) &\sim \pi(\alpha, \mu_\beta) \end{aligned} \tag{3.3}$$

où  $\pi(\cdot)$  est la densité a priori de  $\theta$  qui reflète notre a priori sur les paramètres inconnus.

En suivant cette structure hiérarchique l'équation (3.3), nous pouvons exprimer la fonction de vraisemblance de  $\theta$  comme suit :

$$L(\alpha, \mu_\beta | X_1, X_2, \dots, X_n, \epsilon_1, \dots, \epsilon_n) = \phi(X_1; \alpha, \sigma_\epsilon) \prod_{i=2}^n \phi\left(X_i; \alpha + \beta_i X_{i-1}, \sqrt{\sigma_\epsilon^2 + \sigma_\beta^2 X_{i-1}^2}\right) \tag{3.4}$$

où  $\phi(X; \mu, \sigma)$  désigne la fonction de densité d'une distribution normale avec une moyenne  $\mu$  et un écart type  $\sigma$ .

Par conséquent, la densité a posteriori conjointe des paramètres est donnée par :

$$\pi(\theta | X_1, X_2, \dots, X_n) \propto L(\theta | X_1, X_2, \dots, X_n) \pi(\theta)$$

où  $\pi(\theta)$  est une densité a priori de  $\theta$ .

Pour l'estimation des paramètres du modèle RCAR(1), la distribution a priori de  $\theta = (\alpha, \mu_\beta)^T$  est considérée comme une variable aléatoire continue dans l'ensemble des

nombres réels, suivant une distribution normale.

À partir de la structure hiérarchique de l'équation (3.4), la densité a posteriori conjointe peut être écrite comme :

$$\pi_{\alpha}(\alpha|X_1, X_2, \dots, X_n) \propto \int \pi(\alpha, \mu_{\beta}, \sigma_{\beta}^2, \sigma_{\epsilon}^2|X_1, X_2, \dots, X_n) d\mu_{\beta} d\sigma_{\beta}^2 d\sigma_{\epsilon}^2$$

et

$$\pi_{\mu_{\beta}}(\mu_{\beta}|X_1, X_2, \dots, X_n) \propto \int \pi(\alpha, \mu_{\beta}, \sigma_{\beta}^2, \sigma_{\epsilon}^2|X_1, X_2, \dots, X_n) d\alpha d\sigma_{\beta}^2 d\sigma_{\epsilon}^2$$

La distribution a posteriori conjointe de  $\alpha$  et  $\mu_{\beta}$  est obtenue à partir de la distribution normale, de sorte que les caractéristiques de la distribution permettent souvent de spécifier des bornes inférieures et supérieures pour ces paramètres, ce qui peut être déterminé avec précision.

Pour gérer l'analyse Bayésienne hiérarchique pour le modèle RCAR(1), nous nous intéressons aux propriétés de la densité de  $\theta = (\alpha, \mu_{\beta})^T$ . Obtenir la densité a posteriori conjointe pour  $\theta$  revient à intégrer les coefficients non observés  $\alpha$  et  $\mu_{\beta}$ . En raison de la complexité de la fonction de vraisemblance, nous avons utilisé la méthode MCMC (Gilks et al. 1996)[17] pour générer des échantillons à partir de la distribution a posteriori de  $\theta$ . Nous utiliserons l'échantillonneur de Gibbs (Gelfand et al. 1990)[15], une méthode MCMC largement utilisée, pour obtenir le paramètre à partir de la distribution a posteriori en utilisant le logiciel R avec le package rjags.

Les densités conditionnelles des paramètres dans le modèle RCAR(1) sont :  $\pi_{\alpha}(\alpha|\mu_{\beta}, \sigma_{\beta}^2, \sigma_{\epsilon}^2, \underline{X})$  et  $\pi_{\mu_{\beta}}(\mu_{\beta}|\alpha, \sigma_{\beta}^2, \sigma_{\epsilon}^2, \underline{X})$  comme les densités conditionnelles complètes de  $\alpha$  et  $\mu_{\beta}$  respectivement.

**L'algorithme d'échantillonnage de Gibbs est**

- Initialiser  $\alpha^{(0)}$  et  $\mu_{\beta}^{(0)}$  pour  $k = 1, 2, \dots, m + M$ .
- Tirer  $\alpha^{(k)}$  à partir de  $\pi_{\alpha}(\alpha|\mu_{\beta}^{(k-1)}, \sigma_{\beta}^{2(k-1)}, \sigma_{\epsilon}^{2(k-1)}, \underline{X})$ .
- Tirer  $\mu_{\beta}^{(k)}$  à partir de  $\pi_{\mu_{\beta}}(\mu_{\beta}|\alpha^{(k)}, \sigma_{\beta}^{2(k-1)}, \sigma_{\epsilon}^{2(k-1)}, \underline{X})$ .

où

$m = 2000$  représente un burn-in (période de chauffe).

$M = 10000$  représente le nombre d'échantillons générés après burn-in.

En répétant les étapes d'échantillonnage ci-dessus, nous obtenons une chaîne de Markov en temps discret

$$\{(\alpha^{(k)}, \mu_\beta^{(k)}); k = 1, 2, \dots\}$$

dont la distribution stationnaire est la densité a posteriori conjointe des paramètres. Une chaîne de Markov de paramètres  $\alpha$  et  $\mu_\beta$  est construite en calculant la moyenne des échantillons à partir de la densité a posteriori conjointe en utilisant une distribution standard comme suit :

$$\hat{\alpha}_{Bayes} = \frac{\sum_{k=1}^M \alpha^{(k)}}{M} \quad \text{et} \quad \hat{\mu}_{\beta, Bayes} = \frac{\sum_{k=1}^M \mu_\beta^{(k)}}{M}$$

Pour les valeurs ajustées observées du modèle RCAR(1), nous avons :

$$\hat{X}_t = \hat{\alpha}_{Bayes} + \hat{\mu}_{\beta, Bayes} X_{t-1} \quad , \quad t = 2, 3, \dots, n$$

Pour le problème d'estimation Bayésienne des paramètres pour  $\theta = (\alpha, \mu_\beta)$  dans le modèle RCAR(1), nous utilisons les lois a priori non informatives suivantes :

$$\alpha \sim \mathcal{N}(\mu_0, V_0)$$

$$\mu_\beta \sim \mathcal{N}(\mu_0, V_0)$$

Nous posons  $\sigma_\epsilon^2$  et  $\sigma_\beta^2$  comme étant des constantes, où  $\mu_0 = 0$ ,  $V_0 = 10^6$ .

Lorsque nous utilisons les lois a priori non-informatives et que nous supposons la normalité des coefficients et des erreurs aléatoires, les paramètres du modèle RCAR(1) ont des densités conditionnelles complètes de la forme suivante :

$$\alpha | \mu_\beta, \sigma_\beta^2, \sigma_\epsilon^2, \underline{X} \sim 2\mathcal{N}\left(\frac{b_2}{2b_1}, \frac{1}{b_1}\right)$$

$$\mu_\beta | \alpha, \sigma_\beta^2, \sigma_\epsilon^2, \underline{X} \sim 2\mathcal{N}\left(\frac{d_2}{2d_1}, \frac{1}{d_1}\right)$$

où

$$b_1 = \sum_{t=1}^n \frac{1}{\sigma_\epsilon^2 + \sigma_\beta^2 X_{t-1}^2} + \frac{1}{\sigma_\epsilon^2} + \frac{1}{V_0}$$

$$b_2 = \sum_{t=1}^n \frac{X_t - \mu_\beta X_{t-1}}{\sigma_\epsilon^2 + \sigma_\beta^2 X_{t-1}^2} + \frac{X_0}{\sigma_\epsilon^2} + \frac{\mu_0}{V_0}$$

$$d_1 = \sum_{t=1}^n \frac{X_{t-1}^2}{\sigma_\epsilon^2 + \sigma_\beta^2 X_{t-1}^2} + \frac{1}{V_0}$$

$$d_2 = \sum_{t=1}^n \frac{X_{t-1}(X_t - \alpha)}{\sigma_\epsilon^2 + \sigma_\beta^2 X_{t-1}^2} + \frac{\mu_0}{V_0}$$

### 3.2.3 Étude de simulation

L'objectif de cette étude est d'estimer les paramètres  $\theta = (\alpha, \mu_\beta)$  du modèle RCAR(1) en utilisant les méthodes des moindres carrés (MC), du maximum de vraisemblance (MV) et Bayésienne hiérarchique (Bayes). Les résultats ont montré une comparaison des estimateurs moyens pour des tailles d'échantillons  $n = 100, 300$  et  $500$ .

L'erreur quadratique moyenne (MSE) évalue la différence entre les valeurs estimées et les valeurs simulées. Nous avons également calculé le MSE comme le critère défini ci-dessous :

$$MSE = \frac{\sum_{t=2}^n (X_t - \hat{X}_t)^2}{n - 1}$$

où

- $X_t$  représente les valeurs simulées.
- $\hat{X}_t$  représente les valeurs estimées.

L'étude de simulation est divisée en trois parties. Dans le premier processus, nous avons généré des données  $X_t, t = 1, 2, \dots, t$  à partir du modèle RCAR(1) comme suit :

$$X_t = \alpha + \mu_\beta X_{t-1} + u_t, \quad u_t \sim \mathcal{N}(0, \sigma_u^2), \quad \sigma_u^2 = \sigma_\epsilon^2 + \sigma_\beta^2 X_{t-1}^2$$

Les paramètres du modèle RCAR(1) sont mentionnés dans cinq cas différents, à savoir :

1.  $\alpha = 0.5, \quad \mu_\beta = 1, \quad \sigma_\epsilon^2 = 1, \quad \sigma_\beta^2 = 0$  et  $\mu_\beta^2 + \sigma_\beta^2 = 1$  (non stationnaire)
2.  $\alpha = 0, \quad \mu_\beta = 0.99, \quad \sigma_\epsilon^2 = 1, \quad \sigma_\beta^2 = 0.01$  et  $\mu_\beta^2 + \sigma_\beta^2 = 0.9901$  (tend vers la non-stationnarité)
3.  $\alpha = 0, \quad \mu_\beta = 0.01, \quad \sigma_\epsilon^2 = 1, \quad \sigma_\beta^2 = 0.99$  et  $\mu_\beta^2 + \sigma_\beta^2 = 0.9901$  (tend vers la non-stationnarité)
4.  $\alpha = 0, \quad \mu_\beta = 0.6, \quad \sigma_\epsilon^2 = 1, \quad \sigma_\beta^2 = 0.4$  et  $\mu_\beta^2 + \sigma_\beta^2 = 0.76$  (stationnaire)
5.  $\alpha = 0.5, \quad \mu_\beta = 1, \quad \sigma_\epsilon^2 = 1, \quad \sigma_\beta^2 = 0.99$  et  $\mu_\beta^2 + \sigma_\beta^2 = 1.99$  (explosif)

L'objectif des cinq cas est de visualiser les différentes données en fonction des paramètres différents, et le type de données est essentiel pour ajuster les paramètres de ces méthodes. Les figures (3.1) à (3.3) représentent les échantillons dans les tailles  $n = 100, 300, 500$ .

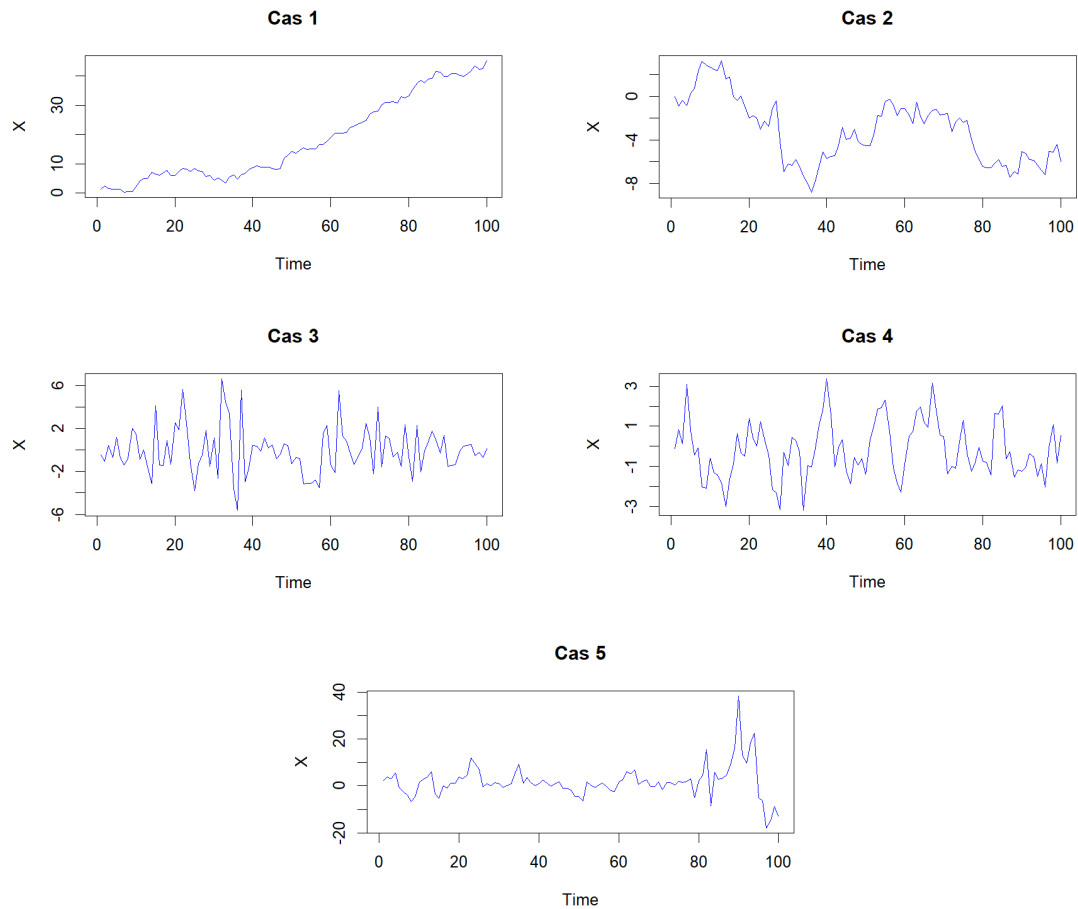


FIGURE 3.1 – Le graphique de la série temporelle pour les données générées pour  $n = 100$ .

**Interprétation :** La figure (3.1) présente les données générées pour une taille d'échantillon de  $n = 100$ . Il est important de noter que :

- Le cas 1 représente des données avec tendance.
- Le cas 2 est illustré comme une marche aléatoire.
- Les cas 3 et 4 semblent également osciller autour de leur moyenne zéro.
- Le cas 5 représente le cas explosif.

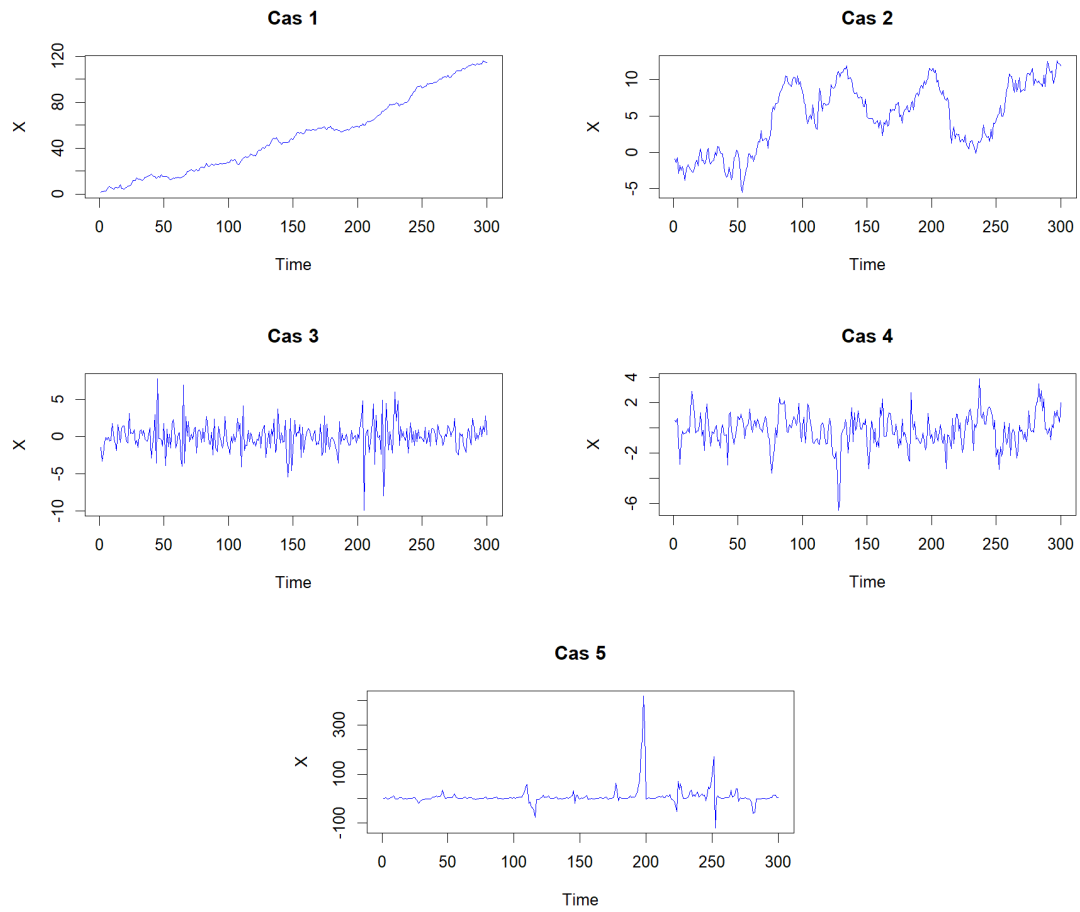


FIGURE 3.2 – Le graphique de la série temporelle pour les données générées pour  $n = 300$ .

**Interprétation :** La figure (3.2) présente les données générées pour  $n = 300$ , démontrant :

- Le processus de tendance et de marche aléatoire dans les cas 1 et 2.
- Un processus stationnaire est représenté dans les cas 3 et 4.
- Un processus explosif est représenté dans les cas 5 .

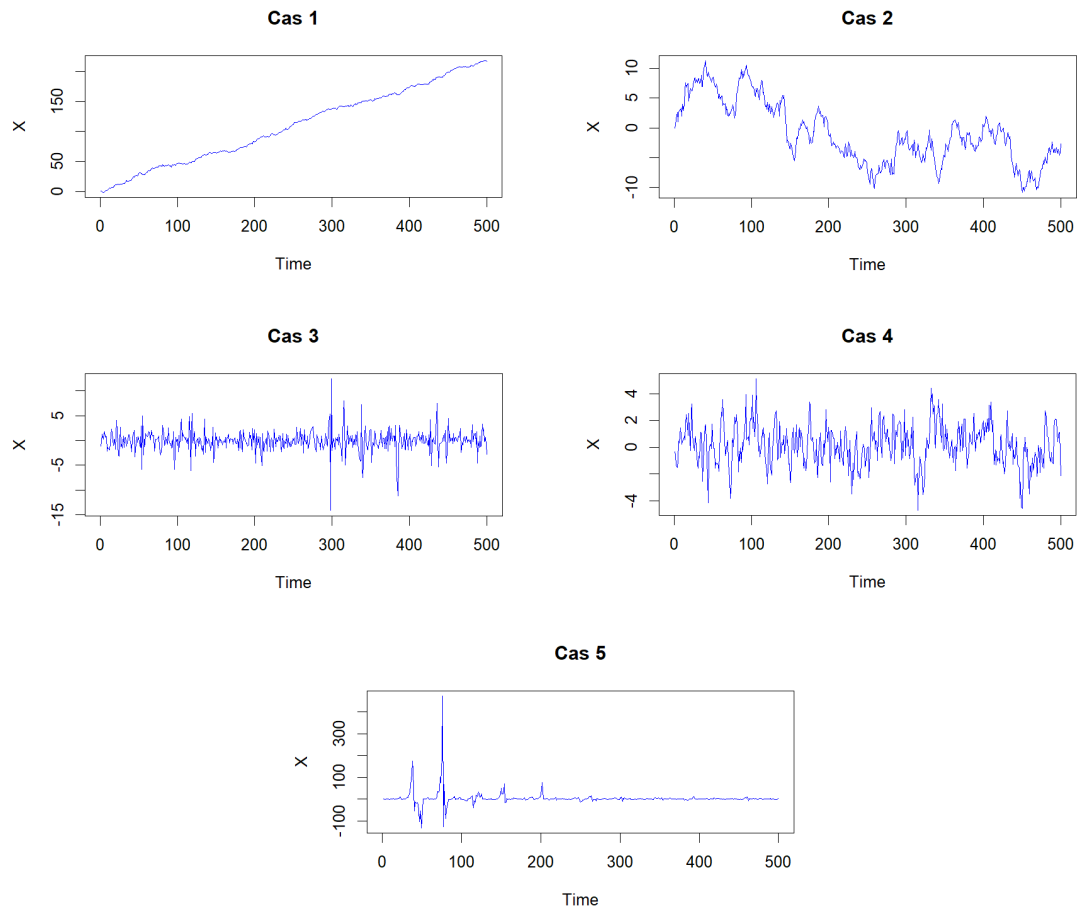


FIGURE 3.3 – Le graphique de la série temporelle pour les données générées pour  $n = 500$ .

**Interprétation :**

- L'impact des grandes tailles d'échantillon  $n = 500$  a également identifié les oscillations dans les cas 3 et 4 comme présenté dans la figure (3.3).

Dans la deuxième partie, nous obtenons l'estimateur  $\hat{\theta}_{MC} = (\hat{\alpha}_{MC}, \hat{\mu}_{\beta,MC})$  à partir de la méthode des moindres carrés,  $\hat{\theta}_{MV} = (\hat{\alpha}_{MV}, \hat{\mu}_{\beta,MV})$  à partir de la méthode du maximum de vraisemblance et  $\hat{\theta}_{Bayes} = (\hat{\alpha}_{Bayes}, \hat{\mu}_{\beta,Bayes})$  à partir de la méthode Bayésienne hiérarchique basée sur la méthode MCMC.

À partir de la méthode des moindres carrés, nous obtenons  $X_t$  à partir de  $\hat{X}_t = \hat{\alpha}_{MC} + \hat{\mu}_{\beta,MC}X_{t-1}$ ,  $t = 2, 3, \dots, n$ . Les valeurs estimées à partir de la méthode du

maximum de vraisemblance sont calculées par  $\hat{X}_t$  à partir de l'analyse Bayésienne est approximée par  $\hat{X}_t = \hat{\alpha}_{MV} + \hat{\mu}_{\beta, MV} X_{t-1}$ ,  $t = 2, 3, \dots, n$ , et  $\hat{X}_t = \hat{\alpha}_{Bayes} + \mu_{\beta, Bayes} X_{t-1}$ ,  $t = 2, 3, \dots, n$ .

Enfin, dans la dernière partie, nous avons simulé des données sur 500 répliques à partir d'un modèle RCAR(1). Nous avons également calculé la moyenne pour chaque paramètre et taille d'échantillon. Pour cela, nous avons calculé la moyenne des MSE (AMSE)<sup>2</sup> pour comparer les estimateurs obtenus par les méthodes précédentes (moindres carrés, du maximum de vraisemblance et Bayésienne) comme le montrent (3.1) à (3.3).

TABLE 3.1 – La moyenne de l'erreur quadratique moyenne (AMSE) de l'estimation des paramètres par la méthode des moindres carrés

Cas	Paramètre	Tailles d'échantillon		
		$n = 100$	$n = 300$	$n = 500$
1	$\alpha = 0.5$	0.5468979	0.522664	0.516744
	$\mu_{\beta} = 1$	0.9942908	0.9958713	0.9981531
	AMSE	0.979956	0.983526	0.9976
2	$\alpha = 0$	0.0029385	-0.014161	0.0000234739
	$\mu_{\beta} = 0.99$	0.9686074	0.9728126	0.9815071
	AMSE	1.438548	1.527216	1.629149
3	$\alpha = 0$	0.0081086	-0.0068064	0.000285
	$\mu_{\beta} = 0.01$	-0.000721	0.003209	0.0057259
	AMSE	1.0608215	1.085601	1.08353
4	$\alpha = 0$	0.004036	-0.0017588	-0.000634743
	$\mu_{\beta} = 0.6$	0.5301467	0.57902	0.59082
	AMSE	1.096348	1.097565	0.098644
5	$\alpha = 0.5$	0.489563	0.528712	0.536795
	$\mu_{\beta} = 1$	0.957998	0.974709	0.9849731
	AMSE	47.77515	52.02487	89.69467

**Interprétation :**

- D'après le tableau (3.1), l'AMSE de la méthode des moindres carrés présente les valeurs minimales pour les cas 2, 3 et 4.

2. Average Mean Square Error (la moyenne de l'erreur quadratique moyenne)

- Lorsque les tailles d'échantillon augmentent, cette méthode est donc cohérente pour l'estimation des paramètres.
- En revanche, le cas 5, considéré comme explosif, affiche une AMSE très élevée.

TABLE 3.2 – La moyenne de l'erreur quadratique moyenne (AMSE) de l'estimation des paramètres par la méthode du maximum de vraisemblance

Cas	Paramètre	Tailles d'échantillon		
		$n = 100$	$n = 300$	$n = 500$
1	$\alpha = 0.5$	0.560159	0.543295	0.521158
	$\mu_\beta = 1$	0.9968926	0.9978951	0.999999
	AMSE	0.98771946	0.995763	0.995624
2	$\alpha = 0$	0.0197952	-0.0024868	0.00296423
	$\mu_\beta = 0.99$	0.9342366	0.9709414	0.972934
	AMSE	1.432453	1.5472064	1.6286721
3	$\alpha = 0$	0.0029821	-0.002523	0.0017460
	$\mu_\beta = 0.01$	0.00020161	0.00523905	-0.0006505
	AMSE	1.076205	1.0941868	1.1845274
4	$\alpha = 0$	-0.1516207	-0.05062976	0.06259691
	$\mu_\beta = 0.6$	0.6304789	0.4812071	0.6822012
	AMSE	2.698073	2.361248	2.29787
5	$\alpha = 0.5$	0.5584173	0.5423054	0.5120381
	$\mu_\beta = 1$	1.100672	0.9688486	1.006477
	AMSE	106.8973	255.0127	267.2161

**Interprétation :**

- La méthode du maximum de vraisemblance donne de bons résultats dans le cas 1, comme le démontrent les données de tendance présentées dans le tableau (3.2).
- L'AMSE a montré des valeurs minimales pour toutes les tailles d'échantillon.

TABLE 3.3 – La moyenne de l’erreur quadratique moyenne (AMSE) de l’estimation des paramètres par la méthode Bayésienne hiérarchique

Cas	Paramètre	Tailles d’échantillon		
		$n = 100$	$n = 300$	$n = 500$
1	$\alpha = 0.5$	0.536627	0.52444025	0.5145789
	$\mu_\beta = 1$	0.9989372	0.9992390	1.002131
	AMSE	1.321617	5.1366578	9.157619
2	$\alpha = 0$	0.0153121	-0.0026055	-0.0007993
	$\mu_\beta = 0.99$	0.97303699	0.9802405	0.9960326
	AMSE	1.4287729	1.5631212	1.850166
3	$\alpha = 0$	0.0045172	-0.0024841	0.00038412
	$\mu_\beta = 0.01$	-0.000932	0.0024384	0.00545186
	AMSE	1.07041	1.097227	1.098250
4	$\alpha = 0$	0.08477534	0.08083943	0.06562123
	$\mu_\beta = 0.6$	0.5397339	0.5624219	0.584967
	AMSE	2.589896	2.5971449	2.596096
5	$\alpha = 0.5$	0.4119895	0.5668433	0.3030206
	$\mu_\beta = 1$	1.019606	1.056783	0.9626697
	AMSE	40.09211	102.5593	234.6891

**Interprétation :**

- D’après le tableau (3.3), les AMSE Bayésiens sont plus élevées que celles des autres méthodes. Il semble que la distribution a priori de la méthode Bayésienne n’affecte pas l’estimation des paramètres dans le modèle RCAR(1).
- Dans la majorité des cas, la méthode des moindres carrés se comporte mieux avec une AMSE minimale. La méthode du maximum de vraisemblance se comporte mieux que les autres méthodes pour les données présentant une tendance, bien qu’elle diffère légèrement de la méthode des moindres carrés. Pour le cas 5, qui est un cas explosif, nous constatons une AMSE très élevée.
- L’auteur Araveeporn Autcha recommande ainsi d’ajuster le modèle RCAR(1) aux données de séries temporelles en utilisant la méthode des moindres carrés, qu’elles soient stationnaires ou non.

### 3.3 Application sur le modèle RCDAR(1,1)

Dans cette section, nous allons présenter le modèle autorégressif dynamique à coefficients aléatoires, noté RCDAR(p,q). Il s'agit d'un cas particulier du modèle RCAR(p). Notre source d'inspiration pour ce modèle provient de l'article intitulé 'Bayesian Analysis of Random Coefficient Dynamic AutoRegressive Model', écrit par Araveeporn Autcha et publié dans le 'Thailand Statistician' en 2012 [3].

#### 3.3.1 Présentation du modèle RCDAR(p,q) et RCDAR(1,1)

Le modèle autorégressif à coefficients aléatoires noté RCAR dans sa forme générale est noté RCAR(p), écrit comme suit :

$$\begin{aligned} X_t &= \alpha + \sum_{i=1}^p \beta_{ti} X_{t-i} + \sigma \epsilon_t, \quad t=1,2,\dots \\ \underline{\beta}_t &= \underline{\mu}_\beta + \Sigma_\beta^{\frac{1}{2}} \underline{u}_t \end{aligned} \quad (3.5)$$

où

$$\begin{aligned} \underline{\beta}_t &= (\beta_{t1}, \dots, \beta_{tp})^T \\ \underline{\mu}_\beta &= (\mu_{\beta 1}, \dots, \mu_{\beta p})^T \end{aligned}$$

Nous supposons que les  $\epsilon_t$  et  $\underline{u}_t$  suivent une distribution indépendante et identiquement distribuée (i.i.d.) avec une moyenne nulle et une variance unitaire, et que  $\underline{u}_t$  est appelé bruit blanc.

De manière empirique, nous pouvons voir que le terme constant du modèle RCAR peut être modélisé en ajoutant les variables exogènes  $y_t$ , comme dans le modèle autorégressif dynamique à coefficients aléatoires (RCDAR) ou modèle RCDAR(p,q) suivant :

$$\begin{aligned} X_t &= y_t + \sum_{i=1}^p \beta_{ti} X_{t-i} + \sigma \epsilon_t, \quad t=1,2,\dots \\ \underline{\beta}_t &= \underline{\mu}_\beta + \Sigma_\beta^{\frac{1}{2}} \underline{u}_t \\ y_t &= \sum_{j=1}^q \eta_j y_{t-j} + \delta_t \end{aligned} \quad (3.6)$$

Dans ce cas, nous allons considérer le modèle RCDAR(1,1) d'ordre un avec  $p = q = 1$  et  $\sigma = 1$  qui peut être écrit comme suit :

$$\begin{aligned} X_t &= y_t + \beta_t X_{t-1} + \epsilon_t, \quad \text{pour } t = 1, 2, \dots \\ \beta_t &= \mu_\beta + \sigma_\beta u_t \\ y_t &= \eta y_{t-1} + \delta_t \end{aligned} \tag{3.7}$$

où

- $\beta_t$  sont des variables aléatoires (i.i.d.) avec une moyenne  $\mu_\beta$  et une variance  $\sigma_\beta^2$ .
- $\epsilon_t$  sont des variables aléatoires (i.i.d.) avec une moyenne nulle et une variance  $\sigma_\epsilon^2$ .
- $\delta_t$  sont des variables aléatoires (i.i.d.) avec une moyenne nulle et une variance  $\sigma_\delta^2$ .
- $\beta_t$ ,  $\epsilon_t$  et  $\delta_t$  sont indépendants.

Pour l'estimation des paramètres du modèle RCDAR(1,1), on observe à partir de l'équation (3.7) que cela consiste  $\eta$ ,  $\mu_\beta$ ,  $\sigma_\beta^2$ ,  $\sigma_\epsilon^2$  et  $\sigma_\delta^2$ , ou définie comme  $\theta = (\eta, \mu_\beta, \sigma_\beta^2, \sigma_\epsilon^2, \sigma_\delta^2)^T$ .

### 3.3.2 Estimation des paramètres d'un modèle RCDAR(1,1) par la méthode Bayésienne hiérarchique

Pour effectuer une analyse Bayésienne pour le modèle RCDAR(1,1), nous proposons le modèle hiérarchique à trois niveaux.

- Le premier niveau concerne la distribution conditionnelle des données  $X_t$  étant donné les observations  $X_{t-1}$  et  $y_t$ , les coefficients  $\eta$ ,  $\beta_t$ ,  $\sigma_\epsilon^2$ .
- Le deuxième niveau consiste en la distribution conditionnelle de  $\beta_t$  étant donné les paramètres  $\mu_\beta$  et  $\sigma_\beta^2$ .
- Enfin, le dernier niveau représente la distribution conditionnelle de  $y_t$  étant donnée les observations  $y_{t-1}$  et les paramètres  $\eta$  et  $\sigma_\delta^2$ .

Par conséquent, étant donné les variables d'échantillon  $X_1, X_2, \dots, X_n$  et les variables exogènes  $y_1, y_2, \dots, y_n$ , nous pouvons exprimer le modèle RCDAR(1,1) dans la structure hiérarchique suivante :

$$\begin{aligned} X_t | X_{t-1}, y_t, \eta, \beta_t, \sigma_\epsilon^2 &\sim \mathcal{N}(y_t + \beta_t X_{t-1}, \sigma_\epsilon^2), \\ \beta_t | \mu_\beta, \sigma_\beta^2 &\sim \mathcal{N}(\mu_\beta, \sigma_\beta^2), \\ y_t | y_{t-1}, \eta &\sim \mathcal{N}(\eta y_{t-1}, \sigma_\delta^2), \end{aligned}$$

$$(\eta, \mu_\beta, \sigma_\beta^2, \sigma_\epsilon^2, \sigma_\delta^2) \sim \pi(\eta, \mu_\beta, \sigma_\beta^2, \sigma_\epsilon^2, \sigma_\delta^2) \quad (3.8)$$

où  $\pi(\cdot)$  est la densité a priori de  $\theta$ , qui est le paramètre inconnu.

En suivant l'équation (3.8), nous pouvons exprimer la fonction de vraisemblance de  $\theta$  comme

$$L(\eta, \mu_\beta, \sigma_\beta^2, \sigma_\epsilon^2, \sigma_\delta^2 | X_1, \dots, X_n, y_1, \dots, y_n) = L(\mu_\beta, \sigma_\beta^2, \sigma_\epsilon^2 | X_1, \dots, X_n, y_1, \dots, y_n) L(\eta, \sigma_\delta^2 | y_1, \dots, y_n)$$

où

$$L(\mu_\beta, \sigma_\beta^2, \sigma_\epsilon^2 | X_1, X_2, \dots, X_n, y_1, \dots, y_n) = \phi(X_1; y_1, \sigma_\epsilon) \prod_{t=2}^n \phi\left(X_t; y_t + \mu_\beta X_{t-1}, \sqrt{\sigma_\epsilon^2 + \sigma_\beta^2 X_{t-1}^2}\right) \quad (3.9)$$

et

$$L(\eta, \sigma_\delta^2 | y_1, \dots, y_n) = \phi(y_1; 0, \sigma_\delta) \prod_{t=2}^n \phi(y_t; \eta y_{t-1} + \sigma_\delta)$$

où  $\phi(X; \mu, \sigma)$  désigne la fonction de densité d'une distribution normale avec une moyenne  $\mu$  et un écart-type  $\sigma$ .

Par conséquent, la densité a posteriori conjointe des paramètres est donnée par :

$$\pi(\theta | X_1, X_2, \dots, X_n, y_1, \dots, y_n) \propto L(\theta | X_1, X_2, \dots, X_n, y_1, \dots, y_n) \pi(\theta)$$

où  $\pi(\theta)$  est une densité a priori de  $\theta$ . À partir de la structure hiérarchique de l'équation (3.8), la densité a posteriori de  $\mu_\beta$  peut être exprimée comme

$$\pi_{\mu_\beta}(\mu_\beta | X_1, X_2, \dots, X_n, y_1, \dots, y_n) \propto \int \pi(\mu_\beta, \sigma_\beta^2, \sigma_\epsilon^2 | X_1, X_2, \dots, X_n, y_1, \dots, y_n) d\sigma_\beta^2 d\sigma_\epsilon^2$$

Pour résoudre cette problématique, il est difficile d'intégrer la densité mentionnée ci-dessus. Par conséquent, ce problème utilise la méthode MCMC. Nous allons effectuer l'échantillonneur de Gibbs une méthode MCMC largement utilisée, pour obtenir les paramètres à partir de la distribution a posteriori en utilisant le logiciel R.

### 3.3.2.1 Choix des densités a priori

Le type d'a priori non informatif, est utilisé à la place d'une distribution a priori subjective lorsque aucune information n'est disponible. En utilisant un priori non informatif, toutes les informations sont obtenues à partir de la densité a posteriori conjointe.

Pour l'estimation des paramètres du modèle RCDAR(1,1) avec  $\theta = (\eta, \mu_\beta, \sigma_\beta^2, \sigma_\epsilon^2, \sigma_\delta^2)^T$ , nous utilisons la distribution uniforme comme un a priori non informatif pour  $\eta$ . La distribution normale est utilisée comme un a priori non informatif pour  $\mu_\beta$ , et la distribution gamma est utilisée comme un a priori non informatif pour  $\sigma_\beta^2$ ,  $\sigma_\epsilon^2$  et  $\sigma_\delta^2$ . Plus précisément, nous avons utilisé les a priori suivants :

$$\begin{aligned}\eta &\sim \mathcal{U}(-0.999, 0.999), \\ \mu_\beta &\sim \mathcal{N}(0, 0.01), \\ \sigma_\beta^2, \sigma_\epsilon^2, \sigma_\delta^2 &\sim \mathcal{IG}(2.001, 10^3).\end{aligned}$$

Il convient de noter que pour  $\eta$ ,  $y_t$  suit un processus autorégressif d'ordre 1, ce qui nous conduit à définir  $|\eta| < 1$ . La variance de la distribution normale pour  $\mu_\beta$  est faible. Nous utilisons la distribution inverse-Gamma pour  $\sigma_\beta^2$ ,  $\sigma_\epsilon^2$  et  $\sigma_\delta^2$ , puis nous choisissons les paramètres de manière à obtenir une variance importante.

### 3.3.2.2 Inférence a posteriori par échantillonnage de Gibbs

Pour effectuer une analyse Bayésienne des données pour le modèle RCDAR(1,1), nous nous intéressons aux propriétés de la densité des paramètres  $\theta = (\eta, \mu_\beta, \sigma_\beta^2, \sigma_\epsilon^2, \sigma_\delta^2)^T$ . Obtenir la densité a posteriori conjointe pour  $\theta$  revient à intégrer les coefficients non observés  $\eta, \mu_\beta, \sigma_\beta^2, \sigma_\epsilon^2$  et  $\sigma_\delta^2$ , ce qui est difficile à réaliser à la fois analytiquement et numériquement, rendant ainsi la construction de la densité a posteriori des paramètres très difficile.

À ce stade, les méthodes MCMC sont de plus en plus utilisées dans les cas où la distribution marginale des variables aléatoires ne peut pas être obtenue soit analytiquement, soit numériquement. Les méthodes MCMC comprennent des algorithmes permettant de construire une chaîne de Markov des paramètres de telle sorte que sa distribution stationnaire soit notre distribution d'intérêt, c'est-à-dire la distribution a posteriori du paramètre d'intérêt.

Nous utilisons l'échantillonneur de Gibbs, qui est la méthode MCMC, pour obtenir des échantillons dépendants de la distribution a posteriori. Plus précisément, nous dérivons les densités conditionnelles  $\pi_\eta(\eta|\mu_\beta, \sigma_\beta^2, \sigma_\epsilon^2, \sigma_\delta^2, \underline{X}, \underline{y})$ ,  $\pi_{\mu_\beta}(\mu_\beta|\eta, \sigma_\beta^2, \sigma_\epsilon^2, \sigma_\delta^2, \underline{X}, \underline{y})$ ,  $\pi_{\sigma_\beta^2}(\sigma_\beta^2|\eta, \mu_\beta, \sigma_\epsilon^2, \sigma_\delta^2, \underline{X}, \underline{y})$ ,  $\pi_{\sigma_\epsilon^2}(\sigma_\epsilon^2|\eta, \mu_\beta, \sigma_\beta^2, \sigma_\delta^2, \underline{X}, \underline{y})$ , et  $\pi_{\sigma_\delta^2}(\sigma_\delta^2|\eta, \mu_\beta, \sigma_\beta^2, \sigma_\epsilon^2, \underline{X}, \underline{y})$  en tant que densités conditionnelles complètes de  $\eta, \mu_\beta, \sigma_\beta^2, \sigma_\epsilon^2$  et  $\sigma_\delta^2$  respectivement, en se basant sur le modèle (3.7).

**L'algorithme d'échantillonnage de Gibbs est**

- Initialiser  $\eta^{(0)}, \mu_\beta^{(0)}, \sigma_\beta^{2(0)}, \sigma_\epsilon^{2(0)}$  et  $\sigma_\delta^{2(0)}$  pour  $k = 1, 2, \dots, m + M$ .
- Tirer  $\eta^{(k)}$  à partir de  $\pi_\eta(\eta|\mu_\beta^{(k-1)}, \sigma_\beta^{2(k-1)}, \sigma_\epsilon^{2(k-1)}, \sigma_\delta^{2(k-1)}, \underline{X}, \underline{y})$ .

- Tirer  $\mu_\beta^{(k)}$  à partir de  $\pi_{\mu_\beta}(\mu_\beta|\eta^{(k)}, \sigma_\beta^{2(k-1)}, \sigma_\epsilon^{2(k-1)}, \sigma_\delta^{2(k-1)}, \underline{X}, \underline{y})$ .
- Tirer  $\sigma_\beta^{2(k)}$  à partir de  $\pi_{\sigma_\beta^2}(\sigma_\beta^2|\eta^{(k)}, \mu_\beta^{(k)}, \sigma_\epsilon^{2(k-1)}, \sigma_\delta^{2(k-1)}, \underline{X}, \underline{y})$ .
- Tirer  $\sigma_\epsilon^{2(k)}$  à partir de  $\pi_{\sigma_\epsilon^2}(\sigma_\epsilon^2|\eta^{(k)}, \mu_\beta^{(k)}, \sigma_\beta^{2(k)}, \sigma_\delta^{2(k-1)}, \underline{X}, \underline{y})$ .
- Tirer  $\sigma_\delta^{2(k)}$  à partir de  $\pi_{\sigma_\delta^2}(\sigma_\delta^2|\eta^{(k)}, \mu_\beta^{(k)}, \sigma_\beta^{2(k)}, \sigma_\epsilon^{2(k)}, \underline{X}, \underline{y})$ .

où

$m$  représente un burn-in de 1000 itérations (période de chauffe).

$M$  le nombre d'échantillons générés lors de 4000 itérations après burn-in.

En répétant les étapes d'échantillonnage ci-dessus, nous obtenons une chaîne de Markov en temps discret, représentée comme suit :

$$\{(\eta^{(k)}, \mu_\beta^{(k)}, \sigma_\beta^{2(k)}, \sigma_\epsilon^{2(k)}, \sigma_\delta^{2(k)}); k = 1, 2, \dots\}$$

dont la distribution stationnaire est la densité a posteriori conjointe des paramètres.

### 3.3.3 Étude de Simulation

Nous avons simulé un échantillon d'observations généré par un modèle RCDAR(1,1) en utilisant l'algorithme de Gibbs. Pour cela les paramètres considérés

$\theta = (\eta, \mu_\beta, \sigma_\beta^2, \sigma_\epsilon^2, \sigma_\delta^2)^T$  afin d'évaluer le comportement de l'analyse Bayésienne.

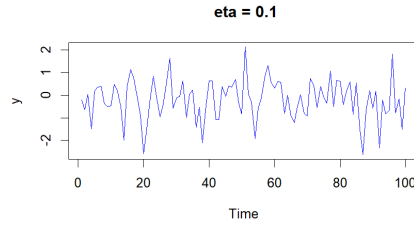
Et nous avons observé le comportement de ce modèle graphiquement en le comparant avec un AR(1). Tout d'abord, nous générons des données  $y_t, t = 1, 2, \dots, n$  à partir du modèle AR(1) en prenant  $\eta = 0.1, 0.5$  et  $0.9$  comme suit :

$$y_t = \eta y_{t-1} + \delta_t : \quad \text{AR}(1)$$

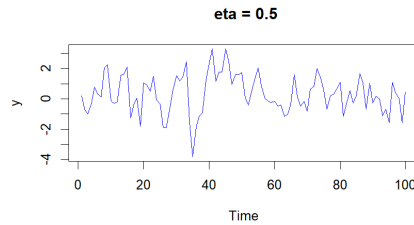
Ainsi, lorsque  $|\eta| < 1$ , les données sont stationnaires, mais lorsque  $|\eta|$  tend vers 1, les données deviennent presque non stationnaires. Dans ce cas, nous considérons  $0 < \eta < 1$ , puisque toutes les autocorrélations sont positives.

Pour illustrer l'implication du modèle AR(1), la figure 3.4 présente les tailles d'échantillon de 100 pour chacun des trois coefficients ( $\eta = 0.1, 0.5$ , et  $0.9$ ).

#Pour eta = 0.1:



#Pour eta = 0.5:



#Pour eta = 0.9:

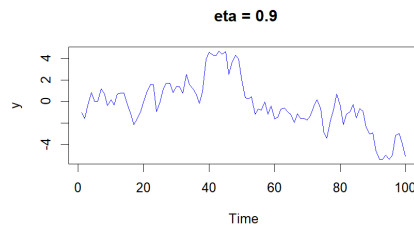


FIGURE 3.4 – Le graphique de la série temporelle des données générées à partir d'un modèle AR(1) pour  $y_t$  ( $\eta = 0.1, 0.5,$  et  $0.9$ ).

**Interprétation :** Dans le modèle AR(1), où la taille de l'échantillon est représentée par  $n$ , les différentes valeurs du coefficient autorégressif  $\eta$  ont un impact sur la stationnarité de la série temporelle générée. Plus précisément :

- Lorsque  $\eta = 0.1$ , la série est stationnaire.
- Lorsque  $\eta = 0.5$ , la série est faiblement stationnaire.
- Lorsque  $\eta = 0.9$ , la série est proche d'un processus non stationnaire.

Ensuite, nous considérons le modèle RCDAR(1,1) dans lequel les valeurs de  $y_t$  sont générées à partir d'un modèle AR(1). Par conséquent, nous obtenons les valeurs de  $X_t$  en termes du modèle RCDAR(1,1).

$$\begin{aligned} X_t &= y_t + \beta_t X_{t-1} + \epsilon_t \\ \beta_t &= \mu_\beta + \sigma_\beta u_t \end{aligned}$$

Dans les figures (3.5) à (3.7), nous présentons la génération de données dans 4 cas différents :

- Cas 1 :  $\sigma^2 = 1$ ,  $\mu_\beta = 0.5$ , et  $\sigma_\beta^2 = 0.25$
- Cas 2 :  $\sigma^2 = 1$ ,  $\mu_\beta = 1$ , et  $\sigma_\beta^2 = 0$
- Cas 3 :  $\sigma^2 = 1$ ,  $\mu_\beta = -0.995$ , et  $\sigma_\beta^2 = 0.01$
- Cas 4 :  $\sigma^2 = 1$ ,  $\mu_\beta = -0.1$ , et  $\sigma_\beta^2 = 0.99$

#Simuler un RCDAR(1,1) pour le cas  $\eta = 0.1$

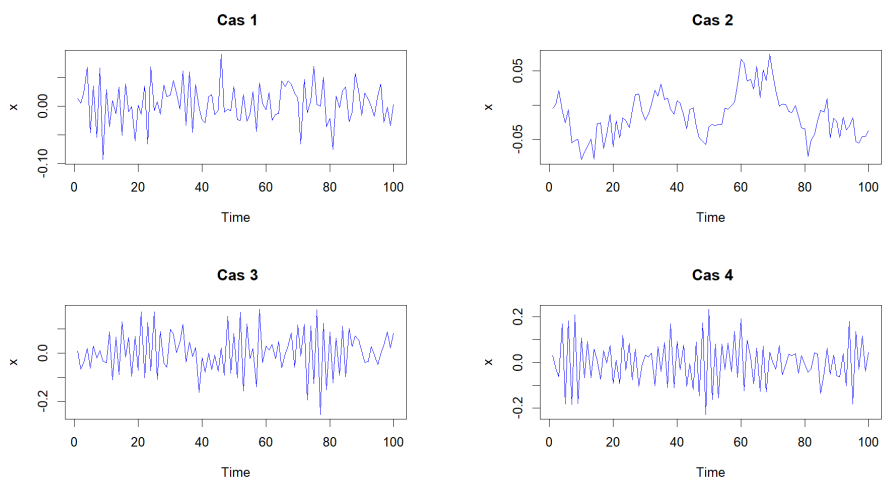


FIGURE 3.5 – Le graphique de la série temporelle des données générées à partir d'un modèle RCDAR(1,1) pour  $X_t$  ( $\eta = 0.1$ ).

#Simuler un RCDAR(1,1) pour le cas  $\eta = 0.5$

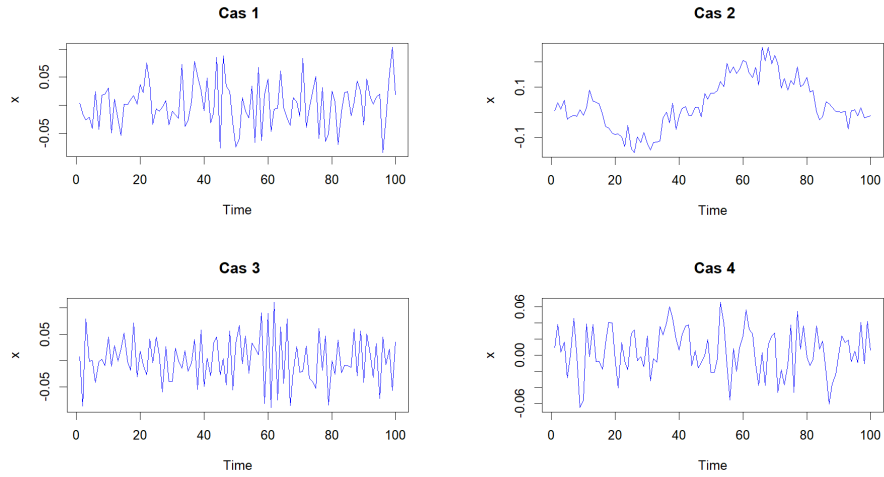


FIGURE 3.6 – Le graphique de la série temporelle des données générées à partir d'un modèle RCDAR(1,1) pour  $X_t$  ( $\eta = 0.5$ ).

#Simuler un RCDAR(1,1) pour le cas  $\eta = 0.9$

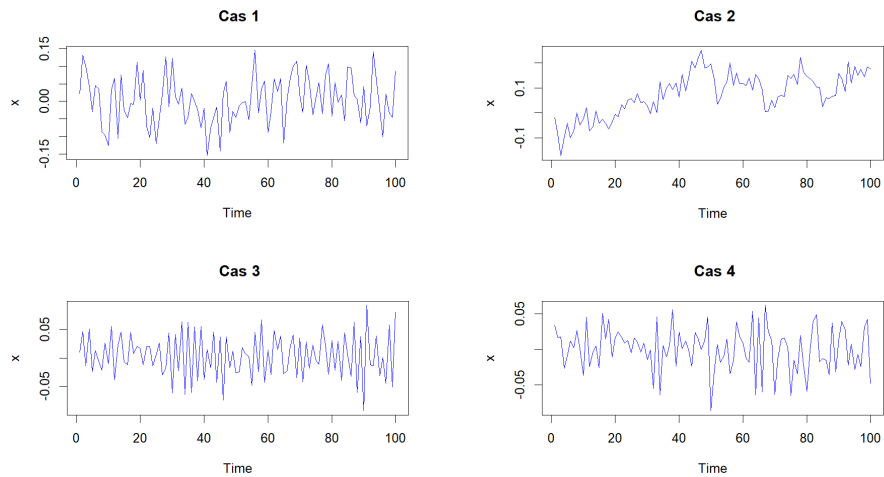


FIGURE 3.7 – Le graphique de la série temporelle des données générées à partir d'un modèle RCDAR(1,1) pour  $X_t$  ( $\eta = 0.9$ ).

**Interprétation :**

- L'auteur Araveporn Autcha, a noté que le cas 1 correspond à un processus stationnaire. Nicholls and Quinn [26] ont établi une condition pour  $\mu_\beta$  et  $\sigma_\beta^2$  : si  $\mu_\beta^2 + \sigma_\beta^2 < 1$ .
- Les cas 2 et 4 démontrent un processus non stationnaire, souvent appelé processus de marche aléatoire, défini par Nicholls and Quinn [26] comme ayant  $\mu_\beta^2 + \sigma_\beta^2 = 1$ . Cependant, le cas 4 dépend de  $\eta$  dans le contexte de  $y_t$ .
- En revanche, le cas 3 présente une tendance à osciller autour de sa moyenne zéro, semblable à un processus stationnaire, mais avec  $\mu_\beta^2 + \sigma_\beta^2 > 1$ .

# Conclusion générale

Dans ce mémoire, nous avons examiné en détail les modèles autorégressifs à coefficients aléatoires d'ordre 1 (RCAR(1)) et nous avons abordé des aspects cruciaux de l'inférence statistique dans ce contexte, en mettant en lumière le modèle RCDAR(1,1). Notre étude a été divisée en trois chapitres distincts, chacun jouant un rôle clé dans notre exploration.

Dans le premier chapitre, nous avons établi les notions générales essentielles. Le deuxième chapitre a été dédié à l'estimation des paramètres sous diverses approches. Quant au troisième chapitre, nous avons simulé les modèles RCAR(1) et RCDAR(1,1) et comparé les estimateurs.

Pour continuer à progresser dans ce domaine de recherche passionnant, plusieurs perspectives pour des travaux futurs se dessinent. Tout d'abord, il serait intéressant d'explorer des extensions des modèles RCAR(1) et RCDAR(1,1) afin de mieux modéliser des phénomènes encore plus complexes et dynamiques. Ensuite, nous pourrions mettre en pratique les méthodes que nous avons développées en les appliquant à des données réelles dans des domaines spécifiques, tels que l'économie, la finance ou d'autres secteurs, en résolvant des problèmes concrets.

Une autre piste de recherche consiste à approfondir l'analyse des résultats de la comparaison des estimateurs dans les deux modèles. Cela nous permettrait d'identifier les scénarios dans lesquels l'un des modèles pourrait être plus approprié que l'autre. De plus, nous pourrions mener une étude approfondie sur la sensibilité des méthodes d'estimation aux paramètres du modèle et aux conditions initiales, en tenant compte de la taille de l'échantillon.

L'optimisation des méthodes d'estimation constitue également un aspect crucial pour améliorer leur efficacité en termes de temps et de calcul. Parallèlement, il serait bénéfique de développer des techniques permettant d'évaluer les incertitudes associées aux estima-

---

tions des paramètres dans les deux modèles, ce qui renforcerait la robustesse de nos résultats.

Enfin, nous pourrions élargir notre comparaison en incluant d'autres modèles de séries chronologiques, offrant ainsi une évaluation plus complète de leurs comportements.

# Bibliographie

- [1] Anděl, J. (1976). *Autoregressive series with random parameters*. *Mathematische Operationsforschung und Statistik*, 7(5), 735-741.
- [2] Araveeporn, A. (2022). *A Comparison of Parameter Estimation Methods for the First-Order of Random Coefficient Autoregressive Model*. *Thailand Statistician*, 20(4), 892-904.
- [3] Araveeporn, A. (2012). *Bayesian Analysis of Random Coefficient Dynamic Autoregressive Model*. *Thailand Statistician*, 10(2), 199-223.
- [4] Atil, L. (2022). *Introduction aux statistiques Bayésiennes et applications*. Cours de 1<sup>ère</sup> année Master, U.M.M.T.O Tizi Ouzou.
- [5] Bayes, T. (1991). *An essay towards solving a problem in the doctrine of chances*. 1763. *MD computing : computers in medical practice*, 8(3), 157-171.
- [6] Billingsley, P. (1961). *The Lindeberg-Levy theorem for martingales*. *Proceedings of the American Mathematical Society*, 12(5), 788-792.
- [7] Billingsley, P. (1995). *Probability and measure*. John Wiley Sons (New York ;Chichester).
- [8] Boreux, J. J., Parent, E., Bernier, J., Bernier, J. (2010). *Pratique du calcul bayésien* (Vol. 118). Springer.
- [9] Brockwell, P. J., Davis, R. A. (2002). *Introduction to time series and forecasting*. New York, NY : Springer New York.
- [10] Brockwell, P. J., Davis, R. A. (2009). *Time series : theory and methods*. Springer science business media.
- [11] Conlisk, J. (1976). *A further note on stability in a random coefficient model*. *International Economic Review*, 759-764.
- [12] Conlisk, J. (1974). *Stability in a random coefficient model*. *International Economic Review*, 529-533.
- [13] Doukhan, P. (2010). *Séries temporelles non linéaires*. Cours de 2<sup>ème</sup> année Master, Université de Cergy-Pontoise.

- 
- [14] Fellag, H. (2022), *Séries chronologiques*. Cours de 1<sup>ère</sup> année Master, U.M.M.T.O Tizi Ouzou.
- [15] Gelfand A, Hills S, Racine PA, Smith A. (1990) *Illustration of Bayesian inference in normal data models using Gibbs sampling*. J Am Stat Assoc. ; 85 : 972-985.
- [16] Geman S, Geman D. (1984). *Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images*. IEEE Tran Pattern Anal Mach Intell. ; 6 : 721-741.
- [17] Gilks W, Richardson S, Spiegelhalter D. (1996). *Markov chain Monte Carlo in practice, interdisciplinary statistics*. Suffolk : Chapman Hall ; pp. 102-103.
- [18] Hastings, W. K. (1970). *Monte Carlo sampling methods using Markov chains and their applications*. Biometrika, 57(1) :97-109.
- [19] H. Tong.(1990). *Nonlinear Time Series*.Oxford University Press.
- [20] Hwang, S. Y., Basawa, I. V. (1993). *Asymptotic optimal inference for a class of nonlinear time series models*. Stochastic processes and their applications, 46(1), 91-113.
- [21] Kass, R. E., Wasserman, L. (1996). *Formal rules for selecting prior distributions : A review and annotated bibliography*. Journal of the American Statistical Association, 435, 1343-1370.
- [22] Lindley, D. V. (1990). *The 1988 Wald memorial lectures : the present position in Bayesian statistics*. Statistical Science, 44-65.
- [23] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). *Equation of state calculations by fast computing machines*. The Journal of Chemical Physics, 21(6) :1087-1092.
- [24] Mohammedi, Y. (2022). *Estimation des paramètres des modèles autorégressifs à coefficients aléatoires*. Mémoire de master en Probabilités et Statistique, U.M.M.T.O Tizi Ouzou.
- [25] Morton A, Finkenstadt BF. (2005). *Discrete time modelling of disease incidence time series by using Markov chain Monte Carlo methods*. J R Stat Ser C Appl Stat. ; 54(3) : 575-594.
- [26] Nicholls, D. F., Quinn, B. G. (2012). *Random Coefficient Autoregressive Models : An Introduction : An Introduction (Vol. 11)*. Springer Science Business Media.
- [27] Parent, E., Bernier, J. (2007). *Le raisonnement bayésien : modélisation et inférence*. Springer Science Business Media.
- [28] Quinn, B. G., Nicholls, D. F. (1981). *The Estimation Of Random Coefficient Autoregressive Models*. Ii. Journal of Time Series Analysis, 2(3), 185-203.
- [29] Quinn, B. G. (1982). *A note on the existence of strictly stationary solutions to bilinear equations*. Journal of Time Series Analysis, 3(4), 249-252.
- [30] Robert, C. P. (2001). *The Bayesian Choice*. New York : Springer.

- [31] Robert, C. P., Casella, G. (2004). *Monte Carlo Statistical Methods*, Springer, New York.
- [32] Wang, Dazhe. (2003). *Frequentist and Bayesian analysis of Random Coefficient Autoregressive models*. Ph.D. dissertation, North Carolina State University.