$N^o d'ordre$: ....

République Algerienne Démocratique et Populaire
Ministère de l'enseignement supérieur et de la Recherche scientifique
Université Mouloud Mammeri de Tizi Ouzou
Faculté des Sciences
Département de Mathématiques

Mémoire de Master

Filière : Mathématiques
Spécialité : Probabilités & Statistique

Présenté Par

Aouidad Dyhia

L'echantillonneur de Gibbs pour l'estimation bayésienne dans un modèle de régression linéaire

Soutenu devant le jury :

Mme Atil Lynda	M.C.A.	UMMTO	Présidente du jury
Mme Ait Mohammed Noura	M.A.A.	UMMTO	Rapporteuse
M Graiche Farid	M.C.B.	UMMTO	examinateur

Année Universitaire : 2021/2022

A mes chers parents, pour tous leurs sacrifices, leur amour, leur tendresse, leur soutien et leurs prières tout au long de mes études. Que ce travail soit l'accomplissement de vos vœux tant allégués, et le fruit de votre soutien infaillible,

Merci d'être toujours là pour moi . . .

Remerciements

Il m'est agréable d'adresser mes premiers remerciements au Dieu le tout puissant pour la persévérance, la santé, l'ouverture d'esprit et l'inspiration dont il m'a gratifié.
Ensuite je tiens à remercier ma promotrice Madame **Ait mohammed Noura**, épouse **Graiche**, qui a été d'un immense soutien pendant toute la durée de ce mémoire. Grâce à son expertise remarquable, elle a été une grande source d'inspiration et de conseils.

J'aimerais aussi remercier Madame Atil Lynda d'avoir présidé le jury et Monsieur Graiche Farid d'avoir examiné ce travail.

Je dédie ce travail à ma mère, qui a oeuvré pour ma réussite, de son amour, de son soutien, et de tous les sacrifices consentis durant toute ma vie.

À mon brave père, qui peut être fier et trouvera ici le résultat de longues années de sacrifices pour m'avoir aidé à avancer dans la vie. Merci pour les valeurs nobles, l'éducation et le soutient permanent venu de toi.

À mon frère, qui m'a soutenue durant ces années, d'avoir été là spécialement pendant les moments difficiles pour m'aider à réaliser ce travail.

Je tiens tout particulièrement à remercier Monsieur **MELLAL lyes**, pour son soutien, encouragement et pour avoir eu la patience de répondre à mes innombrables questions.

Je ne finirais pas sans pour autant remercier tous les enseignants qui m'ont encadrée tout au long de mon cursus et particulièrement ceux du Département des Mathématiques de l'Université Mouloud Mammeri de Tizi-Ouzou, de même que les camarades étudiants et amis du Master pour la collaboration et la discussion tout au long de ce travail.

Table des matières

Notations

$\theta, \rho, \sigma, \alpha, \beta$	Paramètres inconnus
$\hat{\theta}, \hat{\sigma}, \hat{\rho}$	Estimateurs de θ, σ et ρ
ξ	Vecteur des paramètres (θ, σ, ρ)
Θ	Ensemble des paramètres
s^2	Somme des carrés des écarts à la moyenne empirique
D	Espace des décisions
iid	Indépendantes et identiquement distribuée
$\mathbb{N}$	Ensemble des entiers naturels
C	Constante de normalisation
$\mathbb{Z}$	Ensemble des nombres entiers relatifs
$\mathbb{R}$	Ensembles des réels
$\mathbb{R}^m$	Ensemble des vecteurs réels à m dimensions
$\mathbb{R}^k$	Ensemble des vecteurs réels à k dimensions
π	Distribution de probabilité
$\mathbb{E}$	Espérance mathématique
$\mathcal{F}$	Famille des lois sur Θ

MCMC	Markov chain Monte carlo
MCO	Méthode des moindres carrées ordinaire
AR(p)	Processus autoregressif d'ordre p
$N(0, \sigma^2)$	Loi normale centrée
$N_m(\mu, \Sigma)$	Loi normale multivariée de moyenne μ et de matrice de covariance Σ
$t_p(\mu, V, \nu)$	Loi de student p-variée de moyenne μ, de matrice de covariance V et ν degrés de liberté
$\Gamma(k, \theta)$	Loi gamma de paramètres k et θ
I_n	Matrice identité d'ordre n

Introduction générale

« *N*othing in nature is random. A thing appears random only through the incompleteness of our knowledge »
Benedict Spinoza

Dans de nombreuses disciplines, les scientifiques sont confrontés à l'étude de phénomènes aléatoires, qui consiste à recueillir, traiter et interpréter des données dans le but de mener une inférence sur la distribution de probabilité qui est à l'origine de ce phénomène, c'est-à-dire de fournir une analyse (ou une description) d'un phénomène passé, ou une prédiction d'un phénomène à venir de nature similaire. Quand tel est le cas, des techniques statistiques appropriées doivent être employées.

Dans l'analyse des données statistiques, on s'intéresse à l'estimation de plusieurs paramètres, selon le modèle étudié. Le modèle linéaire avec erreurs autocorrélées figure parmi les modèles les plus étudiés, Citons ici les ouvrages fameux de Zellner et Tiao [1964], Parent et Bernier [2007], John et al. [1996], Girard et Parent [2000].
L'analyse du modèle de régression linéaire avec erreurs autocorrélées ne peut pas être menée sous une forme explicite, que l'on adopte le point de vue de la statistique classique ou celui de l'approche bayésienne. Toutefois, comme ce modèle résulte de l'association du modèle linéaire et du modèle autorégressif, il possède une structure conditionnelle adéquate qui permet de réaliser l'inférence bayésienne grâce à l'échantillonnage de Gibbs (voir Gelman et al. [1995] et Gelman et Gelman [1984]).

La méthode Bayésienne est basée sur l'estimation des paramètres issus d'une distribution nommée loi a posteriori, qui mélange l'information portée par les données (via la vraisemblance) et une information a priori des données (avis d'expert, contraintes physiques, ...),voir par exemple Robert [2006] et Robert [2013].

L'analyse bayésienne constitue un développement important de la modélisation des phénomènes complexes, l'utilisation de méthodes bayésiennes peut présenter certains avantages. En effet, ces méthodes utilisent des connaissances antérieures, exprimées sous forme de distribution de probabilité, afin d'apporter une information nouvelle. L'utilisation de lois de distributions offre une certaine souplesse pour modéliser des problèmes de phénomènes aléatoires.

Cependant, certains inconvénients demeurent : tout d'abord, la traduction de connaissances a priori en une loi de probabilité sur les paramètres est loin d'être triviale. De plus, la loi a posteriori s'exprime en général comme une distribution multivariée qui est délicate à manipuler dés que le nombre de paramètres est supérieur à deux ou trois.

Cette difficulté explique pourquoi l'utilisation de l'analyse Bayésienne est assez récente. On peut en effet affirmer qu'une grande part du renouveau des méthodes Bayésiennes est venu du fait que le calcul des distributions a posteriori et prédictives a été rendu possible grâce aux progrés de l'informatique et à l'utilisation systématique de méthodes de simulation basées sur les algorithmes de Gibbs, Metropolis-Hastings et les chaînes de Markov.

L'échantillonnage de Gibbs est applicable lorsque la distribution conjointe ne soit pas explicitement connue ou difficile à échantillonner directement. Mais la distribution conditionnelle de chaque paramètre est supposée être plus facile à déterminer. L'échantillonnage de Gibbs génère chaque paramètre suivant sa distribution conditionnelle aux autres paramètres. La séquence d'échantillons constitue une chaîne de Markov et la distribution stationnaire d'une telle chaîne est en fait la distribution conjointe recherchée.

L'objectif de ce mémoire est d'une part se familiariser avec l'approche bayésienne et d'autre part montrer l'efficacité des méthodes MCMC et plus particulièrement l'échantillonnage de Gibbs pour l'estimation bayésienne des paramètres d'un modéle de régression linéaire avec erreurs autocorrélées .

Ce mémoire est réparti en trois chapitres :

Dans le premier chapitre nous donnons les notions générales et les définitions essentielles dans l'analyse Bayésienne telles que les lois a priori, les lois a posteriori, ...

Le deuxième chapitre est consacré à l'estimation des paramètres et des observations futures d'un modèle de régression linéaire avec erreurs indépendantes ainsi qu'un modèle de régression linéaire avec erreurs autocorrélées. Nous avons estimé les paramètres du modèle de régression à erreurs autocorrélées en utilisant l'échantillonnage de Gibbs. Puis nous avons estimé une observation future en utilisant la densité prédictive a posteriori de cette observation.

Dans le chapitre trois, nous présentons les résultats des simulations pour l'estimation des paramètres d'un modèle de régression à erreurs autocorrélées (AR(1)) et la prédiction d'une valeur future.
Ce mémoire se termine par une conclusion générale et quelques perspectives.

1 Notions générales et défintions

1.1 Introduction

L'analyse Bayésienne, crée par Pierre-Simon de Laplace et Thomas Bayes (1774), a pour première étape l'étude d'une situation et d'identifier une incertitude portée sur un paramètre inconnu θ. Cette incertitude sur θ est modélisée sous la forme d'une distribution, dite a priori, qui apporte une information sur θ pris comme étant une variable aléatoire contrairement à l'analyse fréquentiste qui la considère comme une constante. Cette a priori est actualisée en extrayant de l'information contenue dans les observations, pour obtenir une autre distribution dite distribution a posteriori.

Le schéma ci dessous résume la démarche bayésienne :

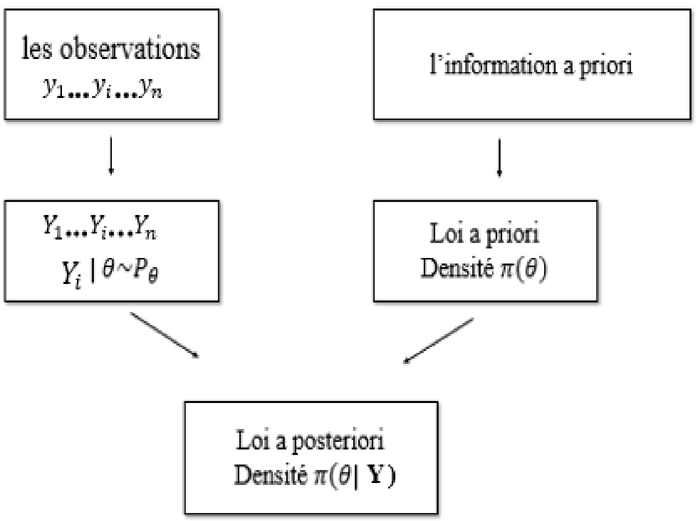

Figure 1.1 La démarche Bayésienne

1.2 Inférence Bayésienne

En général, l'objectif principal de la statistique paramétrique est de faire de l'inférence (estimations, tests d'hypothèses,...) et de la prévision au sujet d'un (ou plusieurs) paramètre d'intérêt θ, qui décrit un phénomène stochastique particulier. Pour cela, nous allons observer le phénomène, recueillir un échantillon de données $y = (y_1, y_2, ..., y_n)$ et calculer la fonction de densité de probabilité conditionnelle des données sachant θ, que l'on note $f(y|\theta)$. Cette densité conjointe, lorsque elle est regardée comme fonction de θ, est appelée fonction de vraisemblance et sera, en général, notée $l(\theta|y)$. Suivant que le paramètre θ soit considéré comme inconnu et déterministe ou comme variable aléatoire, on peut distinguer deux approches de l'inférence statistique : l'approche classique (ou basée sur la fonction de vraisemblance) et l'approche Bayésienne (en référence au théorème de Bayes) .
L'approche statistique Bayésienne qui se vaut une approche cohérente et pratique pour résoudre les problèmes d'inférence statistique, vise à exploiter le plus efficacement possible l'information apportée par y sur le paramètre θ, pour améliorer notre connaissance du paramètre θ et ensuite construire des procédures d'inférence sur θ.

Cette partie, fortement inspirée du livre intitulé : Le choix Bayésien, Principes et pratique, présente l'approche de l'inférence statistique Bayésienne. Pour une compréhension plus détaillée des concepts et des méthodes de la statistique Bayésienne, nous recommandons les références suivantes Bernardo et Smith [2009], Carlin et Louis [1997], Gelman et al. [1995] et Robert [2006].

1.2.1 Modèles Bayésiens

Définition 1.1 *Un modèle statistique Bayésien est constitué d'un modèle statistique paramétrique, $f(y|\theta)$, et d'une distribution a priori pour les paramètres $\pi(\theta)$.*

Théorème de Bayes
Supposons que $y' = (y_1, ...y_n)$ est un vecteur de n observations de distribution $\pi(y|\theta)$. Supposons aussi que le paramètre θ a une distribution de probabilité $\pi(\theta)$, alors

$$\pi(y|\theta)\pi(\theta) = \pi(y, \theta) = \pi(\theta|y)\pi(y) \tag{1.1}$$

Étant donné les observations $(y_1, ..., y_n)$, la distribution conditionnelle de θ est

$$\begin{aligned} \pi(\theta|y) &= \tfrac{\pi(y|\theta)\pi(\theta)}{\pi(y)} \\ &= C\pi(y|\theta)\pi(\theta) \end{aligned} \tag{1.2}$$

Où $C^{-1} = \pi(y) = \mathbb{E}(\pi(y|\theta))$.
La formule (1.1) ou son équivalente (1.2) est appelée la formule de Bayes. Dans cette formule, $\pi(\theta)$, qui exprime ce qu'on sait à propos de θ avant la connaissance des données y, est appelée la distribution a priori de θ. $\pi(\theta|y)$, qui exprime ce qu'on sait à propos de θ après la connaissance des données y, est appelée la distribution a posteriori de θ sachant y. Autrement dit, la densité a posteriori représente une actualisation de l'information a priori au vu de l'information apportée par les observations. La quantité C est la constante de "normalisation" nécessaire pour assurer que la somme de la distribution a posteriori $\pi(\theta|y)$ vaut 1.
$\pi(y|\theta)$ dans la formule (1.2) peut être considérée comme une fonction de θ et non de y. Selon Fisher (1922), elle est appelée la fonction de vraisemblance de θ sachant y qu'on note $l(\theta|y)$. Nous pouvons donc écrire la formule de Bayes comme

$$\begin{aligned} \pi(\theta|y) &= Cl(\theta|y)\pi(\theta) \\ &\propto l(\theta|y)\pi(\theta) \end{aligned} \tag{1.3}$$

où $\propto$ désigne le symbole de proportionnalité.
Donc la distribution a posteriori de θ sachant y est proportionnelle au produit de la loi a priori de θ et la fonction de vraisemblance $l(\theta|y)$.

Par application direct du théorème de Bayes, si θ est supposé être une variable aléatoire de densité a priori $\pi(\theta)$ et si $f(y|\theta) = l(\theta|y)$ est interprétée comme loi de y conditionnellement à θ, la loi de θ conditionnelle à y, $\pi(\theta|y)$ appelée distribution a posteriori de θ, définie par

$$\pi(\theta|y) = \frac{l(\theta|y)\pi(\theta)}{\int_\Theta l(\theta|y)\pi(\theta)d\theta} \tag{1.4}$$

Cette densité est centrale pour l'inférence Bayésienne, elle est suffisante pour déterminer les procédures de décision et, par extension, à conduire toute inférence liée à θ.

1.2.2 Loi a priori

Le choix de loi a priori des paramètres est l'aspect le plus délicat et le plus critiqué en analyse Bayésienne. En fait, ce choix peut avoir différentes motivations et diverses stratégies. Elles peuvent se baser sur des expériences anciennes ou sur une idée, une intuition que le statisticien a sur le mécanisme (physique, économique, biologique, etc) sous-jacent de génération du paramètre θ. Elles peuvent être également motivées par des aspects de calculs. En pratique, il est en effet très rare que l'information a priori disponible soit suffisament précise pour proposer une forme exacte ou même paramètrée pour la distribution a priori sur θ. Ce qui nous amène à faire des approximations. Dans la suite, nous donnons deux approches pour l'approximation des lois a priori, à savoir, les lois a priori conjuguées et les lois a priori non informatives.

Lois a priori conjuguées

Introduite par Raiffa et al. [1961], l'approche a priori conjuguée, peut être justifiée partiellement par un raisonnement d'invariance. En fait, quand l'observation de $y \sim f(y|\theta)$ modifie $\pi(\theta)$ en $\pi(\theta|y)$, l'information transmise par y sur θ est évidemment limitée ; par conséquent, elle ne devrait pas entraîner une modification de toute la structure de $\pi(\theta)$, mais simplement de ses paramètres. En d'autres termes, la modification apportée par les observations doit rester de dimension finie.
Le choix d'un a priori conjugué, bien qu'il soit défendable, est toujours un choix particulier et influence donc, dans une certaine mesure, l'inférence résultante.

Définition 1.2 *Une famille F de lois sur Θ est dite conjuguée si, pour tout π appartenant à cette famille, la loi $\pi(\theta|y)$ appartient également à F.*

Exemple 1.1

$f(y\|\theta)$	$\pi(\theta)$	$\pi(\theta\|y)$
Normale $\mathcal{N}(\theta, \sigma^2)$	*Normale* $\mathcal{N}(\mu, \tau^2)$	$\mathcal{N}(\rho(\sigma^2\mu + \tau^2 y), \rho\sigma^2\tau^2)$ $\rho^{-1} = \sigma^2 + \tau^2$
Poisoon $\mathcal{P}(\theta)$	*Gamma* $\Gamma(\alpha, \beta)$	$\Gamma(\alpha + y, \beta + 1)$
Gamma $\Gamma(\nu, \theta)$	*Gamma* $(\Gamma(\alpha, \beta)$	$\Gamma(\alpha + \nu, \beta + y)$
Binomiale $\mathcal{B}(n, \theta)$	*Beta* $Be(\alpha, \beta)$	$Be(\alpha + y, \beta + n - y)$
Binomiale Négative $Neg(m, \theta)$	*Beta* $Be(\alpha, \beta)$	$Be(\alpha + m, \beta + y)$
Multinomiale $M_k(\theta_1, ..., \theta_k)$	*Dirichlet* $\mathcal{D}(\alpha_1, ..., \alpha_k)$	$\mathcal{D}(\alpha_1 + y_1, ..., \alpha_k + y_k)$
Normale $\mathcal{N}(\mu, 1/\theta)$	*Gamma* $\Gamma(\alpha, \beta)$	$\Gamma(\alpha + 0.5, \beta + (\mu - y)^2/2)$

Table 1.1 Lois a priori conjuguées usuelles.

Familles exponentielles

Les lois a priori conjuguées sont généralement associées à un type particulier de lois d'échantillonnage qui permet toujours leur obtention, ces lois constituent ce qu'on appelle des familles exponentielles.

Définition 1.3 *On appelle famille exponentielle à s-paramètres, toute famille de loi de distribution P_θ dont la densité a la forme suivante :*

$$f(y|\theta) = exp[\Sigma_{i=1}^{s}\eta_i(\theta)T_i(y) - B(\theta)]h(y) \tag{1.5}$$

Où $\eta_i(.)$et $B(.)$ sont des fonctions du paramètre θ et $T_i(.)$ sont des statistiques.

Exemple 1.2
Loi binomiale

$$\begin{aligned} P(Y = y|\theta) &= C_n^y \theta^y (1-\theta)^{n-y} \\ &= C_n^y exp[ylog\theta + (n-y)log(1-\theta)] \\ &= C_n^y exp[ylog[\theta/(1-\theta)] + nlog(1-\theta)] \end{aligned}$$

On a $s = 1$, $\eta_1(\theta) = log(\theta/(1-\theta))$, $T_1(y) = y$, $B(\theta) = nlog(1-\theta)$ et $h(y) = C_n^y$

Remarque 1.1
L'écriture du modèle exponentiel sous la forme canonique est

$$f(y|\theta) = exp[\Sigma_{i=1}^{s}\theta_i T_i(y) - B(\theta)]h(y) \tag{1.6}$$

Ce qui nous donne la forme des lois naturelles conjuguées dans le cas du modèle exponentiel.

Lois conjuguées des familles exponentielles

Proposition 1.1 (Robert [2006])

Soit $f(y|\theta)$ appartenant à une famille exponentielle. Alors une famille de loi a priori conjuguée pour $f(y|\theta)$ est donnée par :

$$\pi(\theta|\mu,\lambda) = K(\mu,\lambda)exp(\theta\mu - \lambda B(\theta)) \tag{1.7}$$

où $K(\mu, \lambda)$ est une constante de normalisation.

La loi a posteriori est de la forme

$$\pi(\theta|y) \propto exp((\mu + y)\theta - (\lambda + 1)B(\theta)) \tag{1.8}$$

Lois a priori non informatives

Lorsque aucune information a priori n'est disponible, l'utilisation des lois a priori conjuguées n'est justifié que par des considération analytique. De plus, dans de telles situations, il est impossible de justifier le choix d'une loi a priori sur des bases subjectives et les hyper-paramètres des lois conjuguées ne peuvent être déterminés que d'une manière arbitraire. On peut alors chercher à utiliser d'autres techniques qui intègre notre ignorance totale sur les paramètres du modèle. De telles lois sont dites, de manière évidente, non informatives. Nous décrivons dans la suite quelques-unes des techniques les plus importantes de construction de lois non informatives. Pour plus de détails concernant ces techniques, voir Robert [2006].

Les lois a priori de Laplace

Historiquement, Laplace fut le premier à utiliser des techniques non informatives. Bien qu'il ne dispose pas d'information sur les paramètres du modèle, il munit ces paramètres d'une loi a priori qui prend en compte son ignorance en donnant la même vraisemblance à chaque valeur du paramètre, soit donc en utilisant une loi uniforme. Son raisonnement, appelé plus tard principe de la raison insuffisante, se fondait sur l'équiprobabilité des évènements élémentaires. Des critiques ont été plus tard avancées sur ce choix. La plus fondamentale, concerne le problème de l'invariance par paramétrisation. Si on passe de $\theta \in \Theta$ à $\nu = g(\theta)$ par une transformation bijective g, l'information a priori ne devrait pas être modifiée, car inexistante. Cependant, si $\pi(\theta) = 1$, la loi a priori sur ν est

$$\pi^*(\eta) = |\frac{d}{d\eta}g^{-1}(\eta)|$$

Par la formule du changement de variable.
Donc $\pi(\eta)$ est le plus souvent non constante.

Lois invariantes

Le fait de formaliser l'absence d'information a priori par une propriété d'invariance est naturelle : si le paramètre $\theta \in \Theta$ représente la masse d'un objet et si l'on possède une information sur θ, la distribution a priori de θ et de $\theta + a$ ne sont certainement pas identiques. Cependant, dire que la distribution de θ et celle de $\theta + a$ sont les mêmes pour tout a exprime certainement l'ignorance sur la valeur de θ.

La loi a priori de Jeffreys
Jeffreys (1946, 1961) propose une approche intrinsèque qui évite effectivement le besoin de prendre en compte une structure d'invariance potentielle, tout en étant souvent compatible lorsque cette structure existe. Les lois a priori non informatives de Jeffreys sont fondées sur l'information de Fisher, donnée par

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

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

$$I(\theta) = -\mathbb{E}_\theta[\frac{\partial^2 log f(y|\theta)}{\partial\theta^2}] \tag{1.9}$$

La loi a priori non informative de Jeffreys est

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

Le choix d'une loi a priori dépendant de l'information de Fisher se justifie par le fait que $I(\theta)$ est largement accepté comme un indicateur de la quantité d'information apportée par le modèle (ou l'obsevation) sur θ (Fisher [1956]).
Dans le cas où θ est un paramètre multidimensionnel, on définit la matrice d'information de Fisher par généralisation de (1.9).
Pour $\theta \in \mathbb{R}^k$, $I(\theta)$ a les éléments suivant :

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

Et la loi non informative de Jeffreys est alors définie par

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

La loi $\pi(\theta)$ vérifie effectivement l'exigence d'invariance par reparamètrisation, puisque, pour une transformation bijective h donnée, nous avons la transformation (jacobienne)

$$I(\theta) = I(h(\theta))(h^{'}(\theta))^2$$

Exemple 1.3
Soit $y \sim N(\mu, \sigma^2)$ avec $\theta = (\mu, \sigma)$ inconnu. Dans ce cas

$$I(\theta) = \mathbb{E}_\theta \begin{pmatrix} 1/\sigma^2 & 2(y-\mu)/\sigma^3 \\ 2(y-\mu)/\sigma^3 & 3(\mu-y)^2/\sigma^4 - 1/\sigma^2 \end{pmatrix} = \begin{pmatrix} 1/\sigma^2 & 0 \\ 0 & 2/\sigma^2 \end{pmatrix}$$

Et la loi non informative de Jeffreys associée est $\pi(\theta) \propto 1/\sigma^2$.

1.2.3 Estimateur de Bayes

En pratique, le principe fondamentale de l'approche Bayésienne est que toute inférence statistique devrait se fonder sur la détermination rigoureuse de trois éléments :

1. La famille des lois des observations $f(y|\theta)$.
2. La distribution a priori des paramètres $\pi(\theta)$
3. Le coût associé aux décisions $L(\delta, \theta)$.

D'un point de vue décisionnel, le modèle statistique inclut trois espaces : Y, espace des observations, Θ, espace des paramètres, et D, espace des décisions (ou espace d'action). En pratique, l'inférence statistique consiste à prendre une décision $d \in D$ par rapport au paramètre $\theta \in \Theta$, fondée sur l'observation $y \in Y$, et il est important de pouvoir comparer les différentes décisions au moyen d'un critére d'evaluation, qui va apparaître sous forme d'un coût.

Définition 1.4 *On appelle coût toute fonction L de $\Theta \times D$ dans $[0, +\infty[$.*

$$L(\theta, \delta) = L(\theta, \delta(y))$$

La fonction L évalue la pénalité résultante de l'emploi de la décision $\delta \in D$ quand le paramètre prend la valeur θ.

Définition 1.5 *Une règle de décision δ est une application de Y dans D ;*

Dans la pratique, la détermination de la fonction de coût est souvent difficile, en particulier parce que les conséquences de chaque action pour chaque valeur de θ sont souvent impossibles à déterminer quand D ou Θ sont de grands ensembles, par exemple quand ils contiennent un nombre infini d'éléments.
La complexité de la détermination de la fonction de coût subjective du décideur incite souvent le statisticien à recours aux fonctions de coût classiques ou canoniques, choisies pour leur simplicité et leur souplesse mathématique (voir Robert [2006]).

Définition 1.6 *Le coût moyen ou risque de la décision δ est défini par*

$$\begin{aligned} R(\theta, \delta) &= \mathbb{E}_\theta[L(\theta, \delta(y)] \\ &= \int_Y L(\theta, \delta(y)) f(y|\theta) dy \end{aligned} \tag{1.10}$$

La fonction δ, de Y dans D est appelée un estimateur (tandis que la valeur $\delta(y)$ est appelée estimation de θ).
Ainsi, l'erreur (coût) est moyennée sur toutes les valeurs de y, proportionnellement à la densité $f(y|\theta)$. Par contre, l'approche Bayésienne de la théorie de la décision intègre sur l'espace Θ, car θ est inconnu, plutôt que de le faire sur

l'espace Y, y étant connu.
Dans le cadre de l'estimation, l'ensemble des décisions D est lui même l'espace des paramètres Θ, et une règle de décision est un estimateur.
Dans l'approche classique, la règle de décision optimale δ doit minimiser $R(\theta,\delta)$ pour tout $\theta \in \Theta$. Mais en fait, il est difficile de trouver une telle décision.

Dans l'approche Bayésienne, pour chaque valeur de y, on cherche la décision $\delta(y)$ qui minimise le coût a posteriori.

Définition 1.7 *Pour une loi a priori donnée, le coût moyen a posteriori est défini par :*

$$\varrho(\pi,\delta|y) = \mathbb{E}^{\pi}(L(\theta,\delta|y) = \int_{\Theta} L(\theta,\delta|y)\pi(\theta|y)d\theta$$

qui moyenne l'erreur (le coût) selon la distribution a posteriori du paramètre θ, conditionnellement à la valeur observée y.

En se donnant une distribution a priori π, il est aussi possible de définir le risque intégré, qui est le risque fréquentiste moyenné sur les valeurs de θ selon leur distribution a priori.

Définition 1.8 *En se donnant une distribution a priori π, le risque intégré est défini par*

$$r(\pi,\delta) = \mathbb{E}^{\pi}[R(\theta,\delta)] = \int_{\Theta}\int_{Y} L(\theta,\delta(y))f(y|\theta)dy\pi(\theta)d\theta$$

qui est le risque fréquentiste moyenné sur les valeurs de θ selon leur distribution a priori π.

Un intérêt particulier de ce concept de risque intégré est qu'il associe un nombre réel à chaque estimateur. Il induit donc un ordre total sur l'ensemble des estimateurs et permet une comparaison directe entre ces estimateurs.

Théorème 1.1 *(Robert [2006])*
Un estimateur minimisant le risque intégré $r(\pi,\delta)$ est obtenu par sélection, pour chaque $y \in Y$, de la valeur $\delta(y)$ qui minimise le coût moyen a posteriori, $\varrho(\pi,\delta|y)$, puisque

$$r(\pi,\delta) = \int_{Y} \varrho(\pi,\delta|y)m(y)dy$$

Définition 1.9 *On définit le risque de Bayes par*

$$\begin{aligned} r(\pi) = r(\pi,\delta^{\pi}) &= E_{\pi}[R(\theta,\delta)] \\ &= \int_{\theta}\int_{Y} L(\theta,\delta(y))f(y|\theta)\pi(\theta)dyd\theta \end{aligned} \tag{1.11}$$

Définition 1.10 *On appelle estimateur de Bayes associé à un coût L et à une distribution a priori π, toute décision δ^{π} qui minimise le risque de Bayes $r(\pi) = r(\pi, \delta^{\pi})$, on a*

$$\delta^{\pi}(y) = \underset{\delta \in D}{argmin} r(\pi, \delta)$$

Exemple 1.4
Dans le cadre de l'estimation, l'ensemble des décisions est l'espace des paramètres Θ, et une règle de décision est un estimateur.

Une fonction de coût couramment utilisée est la fonction de coût quadratique :

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

avec $(\theta, \delta) \in \Theta \times \Theta$
L'estimateur de Bayes δ^{π} du paramètre θ ($\theta \in \mathbb{R}$), associé à une distribution a priori π, est la moyenne a posteriori :

$$\delta^{\pi}(y) = E^{\pi}(\theta|y) = \int_{\Theta} \theta \pi(\theta|y) d\theta$$

En effet comme,

$$E^{\pi}([\theta - \delta(y)]^2|y) = E^{\pi}(\theta^2|y) - 2\delta(y)E^{\pi}(\theta|y) + \delta^2(y)$$

Le minimum du coût a posteriori est effectivement atteint en

$$\delta^{\pi}(y) = E^{\pi}(\theta|y)$$

1.2.4 Fonctions de coût usuelles

Le coût quadratique

Définition 1.11 *La fonction de coût quadratique est la fonction définie par*

$$L(\theta, \delta(y)) = (\theta - \delta(y))^2 \tag{1.12}$$

Proposition 1.2. (Robert [2006])

L'estimateur de Bayes δ^{π} associé à la loi a priori π et au coût quadratique (1.12) est la moyenne a posteriori

$$\delta^{\pi}(y) = E^{\pi}(\theta|y) = \frac{\int_{\Theta} \theta f(\theta|y)\pi(\theta) d\theta}{\int_{\Theta} f(\theta|y)\pi(\theta) d\theta}$$

Corollaire 1.1 (Robert [2006])

Quand $\theta \in R^p$, l'estimateur de Bayes δ^{π} associé à π et au coût quadratique,

$$L(\theta, \delta(y)) = (\theta - \delta)^t Q(\theta - \delta)$$

est la moyenne a posteriori, $\delta^\pi(y) = \mathbb{E}^\pi[\theta|y]$, pour toute matrice $Q_{p\times p}$ symétrique définie positive.
Le coût quadratique est particulièrement intéressant lorsque l'espace des paramètres est borné et le choix d'un coût plus subjectif est impossible. Mais il a donné lieu à de nombreuses critiques, la plus fréquente étant qu'il pénalise trop fortement les grandes erreurs, à cause de sa convexité stricte.

Le coût absolu

Définition 1.12 *La fonction de coût absolu est la fonction définie par*

$$L(\theta, \delta) = |\theta - \delta| \tag{1.13}$$

Ou plus généralement une fonction définie par morceaux

$$L(\theta, \delta) = \begin{cases} k_1(\theta - \delta) & si \quad \theta > \delta \\ k_2(\delta - \theta) & sinon \end{cases} \tag{1.14}$$

Cette fonction evalue le coût d'une décision "δ" quand le paramètre vaut θ. Elle augmente moins rapidement que le coût quadratique et, tout en restant convexe, évite donc de pénaliser trop les grandes erreurs.

Proposition 1.3. (Robert [2006])

Un estimateur de Bayes associé à π et au coût (1.14) est un fractile d'ordre $k_1/(k_1 + k_2)$ de $\pi(\theta|y)$.
En particulier, si $k_1 = k_2$, l'estimateur de Bayes est la médiane a posteriori.

1.3 Méthodes de calcul Bayésien

L'inférence Bayésienne est basée sur le calcul des caractéristiques a posteriori de θ ou de certaines fonctions du type $u(\theta)$ et de leurs espérences mathématiques :

$$U = \int u(\theta)\pi(\theta|y)d\theta \tag{1.15}$$

Dans la pratique, le calcul de (1.15) peut être rendu difficile pour deux raisons :

1. Le calcul explicite de la loi a posteriori, $\pi(\theta|y)$, peut être impossible.
2. Même si $\pi(\theta|y)$ est connu, lorsque l'intégration analytique est impossible, le calcul numérique nécessite parfois un temps de calcul considérable, en particulier lorsque Θ est de grande dimension.

L'objectif dans cette partie est le calcul pratique de $\pi(\theta|y)$ et des intégrales de la forme (1.15).

1.3.1 Méthodes de Monte Carlo par chaînes de Markov

Dans cette section nous allons présenter des méthodes de Monte carlo qui permettent de générer des variables aléatoires suivant approximativement la loi $\pi(\theta|y)$. Les méthodes de Monte carlo par chaînes de Markov, ou méthodes MCMC (Markov chain Monte Carlo), sont des méthodes d'intégration numérique utilisant l'échantionnage à partir de distributions de probabilité.

Une première version de l'algorithme MCMC a été donnée il y a 50 ans par Metropolis et al. [1953] dans un contexte de physique statistique, et généralisé plus tard aux problèmes statistiques par Hastings [1977]. L'idée de base des méthodes est de créer une longue chaîne de Markov, $\{\theta_t\}$ dont les échantillons sont distribués asymptotiquement selon la distribution requise $\pi(\theta|y)$. Dans la suite, nous présenterons les deux types de techniques les plus importantes conçues pour créer des chaînes de Markov de lois stationnaires données, à savoir les algorithmes de Metropolis-Hastings et l'échantillonnage de Gibbs.

Algorithme de Metropolis-Hastings

L'algorithme de Metropolis-Hastings peut être décrit de la façon suivante. Pour une densité donnée $\pi(\theta|y)$, connue à un facteur de normalisation près, et une densité conditionnelle $q(\theta^{'}|\theta, y)$, l'algorithme génère la chaîne $(\theta^{(m)})_m$ comme suit :

- Itération 0 : Initialiser avec une valeur arbitraire $\theta^(0)$
- Itération m : Mettre à jour $\theta^{(m)}$ par $\theta^{(m+1)}(m = 1, 2, ...)$, de la façon suivante
  - **a** Générer $\psi \sim q(\psi|\theta^{(m)}, y)$
  - **b** Poser $\varrho(\theta^{(m)}, \psi) = min\{\frac{\pi(\psi|y)q(\theta^{(m)}|\psi,y)}{\pi(\theta^{(m)}|y)q(\psi|\theta^{(m)},y)}, 1\}$
  - **c** Prendre $\theta^{(m+1)} = \begin{cases} \psi & \text{avec probabilité} \quad \varrho(\theta^{(m)}, \psi) \\ \theta^{(m)} & \text{sinon} \end{cases}$

La convergence théorique de l'algorithme de Metropolis-Hastings est prouvée par Robert et Smith [1993]. La loi de densité $\pi(\theta|y)$ est souvent appelée loi cible ou loi objet, tandis que la loi de densité $q(.|\theta)$ est dite loi de proposition.
Remarque 1.2

q est quelconque mais elle doit vérifier certaines conditions, telles que :

1. q doit être simulable.
2. Le support de q doit couvrir le support de π.
3. q doit être une bonne approximation de π.

L'échantillonneur de Gibbs

L'approche de l'échantillonnage de Gibbs tire son nom des champs aléatoires de Gibbs, où elle a été utilisée pour la première fois par Gelman et Gelman [1984] comme un moyen pour simuler des distributions complexes de grande dimension qui se posent dans le cadre de la restauration d'images et développé plus tard dans le cadre du calcul Bayésien par Gelfand et Smith [1990].

L'échantillonneur de Gibbs est une methode de Monte-carlo par chaîne de Markov (MCMC) permettant d'obtenir des échantillons à partir d'une distribution conjointe par échantillonnage itérative à partir des distribution conditionnelles complètes. Une bonne introduction à l'échantillonneur de Gibbs est donnée par Casella et George [1992]. L'échantillonnage de Gibbs tire profit des structures hiérarchiques d'un modèle, i.e, lorsque celui-ci peut s'écrire sous la forme

$$\pi(\theta_1|y) = \int \pi_1(\theta_1|\theta_2, y)\pi_2(\theta_2|y)d\theta_2 \tag{1.16}$$

L'idée est alors de simuler la loi jointe $\pi_1(\theta_1|\theta_2, y) \times \pi_2(\theta_2|y)$, afin d'obtenir $\pi(\theta_1|y)$ comme loi marginale. Bien entendu, lorsque les deux lois $\pi_1(\theta_1|\theta_2, y)$ et $\pi_2(\theta_2|y)$ sont connues et peuvent être simulées, la génération de θ_1 de $\pi(\theta_1|y)$ est équivalente à la génération de θ_2 de $\pi_2(\theta_2|y)$, puis de θ_1 de $\pi_1(\theta_1|\theta_2, y)$.

La formulation générale de l'algorithme d'échantillonnage de Gibbs pour une loi jointe $\pi(\theta_1, ..., \theta_p|y)$, de lois conditionnelles complètes $\pi(\theta_1|\theta_2, \theta_3, ..., \theta_p, y)$, $\pi(\theta_2|\theta_1, \theta_3, ..., \theta_p, y)$, ..., $\pi(\theta_p|\theta_2, \theta_3, ..., \theta_{p-1}, y)$; supposées disponibles (simulables) ; est exposée ci-dessous.

- Initialisation : Commencer par une valeur arbitraire $(\theta_1^0, ..., \theta_p^0)$.
- Itération $t + 1$: Pour $(\theta_1^{(t)}, ..., \theta_p^{(t)})$ donnée, générer
  - $\theta_1^{(t+1)}$ selon la loi $\pi(\theta_1|\theta_2^{(t)}, \theta_3^{(t)}, ..., \theta_p^{(t)}, y)$,
  - $\theta_2^{(t+1)}$ selon la loi $\pi(\theta_2|\theta_1^{(t+1)}, \theta_3^{(t)}, ..., \theta_p^{(t)}, y)$
  - $\theta_3^{(t+1)}$ selon la loi $\pi(\theta_3|\theta_1^{(t+1)}, \theta_2^{(t+1)}, ..., \theta_p^{(t)}, y)$
  - ...
  - ...
  - $\theta_p^{(t+1)}$ selon la loi $\pi(\theta_p|\theta_1^{(t+1)}, \theta_2^{(t+1)}, ..., \theta_{p-1}^{(t+1)}, y)$,

On peut montrer que le vecteur $\theta^{(t+1)} = (\theta_1^{(t+1)}, \theta_2^{(t+1)}, \theta_3^{(t+1)}, ..., \theta_p^{(t+1)})$, ainsi obtenu converge en loi vers un échantillon provenant de la distribution $\pi(\theta_1, ..., \theta_p|y)$ lorsque t devient grand Gelfand et Smith [1990].

Théorème 1.2 *(Parent et Bernier [2007])*
La suite de vecteurs p-dimensionnels simulés $\{\theta_i\}$ par l'algorithme précédent est une chaîne de Markov dont la distribution invariante, limite ergodique de la chaîne, est la distribution conjointe :

$$\pi(\theta|y) = \pi(\theta_1, \theta_2, ..., \theta p|y)$$

Il en résulte que chaque sous-suite $\{\theta_j^t\}$ a la distribution marginale $\pi(\theta_j|y)$ comme densité limite ergodique.

Ainsi, si on répète tout le processus n fois en parallèle, on obtiendra un échantillon de n vecteurs $\theta_1^{(t+1)}, \theta_2^{(t+1)}, \theta_3^{(t+1)}, ..., \theta_n^{(t+1)}$ indépendants et identiquement distribués selon la loi jointe $\pi(\theta_1, \theta_2, ..., \theta_p|y)$. L'échantillon $\theta_1^{(t+1)}, \theta_2^{(t+1)}, \theta_3^{(t+1)}, ..., \theta_n^{(t+1)}$ peut être utilisé pour l'approximation des densités marginales ou toute autre caractéristique de la loi jointe (Espérances, intervalles HPD crédibles, quantilles, etc) par de simples sommes finies d'expressions connues.
Ainsi, la densité marginale, $\pi(\theta_i|y)$; $i = 1, ..., p$ peut-être approchée comme suit (voir Casella et George [1992]).

$$\pi(\theta_i|y) = \frac{1}{n}\sum_{j=1}^{n} \pi(\theta_i|\theta_{1j}^{(t+1)}, \theta_{2j}^{(t+1)}, ..., \theta_{(i-1)j}^{(t+1)}, \theta_{(i+1)j}^{(t+1)}, ..., \theta_{pj}^{(t+1)}, y)$$

où $\theta_{(j)}^{(t+1)} = (\theta_{1j}^{(t+1)}, \theta_{2j}^{(t+1)}, ..., \theta_{pj}^{(t+1)})$, $j = 1, .., n$.

Pour toute fonction mesurable T de $(\theta_1, ..., \theta_p)$, on a

$$\frac{1}{n}\sum_{j=1}^{n} T(\theta_{1j}^{(t+1)}, \theta_{2j}^{(t+1)}, ..., \theta_{pj}^{(t+1)}|y) \quad \xrightarrow{p.s} \quad \mathbb{E}(T(\theta_1, ..., \theta_p|y))$$

L'algorithme de Gibbs est particulièrement bien adapté au calcul et à l'estimation des modèles Bayésiens, hiérarchiques notament, et il exploite au mieux la structure conditionnelle des équations de modélisation Parent et Bernier [2007]. pour plus de détails sur l'application de l'algorithme de l'échantillonneur de Gibbs au calcul Bayésien général on peut voir Gelfand et Smith [1990].

1.4 Les modèles autorégressifs (AR)

Définition 1.13 *Le processus $\{Y_t, t \geq 0\}$ est un processus autorégressif d'ordre p ($p \geq 0$) si pour tout t, il peut s'écrire sous la forme suivante :*

$$Y_t = \sum_{j=1}^{p} \phi_j Y_{t-j} + \epsilon_t \quad ou \quad \epsilon_t \sim \mathcal{N}(0, \sigma^2) \tag{1.17}$$

Pour $t = 0, 1, 2...$ et $\epsilon_t \perp \epsilon_s$ pour $t \neq s$.

Dans ce cas, on note $Y_t \sim AR(p)$, $\phi = (\phi_1, \phi_2, ..., \phi_p)'$ est le vecteur des paramètres autorégressifs.
Le modèle autorégressif d'ordre p est un modèle de régression pour les series temporelles dans lequel la série est expliquée par ses valeurs passées plutôt que par d'autres variables.

Exemple 1.5
Un processus $AR(1)$ prend la forme suivante :

$$Y_t = \phi Y_{t-1} + \epsilon_t \quad ou \quad \epsilon_t \sim \mathcal{N}(0, \sigma^2) \tag{1.18}$$

1.4.1 Opérateur de retard

Définition 1.14 *L'opérateur de retard B se définit de la manière suivante :*

$$B(Y_t) = Y_{t-1}$$

L'opérateur de retard B peut être utilisé pour décrire la condition nécessaire de la stationnarité du processus autorégressif AR(p).
Remarque 1.3
Pour $n \in \mathbb{N}, B^n(Y_t) = Y_{t-n}$
En utilisant l'opérateur de retard B, l'équation (1.17) peut être écrite sous la forme compacte suivante :

$$\Phi(B)Y_t = \epsilon_t \tag{1.19}$$

Où Φ est le polynôme d'ordre p suivant :

$$\Phi(z) = 1 - \phi_1 z - \phi_2 z^2 - ... - \phi_p z^p$$

$\Phi(z)$ est appelé le polynôme caractéristique du processus et ses racines déterminent si le processus autorégressif est stationnaire ou non.

Théorème 1.3 *Si $Y_t \sim AR(p)$,Y_t est stationnaire si et seulement si le module de toutes les racines du polynôme caractéristique du processus sont supérieures ou égale à 1.*

Exemple 1.6
Considérons le processus AR(1) :

$$Y_t = \phi Y_{t-1} + \epsilon_t, \quad t \geq 1$$

Le processus peut être écrit comme :

$$(1 - \Phi B)Y_t = \epsilon_t$$

Le polynôme caractéristique est donné par :

$$\Phi(z) = (1 - \phi z)$$

La seule racine du polynôme est $z = \frac{1}{\phi}$ $(\phi \neq 0)$.
Le processus AR(1) est stationnaire si et seulement si $|\phi| \leq 1$.

Pour plus de détails sur les caractéristiques du processus AR(p), on peut voir par exemple Box et al. [2015].

1.5 Lois de probabilités usuelles

Nous donnons dans cette partie les lois de probabilités utilisées dans ce mémoire, a savoir la loi Normale, la loi de Student et la loi gamma.

1.5.1 Densité normale multivariée

Soit un vecteur aléatoire Y de moyenne $\mu \in R^m$ et de matrice de covariance Σ d'ordre m définie positive. La variable Y suit la loi normale multivariée de paramètres μ et Σ lorsque sa densité de probabilité peut s'écrire

$$\pi(y) = \frac{1}{(2\pi)^{\frac{m}{2}}|\Sigma|^{\frac{1}{2}}} exp(-\frac{1}{2}(y-\mu)'\Sigma^{-1}(y-\mu)) \tag{1.20}$$

Où $|\Sigma|$ est le déterminant de Σ.

On écrit plus simplement $Y \sim N_m(\mu, \Sigma)$

1.5.2 Loi de Student

Définition 1.15 *Une variable aléatoire Y suit une loi de Student de moyenne θ, de variance σ^2 et de n degrés de liberté si sa densité de probabilité f_Y est définie par :*

$$f_Y(y) = \frac{\Gamma((n+1)/2)S^{-1}}{\Gamma(n/2)(\pi n)^{\frac{1}{2}}} \{1 + \frac{(y-\theta)^2}{nS^2}\}^{-\frac{n+1}{2}}, y \in \mathbb{R}$$

Si $\theta = 0$ et $S = 1$, on dit que Y suit une loi de Student standard de n degrés de liberté et on écrit $Y \sim t(n)$.

On a, pour $\theta \in \mathbb{R}$, $S > 0$, et $n > 0$

$$\int_{-\infty}^{+\infty} \{1 + \frac{(y-\theta)^2}{nS^2}\}^{-\frac{n+1}{2}} dy = \frac{\Gamma(n/2)(\pi n)^{\frac{1}{2}}S}{\Gamma((n+1)/2)} \tag{1.21}$$

Densité de student multivariée

La densité d'une loi de student p-variée est définie comme le mélange d'une loi normale p-variée par une loi gamma particulière. Sa densité s'écrit comme suit :

$$\pi(y) = \frac{|V|^{-\frac{1}{2}}(\nu)^{\frac{p}{2}}\Gamma(\frac{\nu+p}{2})}{\pi^{\frac{p}{2}}\Gamma(\frac{\nu}{2})}[1 + \frac{(y-\mu)'V^{-1}(y-\mu)}{\nu}]^{-(\frac{\nu-p}{2})}, \quad y \in \mathbb{R}^p \qquad (1.22)$$

On dit que Y suit la loi de student p-variée de moyenne μ, de matrice de covariance V et de ν degrés de liberté. On écrit simplement :

$$Y \sim t_p(\mu, V, \nu)$$

1.5.3 Loi Gamma

Définition 1.16 *Une variable aléatoire Y suit une loi Gamma de paramètres k et θ (strictement positifs), ce que l'on note aussi $Y \sim \Gamma(k,\theta)$, si sa fonction de densité de probabilité peut se mettre sous la forme :*

$$f_Y(y) = \frac{y^{k-1}}{\Gamma(k)\theta^k} exp\{-\frac{y}{\theta}\} \qquad (1.23)$$

1.5.4 Loi Inverse-gamma

La distribution gamma inverse est une réciproque de la fonction de densité de probabilité gamma avec des paramètres de forme positifs α, β. La fonction de densité de la loi Gamma inverse est définie par

$$f_Y(y) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{-\alpha-1} \exp\left(\frac{-\beta}{y}\right)$$

2 Estimation Bayésienne des paramètres et des observations futures d'un modèle de régression linéaire.

2.1 Introduction

Parfois nous espérons modéliser une relation entre deux variables X et Y telle que les couples (x_i, y_i), $i = 1, .., n$ sont les données, où la variable prédictive X et la variable réponse Y sont indépendantes. On veut trouver une équation qui décrit cette relation afin de prédire la valeur de Y.
Dans ce chapitre, on va analyser la régression au sens bayésien, on détermine d'abord les lois a posteriori des paramètres inconnus en commençant par le cas où les erreurs sont indépendantes puis le cas où les erreurs sont autocorrélées. Ensuite, on va déterminer la densité prédictive d'une observation future.

2.2 Modèle de régression linéaire à erreurs indépendantes

2.2.1 Modèles de régression

La régression est l'un des outils mathématiques les plus utilisés pour l'analyse des données, elle fournit des méthodes simples pour l'étude des relations entre variables.
L'approche standard de la régression utilise un échantillon de données pour le calcul et l'estimation des paramètres inconnus. Un modèle explicatif est un modèle exprimant une variable Y appelée variable à expliquer (ou réponse), sous forme d'une fonction à une ou plusieurs variables dites variables explicatives notées X_i.

Régression linéaire simple

La régression linéaire simple est une méthode statistique qui permet de résumer et d'étudier la relation entre deux variables continues (quantitatives). Dans un modèle de régression linéaire simple, une variable endogène (dépendante) est expliquée par une seule variable exogène (indépendante) mise sous la forme mathématique suivante :

$$y_i = \theta_0 + \theta_1 x_i + \varepsilon_i \quad , i = 1, ..., n \tag{2.1}$$

avec

θ_0, θ_1 : les paramètres inconnus du modèle.
ε_i : l'erreur aléatoire du modèle.
n : nombre d'observations.

Hypothèses du modèle
Le modèle repose sur les hypothèses suivantes :

1. $\mathbb{E}(\varepsilon_i) = 0$, l'erreur est centrée.
2. $\mathbb{E}(\varepsilon_i^2) = \sigma^2$, la variance de l'erreur est constante.
3. $cov(\varepsilon_i, \varepsilon_j) = 0$, si $\varepsilon_i \neq \varepsilon_j$, les erreurs ne sont pas autocorrélées.
4. $cov(x_i, \varepsilon_i) = 0$, l'erreur n'est pas corrélée avec la variable exogène.
5. La variable exogène n'est pas aléatoire.

Régression linéaire multiple

Le modèle multiple est une généralisation du modèle simple dans lequel figurent plusieurs variables explicatives $x_{i1}, x_{i2}, ..., x_{ip}$.

Le modèle de régression linéaire multiple suppose une relation linéaire entre les variables explicatives et la variable à expliquer, de la forme :

$$\begin{aligned} y_i &= \theta_1 x_{i1} + \theta_2 x_{i2} + ... + \theta_p x_{ip} + \varepsilon_i \\ &= x_i^t \theta + \varepsilon_i, \quad i = 1...n \end{aligned} \tag{2.2}$$

avec $\varepsilon_i \sim N(0, \sigma^2)$, $\theta = (\theta_1, \theta_2, ..., \theta_p)'$ et $x_i^t = (x_{i1}, ..., x_{it})$.
Les paramètres de régression $\theta_1, ..., \theta_p$ caractérisent la relation entre les différentes variables en jeu.
Ce modèle s'écrit sous la forme matriciel suivante

$$Y = X\theta + \varepsilon \tag{2.3}$$

où Y est une matrice $n \times 1$ d'observations, X est une matrice $n \times p$, θ est la matrice $p \times 1$ des paramètres de régression et ε est la matrice $n \times 1$ des erreurs $\varepsilon \sim N(0, \Sigma)$, avec $\Sigma = \sigma^2 I_n$

Le modèle (2.3) s'écrit alors comme suit :

$$\begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ y_n \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & . & . & x_{1p} \\ x_{12} & x_{22} & . & . & x_{2p} \\ . & . & . & . & . \\ . & . & . & . & . \\ x_{n1} & x_{n2} & . & . & x_{np} \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ . \\ . \\ \theta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ \varepsilon_n \end{bmatrix}$$

Afin d'estimer le vecteur θ, nous appliquons la méthode des moindres carrées ordinaire (MCO) qui consiste à minimiser la somme des carrées des erreurs, soit :

$$min \sum_{i=1}^{n} \varepsilon_i^2 = min(\varepsilon^{'} \varepsilon) = min(Y - X\theta)^{'}(Y - X\theta) = minS \tag{2.4}$$

Avec $\varepsilon^{'}$ est la transposé du vecteur ε et $S = (Y - X\theta)^{'}(Y - X\theta)$.

Pour minimiser cette fonction par rapport au vecteur θ nous dérivons S par rapport au même vecteur et on obtient :

$$\frac{\partial S}{\partial \theta} = -2X^{'}Y + 2X^{'}X\theta = 0 \Rightarrow \widehat{\theta} = (X^{'}X)^{-1}X^{'}Y \tag{2.5}$$

L'estimateur de θ a les propriétés suivantes

1. $$E(\widehat{\theta}|\theta, \sigma^2, X) = \theta$$
2. $$Var(\widehat{\theta}|\theta, \sigma^2, X) = \sigma^2(X^{'}X)^{-1}$$
3. $$\widehat{\theta}|\theta, \sigma^2, X \sim N_p(\widehat{\theta}, \sigma^2(X^{'}X)^{-1})$$

La variance σ^2 est estimée par

$$\hat{\sigma}^2 = s^2 = \frac{1}{n-p}\sum_{i=1}^{n}(Y_i - X_i^{'}\widehat{\theta})^2 = \frac{1}{n-p}(Y - X\widehat{\theta})^{'}(Y - X\widehat{\theta}) \tag{2.6}$$

La fonction de vraisemblance

Comme notre modèle est :

$$Y = X\theta + \varepsilon,$$

Et

$$\varepsilon \sim N(0, I\sigma^2)$$

Donc

$$Y \sim N(X\hat{\theta}, I\sigma^2).$$

La fonction de vraisemblance du vecteur Y est

$$L(\theta, \sigma^2 | Y, X) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} exp\left\{-\frac{1}{2\sigma^2}(Y - X\theta)^{'}(Y - X\theta)\right\} \tag{2.7}$$

La forme quadratique apparaissant dans l'exponentielle de la fonction de vraisemblance (2.7) peut être réécrite de la façon suivante :

$$\begin{aligned}(Y - X\theta)^{'}(Y - X\theta) &= (Y - X\widehat{\theta} + X\widehat{\theta} - X\theta)^{'}(Y - X\widehat{\theta} + X\widehat{\theta} - X\theta) \\ &= (Y - X\widehat{\theta})^{'}(Y - X\widehat{\theta}) + (X\widehat{\theta} - X\theta)^{'}(X\widehat{\theta} - X\theta) + 2(Y - X\widehat{\theta})^{'}(X\widehat{\theta} - X\theta) \\ &= (n-p)s^2 + (\widehat{\theta} - \theta)^{'}X^{'}X(\widehat{\theta} - \theta).\end{aligned}$$

On pose $M = (n-p)s^2$, on obtient

$$(Y - X\theta)^{'}(Y - X\theta) = M + (\widehat{\theta} - \theta)^{'}X^{'}X(\widehat{\theta} - \theta).$$

On peut à présent réécrire la vraisemblance :

$$L(\theta, \sigma^2 | Y, X) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} exp\left\{-\frac{1}{2\sigma^2}(M + (\widehat{\theta} - \theta)^{'}X^{'}X(\widehat{\theta} - \theta))\right\} \tag{2.8}$$

2.2.2 La loi a posteriori de θ

L'analyse bayésienne de ce modèle est relativement immédiate si le paramètre σ^2 est connu. On commencera donc la présentation de cas simple, d'autant plus qu'il prépare l'analyse du modèle plus général où σ^2 est inconnu.

Le cas où la variance est connue

Dans ce cas le vecteur θ est uniquement inconnu, donc la loi a posteriori de θ s'écrit comme suit

$$\pi(\theta | Y, X) \propto L(\theta | Y, X)\pi(\theta) \tag{2.9}$$

Où $L(\theta|Y,X)$ est la fonction de vraisemblance qui s'écrit comme suit

$$L(\theta|Y,X) \propto exp\{-\frac{1}{2\sigma^2}(\widehat{\theta}-\theta)^{'}X^{'}X(\widehat{\theta}-\theta)\} \tag{2.10}$$

Supposons que la loi a priori de θ est non informative

$$\pi(\theta) \propto cte \tag{2.11}$$

En remplaçant les formules (2.10) et (2.11) dans (2.9) on obtient la loi a posteriori de θ

$$\pi(\theta|Y,X) \propto exp\{-\frac{1}{2\sigma^2}(\widehat{\theta}-\theta)^{'}X^{'}X(\widehat{\theta}-\theta)\} \tag{2.12}$$

On remarque que c'est une distribution d'une loi normale

$$\theta|Y,X \sim N(\widehat{\theta}, \sigma^2(X^{'}X)^{-1})$$

le cas où la variance est inconnue

La vraisemblance dans ce cas général a été donnée en (2.8) ; on la reproduit ci-dessous :

$$L(\theta,\sigma^2|Y,X) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}}exp\{-\frac{1}{2\sigma^2}(M+(\widehat{\theta}-\theta)^{'}X^{'}X(\widehat{\theta}-\theta))\}$$

Supposons que les paramètres θ et σ^2 sont indépendants de loi a priori non informative :

$$\pi(\theta,\sigma^2) \propto \frac{1}{\sigma^2} \tag{2.13}$$

La loi a posteriori de (θ,σ^2) est donc :

$$\begin{aligned}\pi(\theta,\sigma^2|Y,X) &\propto \pi(\theta,\sigma^2)L(\theta,\sigma^2|Y,X)\\ &\propto \frac{1}{(\sigma^2)^{\frac{n}{2}+1}}exp\{-\frac{1}{2\sigma^2}(M+(\widehat{\theta}-\theta)^{'}X^{'}X(\widehat{\theta}-\theta))\}\end{aligned}$$

En intégrant par rapport à σ^2, on obtient la loi a posteriori de θ :

$$\begin{aligned}\pi(\theta|Y,X) &\propto \int_0^{+\infty}\pi(\theta,\sigma^2|Y,X)d\sigma^2\\ &\propto \int_0^{+\infty}\frac{1}{(\sigma^2)^{\frac{n}{2}+1}}exp\{-\frac{1}{2\sigma^2}(M+(\widehat{\theta}-\theta)^{'}X^{'}X(\widehat{\theta}-\theta))\}d\sigma^2\\ &\propto (M+(\widehat{\theta}-\theta)^{'}X^{'}X(\widehat{\theta}-\theta))^{-\frac{n}{2}}\\ &\propto ((n-p)s^2+(\widehat{\theta}-\theta)^{'}X^{'}X(\widehat{\theta}-\theta))^{-\frac{n}{2}}\end{aligned}$$

Comme $(n-p)s^2 = (Y - X\widehat{\theta})'((Y - X\widehat{\theta})$

La loi a posteriori de θ devient :

$$\begin{aligned}\pi(\theta|Y,X) &\propto [\nu s^2 + (\theta - \widehat{\theta})' X' X(\theta - \widehat{\theta})]^{-\frac{\nu+p}{2}} \\ &\propto [1 + \frac{(\theta - \widehat{\theta})' X' X(\theta - \widehat{\theta})}{\nu s^2}]^{-\frac{\nu+p}{2}}\end{aligned}$$

Avec $\nu = n - p$.
On obtient le noyau d'une loi de student multivariée de moyenne $\widehat{\theta}$ de matrice de covariance $s^2(X'X)^{-1}$ et de ν degré de liberté.

$$\theta|Y,X \sim t_p(\widehat{\theta}, (s^2(X'X)^{-1}, \nu)$$

Pour avoir la loi a posteriori de σ^2, on intègre la loi a posteriori de (θ, σ^2) par rapport a θ, et on aura :

$$\begin{aligned}\pi(\sigma^2|Y,X) &\propto \int_{-\infty}^{+\infty} \pi(\theta, \sigma^2|Y,X) d\theta \\ &\propto \int_{-\infty}^{+\infty} \frac{1}{(\sigma^2)^{\frac{n}{2}+1}} exp\{-\frac{1}{2\sigma^2}(M + (\widehat{\theta} - \theta)' X' X(\widehat{\theta} - \theta))\} d\theta \\ &\propto \int_{-\infty}^{+\infty} \frac{1}{(\sigma^2)^{\frac{n}{2}+1}} exp\{-\frac{1}{2\sigma^2}(\nu s^2 + (\widehat{\theta} - \theta)' X' X(\widehat{\theta} - \theta))\} d\theta \\ &\propto \frac{1}{(\sigma^2)^{(\frac{\nu}{2}+1)}} exp\{-\frac{\nu s^2}{2\sigma^2}\} \\ &\propto \frac{1}{(\sigma^2)^{(\frac{\nu}{2}+1)}} exp\{-\frac{S'}{2\sigma^2}\}\end{aligned}$$

avec $S' = \nu s^2$.

On obtient le noyau d'une loi Gamma-inverse

$$\sigma^2|Y,X \sim \Gamma(\frac{\nu}{2}, \frac{S'}{2})$$

2.3 Modèle de régression linéaire à erreurs autocorrélées

Dans cette section nous nous plaçons dans le cas où les erreurs ε_i ne sont pas indépendantes mais autocorrélées (voir Zellner et Tiao [1964] et Ait Mohammed [2006]).

Soit le modèle :

$$\begin{cases} Y_i = X_i\theta + \mu_i & i = 1..n \\ \mu_i = \rho\mu_{i-1} + \varepsilon_i \end{cases}$$

Où les ε_i sont des variables aléatoires iid selon une loi normale $N(0, \sigma^2)$.
Le modèle (2.14) peut se réécrire sous la forme suivante

$$Y_i = X_i\theta + \rho(y_{i-1} - \theta X_{i-1}) + \varepsilon_i \tag{2.14}$$

Les paramètres du modèle (2.14) sont supposés inconnus. $\theta \in \mathbb{R}$, $\rho \in \mathbb{R}$ et $\sigma > 0$. Nous supposons que le processus AR(1) est stationnaire ($-1 \leq \rho \leq +1$).

Pour $i = 1$ l'équation (2.14) devient $Y_1 = \rho Y_0 + X_1\theta - \rho X_0\theta + \varepsilon_1$ où X_0 et Y_0 sont des quantités observées. Si nous supposons que l'équation (2.14) est représentative de ce qui s'est passé pour $i = 0, -1, -2, ..., -n_0$, où n_0 est inconnu, nous avons, par exemple $Y_0 = X_0\theta + \rho(Y_{-1} - X_{-1}\theta) + \varepsilon_0$ en supposant que $\varepsilon_0 \sim N(0, \sigma^2)$ comme les ε_i, $i = 1, ..., n$.

Comme Y_{-1} et X_{-1} ne sont pas des quantités connues, il est plus simple d'écrire que $Y_0 = X_0\theta + A + \varepsilon_0$ où $A = \rho(Y_{-1} - \theta X_{-1})$ est une fonction de quantités inobservées.
Le vecteur des paramètres du ce modèle est donc $(\theta, \sigma^2, A, \rho)$.
Sous ces hypothèses, Y_0 suit la loi normale $N(\theta X_0 + A, \sigma^2)$.

2.3.1 Lois a posteriori conditionnelles de θ et ρ

La vraisemblance s'écrit :

$$\begin{aligned} L(\theta, \rho, \sigma, A|Y) \propto \frac{1}{\sigma^{-(n+1)}} exp\{-\frac{1}{2\sigma^2}(Y_0 - \theta X_0 - A)^2\} \\ exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - \rho Y_{i-1} - \theta(X_i - \rho X_{i-1}))^2\} \end{aligned} \tag{2.15}$$

Selon la formule de Bayes, la loi a posteriori du vecteur des paramètres $(\theta, \sigma^2, \rho, A)$ est proportionnelle au produit de la vraisemblance et la loi a priori. Comme le dénominateur de la formule de Bayes n'est pas une fonction de θ, la loi a posteriori des paramètres est donc :

$$\pi(\theta, \sigma^2, A, \rho|Y, X) \propto L(\theta, \rho, \sigma^2, A|Y) \times \pi(\theta, \sigma^2, A, \rho) \tag{2.16}$$

On suppose que les paramètres $\theta, \sigma^2, \rho, A$ sont indépendants et de loi a priori non informative :

$$\pi(\theta, \sigma^2, \rho, A) \propto \frac{1}{\sigma} \tag{2.17}$$

En remplaçant (2.15) et (2.17) dans (2.16), on obtient

$$\pi(\theta,\sigma,\rho,A|Y,X) \propto \frac{1}{\sigma^{n+2}} exp\{-\frac{1}{2\sigma^2}[(Y_0-\theta X_0-A)^2 + \sum_{i=1}^{n}(Y_i-\rho Y_{i-1}-\theta(X_i-\rho X_{i-1}))]^2\} \tag{2.18}$$

Pour avoir la loi a posteriori de l'un des paramètres θ, σ, ρ, A, on intègre la loi a posteriori de (θ,σ,ρ,A) par rapport aux trois autres.
Une integration par rapport à A nous donne

$$\begin{aligned}\pi(\theta,\sigma,\rho|Y,X) &\propto \int_{-\infty}^{+\infty} \pi(\theta,\sigma,\rho,A|Y,X)dA \\ &\propto \frac{1}{\sigma^{n+2}} exp\{-\frac{1}{2\sigma^2}[(Y_0-\theta X_0-A)^2 \\ &+ \sum_{i=1}^{n}(Y_i-\rho Y_{i-1}-\theta(X_i-\rho X_{i-1}))]^2\}dA\end{aligned}$$

On remarque que :

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma}} exp\{-\frac{1}{2\sigma^2}[(Y_0-\theta X_0-A)^2\}dA = 1$$

Ce qui implique

$$\pi(\theta,\sigma,\rho|Y,X) \propto \frac{1}{\sigma^{n+1}} exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i-\rho Y_{i-1}-\theta(X_i-\rho X_{i-1}))^2\} \tag{2.19}$$

Une deuxième intégration par rapport à σ nous donne

$$\begin{aligned}\pi(\theta,\rho|Y,X) &\propto \int_{0}^{+\infty} \pi(\theta,\sigma,\rho,|Y,X)d\sigma \\ &\propto \frac{1}{\sigma^{n+1}} exp\{-\frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i-\rho Y_{i-1}-\theta(X_i-\rho X_{i-1}))^2\}d\sigma \\ &\propto [\sum_{i=1}^{n}(Y_i-\rho Y_{i-1}-\theta(X_i-\rho X_{i-1}))^2]^{-\frac{n}{2}} \qquad (2.20) \\ &\propto [\sum_{i=1}^{n}(Y_i-\theta X_i-\rho(Y_{i-1}-\theta X_{i-1}))^2]^{-\frac{n}{2}} \qquad (2.21)\end{aligned}$$

Pour avoir la loi a posteriori conditionnelle de θ on considère le paramètre ρ dans (2.20) comme constante.

En effectuant le changement de variables suivant :

$$\alpha = \sum_{i=1}^{n}(Y_i - \rho Y_{i-1})^2$$
$$\beta = \sum_{i=1}^{n}(Y_i - \rho Y_{i-1})(X_i - \rho X_{i-1})$$
$$\gamma = \sum_{i=1}^{n}(X_i - \rho X_{i-1})^2$$

La loi a posteriori conditionnelle de θ est obtenue comme suit :

$$\begin{aligned}
\pi(\theta|\rho, Y, X) &\propto [\alpha - 2\theta\beta + \gamma\theta^2]^{-\frac{n}{2}} \\
&= [\gamma(\frac{\alpha}{\gamma} - \frac{2\beta\theta}{\gamma} + \theta^2)]^{-\frac{n}{2}} \\
&= \gamma^{-\frac{n}{2}}(\frac{\alpha}{\gamma} - \frac{2\beta\theta}{\gamma} + \theta^2)^{-\frac{n}{2}} \\
&\propto (\frac{\alpha}{\gamma} - \frac{2\beta\theta}{\gamma} + \theta^2)^{-\frac{n}{2}} \\
&= [(\theta - \frac{\beta}{\gamma})^2 - \frac{\beta^2}{\gamma^2} + \frac{\alpha}{\gamma}]^{-\frac{n}{2}} \\
&= (\frac{\alpha}{\gamma} - \frac{\beta^2}{\gamma^2})^{-\frac{n}{2}}[1 + \frac{(\theta - \frac{\beta}{\gamma})^2}{(\frac{\alpha}{\gamma} - \frac{\beta^2}{\gamma^2})}]^{-\frac{n}{2}} \\
&\propto [1 + \frac{(\theta - \frac{\beta}{\gamma})^2}{(n-1)\frac{(\frac{\alpha}{\gamma} - \frac{\beta^2}{\gamma^2})}{(n-1)}}]^{-\frac{n}{2}}
\end{aligned}$$

La loi a posteriori de θ conditionnelle à ρ s'écrit alors :

$$\pi(\theta|\rho, Y, X) \propto [1 + \frac{(\theta - \widehat{\theta}(\rho))^2}{(n-1)K(\rho)}]^{-\frac{n}{2}}$$

On obtient un noyau d'une loi de student de moyenne $\widehat{\theta}(\rho)$, d'une variance $K(\rho)$ et de (n-1) degrés de liberté.

$$\theta|\rho, Y, X \sim t(\widehat{\theta}(\rho), K(\rho), n-1)$$

Avec

$$\widehat{\theta}(\rho) = \frac{\beta}{\gamma} = \frac{\sum_{i=1}^{n}(Y_i - \rho Y_{i-1})(X_i - \rho X_{i-1})}{\sum_{i=1}^{n}(X_i - \rho X_{i-1})^2}$$

Et

$$\begin{aligned} K(\rho) &= \frac{1}{n-1}(\frac{\alpha}{\gamma} - \frac{\beta^2}{\gamma^2}) \\ &= \frac{1}{n-1}\left[\frac{\sum_{i=1}^{n}(Y_i - \rho Y_{i-1})^2}{\sum_{i=1}^{n}(X_i - \rho X_{i-1})^2} - \frac{(\sum_{i=1}^{n}(Y_i - \rho Y_{i-1})(X_i - \rho X_{i-1}))^2}{(\sum_{i=1}^{n}(X_i - \rho X_{i-1})^2)^2}\right] \end{aligned}$$

Pour obtenir la loi a posteriori conditionnelle de ρ, on considère le paramètre θ dans (2.21) comme constant et on pose le changement de variables suivant

$$\begin{aligned} a &= \sum_{i=1}^{n}(Y_i - \theta X_i)^2 \\ b &= \sum_{i=1}^{n}(Y_i - \theta X_i)(Y_{i-1} - \theta X_{i-1}) \\ c &= \sum_{i=1}^{n}(Y_{i-1} - \theta X_{i-1})^2 \end{aligned}$$

La loi a posteriori conditionnelle de ρ est obtenue comme suit :

$$\begin{aligned} \pi(\rho|\theta, Y, X) &\propto [a - 2\rho b + c\rho^2]^{-\frac{n}{2}} \\ &= [c(\frac{a}{c} - \frac{2b\rho}{c} + \rho^2)]^{-\frac{n}{2}} \\ &= c^{-\frac{n}{2}}(\frac{a}{c} - \frac{2b\rho}{c} + \rho^2)^{-\frac{n}{2}} \\ &\propto (\frac{a}{c} - \frac{2b\rho}{c} + \rho^2)^{-\frac{n}{2}} \\ &= [\frac{a}{c} - \frac{b^2}{c^2} + (\rho - \frac{b}{c})^2]^{-\frac{n}{2}} \\ &= (\frac{a}{c} - \frac{b^2}{c^2})^{-\frac{n}{2}}.[1 + \frac{(\rho - \frac{b}{c})^2}{(\frac{a}{c} - \frac{b^2}{c^2})}]^{-\frac{n}{2}} \\ &\propto [1 + \frac{(\rho - \frac{b}{c})^2}{(n-1)\frac{(\frac{a}{c} - \frac{b^2}{c^2})}{(n-1)}}]^{-\frac{n}{2}} \end{aligned}$$

La loi a posteriori de ρ conditionnelle à θ s'écrit alors :

$$\pi(\rho|\theta, Y, X) \propto [1 + \frac{(\rho - \widehat{\rho}(\theta))^2}{(n-1)K(\theta)}]^{-\frac{n}{2}}$$

On obtient un noyau d'une loi de student de moyenne $\widehat{\rho}(\theta)$, de variance $K(\theta)$ et (n-1) degrés de liberté.

$$\rho|\theta, Y, X \sim t(\widehat{\rho}(\theta), K(\theta), n-1)$$

Avec

$$\widehat{\rho}(\theta) = \frac{b}{c} = \frac{\sum_{i=1}^{n}(Y_i - \theta X_i)(Y_{i-1} - \theta X_{i-1})}{\sum_{i=1}^{n}(Y_{i-1} - \theta X_{i-1})^2}$$

Et

$$\begin{aligned} K(\theta) &= \frac{1}{n-1}(\frac{a}{c} - \frac{b^2}{c^2}) \\ &= \frac{1}{n-1}[\frac{\sum_{i=1}^{n}(Y_i - \theta X_i)^2}{\sum_{i=1}^{n}(Y_{i-1} - \theta X_{i-1})^2} - \frac{(\sum_{i=1}^{n}(Y_i - \theta X_i)(Y_{i-1} - \theta X_{i-1}))^2}{(\sum_{i=1}^{n}(Y_{i-1} - \theta X_{i-1})^2)^2}] \end{aligned}$$

2.4 Densité prédictive d'une observation future

La densité prédictive (appelée parfois posterior predictive) est définie comme la densité d'une nouvelle observation y_{n+1} sachant les observations de l'échantillon. Elle se calcule comme la densité de y_{n+1} sachant Y, marginalement par rapport à θ, ρ et σ.

$$\begin{aligned} \pi(y_{n+1}|Y, X, x_{n+1}) &= \int \pi(y_{n+1}, \xi|Y, X, x_{n+1}) d\xi \\ &= \int \pi(y_{n+1}|\xi, Y, X, x_{n+1}) \pi(\xi|Y, X, x_{n+1}) d\xi \\ &= \int \pi(y_{n+1}|\xi) \pi(\xi|Y, X, x_{n+1}) d\xi \end{aligned}$$

Où $\xi = (\theta, \sigma, \rho)'$.
Par conséquent $\pi(y_{n+1}|Y, X, x_{n+1})$ peut être estimé par

$$\hat{\pi}(y_{n+1}|Y, X, x_{n+1}) = \frac{1}{N} \sum_{i=1}^{N} \pi(y_{n+1}|\xi^{(i)}, X, Y, x_{n+1}) \tag{2.22}$$

Où les $\xi^{(i)}$, $i = 1, .., N$, sont les valeurs simulées de ξ par la loi a posteriori et N est le nombre de répétitions.

On a

$$\pi(\xi|Y, X, x_{n+1}) \propto \frac{1}{\sigma^{n+1}} exp\{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (Y_i - \rho Y_{i-1} - \theta(X_i - \rho X_{i-1}))^2\}$$

Et

$$\pi(y_{n+1}|\xi) = \frac{1}{\sqrt{2\pi}\sigma} exp\{\frac{1}{2\sigma^2}(y_{n+1} - \theta x_{n+1} - \rho(y_n - \theta x_n))^2\}$$

Par conséquent

$$\begin{aligned}
\pi(y_{n+1}|Y, X, x_{n+1}) &\propto \int (\pi(y_{n+1}|\xi)\pi(\xi|Y, X, x_{n+1})d\xi \\
&\propto \int \frac{1}{\sigma^{n+2}} exp\{-\frac{1}{2\sigma^2}(y_{n+1} - \theta x_{n+1} - \rho(y_n - \theta x_n))^2\} \\
&exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - \rho Y_{i-1} - \theta(X_i - \rho X_{i-1}))^2\} d\xi
\end{aligned}$$

En intégrant par rapport a σ on obtient :

$$\pi(y_{n+1}|Y, X, x_{n+1}) \propto \left[\sum_{i=1}^{n}(Y_i - \rho Y_{i-1} - \theta(X_i - \rho X_{i-1}))^2 + (y_{n+1} - \theta x_{n+1} - \rho(y_n - \theta x_n))^2\right]^{\frac{-(n+1)}{2}} \tag{2.23}$$

2.4.1 Lois a posteriori conditionnelles de y_{n+1}, θ et ρ

$$\begin{aligned}
\pi(y_{n+1}|Y, X, x_{n+1}) &\propto \left[\sum_{i=1}^{n}(Y_i - \rho Y_{i-1} - \theta(X_i - \rho X_{i-1}))^2 + (y_{n+1} - \theta x_{n+1} - \rho(y_n - \theta x_n))^2\right]^{\frac{-(n+1)}{2}} \\
&\propto \left[1 + \frac{(y_{n+1} - (\theta x_{n+1} - \rho(y_n - \theta x_n)))^2}{\sum_{i=1}^{n}(Y_i - \rho Y_{i-1} - \theta(X_i - \rho X_{i-1}))^2}\right]^{\frac{-(n+1)}{2}} \\
&\propto \left[1 + \frac{(y_{n+1} - (\theta x_{n+1} - \rho(y_n - \theta x_n)))^2}{(n-1)\frac{\sum_{i=1}^{n}(Y_i - \rho Y_{i-1} - \theta(X_i - \rho X_{i-1}))^2}{(n-1)}}\right]^{\frac{-(n+1)}{2}}
\end{aligned}$$

On obtient le noyau d'une loi de student telle que :

$$y_{n+1}|\theta, \rho Y, X, x_{n+1} \sim t_n(\hat{y}_{n+1}, S_1, n)$$

Avec

$$\hat{y}_{n+1} = \theta x_{n+1} - \rho(y_n - \theta x_n)$$

Et

$$S_1 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \rho Y_{i-1} - \theta(X_i - \rho X_{i-1}))^2$$

Pour calculer la loi a posteriori conditionnelle de θ, on effectue le changement de variables suivant pour simplifier les calculs.

$$e_1 = \sum_{i=1}^{n}(Y_i - \rho Y_{i-1})^2 + (y_{n+1} - \rho y_n)^2.$$
$$e_2 = \sum_{i=1}^{n}(Y_i - \rho Y_{i-1})(X_i - \rho X_{i-1}) + (y_{n+1} - \rho y_n)(x_{n+1} - \rho x_n).$$
$$e_3 = \sum_{i=1}^{n}(x_i - \rho x_{i-1})^2 + (x_{n+1} - \rho x_n)^2.$$

On obtient

$$\begin{aligned}\pi(y_{n+1}|Y, X, x_{n+1}) &\propto [e_1 - 2\theta e_2 + \theta^2 e_3]^{\frac{-(n+1)}{2}} \\ &\propto \left[1 + \frac{(\theta - \frac{e_2}{e_3})^2}{(n-1)\frac{(\frac{e_1}{e_3} - \frac{e_2^2}{e_3^2})}{(n-1)}}\right]^{\frac{-(n+1)}{2}}\end{aligned}$$

On obtient le noyau d'une loi de student

$$\theta|y_{n+1}, \rho, Y, X, x_{n+1} \sim t_n(\hat{\theta}, S_2, n)$$

Avec

$$\hat{\theta} = \frac{e_2}{e_3}$$

Et

$$S_2 = \frac{1}{n-1}(\frac{e_1}{e_3} - \frac{e_2^2}{e_3^2})$$

Suivant les mêmes calculs que la loi précédente on obtient la loi a posteriori conditionnelle de ρ

$$\begin{aligned}\pi(\rho|\theta, Y, X, x_{n+1}) &\propto \left[\sum_{i=1}^{n}(Y_i - \theta x_i - \rho(y_{i-1} - \theta X_{i-1}))^2 + (y_{n+1} - \theta x_{n+1} - \rho(y_n - \theta x_n))^2\right]^{\frac{-(n+1)}{2}} \\ &= [q_1 - 2\rho q_2 + \rho^2 q_3]^{\frac{-(n-1)}{2}} \\ &\propto \left[1 + \frac{(\rho - \frac{q_2}{q_3})^2}{(n-1)\frac{(\frac{q_1}{q_3} - \frac{q_2^2}{q_3^2})}{(n-1)}}\right]^{\frac{-(n+1)}{2}}\end{aligned}$$

Avec

$$q_1 = \sum_{i=1}^{n}(Y_i - \theta x_i)^2 + (y_{n+1} - \theta x_{n+1})^2.$$
$$q_2 = \sum_{i=1}^{n}(Y_i - \theta x_i)(y_{i-1} - \theta X_{i-1}) + (y_{n+1} - \theta x_{n+1})(y_n - \theta x_n).$$
$$q_3 = \sum_{i=1}^{n}(y_{i-1} - \theta x_{i-1})^2 + (y_n - \theta x_n)^2.$$

On obtient le noyau d'une loi de student telle que

$$\rho | y_{n+1}, \theta, Y, Xx_{n+1} \sim t_n(\hat{\rho}, S_3, n)$$

Avec

$$\hat{\rho} = \frac{q_2}{q_3}$$

Et

$$S_3 = \frac{1}{n-1}(\frac{q_1}{q_3} - \frac{q_2^2}{q_3^2})$$

Remarque 2.1
De la même manière, on peut calculer la densité prédictive de y_{n+2} sachant les observations y_1, y_2,..., y_n, y_{n+1}.

3 Etude de simulation

3.1 Présentation du modèle

Dans cette section, nous allons présenter un exemple de simulation pour étudier la convergence des estimateurs du paramètre θ du modèle de régression et du paramètre ρ du modèle autorégressif, en utilisant l'algorithme de Gibbs sampler. Nous supposons que le processus AR(1) est stationnaire ($|\rho| \leq 1$)
Nous simulons les observations y, en utilisant le modèle suivant

$$\begin{aligned} y_i &= 4x_i + \mu_i, \qquad i = 1, ..., n, \\ \mu_i &= 0.5\mu_{i-1} + \varepsilon_i, \\ \mu_0 &= 0.5, \end{aligned} \tag{3.1}$$

où les ε_i sont générées par la loi normale centrée et les x_i sont générées par la loi uniforme sur [0,1].
La figure suivante représente le graphe de n=200 observations simulées à partir du modèle (3.1).

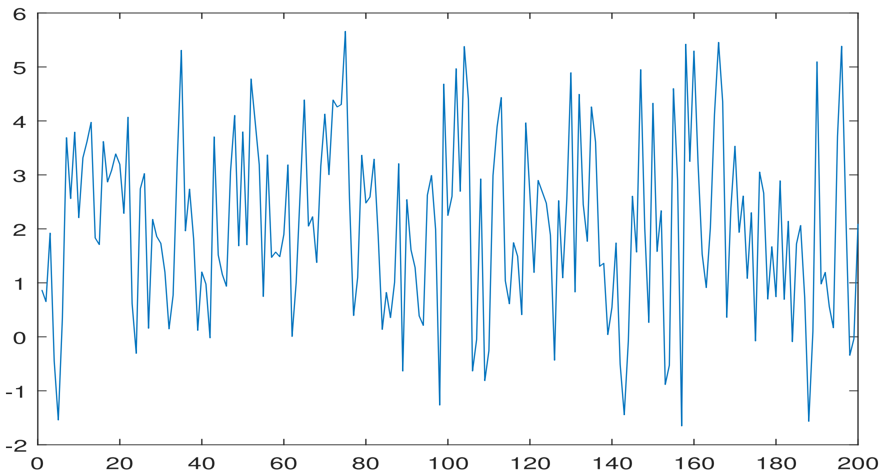

Figure 3.1 Le graphe de la série d'observations y

3.2 Application de l'échantillonneur de Gibbs pour l'estimation des paramètres du modèle

L'algorithme de Gibbs sampler est appliqué pour N=1000 répétitions. Nous approximons la moyenne a posteriori de $\beta \in \{\theta, \rho\}$, en utilisant la fonction de perte quadratique, donnée par $\hat{\beta} = \frac{1}{N}\sum_{i=1}^{N} \beta_i^{(m)}$, avec un "burn-in" période de $m-1$ observations pour atténuer l'influence des valeurs initiales, pour N suites de Gibbs de taille m.

Nous donnons l'écart-type (std), l'erreur quadratique moyenne (rmse) et les intervalles de crédibilités (IC) des paramètres θ et ρ.

$$rmse(\beta) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\beta_i^{(m)} - \beta)^2}$$

$$std(\beta) = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(\hat{\beta} - \beta_i^{(m)})^2}$$

Pour déterminer les IC de β, nous ordonnons $\beta_1^{(m)}, \beta_2^{(m)}, ..., \beta_N^{(m)}$ comme suit : $\beta_{(1)}^{(m)}, \beta_{(2)}^{(m)}, ..., \beta_{(N)}^{(m)}$. L'intervalle de crédibilité à 95% ($\alpha = 0.05$) de β est l'intervalle dont les bornes sont les $\frac{\alpha}{2}$-ème et $(1-\frac{\alpha}{2})$-ème quantiles de la loi a posteriori du paramètre étudié.

Les résultats des estimations sont obtenus pour différentes tailles d'échantillons n=50, n=200 et n=500.

Paramètres	Vraie-valeurs	n=50	n=200	n=500
$\hat{\theta}^{(std)}_{(rmse)}$	4	$3.9394^{(0.3655)}_{(0.3703)}$	$3.9539^{(0.1667)}_{(0.1729)}$	$4.0059^{(0.1040)}_{(0.0041)}$
$IC(\theta)$		$[3.2333, 4.6716]$	$[3.6173, 4.2726]$	$[3.7960, 4.1994]$
$\hat{\rho}^{(std)}_{(rmse)}$	0.5	$0.5665^{(0.1810)}_{(0.1468)}$	$0.5248^{(0.0611)}_{(0.0659)}$	$0.4954^{(0.0399)}_{(0.0401)}$
$IC(\rho)$		$[0.3257, 0.8128]$	$[0.3979, 0.6434]$	$[0.4141, 0.5746]$

Table 3.1 Résultats des estimations des paramètres θ et ρ pour différentes tailles d'échantillons

Le tableau suivant donne les résultats des estimations quand ρ est proche de la zone de non stationnarité ($\rho = 0.95$).

Paramètres	Vraie-valeurs	n=50	n=200	n=500
$\hat{\theta}^{(std)}_{(rmse)}$	4	$4.3146^{(0.4333)}_{(0.5353)}$	$3.9908^{(0.1918)}_{(0.1920)}$	$4.0397^{(0.1133)}_{(0.1200)}$
$IC(\theta)$		[3.4046, 5.1090]	[3.6144, 4.3851]	[3.8320, 4.2579]
$\hat{\rho}^{(std)}_{(rmse)}$	0.95	$0.7960^{(0.0955)}_{(0.1812)}$	$0.9352^{(0.0275)}_{(0.0312)}$	$0.9588^{(0.0127)}_{(0.0155)}$
$IC(\rho)$		[0.6065, 0.9739]	[0.8803, 0.9891]	[0.9334, 0.9835]

Table 3.2 Résultats des estimations des paramètres θ et ρ pour différentes tailles d'échantillons lorsque $\rho = 0.95$

On remarque que la taille n de l'échantillon influence les résultats de l'estimation de la façon suivante :

- Plus n augmente, plus l'estimateur $\hat{\beta}$ s'approche de la vraie valeur du paramètre β.

- Plus n augmente, plus l'amplitude de l'intervalle de crédibilité IC diminue.
- Les paramètres de notre modèle sont bien estimés au sens du rmse.

Les graphes suivants représentent les histogrammes des estimations des paramètres θ et ρ pour les différentes tailles d'échantillons étudiés :

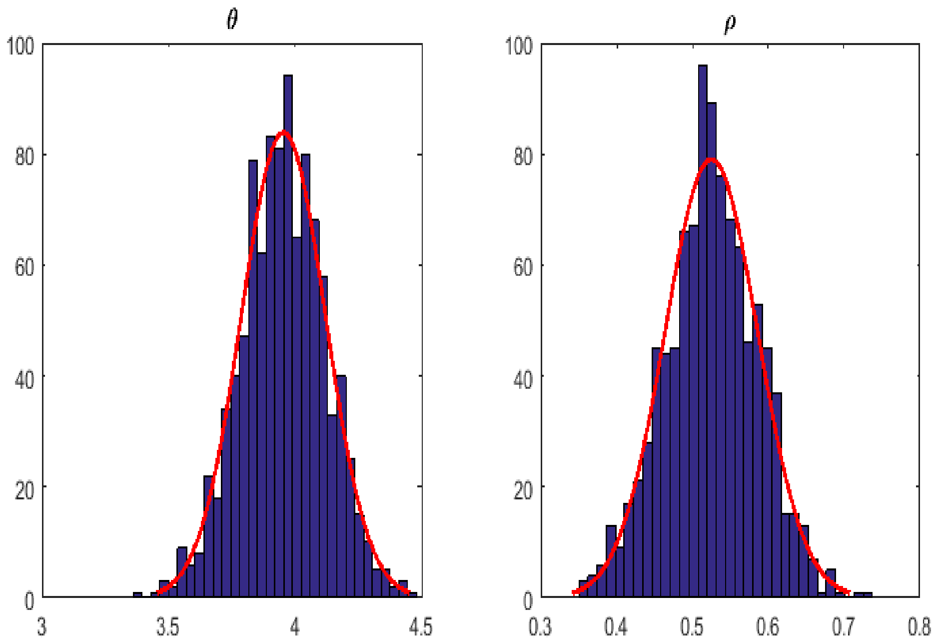

Figure 3.2 Les estimations de θ et ρ pour n=50

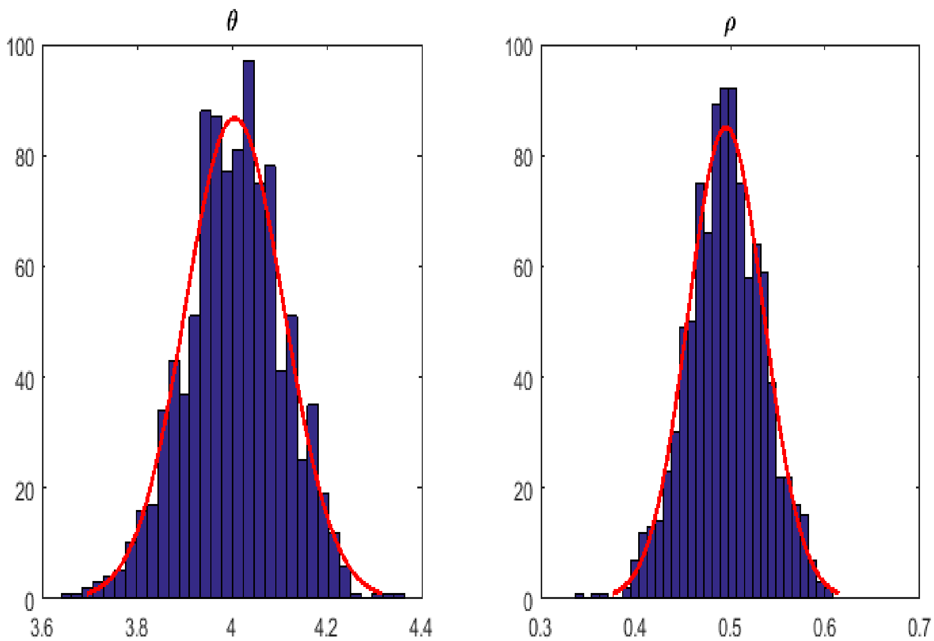

Figure 3.3 Les estimations de θ et ρ pour n=200

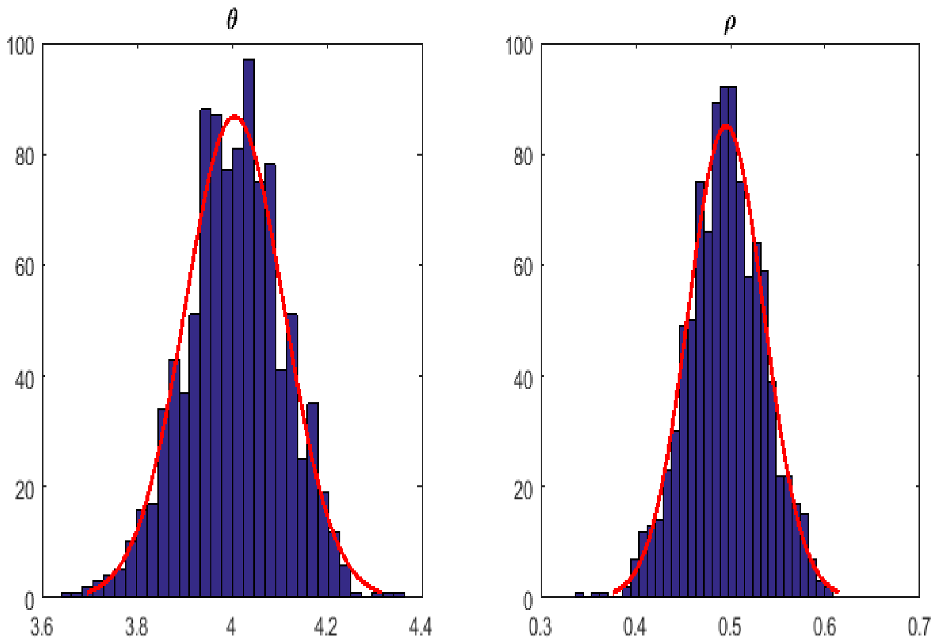

Figure 3.4 Les estimations de θ et ρ pour n=500

Les graphes suivants représentent les histogrammes des estimations des paramètres θ et ρ pour $\rho = 0.95$, pour les différentes tailles d'échantillons étudiés :

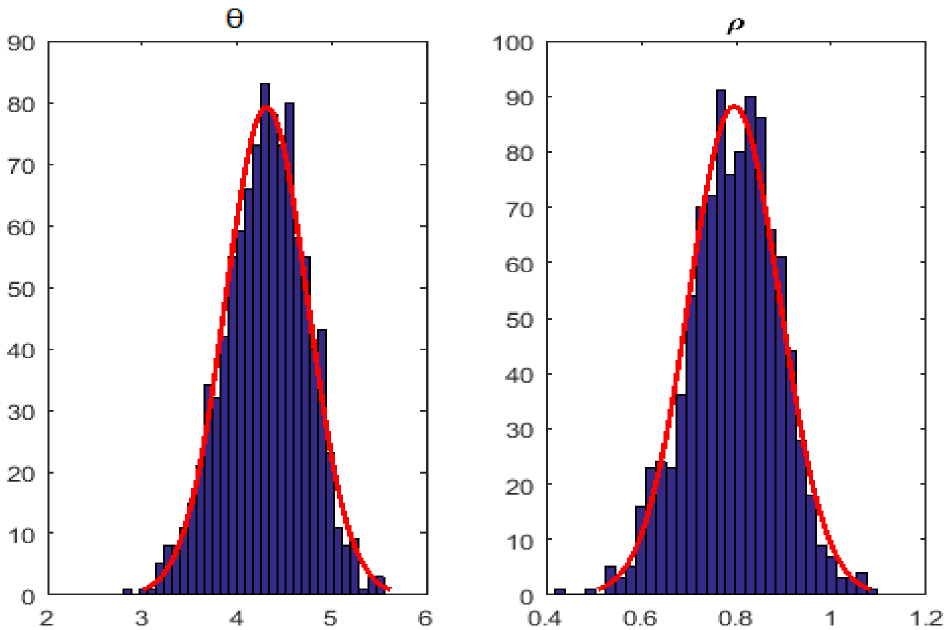

Figure 3.5 Les estimations de θ et ρ pour n=50

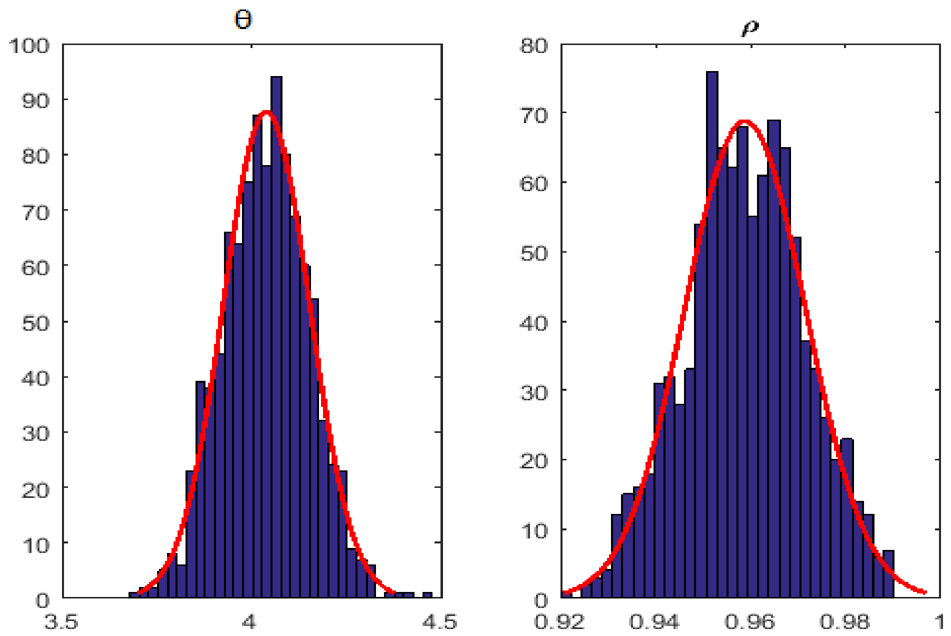

Figure 3.6 Les estimations de θ et ρ pour n=200

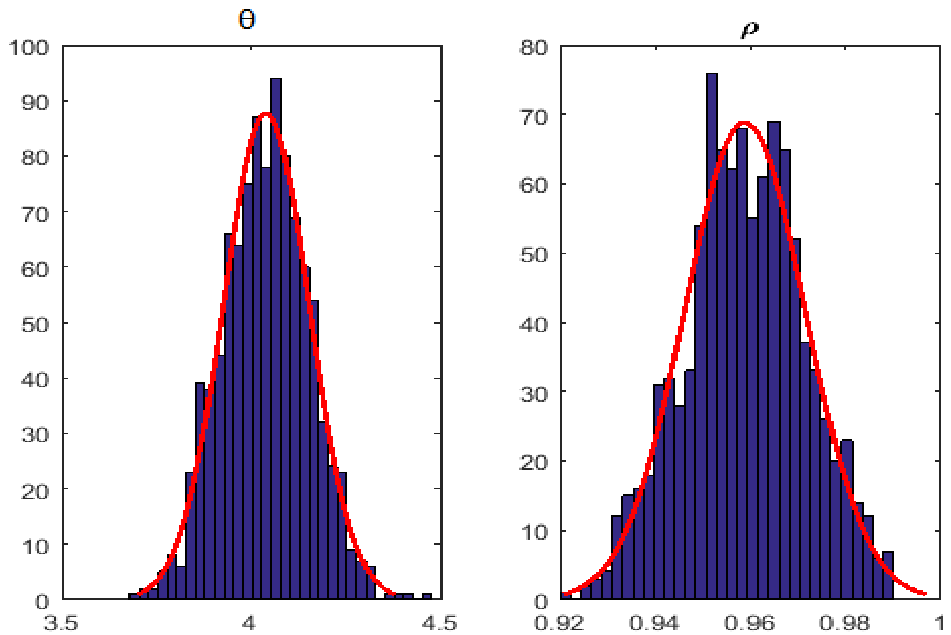

Figure 3.7 Les estimations de θ et ρ pour n=500

On remarque que les données s'ajuste selon une loi normale.

3.3 Application de l'échantillonneur de Gibbs pour la prédiction d'une valeur future

Pour $\theta = 4$, $\rho = 0.5$ et $\varepsilon_i \sim N(0.1)$, la valeur prédite y_{n+1} est estimée comme moyenne a posteriori de la densité prédictive, pour différentes tailles d'échantillon. La vraie valeur de y_{n+1} est obtenue en simulant y_{n+1} plusieurs fois et en prenant la moyenne. On donne aussi le rmse et le std des estimations de la valeur prédite y_{n+1}.

Les résultats de simulation sont donnés dans le tableau (3.2)

	n=50	n=200	n=500
Vraie valeur y_{n+1}	2.0517	1.8958	1.9599
$\hat{y}^{(std)}_{n+1_{(rmse)}}$	$2.0409^{(0.4246)}_{(0.5408)}$	$1.9009^{(0.3463)}_{(0.5114)}$	$1.9627^{(0.3214)}_{(0.4761)}$

Table 3.3 -Les valeurs prédites de y_{n+1} pour différentes tailles d'échantillons

On remarque que les valeurs de std et rmse diminuent quand la taille de l'échantillon n augmente et la valeur prédite est proche de la vraie valeur quelle que soit la taille de l'échantillon.

Conclusion générale

4

Dans ce mémoire, nous avons adopté la méthodologie bayésienne pour estimer les paramètres d'un modèle de régression linéaire dans les deux cas où les résidus sont indépendants et autocorrélés.

On a utilisé l'algorithme de Gibbs Sampler pour estimer le paramètre du modèle de régression linéaire et le paramètre du modèle autorégressif dans les cas où les résidus sont autocorrélés. On a déterminé les lois a posteriori conditionnelles des paramètres de ce modèle et la densité prédictive d'une future observation.

Une étude de simulation a montré que les paramètres du modèle et la valeur future sont bien estimés au sens du rmse.

Plusieurs perspectives peuvent être explorées :

- Étendre le travail au cas d'un modèle autorégressif d'ordre p, $p \geq 2$

- Considérer des lois a priori informatives.

الملخص

في هذه الدراسة ، استخدمنا خوارزمية جيبس لتقدير معاملات نموذج الانحدار الخطي بأخطاء أر (١) والتنبؤ بقيمة مستقبلية . يتم إجراء محاكاة لتوضيح أداء هذه الطريقة.

Résumé

Dans ce travail, nous avons utilisé l'algorithme de Gibbs sampler pour l'estimation des paramètres d'un modèle de régression linéaire à erreurs AR(1) et la prédiction d'une valeur future.
Une étude de simulation est faite pour illustrer la performance de cette méthode.

Abstract

In this study, we used the Gibbs sampler algorithm to estimate the parameters of a linear regression model with AR(1) errors and to predict a future value. A simulation study is carried out to illustrate the performance of this method.

Bibliographie

N. Ait Mohammed. Approche bayésienne dans la détection des observations abberantes. *Thèse de Magister. UMMTO*, 2006.

J. M. Bernardo et A. FM Smith. *Bayesian theory*, volume 405. John Wiley & Sons, 2009.

G. EP Box, G. M. Jenkins, G. C. Reinsel, et M. Ljung, Greta. *Time series analysis : forecasting and control*. John Wiley & Sons, 2015.

B. P. Carlin et T. A Louis. *Bayes and empirical Bayes methods for data analysis*, volume 7. Kluwer Academic Publishers Norwell, MA, USA, 1997.

G Casella et E. George. Explaining the gibbs sampler. *American statistician*, 1992.

S. Chib. Bayes regression with autoregressive errors : A gibbs sampling approach. *Journal of econometrics*, 58(3) :275–294, 1993.

R. Fisher. *Statistical methods and scientific inference*. Oliver and Boyd, 1956.

A. Gelfand, E. Hills, A. Racine-poon, et A.F.M. Smith. Illustration of bayesian inference in normal data models. *Journal of the American Statistical Association*, 85(412) :972–985, 1990.

A. Gelfand et A Smith. Sampling-based approaches to calculate marginal densities. 85(410) :398–409, 1990.

A. Gelman, J. B. Carlin, H. S. Stern, et D. B. Rubin. *Bayesian data analysis*. Chapman and Hall/CRC, 1995.

S Gelman et D. Gelman. Stochastic relaxation, gibbs distributions and the bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*, (6) :721–741, 1984.

W. Gilks, S. Richardson, et D Spiegelhalter. *Markov chain monte carlo in practice.* chapman and hall, New York, 1996.

P. Girard et E. Parent. Analyse bayésienne du modèle linéaire avec erreur autocorrélée par échantillonnage de Gibbs : application à la modélisation d'un procédé agroalimentaire à partir de données recueillies sur ligne. *Revue de Statistique Appliquée*, 48(2) :5–34, 2000. URL http://www.numdam.org/item/RSA_2000__48_2_5_0/.

W. Hastings. Monte carlo sampling methods using markov chains and their application. *Biometrika*, pages 57,79–109, 1977.

D.A. John, Berry K. Ch., et G. A. Zellner. *Bayesian analysis in statistics and econometrics : Essays in honor of Arnold Zellner*, volume 309. John Wiley & Sons, 1996.

N. Metropolis, A. W Rosenbluth, Marshall N. Rosenbluth, A. H Teller, et Edward Teller. Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6) :1087–1092, 1953.

E. Parent et J. Bernier. *Le raisonnement bayésien, modélisation et inférence*. springer-verlag, France, 2007.

H. Raiffa, R. Schlaifer, et al. *Applied statistical decision theory*. Wiley New York, 1961.

C. Robert. *Le choix bayésien, principes et pratique*. springer-verlag, France, 2006.

C. Robert et G Casella. *Monte carlo statistical mrthods*. springer-verlag, France, 2004.

C. P. Robert. Des spécificités de l'approche bayésienne et de ses justifications en statistique inférentielle. *arXiv preprint arXiv :1403.4429*, 2013.

G. Robert et A. Smith. Simple conditions for the convergence of the gibbs sampler and metropolis-hastings algorithms, stochastic processes and their applications. 1993.

R. O. Schlaifer. Probability and statistics for business decisions : An introduction to managerial economics under uncertainty. 1959.

A. Zellner et G. C. Tiao. Bayesian analysis of the regression model with autocorrelated errors. *Journal of the American Statistical Association*, 59(307) : 763–778, 1964.