

N° d'ordre : 07/2017-D/ MT

RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université des Sciences et de la Technologie Houari Boumediene
Faculté des Mathématiques



THÈSE

Présentée pour l'obtention du grade de DOCTEUR EN SCIENCES

EN : MATHÉMATIQUES

SPÉCIALITÉ : PROBABILITÉS ET STATISTIQUE

Par :

SLAMA Abdeldjalil

Sujet :

**ANALYSE BAYÉSIENNE D'UN CHANGEMENT DANS LES
PARAMÈTRES D'UN MODÈLE DE SÉRIE TEMPORELLE
AUTORÉGRESSIF**

Soutenue publiquement le Mercredi 12 Avril 2017 devant le jury composé de :

<i>Mme. H. GUERBYENNE</i>	<i>Professeur</i>	<i>USTHB</i>	<i>Présidente</i>
<i>Mme. H. SAGGOU</i>	<i>Maitre de Conférence/A</i>	<i>USTHB</i>	<i>Directrice de thèse</i>
<i>Mme. K. DJABALLAH</i>	<i>Professeur</i>	<i>USTHB</i>	<i>Examinatrice</i>
<i>M. K. KHALDI</i>	<i>Professeur</i>	<i>Université de Boumerdès</i>	<i>Examineur</i>
<i>Mme. M. OURBIH</i>	<i>Professeur</i>	<i>Université de Tipaza</i>	<i>Examinatrice</i>

A la mémoire des mes parents

A ma Femme et mes enfants

Menna, Amine et Majd

A ma sœur et mes frères

A toute ma famille

Remerciements

Mes premiers remerciements vont à ma directrice de thèse Mme Hafida Saggou, maître de conférence A au département de Probabilités et de statistique, faculté des mathématiques, USTHB, de m'avoir donné l'opportunité de réaliser mon projet de recherche et de m'avoir assuré un encadrement rigoureux tout au long de ces années. Je la remercie également pour ses efforts exceptionnels et pour m'avoir soutenu d'une façon remarquable pendant la réalisation de cette thèse.

J'éprouve une reconnaissance toute particulière pour Mme Guerbyenne Hafida, Professeur au département de Probabilités et de statistique, faculté des mathématiques, USTHB, pour sa disponibilité, ses encouragements et ses précieux conseils. Je tiens à la remercier pour toutes les discussions constructives que nous avons eues. Je la remercie pour avoir accepté de présider le jury de cette thèse.

Mes remerciements vont également à Mme Khadidja Djaballah, Professeur au département de Probabilités et de statistique, faculté des mathématiques, USTHB, à Mme Megdouda Ourbih-Tari, Professeur au Centre universitaire de Tipaza et à M. Khaldi Khaled, professeur au département de mathématiques, faculté des sciences, Université de Boumerdès, pour avoir bien accepté d'être membres de jury de soutenance de cette thèse.

Mes remerciements vont aussi à tous mes collègues du département de Probabilités et de statistique, faculté de mathématiques de l'USTHB ainsi à mes collègues du département de mathématiques et d'informatique de l'université d'Adrar pour leur soutien constant.

Je ne saurais terminer sans remercier tous mes amis qui m'ont soutenu durant ce parcours pour la réalisation de cette thèse.

Résumé

Intitulé : **Analyse Bayésienne d'un changement dans les paramètres d'un modèle de série temporelle autorégressif**

Les problèmes de changement (de rupture) trouvent leurs origines dans beaucoup de domaines expérimentaux et théoriques tels que en hydrologie, métrologie, économie, médecine, etc. Chronologiquement ordonnées, les observations sont collectées dans un intervalle de temps dont on suspect qu'un changement brusque s'est produit dans les paramètres du modèle, et l'objectif est de faire des inférences statistiques (estimations, tests d'hypothèses, etc...) sur la position du point de rupture et sur les autres paramètres du modèle.

Dans ce travail, on présente une analyse Bayésienne d'un changement dans les paramètres d'un modèle autorégressif d'ordre connu p , $AR(p)$. Nous avons considéré deux cas, le premier, un changement dans la moyenne μ et la variance de l'erreur σ^2 , le deuxième cas considéré est un changement à un instant inconnu dans les paramètres d'autorégression et aussi dans la variance de l'erreur σ^2 . Une étude de simulation en utilisant l'algorithme de l'échantillonneur de Gibbs a montré que le test de signification Bayésien pour détecter un changement dans les paramètres d'autorégression du modèle $AR(p)$ est plus puissant que les tests basés sur les sommes partielles des résidus pondérés.

Mots clés : Série temporelle, Modèle Autorégressif, Point de rupture, Analyse Bayésienne, P-values, Échantillonneur de Gibbs.

Abstract

Entitled : **A Bayesian analysis of a change in the parameters of autoregressive time series**

Change point problems arise in a variety of experimental and theoretical fields such as hydrology, metrology, economics, medicine, etc. A change point, which is generally the effect of an external event on the phenomenon of interest, may be represented by a change in the structure of the model or simply by a change of the value of some parameters.

In this work, we consider a Bayesian analysis of a possible change in the parameters of autoregressive time series of known order p , $AR(p)$. An unconditional Bayesian test based on HPD credible sets is determined. From numerical studies via the Gibbs sampler algorithm, we approximated the marginal posterior densities of the change point, of the magnitude of the shift in the autocorrelation parameters, of the variance ratio and the unconditional p-values. A comparative study shows that the Bayesian significance test of change in the parameters of autoregressive time series has a stronger power than the tests based on the partial sums of weighted residuals. Another advantage of the Bayesian significance test over these tests is the possibility of detecting separately the autoregressive parameters which are subject to a change.

Keywords : Autoregressive model, change point, Bayesian analysis, HPD credible set, P-values, Gibbs sampler.

Travaux de recherche :

Les contributions des travaux de recherche réalisés dans le cadre de cette thèse sont 3 publications et 2 communications dans des colloques Internationaux :

Publications :

1. Slama, A. A Bayesian significance test of change for correlated observations, *Discussiones Mathematicae, Probability and Statistics*, 34, 51–63, 2014.
2. Slama, A. A Bayesian significance test of change in the presence of single outlier, *Afrika statistika*, Volume 9(1), pp 685-694, 2014.
3. Slama A. et Saggou H., A Bayesian analysis of a change in the parameters of autoregressive time series, to appear in *Communications in Statistics - Simulation and Computation*, 2016, <http://dx.doi.org/10.1080/03610918.2016.1222423>.

Séminaires et colloques :

1. 8 ième Rencontre d'Analyse Mathématique et ses Applications, RAMA 8, Algiers, Algeria 26th -29th November 2012, www.usthb.dz/RAMA8/
Slama A., Bayesian change point detection for correlated data.
2. Journées Internationales de Statistiques Théoriques et appliquées, JISTA 2013, 24-26 Novembre 2013, Département de probabilités et statistique, Faculté des mathématiques, USTHB, 2013. www.usthb.dz/JISTA13/
Slama A., Bayesian analysis of epidemic change in exponential family.

Table des matières

Introduction générale	1
Introduction	6
1 Notions Générales et définitions	7
1.1 Introduction	7
1.2 Modèles de rupture	7
1.3 Modèle autorégressif $AR(\rho)$	11
1.3.1 Introduction	11
1.3.2 Opérateur de retard	12
1.3.3 Stationnarité	13
1.3.4 Vraisemblance conditionnelle	15
1.4 Inférence Bayésienne	15
1.4.1 Modèles Bayésiens	16
1.4.1.1 Théorème de Bayes	16
1.4.2 Lois a priori	18
1.4.2.1 Lois a priori conjuguées	18
1.4.2.2 Familles exponentielles	19
1.4.2.3 Lois conjuguées des familles exponentielles	19
1.4.2.4 Lois a priori non informatives	20
1.4.3 Estimateur de Bayes	22
1.4.4 Fonctions de coût usuelles	25
1.4.4.1 Coût quadratique	25

TABLE DES MATIÈRES

1.4.4.2	Coût absolu	26
1.4.4.3	Coût 0-1	27
1.4.5	Critères d'optimalités : Minimaxité et admissibilité	28
1.4.5.1	Estimateurs randomisés	28
1.4.5.2	Minimaxité	28
1.4.5.3	Admissibilité	29
1.4.6	Estimateur du maximum a posteriori (MAP)	30
1.4.7	Tests et régions de confiance	30
1.4.7.1	Région α -crédible Bayésien	31
1.4.7.2	Région HPD (highest posteriori density)	31
1.4.7.3	Erreurs et puissance	32
1.4.7.4	Facteur de Bayes	33
1.4.8	p-values	33
1.5	Méthodes de Calcul Bayésien	35
1.5.1	Méthodes classiques d'approximation	36
1.5.1.1	Intégration numérique	36
1.5.1.2	Méthodes de Monte Carlo	37
1.5.2	Méthodes de Monte Carlo par chaînes de Markov	37
1.5.2.1	Algorithmes de Metropolis-Hastings	38
1.5.2.2	Échantillonneur de Gibbs	39
1.6	Lois de probabilités usuelles	41
1.6.1	Loi Normale	42
1.6.2	Loi de Student	42
1.6.3	Loi de Fisher	43
2	Modèle $AR(p)$ avec changement dans la moyenne et la variance	44
2.1	Introduction	44
2.2	Modèle et notations	46
2.3	Analyse Bayésienne	47
2.3.0.1	Loi a priori	47

TABLE DES MATIÈRES

2.3.0.2	Loi a posteriori	48
2.4	Test de signification Bayésien	54
2.5	Étude de simulation	57
2.5.1	Densité a posteriori de m et les p-values de H_0	57
2.5.2	Effet de l'autocorrélation sur le test de signification Bayésien	58
2.6	Conclusion	59
3	Modèle $AR(p)$ avec changement dans les paramètres d'autorégression	60
3.1	Introduction	60
3.2	Modèle et notations	62
3.3	Analyse Bayésienne	63
3.3.0.1	Loi a priori	63
3.3.0.2	Loi a posteriori	64
3.4	Test de signification Bayésien :	67
3.5	Étude de simulation	69
3.5.1	Distributions a posteriori des paramètres	70
3.5.2	Étude de la performance du test	74
3.5.3	Étude de comparaison :	76
3.6	Conclusion	77
	Conclusion générale	79
	Bibliographie	81

Table des figures

3.1	Observations simulées y_t avec $m = 110$	69
3.2	Observations simulées y_t avec $m = 150$	70
3.3	Densité a posteriori de m avec $m = 110$	71
3.4	Distribution a posteriori de δ_1 avec $m = 110$	71
3.5	Distribution a posteriori de δ_2 avec $m = 110$	72
3.6	Distribution a posteriori de τ avec $m = 110$	72
3.7	Distribution a posteriori de m avec $m = 150$	73
3.8	Distribution a posteriori de δ_1 avec $m = 150$	73
3.9	Distribution a posteriori de δ_2 avec $m = 150$	73
3.10	Distribution a posteriori de τ avec $m = 150$	73

Liste des tableaux

2.1	Densité a posteriori marginale de m et les p-values conditionnelles sachant m de H_{01} et de H_{02} au voisinage du vrai point de rupture $m = 34$	57
2.2	p-values inconditionnelles et conditionnelle sachant m et ϕ_1 de H_{01} et H_{02}	58
2.3	Les p-values conditionnelles sachant $\rho = 0$ et inconditionnelles des sous hypothèses H_{01} et de H_{02} pour différentes valeurs de ϕ_1	58
3.1	Densité a posteriori de m , les p-values conditionnelles et inconditionnelles de H_{011} , H_{012} et H_{02} pour $m = 110$	71
3.2	Densité a posteriori de m , les p-values conditionnelles et inconditionnelles de H_{011} , H_{012} et H_{02} pour $m = 150$	72
3.3	Estimateurs a posteriori de m , δ_1 , δ_2 et de τ lorsau $m = 110$	73
3.4	Estimateurs a posteriori de m , δ_1 , δ_2 et de τ lorsque $m = 150$	74
3.5	Mode a posteriori de m et la p-value inconditionnelle de H_{011} , H_{012} et de H_{011} pour différentes valeurs de m , ϕ_1 , ϕ_2 , ψ_1 , ψ_2 , σ_1 et de σ_2	75
3.6	Taux de rejet de H_{011} , H_{012} et de H_{02} au niveau de signification de 5% pour 1000 échantillons avec différentes valeurs de δ_1 , δ_2 et de τ	75
3.7	Puissances empiriques des tests T_{nR}^* , $T_{nR}^*(\epsilon_0)$, $T_{nR}^*(q_0)$, T_{nP}^* , $T_{nP}^*(\epsilon_0)$, $T_{nP}^*(q_0)$ et le test de signification Bayésien de H_{01} pour un niveau de signification égal à 0.05.	77

Notations

$AR(p)$	Processus autorégressif d'ordre p
$F, F(n, m)$	Loi de Fisher de n et m degrés de liberté
$T(n)$	Loi de Student à n degrés de liberté
$N(\mu, \sigma^2)$	Loi gaussienne (normale) d'espérance μ et de variance σ^2
H_0	Hypothèse nulle
H_1	Hypothèse alternative
$\pi(\theta)$	Loi a priori sur θ
$L(\theta, \delta)$	Fonction de coût de δ en θ
$l(\theta y)$	Fonction de vraisemblance
iid	indépendantes et identiquement distribuées
$\pi(\theta y)$	La loi a posteriori sur θ
\propto	Proportionnelle
α	niveau de signification
n	Taille de l'échantillon
(Y_1, Y_2, \dots, Y_n)	suite de n variables aléatoires
(y_1, y_2, \dots, y_n)	suite de n observations
$I(\theta)$	Information de Fisher

Introduction générale

Le monde est en changement permanent. Une prise de conscience de ces changements peut aider les gens à éviter les pertes inutiles et d'exploiter les transitions en leurs faveurs. Du point de vue statistique, la perturbation de l'homogénéité stochastique des données pourrait être un signe de certains changements (pannes, dysfonctionnements, etc.) dans le processus observé. En fait et dans de nombreuses situations pratiques, le statisticien est confronté au problème de la détection de l'emplacement et le nombre de points de changement (de rupture) dans les données statistiques. Beaucoup des problèmes pratiques de détection de changement peuvent être trouvés dans de nombreuses disciplines, dont on peut citer ;

Exemple 0.1. Analyse du marché Boursier : Les rapports journaliers du marché boursiers américains montrent que le prix de l'action pour toute entreprise varie chaque jour. Bien que le changement d'un prix de l'action est normal selon la théorie économique, il y a des changements qui sont anormaux et méritent une attention particulière de l'investisseur. On peut par exemple se poser la question de savoir si l'affaire du Watergate a provoqué un changement significatif dans le marché boursier américain.

Exemple 0.2. Contrôle statistique d'un processus : En contrôle statistique d'un processus, parfois on est préoccupé par d'éventuels changements dans sa variance. Lorsque la valeur de la variance est supérieure à une certaine valeur bien déterminée, le processus doit être arrêté et déclaré "hors de contrôle".

Exemple 0.3. Analyse des données géologiques : Les mesures sur des échantillons carottés obtenues à partir de différents sites géologiques sont très importants pour un géologue. S'il existe un changement significatif dans les mesures du même type de carotte obtenue

INTRODUCTION GÉNÉRALE

à partir de différents sites, le géologue pourrait s'intéresser à savoir sur quel site a eu le changement, et par conséquent dans quel site il y a le pétrole, l'or, ou d'autres mines souterraines.

Exemple 0.4. Contrôle de qualité : Dans un processus continu de production, on prévoit que la qualité des produits demeure stable. Cependant, pour plusieurs raisons, le processus pourrait échouer à produire des produits de la même qualité. On veut donc savoir s'il y a un point où la qualité des produits commence à se détériorer.

Exemple 0.5. Taux de mortalité par accident de la circulation routière : En juillet 2009, une loi relative à l'organisation, la sécurité et la police de la circulation routière a été promulguée en Algérie. La promulgation de cette loi pourrait avoir un effet sur la réduction de taux de mortalité sur les routes. On peut, par conséquent, étudier la variation du taux de mortalité par accident de la circulation avant et après la promulgation de la loi.

Parmi les exemples mentionnés ci-dessus, les exemples 0.1 et 0.2, sont des problèmes de détection de point de changement appelés en ligne ou séquentiel, les autres exemples rentrent dans la catégorie de la détection de point de changement hors ligne ou analyse rétrospective d'un point de changement.

Depuis les travaux de Page [81, 82] dans le cadre du contrôle de la qualité d'une production d'objets, le problème de détection de changement est devenu un sujet important en statistique et une attention considérable a été accordée à ce problème sous différentes approches. Dans le cadre non paramétrique, le problème a été étudié par de nombreux auteurs, tels que Bhattacharya [8], Dumbgen [33], Carlstein [19], Pettitt [84], Yao [111], Brodsky et Darkhovsky [15]. Huskova et al. [63, 64] ont utilisé une approche basée sur les sommes partielles des résiduelles pondérées pour la détection d'un changement dans les paramètres de processus autorégressif, la variance est supposée constante. Pour plus de détails sur les techniques non paramétriques pour la détection des points de changement nous renvoyons à Csorgo and Horvath [30, 29] et leurs références ;

L'approche classique basée sur la fonction de vraisemblance a été largement utilisée dans la littérature pour la détection et l'estimation du point de rupture. Citons, par exemple, les travaux de Quandt [89, 90], Hinkley [54], Sen and Srisvastava [102, 101],

INTRODUCTION GÉNÉRALE

Hawkins [51], Horváth [59], Siegmund [103, 104], Worsley [116, 117, 118], Kim.H,J [68] et de Bai [4]. Quandt [89, 90] a utilisé l'estimateur du maximum de vraisemblance pour estimer les paramètres d'un modèle de régression linéaire à deux phases. Sen et Srivastava [101] (1975) ont comparé les statistiques de type Bayésien contre les statistiques basées sur la méthode du maximum de vraisemblance (ML) pour détecter un changement dans la moyenne des variables aléatoires normales. Ils ont constaté que lorsque la valeur initiale de la moyenne est inconnue et pour des petits changements, le test de type bayésien est plus puissant que les tests du maximum de vraisemblance. Kim.H,J [68] a considéré le test de rapport de vraisemblance généralisé pour détecter un changement dans la moyenne d'une suite de variables aléatoires non indépendantes. Il a examiné l'effet de la corrélation sur la performance du test de rapport de vraisemblance déterminé sous l'hypothèse d'indépendance et il a suggéré une approximation analytique des p-values du test de rapport de vraisemblance. Kim et Siegmund [72] ont considéré le test du rapport de vraisemblance pour un point de changement dans un modèle de régression simple et ont déterminé une approximation analytique du niveau de signification de test. Les propriétés analytiques du test du rapport de vraisemblance ont été étudiées par Kim H. J. [73].

Dans le cadre Bayésien, le problème de détection des points de rupture a été étudié par plusieurs auteurs. Les premiers travaux sont dus à Chernoff et Zacks [24] où ils se sont intéressés à la détection d'un changement dans la moyenne d'une suite d'observations i.i.d et normalement distribuées et un test Bayésien pour détecter un changement dans la moyenne est proposé. Les propriétés du test ont été étudiées plus tard par Kander et Zacks [69] pour le cas le plus général d'une suite de variables aléatoires issue d'une famille exponentielle à un seul paramètre. Ilmakunnas et Tsurumi [65] ont étudié un modèle de régression gaussien à erreurs autocorrélés où le coefficient d'autocorrélation change à un point supposé connu, ils se sont intéressés au paramètre de translation pour lequel ils déterminent un intervalle de confiance conditionnel.

Menzeffricke [77] a obtenu les distributions a posteriori du point de changement et du rapport de variance d'une suite de variables aléatoires gaussiennes. Guerbyenne [47] a étendu les travaux de Manzericke [77] au cas du modèle gaussien à erreurs autocorrélées avec changement dans la variance à un instant inconnu, les densités a posteriori du point

INTRODUCTION GÉNÉRALE

de changement, du coefficient de translation et de rapport des variances ont été obtenues dans les trois cas suivants :

1. Les moyennes sont distinctes et connues
2. Les moyennes sont égales et inconnues
3. Les moyennes sont distinctes et inconnues

Hsu [60] a déterminé deux tests pour un changement dans la variance d'une séquence de variables aléatoires indépendantes. Aussi, pour une séquence de variables aléatoires indépendantes et normalement distribuées, Abraham et Wei [1] ont déterminé les distributions a posteriori des paramètres d'une série temporelle ARMA. En supposant que le point de rupture est connu, Broemeling et Tsurumi [13] ont considéré un modèle linéaire avec erreurs corrélées. Ng [80] a analysé un modèle linéaire dans lequel la moyenne et la précision changent à un instant inconnu. Ils ont déterminé les distributions a posteriori du point de rupture et de rapport des précisions.

Le problème de rupture dans les modèles de régression a été considéré par Ferreira [35], Holbert et Broemeling [57], Holbert [58], Chin Choy et Broemeling [28], et Salazar et al. [99]. Récemment, Chaturvedi et Shrivastava [21] ont donné une analyse bayésienne d'un modèle de régression linéaire avec des changements structurels dans les paramètres de régression et de la précision avec l'hypothèse d'une distribution a priori mélange pour les paramètres considérés dans le modèle. Il y a beaucoup d'autres papiers qui sont apparus dans la littérature en utilisant l'approche non bayésienne pour la détection de rupture dans les modèles de régression, citons par exemple Kim H.J. [73], Kim et Siegmund [72], Bai [3] et Gurevich et Vexler [48].

Kim.D [71], pour un modèle de régression où les erreurs sont indépendantes, a développé un test de signification Bayésien pour tester la stationnarité des paramètres par la détermination des densités a posteriori du coefficient de translation et du rapport des variances. Ce test, en fait, ne suppose pas que la variance soit constante comme avait supposé Chow [27], ou que le point de changement soit fixe (Problème de Behrens-Fisher). Kim.D [71], par des méthodes de calcul numérique, a montré que ce test est plus puissant que les tests de Cusum et Cusum des carrés développés par Brown, Durbin et Evans [16]. Ghorbanzadeh et Lounes [45] ont proposé un test bayésien basé sur les régions HPD

INTRODUCTION GÉNÉRALE

crédibles pour détecter un changement dans les paramètres d'une séquence de variables aléatoires indépendantes issues d'une famille exponentielle. Slama [106] a étudié l'impact d'une contamination sur la performance du test de signification Bayésien pour la détection de rupture dans un échantillon gaussien en présence d'une observation aberrante.

Par ailleurs, le problème de détection de ruptures multiples a été largement étudié, on peut citer, entre autres, les travaux de Inclan et Tiao [66], Wang and Zivot [114], Chib [26], Fan et al. [34], Lai et Xing [74], Jandhyala et al. [67].

Pour les modèles autorégressifs, de nombreux articles sont publiés sur la détection et l'estimation du point de changement dans les paramètres du modèle $AR(p)$. Davies et al. [31] ont étudié le comportement asymptotique d'une statistique du rapport de vraisemblance de type Gaussien pour tester un changement dans les paramètres d'un processus autorégressif d'ordre p . Huskova et al. [63, 64] ont utilisé une approche basée sur les sommes partielles de résidus pondérés (méthodes asymptotique et bootstrap). Gombay [39] a utilisé un test basé sur le vecteur score. Venkatesan et Arumugam [112] ont considéré le problème des changements graduels dans les paramètres d'un modèle de série chronologique autorégressif. Barbieri et Conigliani [2] ont adopté l'approche des facteurs de Bayes pour détecter un changement dans la moyenne d'un modèle autorégressif stationnaire. Kezim et Abdelli [70] ont proposé une analyse bayésienne d'un processus autorégressif d'ordre un ($AR(1)$) avec un changement à la fois dans la variance et dans les coefficients d'autocorrélation à un instant inconnu.

Notre travail est consacré à la détection et l'estimation du point de rupture ainsi que à l'estimation des autres paramètres d'un modèle de série temporelle autorégressif d'ordre connu, $AR(p)$. L'adoption de l'approche Bayésienne et l'utilisation de l'algorithme de l'échantillonneur de Gibbs (Gibbs sampler) ont permis de résoudre les difficultés associées à la détection de ruptures et à l'estimation des paramètres d'un modèle de changement (change point models), tels que l'approximation des intégrales complexes de grande dimension.

Cette thèse comprend trois chapitres,

Le premier chapitre est consacré à la présentation des définition et les notions générales nécessaires à la bonne compréhension de la notion de rupture dans les modèles de série

INTRODUCTION GÉNÉRALE

temporelles autorégressif : l'approche de l'inférence bayésienne, les modèles de rupture, le modèle autorégressif d'ordre p , méthode de calcul bayésien et lois de probabilités usuelles.

Dans le chapitre deux, nous donnons le test de signification Bayésien pour un changement dans la moyenne et la variance d'un modèle autorégressif d'ordre connu p est présenté. Aussi nous examinons l'effet de l'autocorrélation des observations sur le test de signification bayésien pour détecter un changement dans la moyenne.

Dans le chapitre trois, nous présentons une analyse Bayésienne d'un modèle autorégressif d'ordre connu p avec un changement dans les paramètres d'autorégression aussi bien dans la variance à un instant inconnu m . Aussi, nous développons un test de signification Bayésien inconditionnel basé sur la construction des régions HPD crédibles et le calcul des p-values.

Enfin, nous terminons par une conclusion générale et l'énumération de quelques perspectives de recherche.

Chapitre 1

Notions Générales et définitions

1.1 Introduction

Dans le présent chapitre nous donnons quelques définitions de différentes notions présentées dans cette thèse. Nous présentons des notions générales relatives aux problèmes de rupture, les modèle de série temporelles autorégressif d'ordre connu, l'analyse statistique Bayésienne et les méthodes de calcul Bayésien.

1.2 Modèles de rupture

Soit (Y_1, Y_2, \dots, Y_n) une suite de n variables aléatoires de fonction de distribution de probabilité F_1, F_2, \dots, F_n respectivement. On dit que les observations (y_1, y_2, \dots, y_n) sont issues d'un modèle de rupture à l'instant m si les premières observations (y_1, y_2, \dots, y_m) suivent une loi commune F_0 et que les dernières observations $(y_{m+1}, y_{m+2}, \dots, y_n)$ suivent une autre loi commune F_1 . Le point m est appelé le point le point de rupture ou de changement, souvent supposé inconnu. Le but en fait, est de faire des inférences statistiques (estimations, tests d'hypothèses, etc...) sur la position du point de rupture ainsi que sur les autres paramètres du modèle.

Définition 1.1. Au sens large, le terme de "rupture" correspond à une modification brusque à un instant donné dans une (ou les) caractéristique (s) d'une séquence, d'un processus ou

d'un champ aléatoire.

Le problème de point rupture peut être considéré comme l'un des problèmes centraux de l'inférence statistique, reliant conjointement la théorie du contrôle statistique, la théorie de l'estimation et de tests des hypothèses, des approches classiques et bayésiennes, et de taille d'échantillon fixe et procédures séquentielles. A l'origine, un problème de rupture est caractérisé par : Lorsqu' un processus est sous contrôle, les observations sont distribuées selon F_0 , à un point inconnu m le processus saute "hors de contrôle" et les observations suivantes (celles après m) sont distribuées selon F_1 . L'objectif est alors de déclencher "dès que possible" une alarme après que le processus saute hors de contrôle.

Historiquement, le problème de détection des points de rupture a commencé à émerger dans la décade 1920-1930. Motivé par des travaux sur le contrôle de la qualité initiés par Shewhart [100], où Il a développé des méthodes de contrôle statistique de la qualité dans les fabrications industrielles de grandes séries. Mais ce n'est que à la fin des années 1950 que le problème des points de rupture a été formulé dans un sens mathématique précis par A.N.Kolmogorov (voir [15]). Les premiers travaux concernant l'estimation et la détection de point de rupture sont dus à Page [81, 82] et à Girshick et Rubin [46] qui sont apparus dans les années 1950 dans le cadre de contrôle industriel de qualité (La carte de contrôle de Shewhart, test Cusum). Afin de détecter un changement dans la moyenne d'une suite normale, i.e. de μ_0 à $\mu_1 > \mu_0$, Page [81] a proposé que le processus est hors de contrôle dès que la quantité $S_n - \min_{1 \leq k \leq n} S_k$ devient grande, où, $S_k = \sum_{i=1}^k (y_i - \mu^*)$, $\mu_0 < \mu^* < \mu_1$ est choisie de façon appropriée. Cette procédure est appelée procédure de Cusum (Cumulative sum). Girshick et Rubin [46] ont proposé pour le même problème un test quasi-Bayésien. L'approche Bayésienne a été considérée dans Chernoff et Zacks [24] et dans Kander et Zacks [69]. L'approche non paramétrique a été considérée plus tard dans Bhattacharya et Johnson [8] et Bhattacharya et Frierson [9].

Le problème de point de rupture a ensuite été intensivement étudié dans les travaux de Lorden [75], Pollak [86], Hinkley [53, 54, 55, 56], Zacks [119] et d'autres. Aujourd'hui, ce domaine de recherche ressemble à une grande famille de problèmes mathématiques reflétant différentes approches. Pour une revue plus générale sur les problèmes de rupture on peut voir Chen et Gupta [23], Csorgo, M., Horvath [30] , Basseville et Nikiforov [5] et

Brodsky et Darkhovsky [15].

Les classifications suivantes de problèmes de point de rupture peuvent être proposées[15]:

1. Suivant la méthode d'acquisition des données :

Il existe deux possibilités fondamentales : Soit l'ensemble des observations (y_1, y_2, \dots, y_n) sont disponibles au moment où l'hypothèse d'homogénéité des observations est vérifiée, on parle alors d'un problème de détection de point de rupture hors ligne ou analyse rétrospective de point de rupture. Soit par contre, les observations sont obtenues séquentiellement, i.e. un par un. Cette situation est assez courante dans les systèmes de contrôle et de diagnostique. Dans cette situation, le problème de détection de point de rupture est dit en ligne ou séquentiel ou encore en temps réel.

2. Suivant la complétude de l'information statistique a priori :

Selon des informations statistiques a priori, on peut distinguer entre les méthodes paramétriques, semi paramétriques et non paramétriques de détection de ruptures. Les méthodes paramétriques sont basées sur la complète disponibilité des informations a priori sur le modèle, i.e. le modèle probabiliste des données.

3. Suivant les caractéristiques des données :

En fonction des caractéristiques des données, on peut distinguer les problèmes de rupture pour les processus aléatoires et les problèmes de rupture pour les champs aléatoires. En fonction du domaine discret ou continu du processus (ou champ) aléatoire, on peut distinguer entre les problèmes de rupture pour des processus (champs) aléatoire discrets et continus.

Aussi, par le caractère de dépendance statistique entre les observations, nous pouvons formuler des problèmes de rupture pour des séquences (champs) aléatoires avec des observations indépendantes, et des problèmes de rupture pour les variables aléatoires dépendantes à une dimension et multidimensionnelle.

4. Suivant le type de rupture :

Dans le cadre classique de points de rupture, les changements sont brusques : les propriétés stochastiques sont constantes pendant un certain temps, puis sautent soudainement à une autre valeur. Dans un certain nombre de situations, cepen-

dant, les changements se produisent progressivement plutôt que brusquement : les propriétés sont (approximativement) constantes pendant un certain temps, puis atteignent un point où ils commencent lentement à changer. Ce type de rupture est appelé changement graduel dans les caractéristiques probabilistes des observations. Aussi, en fonction du nombre de points de rupture, on peut distinguer entre les problèmes de rupture avec un seul ou multiple points de rupture, on peut trouver aussi les problèmes de rupture épidémique comme cas particulier des problèmes de rupture multiple.

D'une manière générale, le problème de détection de rupture dans les caractéristiques probabilistes des observations consiste à tester l'hypothèse suivante :

$$H_0 : F_1 = F_2 = \dots = F_n$$

contre l'hypothèse alternative

$$\begin{aligned} H_1 : F_1 = F_2 = \dots = F_{k_1} \neq F_{k_1+1} = F_{k_1+2} = \dots = F_{k_2} \neq F_{k_2+1} \\ = \dots = F_{k_q} \neq F_{k_q+1} = \dots = F_n. \end{aligned}$$

Où $1 < k_1 < k_2 < \dots < k_q < n$, q est le nombre inconnu des points de rupture, et k_1, k_2, \dots, k_q sont leurs positions.

L'inférence statistique dans les modèles de rupture consiste en la détection et l'estimation du (ou des) point (s) où il y a lieu le changement ainsi qu'à l'estimation des autres paramètres du modèle.

Dans le cadre paramétrique, on suppose que les distributions F_1, F_2, \dots, F_n appartiennent à la même famille de distribution paramétrique $F_\theta, \theta \in \mathbb{R}^p$, le problème de détection de changement (de rupture) est alors de tester l'hypothèse nulle sur les paramètres $\theta_i, \theta_i \in \mathbb{R}^p, i = 1, \dots, n$:

$$H_0 : \theta_1 = \theta_2 = \dots = \theta_n = \theta \quad (\text{inconnu})$$

contre l'hypothèse alternative

$$\begin{aligned} H_1 : \theta_1 = \theta_2 = \dots = \theta_{k_1} \neq \theta_{k_1+1} = \theta_{k_1+2} = \dots = \theta_{k_2} \neq \theta_{k_2+1} \\ = \dots = \theta_{k_q} \neq \theta_{k_q+1} = \dots = \theta_n. \end{aligned}$$

où q, k_1, k_2, \dots, k_q sont à estimer. Si $q = 1$ c'est un problème d'un seul point de rupture, et il est équivalent à :

$$H_0 : m = n \quad \text{contre} \quad H_0 : m \neq n$$

. Dans le présent travail, nous considérons l'approche paramétrique Bayésienne pour la détection rétrospective d'un seul point de rupture dans un modèle de série temporelle autorégressif.

1.3 Modèle autorégressif $AR(p)$

1.3.1 Introduction

Une série chronologique ou une série temporelle est une suite d'observations représentant l'évolution d'une quantité spécifique au cours du temps. En pratique, les séries temporelles sont présentes dans des domaines aussi variés que l'économétrie, la médecine, la métrologie ou la démographie. Les modèles mathématiques de séries temporelles sont utilisés pour une tentative de donner un sens aux phénomènes ou données représentés par une série temporelle. Les modèles de séries temporelles sont adoptés, entre autres, pour l'étude de la prévision, le contrôle ou pour la détection de point(s) de rupture (changement).

Il existe de nombreux modèles utilisés pour les séries chronologiques, les plus utilisés sont, les modèles autorégressifs (AR), les modèles intégrés (I) et les modèles à moyenne mobiles (MA). Ces modèles sont souvent entrelacés pour générer de nouveaux modèles. Par exemple, les modèles autorégressifs à moyenne mobiles (ARMA) qui combinent le modèle (AR) et le modèle (MA). Un autre exemple des modèles de séries temporelles est le modèle autorégressif intégré à moyenne mobile (ARIMA), qui combine les trois modèles mentionnés précédemment. Le modèle de série temporelle considéré dans ce travail est le modèle autoregressif d'ordre p , $AR(p)$.

Définition 1.2. La suite $\{Y_t, t \in \mathbb{Z}\}$ est un processus autorégressif d'ordre p ($p > 0$) si pour tout t , il peut s'écrire sous la forme suivante :

$$Y_t = \sum_{j=1}^p \phi_j Y_{t-j} + \epsilon_t, \quad \text{où } \epsilon_t \sim N(0, \sigma^2) \quad (1.1)$$

Pour $t = 0, \pm 1, \pm 2, \dots$,. Avec $\epsilon_t \perp \epsilon_s$ pour $t \neq s$. p est un entier non négatif.

Dans ce cas, on note $Y_t \sim AR(p)$, le paramètre $\phi = (\phi_1, \phi_2, \dots, \phi_p)$ est le vecteur des paramètres autorégressifs.

Un modèle autorégressif d'ordre p est un modèle de régression pour séries temporelles Y_t dans lequel la série est expliquée par ses valeurs précédentes Y_{t-i} .

Remarque 1.1. Le processus $\{Y_t, t \in \mathbb{Z}\}$ est dit un processus autorégressif de moyenne μ si $Y_t - \mu$ est un processus $AR(p)$.

Exemple 1.1. Un processus $AR(1)$ prend la forme suivante :

$$Y_t = \phi Y_{t-1} + \epsilon_t, \quad \text{où } \epsilon \sim N(0, \sigma^2)$$

1.3.2 Opérateur de retard

Définition 1.3. L'opérateur de retard B est défini par :

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

Remarque 1.2. Pour $n \in \mathbb{N}$, $B^n(Y_t) = Y_{t-n}$

En utilisant l'opérateur de retard B , l'équation (1.1) peut être écrite sous la forme compacte suivante :

$$\phi(B)Y_t = \epsilon_t \quad (1.2)$$

où ϕ est le polynôme d'ordre p suivant

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p$$

appelé le polynôme caractéristique associé au modèle $AR(p)$.

Le polynôme $\phi(z)$ est utilisé pour donner une condition de stationnarité au second ordre du processus $AR(p)$.

1.3.3 Stationnarité

La stationnarité d'une série temporelle est liée à ses propriétés statistiques relativement au temps. Dans le sens le plus strict, une série temporelle stationnaire montre un « comportement statistique » identique et semblable dans le temps et cela est souvent caractérisé comme une distribution de probabilité constante dans le temps.

Définition 1.4. Soit $\{Y_t : t \in \mathbb{Z}\}$ une suite de variables aléatoires telle que $V(Y_t) < \infty$ pour tout $t \in \mathbb{Z}$, on appelle fonction d'autocovariance de Y_t la fonction $\gamma(\cdot, \cdot)$ définie par

$$\gamma(r, s) = \text{Cov}(Y_r, Y_s) = E(Y_r - E(Y_r))(Y_s - E(Y_s)), r, s \in \mathbb{Z}$$

Définition 1.5. (Stationnarité)

Une suite $\{Y_t : t \in \mathbb{Z}\}$ de variables aléatoires est dite (faiblement) stationnaire (ou stationnaire du second ordre) si elle vérifie les propriétés suivantes :

1. $\forall t \in \mathbb{Z}, E[Y_t] = \mu < \infty$, indépendant de t ,
2. $\forall t \in \mathbb{Z}, E[Y_t^2] < \infty$,
3. $\gamma(r, s) = \gamma(r + t, s + t), \forall r, s, t \in \mathbb{Z}$,

La stationnarité telle qu'elle vient d'être définie est souvent désignée dans la littérature comme stationnarité faible, stationnarité de covariance, stationnarité au sens large ou stationnarité de second ordre.

Remarque 1.3. Si $\{Y_t : t \in \mathbb{Z}\}$ est stationnaire, alors $\gamma(r, s) = \gamma(r - s, 0)$ pour tout $r, s \in \mathbb{Z}$. on peut alors redéfinir la fonction d'autocovariance d'une suite stationnaire Y_t comme une fonction d'une seule variable par

$$\gamma(h) = \gamma(h, 0) = \text{Cov}(Y_{t+h}, Y_t) \quad \text{pour tous } t, h \in \mathbb{Z}$$

et la fonction d'autocorrélation ρ de Y_t est définie par

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \text{Corr}(Y_{t+h}, Y_t), \quad \text{pour tous } t, h \in \mathbb{Z}$$

Une autre notion importante de la stationnarité est donnée par la définition suivante :

Définition 1.6. (stationnarité au sens strict)

Une suite $\{Y_t : t \in \mathbb{Z}\}$ de variables aléatoires est dite stationnaire au sens strict (au sens fort) si les distributions conjointes de $(Y_{t_1}, \dots, Y_{t_k})$ et de $(Y_{t_1+h}, \dots, Y_{t_k+h})$ sont les mêmes pour tout entier positif k et pour tous $t_1, \dots, t_k, h \in \mathbb{Z}$.

Remarque 1.4. La définition (1.6) est équivalente à que (Y_1, \dots, Y_k) et $(Y_{1+h}, \dots, Y_{k+h})$ ont la même distribution conjointe pour tous entiers k et h .

Remarque 1.5. La stationnarité au sens strict implique la stationnarité (au sens faible).

La stationnarité du processus $AR(p)$ impose des conditions strictes sur les valeurs possibles des paramètres du modèle. Le théorème suivant donne la condition imposée sur les paramètres du processus $AR(p)$ pour qu'il soit stationnaire.

Théorème 1.1. *Un processus autorégressif $AR(p)$ est stationnaire si et seulement si son polynôme caractéristique $\phi(z)$ est tel que*

$$\phi(z) \neq 0 \quad \text{avec } z \in \mathbb{C} \quad \text{tel que } |z| \leq 1.$$

en d'autre terme, si et seulement si le module des toutes les racines du polynôme caractéristique du processus sont supérieures à 1.

La démonstration de ce théorème est donnée dans Box et Jenkins [10] pages 53-54.

Exemple 1.2. Considérons le processus $AR(1)$,

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

le processus peut être écrit comme

$$(1 - \phi B)Y_t = \varepsilon_t$$

. le polynôme caractéristique est donné par :

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

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

Pour plus de détails sur les caractéristiques du processus $AR(p)$ on peut voir Box et Jenkins [10], Fuller [38], Priestley [?] et Brockwell et Davis [12].

1.3.4 Vraisemblance conditionnelle

Soient (X, Y) un couple de variables aléatoires de densité conjointe $f_{X,Y}(x, y | \theta)$. La densité conjointe de (X, Y) peut être écrite comme le produit de la densité conditionnelle de $(Y | X)$ et de la densité marginale de X , comme

$$f_{X,Y}(y, x | \theta) = f_{Y|X}(y | x, \theta) f_X(x | \theta).$$

Définition 1.7. Soit y_1, \dots, y_n n observations iid supposés suivre la distribution $f_{Y|X}$, on appelle la vraisemblance conditionnelle de Y sachant X la quantité

$$l(y | \theta, X) = \prod_{i=1}^n f_{Y|X}(y_i | X, \theta).$$

La quantité $l(y | \theta, X)$ est vue comme étant une fonction de θ une fois qu'on a observé y .

Si on dispose des observations (y_1, \dots, y_n) d'un processus stationnaire Y_t supposé suivre un modèle $AR(p)$, c'est à dire vérifiant

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

La vraisemblance conditionnelle de (y_1, \dots, y_n) est donnée par

$$l(y | \sigma, \phi_1, \dots, \phi_p) \propto \frac{1}{\sigma} \exp \left\{ -\frac{1}{\sigma^2} \sum_{t=1}^n \left(y_t - \sum_{i=1}^p \phi_i y_{t-i} \right)^2 \right\}.$$

La vraisemblance conditionnelle est basée sur l'hypothèse que les observations $(y_0, y_{-1}, \dots, y_{1-p})$ sont disponibles (supposées fixes)[92]. De plus, lorsque la taille de l'échantillon n est suffisamment grand, les premières observations initiales ont un effet négligeable sur la vraisemblance totale [49].

1.4 Inférence Bayésienne

En général, l'objectif principal de la statistique paramétrique est de faire des inférences (estimations, tests d'hypothèses,..) 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 $y = (y_1, y_2, \dots, y_n)$ et calculer la fonction de

densité ou probabilité conditionnelle des données sachant θ , que l'on note $f(y|\theta)$. Cette densité conjointe, lorsque elle est vue comme fonction de θ , est appelée la 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 de Thomas Bayes).

L'approche statistique Bayésienne qui se veut 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 . Elle se distingue des approches dites " classiques " par la construction et l'utilisation d'une loi de probabilité sur l'ensemble des paramètres objet d'une inférence statistique.

Cette partie, fortement inspirée de : Le choix bayésien, Principes et pratique (Robert, 2006) 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 [7]; Carlin et Louis [17]; Gelman et al. [42] et Robert [96].

1.4.1 Modèles Bayésiens

Définition 1.8. On appelle modèle statistique Bayésien, la donnée d'un modèle statistique paramétrique, $(\mathcal{Y}, \mathcal{B}, \{P_\theta, \theta \in \Theta\})$, où \mathcal{Y} désigne l'espace des données, Θ est l'espace des paramètres qui est muni d'une probabilité π tel que $(\Theta, \mathcal{A}, \pi)$ est un espace probabilisé, π est appelée loi a priori de θ . $\{P_\theta, \theta \in \Theta\}$ est une famille de probabilités définies sur la tribu \mathcal{B} de parties de \mathcal{Y} .

1.4.1.1 Théorème de Bayes

Supposons que $y = (y_1, \dots, y_n)$ 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

$$\pi(y|\theta)\pi(\theta) = \pi(y, \theta) = \pi(\theta|y)\pi(y). \quad (1.3)$$

Étant donné les observations (y_1, \dots, y_n) , la distribution conditionnelle de θ est

$$\begin{aligned} \pi(\theta|y) &= \frac{\pi(y|\theta)\pi(\theta)}{\pi(y)}, \\ &= c\pi(y|\theta)\pi(\theta), \end{aligned} \quad (1.4)$$

où, $c^{-1} = \pi(y) = E(\pi(y|\theta))$ est l'espérance mathématiques de $\pi(y|\theta)$ par rapport à la distribution $\pi(\theta)$.

La formule (1.3) ou son équivalente (1.4) est appelée la formule de Bayes (ou théorème 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 l'intégrale ou la somme de la distribution a posteriori $\pi(\theta | y)$ vaut 1.

Étant donnée les données y , $\pi(y|\theta)$ dans la formule (1.4) peut être considérée comme une fonction de θ et non de y . Suivant Fisher (1922), elle est appelée la fonction de vraisemblance de θ sachant y qui peut être écrite $l(\theta | y)$. Nous pouvons donc écrire la formule de Bayes comme

$$\begin{aligned} \pi(\theta|y) &= c l(\theta|y)\pi(\theta) \\ &\propto l(\theta|y)\pi(\theta) \end{aligned} \quad (1.5)$$

c'est-à-dire la distribution a posteriori de θ sachant y est proportionnelle au produit de la loi a priori de θ et la fonction de vraisemblance $l(\theta|y)$.

La fonction de vraisemblance $l(\theta|y)$ joue un rôle très important dans la formule de Bayes. Elle est la fonction par laquelle les données y modifient la connaissance préalable de θ ; elle peut donc être considérée comme représentante de l'information à propos de θ provenant des données y .

Par application directe 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 conditionnelle de

y conditionnellement à θ , la loi de θ conditionnelle à y , $\pi(\theta|y)$ appelée distribution a posteriori de θ , est définie par

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

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

1.4.2 Lois a priori

Le choix de la 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. Il peut 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 suffisamment précise pour proposer une forme exacte ou même paramétrée 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.

1.4.2.1 Lois a priori conjuguées

Introduite par Raiffa et Schlaifer [91], l'approche des lois a priori conjuguées, peut être justifiée partiellement par un raisonnement d'invariance. En fait, quand l'observation de $y \sim f(y|\theta)$ vient modifier $\pi(\theta)$ 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'une loi a priori conjuguée, bien qu'il soit défendable, est toujours un choix particulier et il a une influence, dans une

certaine mesure, sur l'inférence résultante. De plus, il peut obliger à ignorer une partie de l'information a priori si cette dernière n'est pas complètement compatible avec la structure de la loi a priori conjuguée. Une analyse bayésienne fondée sur les lois conjuguées devrait donc être suivie d'une analyse de robustesse.

Définition 1.9. Une famille \mathcal{F} de lois sur Θ est dite conjuguée si, pour toute π appartenant à cette famille, la loi $\pi(\theta | y)$ appartient également à \mathcal{F} .

1.4.2.2 Familles exponentielles

Un type particulier de lois de probabilité qui permet une détermination directe de famille de lois conjuguées est la famille exponentielle.

Définition 1.10. Soient μ une mesure σ -finie sur \mathcal{Y} , Θ l'espace des paramètres, C et h des fonctions respectivement de \mathcal{Y} et Θ dans \mathbb{R}^+ , et R et T des fonctions de Θ et \mathcal{Y} dans \mathbb{R}^k . La famille des distributions de densité (par rapport à μ)

$$f(y | \theta) = C(\theta)h(y) \exp\{\langle R(\theta), T(y) \rangle\} \quad (1.6)$$

est dite famille exponentielle de dimension k . Dans le cas particulier où $\Theta \subset \mathbb{R}^k$, $\mathcal{Y} \subset \mathbb{R}^k$ et

$$f(y|\theta) = C(\theta)h(y) \exp\{\theta \cdot y\}, \quad (1.7)$$

la famille est dite naturelle.

De nombreuses lois de probabilités usuelles (discrètes et continues) appartiennent à des familles exponentielles telles que la loi Binomiale, loi de Poisson, loi normale, loi Gamma, etc.

1.4.2.3 Lois conjuguées des familles exponentielles

Soit $f(y | \theta) = h(y)e^{\theta \cdot y - \psi(\theta)}$, loi générique d'une famille exponentielle. Cette loi admet alors une famille de lois conjuguée,

Proposition 1.2. (voir C. P. Robert [96])

Une famille conjuguée pour $f(y | \theta)$ est donnée par

$$\pi(\theta \mid \mu, \lambda) = K(\mu, \lambda) e^{\theta \cdot \mu - \psi(\theta)}$$

où $K(\mu, \lambda)$ est la constante de normalisation de la densité. La loi a posteriori correspondante est $\pi(\theta \mid \mu + y, \lambda + 1)$.

Des exemples de lois a priori conjuguées pour quelques familles exponentielles usuelles peuvent être trouvées dans Christian P. Robert [96].

1.4.2.4 Lois a priori non informatives

Lorsque aucune information a priori n'est disponible, l'utilisation des lois a priori conjuguées n'est justifiée que par des considérations analytiques. 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 intègrent 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, nous renvoyons à Christian P. Robert [96].

Lois a priori de Laplace

Historiquement, Laplace fut le premier à utiliser des techniques non informatives. Bien que ne disposant 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, en utilisant donc une loi uniforme. Son raisonnement, appelé plus tard principe de la raison insuffisante, se fondait sur l'équiprobabilité des événements élémentaires. Des critiques ont été plus tard avancées sur ce choix, dont la plus fondamentale concerne le problème de l'invariance par paramétrisation. Si on passe de $\theta \in \Theta$ à $\eta = g(\theta)$ 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 η est

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

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

Pour aboutir à une notion plus intrinsèque (donc plus acceptable) des lois a priori non informatives il faut proposer une définition invariante par reparamétrisation (Voir chapitre 6 de J.O. Berger [6] et chapitre 9 de C.P Robert [96] pour plus de détails sur le principe d'invariance et les lois non informatives).

L'approche invariante n'est que partiellement satisfaisante, car elle implique la référence à une structure d'invariance, qui peut être parfois choisie de plusieurs manières, ne pas exister, ou être sans intérêt pour le décideur [96].

Loi a priori de Jeffreys

Jeffreys (1946, 1961) a proposé 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

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

dans le cas unidimensionnel. Sous certaines conditions de régularité, cette information est aussi égale

$$I(\theta) = -\mathbb{E}_\theta \left[\frac{\partial^2 \log f(y | \theta)}{\partial \theta^2} \right] \tag{1.8}$$

La loi a priori non informative de Jeffreys est

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

Le choix d'une loi a priori dépendante de l'information de Fisher se justifie par le fait que $I(\theta)$ est largement acceptée comme un indicateur de la quantité d'information apportée par le modèle (ou l'observation) sur θ (Fisher,[36]).

Dans le cas où θ est un paramètre multidimensionnel, on définit la matrice d'information de Fisher par généralisation de (1.8). Pour $\theta \in \mathbb{R}^k$, $I(\theta)$ a les éléments suivants :

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

et la loi non informative de Jeffreys est alors définie par

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

la loi $\pi(\theta)$ vérifie effectivement l'exigence d'invariance par reparamétrisation, 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

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

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

1.4.3 Estimateur de Bayes

En pratique, le principe fondamental 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 : \mathcal{Y} , espace des observations, Θ , espace des paramètres, et \mathcal{D} , espace des décisions (ou espace d'actions). En pratique, l'inférence statistique consiste à prendre une décision $d \in \mathcal{D}$ par rapport au paramètre $\theta \in \Theta$, fondée sur l'observation $y \in \mathcal{Y}$, et il est important de pouvoir comparer les différentes décisions au moyen d'un critère d'évaluation, qui va apparaître sous forme d'un coût.

Définition 1.11. On appelle coût toute fonction L de $\Theta \times \mathcal{D}$ dans $[0, +\infty[$.

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

Définition 1.12. Une règle de décision d est une application de \mathcal{Y} dans \mathcal{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 \mathcal{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 recourir aux fonctions de coût classiques ou canoniques, choisies pour leur simplicité et leur souplesse mathématique [96].

Définition 1.13. Le coût moyen (ou risque fréquentiste) de la décision δ est défini par

$$\begin{aligned} R(\theta, \delta) &= \mathbb{E}_\theta[L(\theta, \delta(y))] \\ &= \int_{\mathcal{Y}} L(\theta, \delta(y)) f(y|\theta) dy \end{aligned} \tag{1.9}$$

La fonction δ , de \mathcal{Y} dans \mathcal{D} est appelée un estimateur de θ et 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 intègre sur l'espace Θ , car θ est inconnu, plutôt que de le faire sur l'espace \mathcal{Y} , y étant connu.

Dans le cadre de l'estimation, l'ensemble des décisions \mathcal{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 1.14. Pour une loi a priori π donnée, le coût moyen a posteriori est défini par :

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

qui est la moyenne de l'erreur (c'est-à-dire 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 1.15. En se donnant une distribution a priori π , le risque intégré est défini par

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

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, et non une fonction de θ . Il induit donc un ordre total sur l'ensemble des estimateurs et permet une comparaison directe entre ces estimateurs.

Théorème 1.3. (C. P. Robert [96])

Un estimateur minimisant le risque intégré $r(\pi, \delta)$ est obtenu par sélection, pour chaque $y \in \mathcal{Y}$, de la valeur $\delta(y)$ qui minimise le coût moyen a posteriori, $\varrho(\pi, \delta|y)$, puisque

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

où $m(y) = \int f(y|\theta)\pi(\theta)d\theta$

Définition 1.16. On appelle risque de Bayes la quantité définie par

$$\begin{aligned} r(\pi) = r(\pi, \delta^\pi) &= \mathbb{E}^\pi[R(\theta, \delta)] \\ &= \int_{\Theta} \int_{\mathcal{Y}} L(\theta, \delta(y))f(y|\theta)\pi(\theta)dyd\theta \end{aligned} \quad (1.10)$$

Définition 1.17. Un estimateur de Bayes associé à une distribution a priori π et à une fonction de coût L est un estimateur δ^π minimisant $r(\pi, \delta)$. Pour chaque $y \in \mathcal{Y}$, ce dernier est donné par :

$$\delta^\pi(y) = \arg \min_d \varrho(\pi, d|y)$$

Notons que le Théorème (1.3) fournit un outil constructif pour la détermination des estimateurs de Bayes. Aussi, d'un point de vue strictement bayésien, et comme nous connaissons la valeur observée de y , alors seul le coût moyen a posteriori $\rho(\pi, d|y)$ est utilisé pour déterminer les estimateurs de Bayes.

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 :

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

avec $(\theta, a) \in \Theta \times \Theta$

L'estimateur de Bayes δ^π du paramètre $\theta (\in \mathbb{R})$, 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 par

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

1.4.4 Fonctions de coût usuelles

Les choix les plus classiques de fonctions de coût sont les coûts suivants :

1.4.4.1 Coût quadratique

Définition 1.18. La fonction de coût quadratique est la fonction définie par :

$$L(\theta, \delta(y)) = (\theta - \delta(y))^2 \tag{1.11}$$

Proposition 1.4. (voir C. P. Robert [96])

L'estimateur de Bayes δ^π associé à la loi a priori π et au coût quadratique (1.11) 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 1.5. (voir C. P. Robert [96])

L'estimateur de Bayes δ^π associé à π et au coût quadratique pondéré

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

où $\omega(\theta)$ est une fonction positive, est

$$\delta^\pi(y) = \frac{E^\pi[\omega(\theta)\theta|y]}{E^\pi[\omega(\theta)|y]}$$

Corollaire 1.6. (voir C. P. Robert [96])

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

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

est égal la moyenne a posteriori, $\delta^\pi(y) = E^\pi[\theta | y]$, 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, du fait de sa convexité stricte.

1.4.4.2 Coût absolu

Définition 1.19. La fonction de coût absolue est la fonction définie par

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

Ou plus généralement une fonction définie par morceaux

$$L(\theta, \delta(y)) = \begin{cases} k_1(\theta - \delta(y)) & \text{si } \theta > \delta(y) \\ k_2(\delta(y) - \theta) & \text{sinon} \end{cases} \tag{1.13}$$

Cette fonction augmente moins rapidement que le coût quadratique et, tout en restant convexe, évite donc de pénaliser trop les plus grandes erreurs.

Proposition 1.7. (voir C. P. Robert [96])

L'estimateur de Bayes associé à la loi a priori π et à la fonction de coût linéaire par morceaux (1.13) est le fractile d'ordre $k_1/(k_1 + k_2)$ de $\pi(\theta | y)$.

En particulier, si $k_1 = k_2$, soit dans le cas du coût absolu, on obtient la médiane a posteriori.

1.4.4.3 Coût 0-1

Définition 1.20. On appelle coût 0-1, l'application L définie par

$$L(\theta, \delta(y)) = \begin{cases} 0 & \text{si la décision est bonne} \\ 1 & \text{sinon} \end{cases}$$

Ce coût est essentiellement utilisé dans l'approche classique des tests d'hypothèse, proposée par Neyman et Pearson. C'est un exemple typique d'un coût non quantitatif. En effet, pour ce coût, la pénalité associée à un estimateur $\delta(y)$ est 0 si la réponse est correcte et 1 sinon.

Exemple 1.5. Soit une partition de Θ en Θ_0 et Θ_1 . Soit le test de l'hypothèse nulle $H_0 : \theta \in \Theta_0$ contre l'hypothèse alternative $H_1 : \theta \notin \Theta_0$. Alors $\mathcal{D} = \{0, 1\}$, où 1 représente l'acceptation de H_0 et 0 son rejet. (En d'autres termes, la fonction de θ estimée est $I_{\Theta_0}(\theta)$).

On définit donc la décision de la manière suivante :

- $\delta(y) = 1$: on accepte H_0 ,
- $\delta(y) = 0$: on rejette H_0 ,

La fonction de coût 0 – 1 s'écrit donc :

$$L(\theta, \delta(y)) = \begin{cases} 1 - \delta(y) & \text{si } \theta \in \Theta_0 \\ \delta(y) & \text{sinon} \end{cases} \quad (1.14)$$

et le risque associé est

$$R(\theta, \delta) = E_\theta[L(\theta, \delta(y))] = (1 - \delta(y))P_\theta(\theta \in \Theta_0|y) + \delta(y)P_\theta(\theta \in \Theta_1|y)$$

Et donc pour $\delta(y) = 0$, on a $R(\theta, \delta) = P_\theta(\theta \in \Theta_0|y)$,

et pour $\delta(y) = 1$, on a $R(\theta, \delta) = P_\theta(\theta \in \Theta_1|y)$,

Ainsi, $\delta(y) = 1 \Leftrightarrow P^\theta(\theta \in \Theta_1|y) \leq P^\theta(\theta \in \Theta_0|y)$. C'est-à-dire que l'estimation permet d'accepter H_0 si c'est l'hypothèse la plus probable a posteriori.

1.4.5 Critères d'optimalités : Minimaxité et admissibilité

1.4.5.1 Estimateurs randomisés

Définition 1.21. Un estimateur randomisé δ^* est une décision de l'espace \mathcal{D}^* , l'espace des distributions de probabilité sur l'espace de décision \mathcal{D} .

Utiliser un estimateur randomisé δ^* signifie que l'action est générée selon la distribution de densité de probabilité $\delta^*(y, \cdot)$, une fois que l'observation y a été recueillie. Ainsi, on définit le coût de l'estimateur randomisé δ^* et le risque randomisés par :

$$L^*(\theta, \delta^*) = \int_{\mathcal{D}} L(\theta, \delta) \delta^*(y, a) da \quad \text{et} \quad R^*(\theta, \delta^*) = E_\theta[L^*(\theta, \delta^*)]$$

Théorème 1.8. (voir C. P. Robert [96])

Pour toute distribution a priori π sur θ , le risque de Bayes pour l'ensemble des estimateurs randomisés est le même que celui pour l'ensemble des estimateurs non randomisés, soit

$$\inf_{\delta \in \mathcal{D}} r(\pi, \delta) = \inf_{\delta^* \in \mathcal{D}^*} r(\pi, \delta^*) = r(\pi)$$

1.4.5.2 Minimaxité

Le critère de minimaxité vise à minimiser le coût moyen dans le cas le moins favorable.

Définition 1.22. On appelle risque minimax associé à la fonction de coût L la valeur

$$\bar{R} = \inf_{\delta \in \mathcal{D}^*} \sup_{\theta} R(\theta, \delta) = \inf_{\delta \in \mathcal{D}^*} \sup_{\theta} E_\theta[L(\theta, \delta(y))]$$

et estimateur minimax tout estimateur (éventuellement randomisé) δ_0 tel que

$$\sup_{\theta} R(\theta, \delta_0) = \bar{R}.$$

L'estimateur minimax correspond au point de vue " faire le mieux dans le pire des cas".

Lemme 1.9. (voir C. P. Robert [96])

Le risque de Bayes est toujours plus petit que le risque minimax,

$$\underline{R} = \sup_{\pi} r_{\pi}(\pi) = \sup_{\pi} \inf_{\delta \in \mathcal{D}} r(\pi, \delta) \leq \bar{R} = \inf_{\delta \in \mathcal{D}^*} \sup_{\theta} R(\theta, \delta).$$

La première valeur est dite risque maximin et une distribution π^* telle que $r(\pi^*) = \underline{R}$ est appelée distribution a priori la moins favorable, quand de telles distributions existent.

Lemme 1.10. *Si δ_0 est un estimateur de Bayes pour π_0 et si $R(\theta, \delta_0) \leq r(\pi_0, \theta)$ pour tout θ dans le support de θ , δ_0 est minimax et π_0 est la distribution la moins favorable.*

1.4.5.3 Admissibilité

Définition 1.23. Un estimateur δ_0 est inadmissible s'il existe un estimateur δ_1 qui domine δ_0 , c'est-à-dire tel que pour tout θ

$$R(\theta, \delta_0) \geq R(\theta, \delta_1)$$

et, pour au moins une valeur θ_0 du paramètre,

$$R(\theta_0, \delta_0) > R(\theta_0, \delta_1)$$

Sinon, δ_0 est dit admissible.

Le critère d'inadmissibilité est particulièrement intéressant pour son aspect réductrice. Du moins en théorie, il semble logique de soutenir que les estimateurs inadmissibles ne devraient pas être considérés, puisqu'ils peuvent être améliorés uniformément.

Proposition 1.11. (voir C. P. Robert [96])

S'il existe un unique estimateur minimax, cet estimateur est admissible.

La réciproque est fausse.

Proposition 1.12. (voir C. P. Robert [96])

Si δ_0 est admissible de risque constant, δ_0 est l'unique estimateur minimax.

la proposition suivante donne une condition suffisante d'admissibilité des estimateurs bayésiens.

Proposition 1.13. (voir C. P. Robert [96])

Si la distribution a priori π est strictement positive sur θ , de risque de Bayes fini, et la fonction de risque $R(\theta, \delta)$ est une fonction continue de θ pour tout δ , l'estimateur de Bayes δ^π est admissible.

Proposition 1.14. (voir C. P. Robert [96])

Si l'estimateur de Bayes associé à une loi a priori π est unique alors il est admissible.

Proposition 1.15. (voir C. P. Robert [96])

Si un estimateur de Bayes, δ^π , associé à une loi a priori (propre ou impropre) π , est tel que le risque de Bayes,

$$r(\pi) = \int_{\Theta} R(\theta, \delta^\pi) \pi(\theta) d\theta,$$

soit fini, δ^π est admissible.

1.4.6 Estimateur du maximum a posteriori (MAP)

Définition 1.24. On appelle estimateur du maximum a posteriori (MAP) tout estimateur $\delta^\pi(y)$ de θ qui maximise $\pi(\theta|y)$, i.e,

$$\delta^\pi(y) \in \underset{\theta}{\operatorname{argmax}} \pi(\theta|y)$$

L'estimateur MAP d'un paramètre θ est obtenu en cherchant le mode de la distribution de probabilité a posteriori du θ .

L'estimateur du maximum a posteriori est un estimateur associé au coût 0-1 (voir exemple (1.5), et il est asymptotiquement équivalent aux estimateurs du maximum de vraisemblance classiques, du fait que lorsque la taille de l'échantillon tend vers l'infini, l'information contenue dans cet échantillon devient prédominante par rapport à l'information fixe apportée par la loi a priori π [96].

1.4.7 Tests et régions de confiance

Soit un modèle statistique $f(y | \theta)$ avec $\theta \in \Theta$. Étant donné un sous ensemble d'intérêt de Θ , Θ_0 , qui se réduit parfois à un singleton $\{\theta_0\}$, la question posée est : la vraie valeur du paramètre θ appartient-elle à Θ_0 , ce qu'on appelle tester l'hypothèse

$$H_0 : \theta \in \Theta_0$$

appelée hypothèse nulle.

Dans la perspective de Neyman-Pearson, le problème de test est formalisé à l'aide d'un espace de décision \mathcal{D} restreint à {oui, non} ou, d'une manière équivalente, à $\{1, 0\}$, ce qui revient à estimer $I_{\Theta_0}(\theta)$ uniquement par des estimateurs à valeurs dans $\{1, 0\}$

Lorsque $\Theta_0 \neq \Theta_1$, on définit l'hypothèse alternative pour laquelle on veut tester H_0 par

$$H_1 : \theta \in \Theta_1$$

Avec l'hypothèse

$$\pi(\theta \notin \Theta_0 \cup \Theta_1) = 0$$

Ainsi, toute procédure de test φ apparaît comme un estimateur de $I_{\Theta_0}(\theta)$ et nous n'avons besoin que d'une fonction de coût $L(\theta, \varphi)$ pour construire des estimateurs de Bayes. Par exemple, la fonction de coût proposée par Neyman et Pearson est le coût "0 - 1"

$$L(\theta, \varphi) = \begin{cases} 1 & \text{si } \varphi \neq I_{\Theta_0}(\theta) \\ 0 & \text{sinon} \end{cases}$$

Pour ce coût, la solution bayésienne est

$$\varphi^\pi(y) = \begin{cases} 1 & \text{si } P^\pi(\theta \in \Theta_0 | y) > P^\pi(\theta \in \Theta_0^c | y) \\ 0 & \text{sinon} \end{cases}$$

1.4.7.1 Région α -crédible Bayésien

Définition 1.25. Pour une loi a priori π , une région C de Θ est dite α -crédible pour θ si et seulement si

$$P^\pi(\theta \in C | y) > 1 - \alpha$$

1.4.7.2 Région HPD (highest posteriori density)

Définition 1.26. Une région HPD pour θ est la région C^π définie par

$$C^\pi = \{\theta, \pi(\theta | y) > h_\alpha\}$$

où h_α est la plus grandes valeur telle que

$$P^\pi(\theta \in C^\pi | y) \geq 1 - \alpha$$

Considérer les régions HPD est motivé par le fait qu'elles sont de volume minimal parmi les régions α -crédibles, par conséquent, elles peuvent être perçues comme des solutions optimales dans un cadre de décision.

1.4.7.3 Erreurs et puissance

Lorsque on définit une règle de décision, on doit regarder quelle est la probabilité de faire des erreurs quand on prend l'une ou l'autre décision en fonction de la valeur de $\theta \in \Theta$. On peut faire deux types d'erreur :

1. On fait une erreur de première espèce, α , quand on rejette H_0 à tort (où d'accepter à tort l'hypothèse H_1).
2. On fait une erreur de seconde espèce, β , quand on accepte H_0 à tort. C'est à dire accepter H_0 alors que $\theta \in \Theta_1$.

Ce qui nous donne le tableau :

Décision Réalité	H_0 est vraie	H_0 est fausse
Non rejet de H_0	Cohérent	Erreur type II : risque β
Rejet de H_0	Erreur type I : risque α	Cohérent

Définition 1.27. La quantité α est appelée niveau de signification du test. C'est la probabilité de rejeter H_0 lorsque elle est vraie. i.e :

$$\alpha = P(\text{rejeter } H_0 | H_0 \text{ est vraie})$$

La quantité $1 - \alpha$ représente le degré de confiance du test.

Définition 1.28. La quantité $1 - \beta$ est appelée la puissance du test. C'est-à-dire la probabilité de rejeter H_0 lorsque elle est fausse. i.e

$$\begin{aligned} \text{Puissance du test} &= P(\text{rejeter } H_0 | H_0 \text{ est fausse}) \\ &= 1 - P(\text{accepter } H_0 | H_0 \text{ est fausse}) \\ &= 1 - \beta \end{aligned}$$

1.4.7.4 Facteur de Bayes

Définition 1.29. Le facteur de Bayes est le rapport des probabilités a posteriori des hypothèses nulle et alternative sur le rapport des probabilités a priori de ces mêmes hypothèses, soit

$$B_{01}^{\pi} = \frac{P(\theta \in \Theta_0 | Y)}{P(\theta \in \Theta_1 | Y)} \bigg/ \frac{\pi(\theta \in \Theta_0)}{\pi(\theta \in \Theta_1)}$$

Ce rapport évalue la modification de la vraisemblance de l'ensemble Θ_0 par rapport à celle de l'ensemble Θ_1 due à l'observation et peut se comparer naturellement à 1.

Dans le cas particulier où $\Theta_0 = \{\theta_0\}$ et $\Theta_1 = \{\theta_1\}$, le facteur de Bayes se simplifie et devient le rapport de vraisemblance classique

$$B_{01}^{\pi} = \frac{f(y | \theta_0)}{f(y | \theta_1)}$$

En général, le facteur de Bayes dépend de l'information a priori, mais il est souvent proposé comme réponse bayésienne " objective ", car il élimine partiellement l'influence du modèle a priori et il souligne le rôle des observations. De ce fait, il peut être perçu comme un rapport de vraisemblance bayésien, car, si π_0 est la loi a priori sous H_0 et π_1 , la loi a priori sous H_1 , le facteur de Bayes B_{01}^{π} peut s'écrire

$$B_{01}^{\pi} = \frac{\int_{\Theta_0} f(y | \theta_0) \pi_0(\theta) d\theta}{\int_{\Theta_1} f(y | \theta_1) \pi_1(\theta) d\theta}$$

ce qui revient donc à remplacer les vraisemblances par des marginales sous les deux hypothèses.

Pour une présentation plus détaillée des méthodes de l'inférence Bayésienne et de la théorie de la décision avec des exemples on pourra consulter J. Berger[6] et C. Robert[96] et leurs références.

1.4.8 p-values

Les fréquentistes (et praticiens) ont tenté de compenser les inconvénients de l'approche de Neyman-Pearson en supprimant le niveau de signification α et en proposant

une réponse prenant ses valeurs dans $[0, 1]$ et, de façon plus importante, dépendant des observations de manière plus adaptative qu'une acceptation ou un rejet établis en comparant $T(Y_1, Y_2, \dots, Y_n)$ (la statistique du test) à un seuil donné. La notion suivante de la p-value a été introduite pour la première fois par Fisher [36].

Définition 1.30. La p-value associée à un test est le niveau de signification α le plus petit pour lequel l'hypothèse nulle est rejetée.

Supposons que pour $\alpha \in (0, 1)$, nous avons un test de taille α avec région de rejet R_α . Alors

$$\text{p-value} = \inf\{\alpha : T(Y_1, Y_2, \dots, Y_n) \in R_\alpha\}$$

où $T(Y_1, Y_2, \dots, Y_n)$ est la statistique du test.

Typiquement, la p-value est une mesure de l'évidence empirique contre H_0 , i.e. elle est considérée comme une mesure d'à quel point les données plaident contre l'hypothèse nulle : plus la p-value est faible, plus l'évidence contre H_0 est forte. Les niveaux suivants sont souvent utilisés par les chercheurs [115] :

p-value	L'évidence
$< .01$	Très fortes évidences contre H_0
$.01 - .05$	Fortes évidences contre H_0
$.05 - .10$	Faibles évidences contre H_0
$> .1$	Peu ou pas d'évidences contre H_0

Théorème 1.16. (Wasserman [115]) Supposons que pour tester $H_0 : \theta \in \Theta_0$ contre $H_1 : \theta \notin \Theta_0$, un test de niveau α est défini par

$$\text{Rejeter } H_0 \text{ si et seulement si } T(Y_1, Y_2, \dots, Y_n) \geq c_\alpha$$

Alors

$$\text{p-value} = \sup_{\theta_0 \in \Theta_0} P_{\theta_0}(T(Y_1, Y_2, \dots, Y_n) \geq T(y_1, y_2, \dots, y_n))$$

où $y^n = (y_1, y_2, \dots, y_n)$ est la valeur observée de $Y^n = (Y_1, Y_2, \dots, Y_n)$. Si $\Theta_0 = \{\theta_0\}$ alors

$$\text{p-value} = P_{\theta_0}(T(Y_1, Y_2, \dots, Y_n) \geq T(y_1, y_2, \dots, y_n))$$

Ainsi, la p-value est la probabilité (sous H_0) d'observer une valeur de la statistique du test supérieure ou égale à que ce qui a été observé.

Généralement, on compare la p-value à un seuil préalablement défini (α), si la p-value est inférieure à ce seuil, on rejette l'hypothèse nulle en faveur de l'hypothèse alternative, et le test est déclaré « statistiquement significatif ».

1.5 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 espérances mathématiques :

$$U = \int u(\theta)\pi(\theta | y)d\theta \quad (1.15)$$

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 ; et
2. même si $\pi(\theta | y)$ est connu et lorsque l'intégration analytique est impossible, le calcul numérique nécessite parfois un temps de calcul considérable, en particulier lorsque Θ est de grande dimensions.

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

Le cas particulier et important du calcul des lois marginales se formule également dans ce cadre. Supposons θ vectoriel, on considère $\theta = (\theta_1, \theta_2)$, fixons la première partie des composantes à une valeur donnée $\theta_1 = t$; on peut écrire d'après les règles opératoires usuelles sur les densités conditionnelles :

$$\pi(t | y) = \int \pi(t, \theta_2 | y)\pi(\theta_2 | y)d\theta_2$$

En posant $u_t(\theta_2) = \pi(t, \theta_2 | y)$ on voit que :

$$\pi(t | y) = \int u_t(\theta_2)\pi(\theta_2 | y)d\theta_2$$

Donc, pour calculer la densité marginale de θ_1 au point $\theta_1 = t$ on peut effectuer le même calcul que (1.15) d'une espérance mathématique de la conditionnelle $u_t(\theta_2) =$

$\pi(t, \theta_2 | y)$ (dite conditionnelle complète de la composante θ_1) par rapport à la composante complémentaire θ_2 .

1.5.1 Méthodes classiques d'approximation

Cette section présente brièvement quelques techniques classiques qui peuvent faciliter les calculs bayésiens ;

1.5.1.1 Intégration numérique

A partir de la simple méthode de Simpson, plusieurs approches ont été conçues en Mathématiques appliquées pour l'approximation numérique d'intégrales. Par exemple, la quadrature polynomiale est censée approcher les intégrales liées à des distributions proches de la loi normale [96], L'approximation de base est donnée par

$$\int_{-\infty}^{+\infty} e^{-t^2/2} f(t) dt \approx \sum_{i=1}^n \omega_i f(t_i)$$

où

$$\omega_i = \frac{2^{n-1} n! \sqrt{n}}{n^2 [H_{n-1}(t_i)]^2}$$

et t_i est le i -ème zéro du n -ème polynôme d'Hermite, $H_n(t)$.

D'autres approximations d'intégrales reliées à la méthode précédente sont disponibles, qui reposent sur différentes bases orthogonales classiques ou les ondelettes. Ces méthodes qui exigent généralement que la fonction f soit suffisamment régulière, deviennent très instables (et donc peu fiables) lorsque la dimension de θ augmente, et que quelle que soit la méthode d'intégration numérique utilisée, sa précision diminue dramatiquement lorsque la dimension de Θ augmente. De façon plus spécifique, l'erreur associée aux méthodes numériques se comporte comme une puissance de la dimension de Θ . Par conséquent, la plupart des méthodes standard ne devraient pas être utilisées pour l'intégration en dimension supérieure à 4 et il est préférable d'utiliser une méthode de simulation [96].

1.5.1.2 Méthodes de Monte Carlo

La méthode de Monte Carlo, introduite par Metropolis et Ulam [78] et Von Neumann [113] est utilisée pour approximer des intégrales de la forme (1.15), en tirant d'avantage qu'elle peut s'écrire sous la forme :

$$U = \int u(\theta)f(y | \theta)\pi(\theta)d\theta \quad (1.16)$$

du fait que $\pi(\theta)$ soit une densité de probabilité (en supposant qu'il s'agisse d'une loi a priori propre). S'il est possible de produire des variables aléatoires $\theta_1, \dots, \theta_m$ de $\pi(\theta)$, la moyenne

$$\frac{1}{m} \sum_{i=1}^m u(\theta_i)f(y | \theta_i) \quad (1.17)$$

converge (presque sûrement) vers (1.16) lorsque m tend vers $+\infty$, selon la Loi des Grands Nombres.

Plus généralement, si h est une densité de probabilité telle que $\text{supp}(h)$ inclut le support de $u(\theta)f(y | \theta)\pi(\theta)$, l'intégrale (1.16) peut aussi être représentée comme une espérance en h , à savoir

$$\int \frac{u(\theta)f(y | \theta)\pi(\theta)}{h(\theta)} h(\theta)d\theta$$

Cette représentation conduit à la méthode de Monte Carlo avec fonction d'importance h : générer $\theta_1, \dots, \theta_m$ selon h et approcher (1.16) par

$$\frac{1}{m} \sum_{i=1}^m u(\theta_i)\omega_i(\theta_i)$$

avec les poids $\omega_i(\theta_i) = f(y | \theta_i)\pi(\theta_i)/h(\theta_i)$. De nouveau, par la Loi des Grands Nombres, cette approximation converge presque certainement vers (1.16).

1.5.2 Méthodes de Monte Carlo par chaînes de Markov

Dans cette section nous allons présenter des méthodes de Monte carlo qui permet de générer des variables aléatoires suivant approximativement la loi de probabilité a posteriori $\pi(\theta | y)$ lorsque cette loi n'est pas disponible (ne peut pas être simulée directement). Contrairement aux méthodes de Monte carlo classiques, ces méthodes évitent l'appel à la

fonction d'importance, h , souvent difficile à déterminer. Les méthodes de Monte Carlo par chaînes de Markov, ou méthodes MCMC (Markov chain Monte Carlo), sont des méthodes d'intégration numérique utilisant l'échantillonnage 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. [79] dans un contexte de physique statistique, et généralisée plus tard aux problèmes statistiques par Hastings [50]. L'idée de base de ces méthodes est de générer une chaîne de Markov $\{\theta_t\}$ dont les échantillons sont distribués asymptotiquement selon la distribution requise $\pi(\theta | y)$.

Définition 1.31. [97]

On appelle algorithme MCMC (pour Monte Carlo Markov Chain) toute méthode produisant une chaîne de Markov ergodique de loi stationnaire la distribution d'intérêt.

Dans la suite, nous présenterons les deux types de techniques les plus importantes conçues pour créer des chaînes de Markov de loi stationnaire donnée, à savoir les algorithmes de Metropolis-Hastings et l'échantillonnage de Gibbs.

1.5.2.1 Algorithmes de Metropolis-Hastings

La méthode de Metropolis-Hastings est historiquement la première des méthodes MCMC. elle a été développée par Metropolis et al. [79] au départ pour la physique particulaire, et généralisé par Hastings ([50]) dans un cadre plus statistique. 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 $\theta^{(m+1)}$ ($m = 1, 2, \dots$), de la façon suivante :
 - a) Générer $\xi \sim q(\xi | \theta^{(m)}, y)$
 - b) Poser $\rho(\theta^{(m)}, \xi) = \min\left\{\frac{\pi(\xi | y)q(\theta^{(m)} | \xi, y)}{\pi(\theta^{(m)} | y)q(\xi | \theta^{(m)}, y)}, 1\right\}$
 - c) Prendre $\theta^{(m+1)} = \begin{cases} \xi & \text{Avec probabilité } \rho(\theta^{(m)}, \xi) \\ \theta^{(m)} & \text{Sinon} \end{cases}$

La convergence théorique de l'algorithme de Metropolis-Hastings est prouvée par Roberts et Smith[93]. La loi de densité $\pi(\theta | y)$ est souvent appelée loi cible ou loi objet, tandis que la loi de densité $q(\cdot | \theta)$ est dite loi de proposition.

Remarque 1.6. q est quelconque mais elle doit vérifi e certaines conditions, telles que :

1. Doit  tre simulable rapidement,
2. Le support de q doit couvrir le support de π ,
3. q doit  tre une bonne approximation de π ,

1.5.2.2  chantillonneur de Gibbs

L'approche de l' chantillonnage de Gibbs tire son nom des champs al atoires de Gibbs, o  elle a  t  utilis e pour la premi re fois par Geman et Geman [43] comme un moyen pour simuler des distributions complexes de grande dimension qui interviennent dans le cadre de la restauration d'images. L'approche a  t  d velopp e plus tard dans le cadre du calcul Bay sien par Gelfand et Smith [41]. L' chantillonneur de Gibbs est une m thode de Monte-Carlo par cha ne de Markov (MCMC) permettant d'obtenir des  chantillons   partir d'une distribution conjointe par  chantillonnage it ratif en utilisant des distributions conditionnelles compl tes. Une bonne introduction   l' chantillonneur de Gibbs est donn e par Casella et George [20]. L' chantillonnage de Gibbs tire profit des structures hi rarchiques d'un mod le,i.e, lorsque celui-ci peut s' crire sous la forme

$$\pi(\theta_1 | y) = \int \pi_1(\theta_1 | \theta_2, y)\pi_2(\theta_2 | y)d\theta_2 \quad (1.18)$$

L'id e est alors de simuler la loi jointe $\pi_1(\theta_1 | \theta_2, y)\pi_2(\theta_2 | y)$ afin d'obtenir $\pi(\theta_1 | y)$ comme loi marginale de θ_1 . Bien entendu, lorsque les deux lois $\pi_1(\theta | \theta_2, y)$ et $\pi_2(\theta_2 | y)$ sont connues et peuvent  tre simul es, la g n ration de θ_1 de $\pi(\theta_1 | y)$ est  quivalente   la g n ration de θ_2 de $\pi_2(\theta_2 | y)$, puis de θ_1 de $\pi_1(\theta_1 | \theta_2, y)$.

La formulation g n rale de l'algorithme d' chantillonnage de Gibbs pour une loi jointe $\pi(\theta_1, \dots, \theta_p | y)$, de lois conditionnelles compl tes $\pi(\theta_1 | \theta_2, \theta_3, \dots, \theta_p, y)$, $\pi(\theta_2 | \theta_1, \theta_3, \dots, \theta_p, y)$, ..., $\pi(\theta_p | \theta_2, \theta_3, \dots, \theta_{p-1}, y)$; suppos es disponibles (simulables); est expos e ci-dessous.

- Initialisation : Commencer par une valeur arbitraire $\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_p^{(0)})$.
- Itération $t + 1$: pour $(\theta_1^{(t)}, \dots, \theta_p^{(t)})$ donné, générer
 1. $\theta_1^{(t+1)}$ selon la loi $\pi(\theta_1 | \theta_2^{(t)}, \theta_3^{(t)}, \dots, \theta_p^{(t)}, y)$,
 2. $\theta_2^{(t+1)}$ selon la loi $\pi(\theta_2 | \theta_1^{(t+1)}, \theta_3^{(t)}, \dots, \theta_p^{(t)}, y)$,
 3. $\theta_3^{(t+1)}$ selon la loi $\pi(\theta_3 | \theta_1^{(t+1)}, \theta_2^{(t+1)}, \dots, \theta_p^{(t)}, y)$,
 4. ...
 5. $\theta_p^{(t+1)}$ selon la loi $\pi(\theta_p | \theta_1^{(t+1)}, \theta_2^{(t+1)}, \dots, \theta_{p-1}^{(t+1)}, y)$,

Sous des conditions de régularité, le vecteur $\theta^{(t+1)} = (\theta_1^{(t+1)}, \theta_2^{(t+1)}, \theta_3^{(t+1)}, \dots, \theta_p^{(t+1)})$ ainsi obtenu converge en loi vers un échantillon provenant de la distribution $\pi(\theta_1, \dots, \theta_p | y)$ lorsque t devient grand et quelle que soit la valeur initiale $\theta^{(0)}$.

Théorème 1.17. [87] *La suite de vecteurs p -dimensionnels simulés $\{\theta^{(i)}\}$ par l'algorithme précédent est une chaîne de Markov dont la distribution p -dimensionnelle invariante, limite ergodique de la chaîne, est la distribution conjointe :*

$$\pi(\theta | y) = \pi(\theta_1, \theta_2, \dots, \theta_p | y)$$

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

Ainsi, si on répète tout le processus n fois en parallèle (les répétitions), on obtiendra un échantillon de n vecteurs $\theta_{(1)}^{(t+1)}, \theta_{(2)}^{(t+1)}, \dots, \theta_{(n)}^{(t+1)}$ indépendamment et identiquement distribués selon la loi jointe $\pi(\theta_1, \theta_2, \dots, \theta_p | y)$. L'échantillon $\theta_{(1)}^{(t+1)}, \theta_{(2)}^{(t+1)}, \dots, \theta_{(n)}^{(t+1)}$ peut être utilisé pour l'approximation des densités marginales ou toute autre caractéristique de la loi jointe (Espérances, Intervalles HPD crédibles, Quantiles, etc.) par de simples sommes finies d'expressions connues.

Ainsi, la densité marginale $\pi(\theta_i | y); i = 1, \dots, p$; peut être approcher par [20] :

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

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

pour toute fonction mesurable g de $(\theta_1, \dots, \theta_p)$, on a

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

La notation $E(g(\theta|y))$ est assez générale pour inclure la moyenne a posteriori $E[\theta|y]$ (avec $g(\theta) = \theta$), la matrice de covariances $V(\theta|y)$ (où $g(\theta) = (\theta - [\theta|y])^2$) ou encore la probabilité $P[\theta \in A|y]$ (où $g(\theta) = I_A(\theta)$), la fonction indicatrice de A).

La vitesse de convergence, c'est-à-dire le taux de décroissance de la distance entre la loi de $\theta^{(k)}$ et sa limite a été étudiée d'un point de vue théorique [96]. En pratique, le problème de déterminer le nombre de simulations à effectuer pour atteindre la convergence dépend de l'application particulière considérée et ne peut pas être spécifié à l'avance [40]. Pour plus de détails sur les critères de convergence des algorithmes MCMC nous renvoyons le lecteur à consulter les livres de Robert et Casella [95], Robert [96] et Robert [97].

L'algorithme de Gibbs est particulièrement bien adapté au calcul et à l'estimation des modèles bayésiens, hiérarchiques notamment, et il exploite au mieux la structure conditionnelle des équations de modélisation [87]. Pour plus de détail sur l'application de l'algorithme de l'échantillonneur de Gibbs au calcul Bayésien générale on peut voir Gelfand et Smith [41] et Gelfand et al. [40].

Plus de détail sur les méthodes MCMC peut être trouvés dans les ouvrages de Casella et George [20], Gelfand et Smith [41], Gilks et al. [44] et Robert et Casella [95].

Les méthodes de simulation de Monte Carlo par chaîne de Markov (MCMC) sont de plus en plus utilisées dans le calcul statistique pour estimer les caractéristiques de la fonction de vraisemblance et les distributions de probabilités a posteriori.

1.6 Lois de probabilités usuelles

Nous donnons dans cette partie les lois de probabilités utilisées dans cette thèse, à savoir loi normale, loi de student et loi de Fisher ;

1.6.1 Loi Normale

Définition 1.32. On dit qu'une variable aléatoire réelle Y suit une loi normale (ou gaussienne, loi de Laplace-Gauss) centrée réduite si sa loi admet pour densité

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}x^2\right], x \in \mathbb{R}$$

Et on écrit $Y \sim N(0,1)$, pour dire Y suit une loi normale centrée réduite.

Définition 1.33. Pour $\sigma > 0$ et $m \in \mathbb{R}$, on dit qu'une variable aléatoire réelle Y suit une loi gaussienne (Normale) $N(\theta; \sigma^2)$ si Y vérifie l'une des deux propriétés équivalentes suivantes :

1. $Y = \sigma X + \theta$ où X suit une loi gaussienne centrée réduite (i.e. $N(0,1)$);
2. la densité de Y est

$$f_X(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(x - \theta)^2\right], x \in \mathbb{R}$$

Et on écrit $Y \sim N(\theta, \sigma^2)$, avec $E(Y) = \theta$, $\text{var}(Y) = \sigma^2$.

Sommes de variables gaussiennes indépendantes. Supposons que Y suit la loi $N(\theta, \sigma^2)$, Y' suit la loi $N(\theta', \sigma'^2)$, et Y et Y' sont indépendantes. Alors $Y + Y'$ suit la loi $N(\theta + \theta', \sigma^2 + \sigma'^2)$.

1.6.2 Loi de Student

Définition 1.34. on dit Une variable aléatoire Y suit une loi de Student de moyenne θ et de variance σ^2 et de n degrés de liberté si sa densité de probabilité f_Y est définie par :

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

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

Remarque 1.7. On a, pour $\theta \in \mathbb{R}$, $S > 0$, et $n > 0$

$$\int_{-\infty}^{+\infty} \left\{1 + \frac{(y - \theta)^2}{nS^2}\right\}^{-\frac{n+1}{2}} dy = \frac{\Gamma(n/2)(\pi n)^{\frac{1}{2}} S}{\Gamma((n+1)/2)} \quad (1.19)$$

1.6.3 Loi de Fisher

Définition 1.35. On dit qu'une variable aléatoire réelle Y suit une loi de Fisher ou loi de Fisher-Snedecor avec (n_1, n_2) degrés de liberté si la densité de probabilité de Y est donnée par

$$f_Y(y) = \begin{cases} \frac{\Gamma((n_1 + n_2)/2)}{\Gamma(n_1/2)\Gamma(n_2/2)} \left(\frac{n_1}{n_2}\right)^{n_1/2} y^{n_2/2-1} \left(1 + \frac{n_1}{n_2}y\right)^{-(n_1+n_2)/2} & \text{si } y \geq 0 \\ 0 & \text{sinon} \end{cases} \quad (1.20)$$

et on écrit $Y \sim F(n_1, n_2)$. Son espérance n'existe que si $n_2 \geq 3$ et vaut $\frac{n_2}{n_2-2}$. Sa variance n'existe que si $n_2 \geq 5$ et vaut $\frac{2n_2^2(n_1+n_2-2)^2}{n_1(n_2-2)} (n_2 - 4)$.

Remarque 1.8. Pour $y \geq 0$, on a

$$\int_0^{+\infty} y^{n_2/2-1} \left(1 + \frac{n_1}{n_2}y\right)^{-(n_1+n_2)/2} dy = \frac{\Gamma(n_1/2)\Gamma(n_2/2)}{\Gamma((n_1 + n_2)/2)} \left(\frac{n_2}{n_1}\right)^{n_1/2} \quad (1.21)$$

Proposition 1.18. Si F suit une loi de Fisher $F(n_1, n_2)$ alors $1/F$ suit une loi de Fisher $F(n_2, n_1)$.

Chapitre 2

Modèle $AR(p)$ avec changement dans la moyenne

2.1 Introduction

L'existence d'un changement, non pris en considération, dans les paramètres d'une série temporelle autorégressive peut conduire à des mauvaises prévisions et de mauvaises analyses de la série. Il existe différentes méthodes pour tester si un changement est survenu au cours de la période observée des séries temporelles autorégressives, voir Tsay [109], Davies et al. [31] et Gombay [39] pour des références.

Dans le cadre Bayésien, Le modèle autorégressif $AR(p)$ avec changement dans la moyenne a été considéré par Barbieri et Conigliani [2] où ils ont adopté une approche utilisant les facteurs de Bayes et le facteur de Bayes fractionnaire. Le problème de l'identification d'un modèle autorégressif et la détection d'un changement éventuel de niveau ont été considérés par l'adoption de l'approche bayésienne et de supposer que l'information a priori sur les paramètres des différents modèles est faible.

Slama [107] a présenté un test de signification bayésien pour tester un changement dans les paramètres du modèle suivant :

$$(1) \begin{cases} y_i - \mu_1 = \rho(y_{i-1} - \mu_1) + e_i & i = 1, \dots, m \\ y_{m+1} - \mu_2 = \rho(y_m - \mu_1) + e_{m+1} \\ y_i - \mu_2 = \rho(y_{i-1} - \mu_2) + e_i & i = m+2, \dots, n \end{cases}$$

où

$$e_i \xrightarrow{i.i.d} \begin{cases} N(0, \sigma_1^2) & , i = 1, \dots, m \\ N(0, \sigma_2^2) & , i = m+1, \dots, n \end{cases}$$

$\mu_1, \mu_2 \in \mathbb{R}$ et $\sigma_1^2, \sigma_2^2 \in \mathbb{R}_+^*$, m est le point de rupture supposé inconnu, y_0 est une valeur initiale supposée donnée, ρ est le coefficient de corrélation supposé inconnu et vérifie $|\rho| < 1$,

Deux cas ont été considérés :

1. μ_1 est connu,
2. μ_1 est inconnu.

Par hypothèse que les paramètres sont indépendants, la loi a priori du paramètre $\theta = (m, \rho, \mu_1, \mu_2, r_1, r_2)$, ($r_i = \frac{1}{\sigma_i^2}$, $i = 1, 2$), est choisie comme suit :

Une loi a priori uniforme sur $\{1, \dots, n-2\}$ pour m , une loi uniforme sur $] -1, +1[$ pour ρ et pour (μ_1, μ_2, r_1, r_2) une loi a priori non informative de Jeffreys, ce qui donne :

$$\pi(\theta) \propto \frac{1}{r_1 r_2}, \quad (2.1)$$

où $m \in \{1, \dots, n-2\}$, $\mu_1, \mu_2 \in \mathbb{R}$, $r_1, r_2 \in \mathbb{R}_+^*$ et $\rho \in] -1, +1[$.

Les lois de probabilité a posteriori conditionnelles du point de rupture m , du coefficient de corrélation ρ , du paramètre de translation δ et du rapport des variances τ ont été déterminées. Une étude de simulation a montré que le test de signification bayésien est sensible à l'autocorrélation des données.

Dans ce chapitre, nous proposons un test Bayésien basé sur la construction des régions HPD crédibles pour détecter un changement dans la moyenne des observations autocorrélées. Les observations sont considérées issues d'un processus autorégressif d'ordre connu p avec changement dans la moyenne et la variance des erreurs à un instant inconnu m . En fait, le présent chapitre est une généralisation du travail de Slama [107] au cas où les erreurs suivent un processus autorégressif d'ordre p , $p \geq 1$.

2.2 Modèle et notations

Supposons que nous avons les observations $y = (y_1, \dots, y_n)$ issues d'un processus stochastique $\{Y_t, t \in \mathbb{Z}\}$. On considère le modèle autorégressif d'ordre fixe p , ($AR(p)$) :

$$Y_t - \mu_t = \sum_{i=1}^p \phi_i (Y_{t-i} - \mu_{t-i}) + \epsilon_t$$

où $\mu_t = \mathbb{E}(Y_t)$ et ϵ_t sont des variables aléatoires indépendantes et identiquement distribuées suivant une loi normale de moyenne 0 et de variance σ^2 , i.e. ϵ_t (*iid*) $\rightsquigarrow N(0, \sigma^2)$.

Dans ce travail, on considère le modèle avec changement dans la moyenne μ_t et dans la variance σ^2 à un instant inconnu m , c'est-à-dire :

$$\begin{cases} \mathbb{E}(Y_t) = \mu_1 \text{ et } \sigma^2 = \sigma_1^2, & \text{pour } t = 1, \dots, m \\ \mathbb{E}(Y_t) = \mu_2 \text{ et } \sigma^2 = \sigma_2^2 & \text{pour } t = m + 1, \dots, n. \end{cases}$$

$$\mu_1, \mu_2 \in \mathbb{R}, \sigma_1, \sigma_2 \in \mathbb{R}_+.$$

D'une façon équivalente, nous avons le modèle suivant :

$$\begin{cases} Y_t - \mu_1 = \sum_{i=1}^p \phi_i (Y_{t-i} - \mu_1) + \epsilon_t, & t = 1, \dots, m \\ Y_t - \mu_2 = \sum_{i=1}^p \phi_i (Y_{t-i} - \gamma_{t-i} \mu_1 - (1 - \gamma_{t-i}) \mu_2) + \epsilon_t, & t = m + 1, \dots, m + p \\ Y_t - \mu_2 = \sum_{i=1}^p \phi_i (Y_{t-i} - \mu_2) + \epsilon_t, & t = m + p + 1, \dots, n \end{cases} \quad (2.2)$$

où γ_t est la fonction indicatrice telle que $\gamma_{t-i} = 1$ if $t - i \leq m$ et $\gamma_{t-i} = 0$ si $t - i > m$.

et

$$\epsilon_t \xrightarrow{iid} \begin{cases} N(0, \sigma_1^2) & \text{pour } i = 1, \dots, m \\ N(0, \sigma_2^2) & \text{pour } i = m + 1, \dots, n \end{cases}$$

On suppose que les racines du polynôme caractéristique du modèle sont à l'extérieur du cercle unité, i.e. le vecteur paramètre $\phi^p = (\phi_1, \dots, \phi_p)$ se situe dans la région de stationnarité $\Phi_p = \{\phi^p : (\varphi(z) = 0) \Rightarrow |z| > 1\}$, où $\varphi(z) = 1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p$ est le polynôme caractéristique du modèle $AR(p)$. Les paramètres μ_1, μ_2 ($\mu_1, \mu_2 \in \mathbb{R}$), ϕ_i , ($i = 1, \dots, p$), σ_1, σ_2 ($\sigma_1, \sigma_2 > 0$) sont supposés inconnus, et $m \in \{1, \dots, n - 2\}$ est le point de rupture supposé aussi inconnu.

On a l'ensemble des paramètres $\theta = (m, \phi^p, \mu_1, \mu_2, r_1, r_2)$ où $r_i = 1/\sigma_i^2$. La fonction de vraisemblance conditionnelle du modèle (2.2) basée sur les observations $y = (y_1, y_2, \dots, y_n)$ est alors

$$\begin{aligned}
 l(\theta | y) \propto & r_1^{\frac{m}{2}} r_2^{\frac{n-m}{2}} \exp \left\{ -\frac{r_1}{2} \left[\sum_{t=1}^m (y_t - \mu_1 - \sum_{i=1}^p \phi_i (y_{t-i} - \mu_1)) \right]^2 \right\} \\
 & \exp \left\{ -\frac{r_2}{2} \left[\sum_{t=m+1}^{m+p} (y_t - \mu_2 - \sum_{i=1}^p \phi_i (y_{t-i} - \gamma_{t-i} \mu_1 - (1 - \gamma_{t-i}) \mu_2)) \right]^2 \right\} \\
 & \exp \left\{ -\frac{r_2}{2} \left[\sum_{t=m+p+1}^n (y_t - \mu_2 - \sum_{i=1}^p \phi_i (y_{t-i} - \mu_2)) \right]^2 \right\}.
 \end{aligned} \tag{2.3}$$

où les quantités initiales (y_{-p+1}, \dots, y_0) sont supposées données.

On considère les notations suivantes :

$$\begin{aligned}
 \phi^p &= (\phi_1, \phi_2, \dots, \phi_p), & \phi^{(j)} &= (\phi_1, \dots, \phi_{j-1}, \phi_{j+1}, \dots, \phi_p), \\
 \phi(p) &= 1 - \sum_{i=1}^p \phi_i,
 \end{aligned} \tag{2.4}$$

$$SS_1(m, \phi^p, \mu_1) = \sum_{t=1}^m \left(y_t - \mu_1 - \sum_{i=1}^p \phi_i (y_{t-i} - \mu_1) \right)^2, \tag{2.5}$$

$$a(m, \phi^p) = \sum_{t=m+1}^{m+p} \left(1 - \sum_{i=1}^p \phi_i (1 - \gamma_{t-i}) \right)^2 + (n - m - p) \left(1 - \sum_{i=1}^p \phi_i \right)^2, \tag{2.6}$$

$$\begin{aligned}
 SS_2(m, \phi^p, \mu_1) &= \sum_{t=m+1}^n (y_t - \mu_1 - \sum_{i=1}^p \phi_i (y_{t-i} - \mu_1))^2 \\
 &\quad - \frac{1}{a(m, \phi^p)} \left[\sum_{t=m+1}^{m+p} (y_t - \mu_1 - \sum_{i=1}^p \phi_i (y_{t-i} - \mu_1)) \left(1 - \sum_{i=1}^p \phi_i (1 - \gamma_{t-i}) \right) \right. \\
 &\quad \left. + \sum_{t=m+p+1}^n (y_t - \mu_1 - \sum_{i=1}^p \phi_i (y_{t-i} - \mu_1)) \left(1 - \sum_{i=1}^p \phi_i \right) \right]^2,
 \end{aligned} \tag{2.7}$$

2.3 Analyse Bayésienne

2.3.0.1 Loi a priori

Comme l'information a priori sur $\theta' = (\mu_1, \mu_2, r_1, r_2)$ est souvent vague ou diffuse, nous allons utiliser une loi a priori non informative de Jeffreys pour θ' . Supposons que les lois

a priori du point de rupture m et de $\phi^{(p)}$ sont données par :

$$\pi(m) \propto \frac{1}{n-2}; \quad m \in \{1, \dots, n-2\}, \quad \pi(\phi^{(p)}) \propto \text{constant dans } \Phi_p$$

Les paramètres m , $\phi^{(p)}$ et θ' sont supposés indépendants. La distribution a priori de θ est donnée alors par :

$$\pi(\theta) \propto \frac{1}{r_1 r_2} \quad (2.8)$$

où $m \in \{1, \dots, n-2\}$, $r_1, r_2 \in \mathbb{R}_+$, $\mu_1, \mu_2 \in \mathbb{R}$ et $\phi_i \in \Phi^{(p)}$ pour $i = 1, \dots, p$

2.3.0.2 Loi a posteriori

La distribution a posteriori de θ , obtenue par combinaison de (2.3) et (2.8) est donnée par

$$\begin{aligned} \pi(\theta | y) \propto & r_1^{\frac{m}{2}-1} r_2^{\frac{n-m}{2}-1} \exp \left\{ -\frac{r_1}{2} \left[\sum_{t=1}^m (y_t - \mu_1 - \sum_{i=1}^p \phi_i (y_{t-i} - \mu_1)) \right]^2 \right\} \\ & \exp \left\{ -\frac{r_2}{2} \left[\sum_{t=m+1}^{m+p} (y_t - \mu_2 - \sum_{i=1}^p \phi_i (y_{t-i} - \gamma_{t-i} \mu_1 - (1 - \gamma_{t-i}) \mu_2)) \right]^2 \right\} \\ & \exp \left\{ -\frac{r_2}{2} \left[\sum_{t=m+p+1}^n (y_t - \mu_2 - \sum_{i=1}^p \phi_i (y_{t-i} - \mu_2)) \right]^2 \right\}. \end{aligned} \quad (2.9)$$

L'hypothèse nulle H_0 exprimant qu'il n'y a pas de changement dans les paramètres du modèle (2.2) est,

$$\delta = \mu_2 - \mu_1 = 0 \quad \text{et} \quad \tau = \sigma_2^2 / \sigma_1^2 = 1$$

et l'hypothèse alternative H_1 est donnée par :

$$\delta = \mu_2 - \mu_1 \neq 0 \quad \text{ou} \quad \tau = \sigma_2^2 / \sigma_1^2 \neq 1$$

Pour le test de signification Bayésien, les distributions a posteriori de δ et de τ sont nécessaires pour obtenir les régions HPD crédibles (Highest Posterior Density) de δ et de τ . Le théorème suivant donne les distributions a posteriori conditionnelles des paramètres δ , τ , ϕ^p , μ_1 et de m :

Théorème 2.1. 1. Sachant m, ϕ^p, μ_1 et τ , la distribution a posteriori conditionnelle de δ est :

$$\pi(\delta | m, \phi^p, \mu_1, \tau, y) \propto \left\{ 1 + \frac{a(m, \phi^p) (\delta - \widehat{\delta}(m, \phi^p, \mu_1))^2}{(n-1)S_2^2(m, \phi^p, \tau)} \right\}^{-\frac{n}{2}}, \quad (2.10)$$

où

$$\widehat{\delta}(m, \phi^p, \mu_1) = \frac{\sum_{t=m+1}^{m+p} (y_t - \mu_1 - \sum_{i=1}^p \phi_i (y_{t-i} - \mu_1)) (1 - \sum_{i=1}^p \phi_i (1 - \gamma_{t-i}))}{a(m, \phi^p)} + \frac{\sum_{t=m+p+1}^n (y_t - \mu_1 - \sum_{i=1}^p \phi_i (y_{t-i} - \mu_1)) (1 - \sum_{i=1}^p \phi_i)}{a(m, \phi^p)},$$

$$S_2^2(m, \phi^p, \mu_1, \tau) = \frac{\tau SS_1(m, \phi^p, \mu_1) + SS_2(m, \phi^p, \mu_1)}{(n-1)},$$

qui est une distribution de Student t avec moyenne $\widehat{\delta}(m, \phi^p, \mu_1)$, de précision $\frac{a(m, \phi^p)}{S_2^2(m, \phi^p, \tau)}$, et de $(n-1)$ degrés de liberté. D'une manière équivalente, la quantité

$$T(\delta) = \frac{a^{\frac{1}{2}}(m, \phi^p) (\delta - \widehat{\delta}(m, \phi^p, \mu_1))}{S_2(m, \phi^p, \tau)}, \quad (2.11)$$

est distribuée a posteriori et conditionnellement m, ϕ^p, μ_1 et τ suivant une loi de student standard avec $(n-1)$ degrés de liberté.

2. Sachant m, ϕ^p et μ_1 , la distribution a posteriori conditionnelle de τ est :

$$\pi(\tau | m, \phi^p, \mu_1, y) \propto \tau^{\frac{n}{2}-1} \left\{ \tau SS_1(m, \phi^p, \mu_1) + SS_2(m, \phi^p, \mu_1) \right\}^{-\frac{(n-1)}{2}}. \quad (2.12)$$

et la quantité :

$$F(\tau) = \frac{SS_1(m, \phi^p, \mu_1)/m}{SS_2(m, \phi^p, \mu_1)/(n-m-1)} \tau, \quad (2.13)$$

est distribuée a posteriori et conditionnellement à m, ϕ^p et μ_1 suivant une loi de Fisher avec $(m, n-m-1)$ degrés de liberté.

3. Sachant m, ϕ^p et τ , la distribution a posteriori conditionnelle de μ_1 est :

$$\pi(\mu_1 | m, \phi^p, \tau, Y) \propto \left\{ 1 + \frac{b(m, \phi^p, \tau) (\mu_1 - \widehat{\mu}_1(m, \phi^p, \tau))^2}{(n-2)S_3^2(m, \phi^p, \tau)} \right\}^{-\frac{(n-1)}{2}}, \quad (2.14)$$

où,

$$\widehat{\mu}_1(m, \phi^p, \tau) = \frac{c(m, \phi^p, \tau)}{b(m, \phi^p, \tau)}, \quad (2.15)$$

$$S_3^2(m, \phi^p, \tau) = \frac{1}{(n-2)} \left[d(m, \phi^p) - \frac{c^2(m, \phi^p, \tau)}{b(m, \phi^p, \tau)} \right], \quad (2.16)$$

avec,

$$b(m, \phi^p, \tau) = \phi^2(p)(n - m + \tau m) - \frac{\phi(p)}{a(m, \phi^p)} \left[(n - m - p)\phi(p) + \sum_{t=m+1}^{m+p} (1 - \sum_{i=1}^p \phi_i(1 - \gamma_{t-i})) \right]^2, \quad (2.17)$$

$$d(m, \phi^p) = \tau \sum_1^m (y_t - \sum_{i=1}^p \phi_i y_{t-i})^2 + \sum_{m+1}^n (y_t - \sum_{i=1}^p \phi_i y_{t-i})^2 - \frac{1}{a(m, \phi^p)} \left[\sum_{m+1}^{m+p} (y_t - \sum_{i=1}^p \phi_i y_{t-i})(1 - \sum_{i=1}^p \phi_i(1 - \gamma_{t-i})) + \sum_{m+p+1}^n (y_t - \sum_{i=1}^p \phi_i y_{t-i})\phi(p) \right]^2, \quad (2.18)$$

$$c(m, \phi^p, \tau) = \tau \phi(p) \sum_1^m (y_t - \sum_{i=1}^p \phi_i y_{t-i}) + \phi(p) \sum_{m+1}^n (y_t - \sum_{i=1}^p \phi_i y_{t-i}) - \frac{\phi(p)}{a(m, \phi^p)} \left[(n - m - p)\phi(p) + \sum_{m+1}^{m+p} (1 - \sum_{i=1}^p \phi_i(1 - \gamma_{t-i})) \right] \times \left[\sum_{m+1}^{m+p} (y_t - \sum_{i=1}^p \phi_i y_{t-i})(1 - \sum_{i=1}^p \phi_i(1 - \gamma_{t-i})) + \sum_{m+p+1}^n (y_t - \sum_{i=1}^p \phi_i y_{t-i})\phi(p) \right], \quad (2.19)$$

qui est la distribution Student t de moyenne $\widehat{\mu}_1(m, \phi^p, \tau)$, de précision $\frac{b(m, \phi^p, \tau)}{S_3^2(m, \phi^p, \tau)}$, et de $(n-2)$ degrés de liberté.

4. Pour $j = 1, \dots, p$, sachant $m, \phi^{(j)}, \mu_1, \delta$ et τ , la distribution a posteriori conditionnelle de ϕ_j est :

$$\pi(\phi_j | m, \phi^{(j)}, \mu_1, \delta, \tau, Y) \propto \left\{ 1 + \frac{e(m, \mu_1, \delta, \tau)}{(n-1)S_4^2(m, \phi^{(j)}, \mu_1, \delta, \tau)} \times [\phi_j - \widehat{\phi}_j(m, \phi^{(j)}, \mu_1, \delta, \tau)]^2 \right\}^{-\frac{(n-1)}{2}}, \quad (2.20)$$

où,

$$\widehat{\phi}_j(m, \mu_1, \delta, \phi^{(j)}) = \frac{e_1(m, \phi^{(j)}, \mu_1, \delta, \tau)}{e(m, \mu_1, \delta, \tau)}, \quad (2.21)$$

avec,

$$\begin{aligned} e_1(m, \phi^{(j)}, \mu_1, \delta, \tau) = & \tau \sum_{t=1}^m (y_{t-j} - \mu_1) \left(y_t - \mu_1 - \sum_{i \neq j} (y_{t-i} - \mu_1) \right) + \\ & + \sum_{t=m+1}^{m+p} (y_{t-j} - \gamma_{t-j} \mu_1 - (1 - \gamma_{t-j})(\delta + \mu_1)) \times \\ & \left(y_t - \mu_1 - \delta - \sum_{i \neq j} (y_{t-i} - \gamma_{t-i} \mu_1 - (1 - \gamma_{t-i})(\delta + \mu_1)) \right) + \\ & + \sum_{t=m+p+1}^n (y_{t-j} - \delta - \mu_1) \left(y_t - \delta - \mu_1 - \sum_{i \neq j} \phi_i (y_{t-i} - \delta - \mu_1) \right), \end{aligned} \quad (2.22)$$

et

$$\begin{aligned} e(m, \mu_1, \delta, \tau) = & \tau \sum_{i=1}^m (y_{t-j} - \mu_1)^2 + \sum_{t=m+1}^{m+p} (y_{t-j} - \gamma_{t-j} \mu_1 - (1 - \gamma_{t-j})(\delta + \mu_1))^2 \\ & + \sum_{t=m+p+1}^n (y_{t-j} - \delta - \mu_1)^2, \end{aligned} \quad (2.23)$$

$$\begin{aligned} S_4^2(m, \phi^{(j)}, \mu_1, \delta, \tau) = & \frac{1}{n-1} \left\{ \left(y_t - \mu_1 - \sum_{i \neq j} (y_{t-i} - \mu_1) \right)^2 + \right. \\ & + \left(y_t - \mu_1 - \delta - \sum_{i \neq j} (y_{t-i} - \gamma_{t-i} \mu_1 - (1 - \gamma_{t-i})(\delta + \mu_1)) \right)^2 + \\ & \left. + \left(y_t - \delta - \mu_1 - \sum_{i \neq j} \phi_i (y_{t-i} - \delta - \mu_1) \right)^2 \right\}, \end{aligned}$$

qui est la distribution de student t avec moyenne $\widehat{\phi}_j(m, \mu_1, \delta, \phi^{(j)})$, de précision $\frac{e(m, \mu_1, \delta, \tau)}{S_4^2(m, \phi^{(j)}, \mu_1, \delta, \tau)}$, et de $(n-1)$ degrés de liberté.

5. Sachant ϕ^p , μ_1 et τ la distribution a posteriori conditionnelle de m est :

$$\pi(m | \phi^p, \mu_1, \tau, y) \propto a(m, \phi^p)^{-\frac{1}{2}} \tau^{\frac{m-1}{2}} \left\{ \tau SS_1(m, \phi^p, \mu_1) + SS_2(m, \phi^p, \mu_1) \right\}^{-\frac{n-1}{2}}. \quad (2.24)$$

Preuve :

Détermination de la distribution conjointe de δ , τ , ϕ^p , μ_1 et de m :

Par transformation bijective de l'ensemble des paramètres $\Theta = (m, \phi^{(p)}, \mu_1, \mu_2, r_1, r_2)$ à $\Phi = (m, \phi^{(p)}, \mu_1, \delta, \tau)$, on trouve la distribution a posteriori de Φ ; qui est,

$$\begin{aligned} \pi(\Phi | y) &= \int_{r_2} \pi(m, \phi^{(p)}, \mu_1, \delta + \mu_1, \tau r_2, r_2 | y) |r_2| dr_2 \\ &\propto \tau^{\frac{m}{2}-1} \left\{ \tau \sum_{t=1}^m \left[y_t - \mu_1 - \sum_{i=1}^p \phi_i (y_{t-i} - \mu_1) \right]^2 \right. \\ &+ \sum_{t=m+1}^{m+p} \left[y_t - \delta - \mu_1 - \sum_{i=1}^p \phi_i (y_{t-i} - \gamma_{t-i} \mu_1 - (1 - \gamma_{t-i})(\delta + \mu_1)) \right]^2 \\ &\left. + \sum_{t=m+p+1}^n \left[y_t - \delta - \mu_1 - \sum_{i=1}^p \phi_i (y_{t-i} - \delta - \mu_1) \right]^2 \right\}^{-\frac{n}{2}} \end{aligned} \quad (2.25)$$

1. La distribution a posteriori conditionnelle de δ :

La distribution a posteriori conjointe donnée dans (2.25) peut être écrite sous la forme :

$$\begin{aligned} \pi(m, \phi^{(p)}, \mu_1, \delta, \tau | y) &\propto \tau^{\frac{m}{2}-1} \left\{ \tau SS_1(m, \phi^p, \mu_1) + SS_2(m, \phi^p, \mu_1) \right. \\ &\left. + a(m, \phi^p) \left(\delta - \widehat{\delta}(m, \phi^p, \mu_1) \right)^2 \right\}^{-\frac{n}{2}}. \end{aligned} \quad (2.26)$$

Par application du théorème de Bayes, la distribution a posteriori conditionnelle de δ est obtenue comme donnée dans (2.10).

2. La distribution a posteriori conditionnelle de τ :

Par intégration par rapport à δ , on obtient la distribution a posteriori conjointe de m, ϕ^p, μ_1 et de τ :

$$\pi(m, \phi^p, \mu_1, \tau | y) \propto a(m, \phi^p)^{-\frac{1}{2}} \tau^{\frac{m}{2}-1} \left\{ \tau SS_1(m, \phi^p, \mu_1) + SS_2(m, \phi^p, \mu_1) \right\}^{-\frac{n-1}{2}}. \quad (2.27)$$

Par application du théorème de Bayes, la distribution a posteriori conditionnelle de τ est donnée comme dans (2.12).

3. Distribution a posteriori conditionnelle de μ_1 :

Par un calcul simple, La distribution a posteriori conjointe donnée dans (2.27) peut

être écrite sous la forme :

$$\pi(m, \phi^p, \mu_1, \tau | y) \propto a(m, \phi^p)^{-\frac{1}{2}} \tau^{\frac{m}{2}-1} \left\{ d(m, \phi^p) - \frac{c^2(m, \phi^p, \tau)}{b(m, \phi^p, \tau)} + b(m, \phi^p, \tau) (\mu_1 - \widehat{\mu}_1(m, \phi^p, \tau))^2 \right\}^{-\frac{n-1}{2}}.$$

où $a(m, \phi^p)$, $d(m, \phi^p)$, $c(m, \phi^p, \tau)$, $b(m, \phi^p, \tau)$ et $\widehat{\mu}_1(m, \phi^p, \tau)$ sont données dans (2.6), (2.18), (2.19), (2.17) et (2.15) respectivement.

Par application du théorème de Bayes et l'intégrale (1.19), la distribution a posteriori conditionnelle de μ_1 est donnée par (2.14).

4. Distribution a posteriori conditionnelle de ϕ_i , $i = 1, \dots, p$:

Par un calcul simple, la distribution a posteriori conjointe donnée dans (2.25) peut être écrite sous la forme :

$$\begin{aligned} \pi(m, \phi^p, \mu_1, \delta, \tau | y) \propto & a(m, \phi^p)^{-\frac{1}{2}} \tau^{\frac{m}{2}-1} \left\{ (y_t - \mu_1 - \sum_{i \neq j} (y_{t-i} - \mu_1))^2 \right. \\ & + (y_t - \mu_1 - \delta - \sum_{i \neq j} (y_{t-i} - \gamma_{t-i} \mu_1 - (1 - \gamma_{t-i})(\delta + \mu_1)))^2 \\ & + (y_t - \delta - \mu_1 - \sum_{i \neq j} \phi_i (y_{t-i} - \delta - \mu_1))^2 \\ & \left. + e(m, \mu_1, \delta, \tau) \left[\phi_j - \widehat{\phi}_j(m, \phi^{(j)}, \mu_1, \delta, \tau) \right]^2 \right\}^{-\frac{n-1}{2}}. \end{aligned}$$

où $\widehat{\phi}_j(m, \phi^{(j)}, \mu_1, \delta, \tau)$ et $e(m, \mu_1, \delta, \tau)$ sont données dans (2.21) et (2.23) respectivement.

Par application du théorème de Bayes et l'intégrale (1.19), la distribution a posteriori conditionnelle de ϕ_i , pour $i = 1, \dots, p$ est donnée comme en (2.20).

5. La distribution a posteriori conditionnelle du point de rupture m :

A partir de la distribution a posteriori conjointe de m , ϕ^p , μ_1 et de τ donnée dans (2.27) et l'application du théorème de Bayes on obtient la distribution a posteriori conditionnelle du point de rupture m telle que donnée dans (2.24).

Remarque 2.1. Comme les degrés de liberté m et $n - m - 1$ de la distribution de Fisher F sont supérieures ou égal à 1, ce qui impose au point de rupture m d'appartenir à l'ensemble $\{1, n - 2\}$.

La distributions a posteriori marginale du point de rupture m est donnée par

$$\pi(m|y) = \int_{\phi^p} \int_{\mu_1} \int_{\tau} \pi(m, \phi^p, \mu_1, \tau | y) d\tau d\mu_1 d\phi^p \quad (2.28)$$

Les distributions a posteriori inconditionnelles de $T(\delta)$ et de $F(\tau)$ sont données, respectivement, par,

$$\begin{aligned} \pi(T(\delta) | y) &= \sum_m \left\{ \int_{\phi^p} \left[\int_{\mu_1} \left(\int_{\tau} \pi(T(\delta) | m, \phi^p, \mu_1, \tau, y) \pi(\tau | m, \phi^p, \mu_1, y) d\tau \right) \right. \right. \\ &\quad \left. \left. \times \pi(\mu_1 | m, \phi^p, y) d\mu_1 \right] \pi(\phi^p | m, y) d\phi^p \right\} \pi(m | y). \end{aligned} \quad (2.29)$$

$$\begin{aligned} \pi(F(\tau) | y) &= \sum_m \left\{ \int_{\phi^p} \left[\int_{\mu_1} \pi(F(\tau) | m, \phi^p, \mu_1, \tau, y) \pi(\mu_1 | m, \phi^p, y) d\mu_1 \right] \right. \\ &\quad \left. \times \pi(\phi^p | m, y) d\phi^p \right\} \pi(m | y). \end{aligned} \quad (2.30)$$

Pour $p = 1$, sachant ϕ_1 , les distributions a posteriori conditionnelles de $T(\delta)$ et de $F(\tau)$ sont données, respectivement, par,

$$\begin{aligned} \pi(T(\delta) | \phi_1, y) &= \sum_m \left[\int_{\mu_1} \left(\int_{\tau} \pi(T(\delta) | m, \phi_1, \mu_1, \tau, y) \pi(\tau | m, \phi_1, \mu_1, y) d\tau \right) \right. \\ &\quad \left. \times \pi(\mu_1 | m, \phi_1, y) d\mu_1 \right] \pi(m | y). \end{aligned} \quad (2.31)$$

$$\pi(F(\tau) | \phi_1, y) = \sum_m \left[\int_{\mu_1} \pi(F(\tau) | m, \phi_1, \mu_1, \tau, y) \pi(\mu_1 | m, \phi_1, y) d\mu_1 \right] \times \pi(m | y). \quad (2.32)$$

2.4 Test de signification Bayésien

L'hypothèse nulle H_0 peut être divisée en deux sous hypothèses $H_{01} : \delta = 0$ et $H_{02} : \tau = 1$, et H_o sera rejetée si l'une des sous hypothèses est rejetée. Comme $T(\delta)$ et $F(\tau)$ sont conditionnellement indépendantes, on définit séparément les régions HPD crédibles de $T(\delta)$ et de $F(\tau)$ basées sur les distributions a posteriori conditionnelles. Ces régions HPD crédibles seront utilisées pour définir les p-values inconditionnelles et par conséquent un test de signification Bayésien inconditionnel.

Sachant m, ϕ^p, μ_1 et τ , le $(1 - \alpha)$ -HPD crédible région de $T(\delta)$ est définie par :

$$C_\delta = \{T(\delta)/|T(\delta)| < t_{\alpha/2}(n-1)\}$$

où $t_{\alpha/2}(n-1)$ est le quantile d'ordre $(1 - \alpha/2)$ de la distribution de Student t avec $(n-1)$ degrés de liberté. Par conséquent, sachant m, ϕ^p, μ_1 et τ , la règle de décision pour H_{01} , est de rejeter H_{01} si $T(0) \in \overline{C}_\delta$, où \overline{C}_δ est le complémentaire de C_δ représentant la région de rejet de H_{01} .

La p-value inconditionnelle de H_{01} calculée à partir de (2.29) est donc :

$$\begin{aligned} P_{\delta=0|y} &= 2 \sum_m \left\{ \int_{\phi^p} \left[\int_{\mu_1} \left(\int_{\tau} \{1 - \mathcal{T}_{n-1}(|T(0)|)\} \pi(\tau|m, \phi^p, \mu_1, y) d\tau \right) \right. \right. \\ &\quad \left. \left. \times \pi(\mu_1|m, \phi^p, y) d\mu_1 \right] \pi(\phi^p|m, y) d\phi^p \right\} \pi(m|y), \\ &= 2 E_m E_{\phi^p} E_{\mu_1} E_{\tau} \{1 - \mathcal{T}_{n-1}(|T(0)|)\}. \end{aligned} \quad (2.33)$$

où \mathcal{T}_{n-1} est la fonction cumulative de la distribution de Student standard avec $(n-1)$ degrés de liberté. Les espérances mathématiques E_m, E_{μ_1} et E_{τ} sont pris par rapport à m, μ_1 et τ , respectivement. E_{ϕ^p} pour indiquer l'espérance mathématique par rapport à $\phi_1, \phi_2, \dots, \phi_p$ respectivement. Notre test est alors de rejeter H_{01} , si $P_{\delta=0|y}$ est inférieure à α .

De même, la p-value inconditionnelle de H_{02} est donnée par :

$$\begin{aligned} P_{\tau=1|y} &= 2 \sum_m \left\{ \int_{\phi^p} \left[\int_{\mu_1} \{1 - \mathcal{F}_{m, n-m-1}[\max(F(1), 1|F(1))]\} \right. \right. \\ &\quad \left. \left. \times \pi(\mu_1|m, \phi^p, y) d\mu_1 \right] \pi(\phi^p|m, Y) d\phi^p \right\} \pi(m|y), \\ &= 2 E_m E_{\phi^p} E_{\mu_1} \{1 - \mathcal{F}_{m, n-m-1}[\max(F(1), 1|F(1))]\}. \end{aligned} \quad (2.34)$$

où $\mathcal{F}_{m, n-m-1}$ est la densité cumulative de la distribution de Fisher F avec $(m, n - m - 1)$ degrés de liberté, Aussi, le test est de rejeter H_{02} , si $P_{\tau=1|y}$ est inférieure de α .

Les p-values conditionnelles sachant m des sous hypothèses H_{01} et H_{02} sont données respectivement par :

$$\begin{aligned} P_{\delta=0|m, y} &= 2 \int_{\phi^p} \left[\int_{\mu_1} \left(\int_{\tau} \{1 - \mathcal{T}_{n-1}(|T(0)|)\} \pi(\tau|m, \phi^p, \mu_1, y) d\tau \right) \right. \\ &\quad \left. \times \pi(\mu_1|m, \phi^p, y) d\mu_1 \right] \pi(\phi^p|m, y) d\phi^p, \\ &= 2 E_{\phi^p} E_{\mu_1} E_{\tau} \{1 - \mathcal{T}_{n-1}(|T(0)|)\}. \end{aligned} \quad (2.35)$$

et

$$\begin{aligned}
 P_{\tau=1|m,y} &= 2 \int_{\phi^p} \left[\int_{\mu_1} \left\{ 1 - \mathcal{F}_{m,n-m-1}[\max(F(1), 1|F(1))] \right\} \right. \\
 &\quad \left. \times \pi(\mu_1|m, \phi^p, y) d\mu_1 \right] \pi(\phi^p|m, y) d\phi^p, \\
 &= 2E_{\phi^p} E_{\mu_1} \left\{ 1 - \mathcal{F}_{m,n-m-1}[\max(F(1), 1|F(1))] \right\}.
 \end{aligned} \tag{2.36}$$

Les p-values conditionnelles sachant ϕ_1 des sous hypothèses H_{01} et H_{02} sont données respectivement par :

$$\begin{aligned}
 P_{\delta=0|\phi_1,y} &= 2 \sum_m \left\{ \int_{\phi^p} \left[\int_{\mu_1} \left(\int_{\tau} \{ 1 - \mathcal{T}_{n-1}(|T(0)|) \} \pi(\tau|m, \phi^p, \mu_1, y) d\tau \right) \right. \right. \\
 &\quad \left. \left. \times \pi(\mu_1|m, \phi^p, y) d\mu_1 \right] \pi(\phi^p|m, y) d\phi^p \right\} \pi(m|y), \\
 &= 2E_m E_{\phi^p} E_{\mu_1} E_{\tau} \left\{ 1 - \mathcal{T}_{n-1}(|T(0)|) \right\}.
 \end{aligned} \tag{2.37}$$

et

$$\begin{aligned}
 P_{\tau=1|\phi_1,y} &= 2 \sum_m \left\{ \int_{\phi^p} \left[\int_{\mu_1} \left\{ 1 - \mathcal{F}_{m,n-m-1}[\max(F(1), 1|F(1))] \right\} \right. \right. \\
 &\quad \left. \left. \times \pi(\mu_1|m, \phi^p, y) d\mu_1 \right] \pi(\phi^p|m, y) d\phi^p \right\} \pi(m|y), \\
 &= 2E_m E_{\phi^p} E_{\mu_1} \left\{ 1 - \mathcal{F}_{m,n-m-1}[\max(F(1), 1|F(1))] \right\}.
 \end{aligned} \tag{2.38}$$

Les p-values conditionnelles sachant m et ϕ_1 des sous hypothèses H_{01} et H_{02} sont données respectivement par :

$$\begin{aligned}
 P_{\delta=0|m,\phi_1,y} &= 2 \int_{\mu_1} \left(\int_{\tau} \{ 1 - \mathcal{T}_{n-1}(|T(0)|) \} \pi(\tau|m, \phi^p, \mu_1, y) d\tau \right) \\
 &\quad \times \pi(\mu_1|m, \phi^p, y) d\mu_1, \\
 &= 2E_{\mu_1} E_{\tau} \left\{ 1 - \mathcal{T}_{n-1}(|T(0)|) \right\}.
 \end{aligned} \tag{2.39}$$

et

$$\begin{aligned}
 P_{\tau=1|m,y} &= 2 \int_{\mu_1} \left\{ 1 - \mathcal{F}_{m,n-m-1}[\max(F(1), 1|F(1))] \right\} \\
 &\quad \times \pi(\mu_1|m, \phi^p, y) d\mu_1, \\
 &= 2E_{\mu_1} \left\{ 1 - \mathcal{F}_{m,n-m-1}[\max(F(1), 1|F(1))] \right\}.
 \end{aligned} \tag{2.40}$$

Les quantités (2.28), (2.33), (2.34), (2.35), (2.36), (2.37), (2.38), (2.39), (2.40) sont évaluées numériquement par l'algorithme de l'échantillonneur de Gibbs (gibbs Sampler) en utilisant les distributions a posteriori conditionnelles données dans le théorème 2.1.

2.5 Étude de simulation

Une étude de simulation est utilisée pour étudier l'effet de l'autocorrélation des données sur le test de signification Bayésien basé sur les régions HPD crédibles

Nous avons simulé des échantillons à partir du modèle (2.2) pour $\rho = 1$, $\phi_1 = 0.5$, $\mu_1 = 0.5$, $\mu_2 = 2.0$, $\sigma_1^2 = 0.5$, $\sigma_2^2 = 1.0$, $m = 34$, $y_0 = 0.5$ et pour $n = 70$.

A partir de ces observations, par application de l'algorithme de l'échantillonneur de Gibbs avec 5000 répétitions, la densité a posteriori marginale du point de changement et des p-values conditionnelles et inconditionnelles des sous hypothèses $H_{01} : \delta = 0$ et de $H_{02} : \tau = 1$. Les résultats sont donnés dans le tableau (3.1).

2.5.1 Densité a posteriori de m et les p-values de H_0

m	$\pi(m y)$	$P_{\delta=0 m,y}$	$P_{\tau=1 m,y}$
30	0.023	0.002	0.046
31	0.016	0.017	0.031
32	0.075	0.000	0.032
33	0.114	0.004	0.028
34	0.245	0.001	0.028
35	0.033	0.014	0.063
36	0.002	0.109	0.127
37	0.001	0.206	0.103
38	0.002	0.285	0.081

TABLE 2.1 – Densité a posteriori marginale de m et les p-values conditionnelles sachant m de H_{01} et de H_{02} au voisinage du vrai point de rupture $m = 34$.

Table 2.1, montre que le mode a posteriori de $\pi(m|y)$ est la vraie valeur de $m = 34$, et les sous hypothèses H_{01} et H_{02} sont rejetées conditionnellement à m au niveau de signification $\alpha = 0.05$.

Les p-values inconditionnelles et conditionnelles à m et ρ des sous hypothèses H_{01}

et H_{02} données dans la Table (2.2) montrent que le test de signification Bayésien rejette inconditionnellement les sous hypothèses H_{01} et H_{02} au niveau de signification $\alpha = 0.10$ et $\alpha = 0.05$ respectivement. Par conséquent, l'hypothèse H_0 est rejetée inconditionnellement au niveau de signification $\alpha = 0.05$.

$P_{\delta=0 y}$	$P_{\tau=1 y}$	$P_{\delta=0 m,\phi_1,y}$	$P_{\tau=1 m,\phi_1,y}$
0.065	0.0433	0.0000	0.0379

TABLE 2.2 – p-values inconditionnelles et conditionnelle sachant m et ϕ_1 de H_{01} et H_{02} .

2.5.2 Effet de l'autocorrélation sur le test de signification Bayésien

Pour examiner l'effet de l'autocorrélation sur le test de signification Bayésien basé sur les régions HPD crédibles sous l'hypothèse de l'indépendance, on a simulé des observations à partir du modèle (2.2) avec $\mu_1 = 0,5$, $\mu_2 = 1,5$, $\sigma_1^2 = 0,5$, $\sigma_2^2 = 1,0$ et $m = 34$ pour des valeurs de ϕ_1 à partir de $-0,7$ à $0,7$ avec un pas de $0,2$. Il est observé que les p-values déterminées sous l'hypothèse de l'indépendance des observations $P_{\delta=0/\rho=0,y}$ et $P_{\tau=1/\rho=0,y}$ sous-estiment les vrais p-values $P_{\delta=0/y}$ et $P_{\tau=1/y}$ quand le coefficient de corrélation est positif, et les surestiment lorsque le coefficient de corrélation est négatif (table (2.3)).

ϕ_1	$P_{\delta=0/y}$	$P_{\tau=1/y}$	$P_{\delta=0/\phi_1=0,y}$	$P_{\tau=1/\phi_1=0,y}$
-0.7	8.10^{-13}	0,035	$5,4.10^{-3}$	0,177
-0.5	5.10^{-11}	0,028	$7,9.10^{-5}$	0,253
-0.3	$2,7.10^{-10}$	0,027	$1,4.10^{-6}$	0,146
0.0	$9,9.10^{-7}$	0,026	$2,9.10^{-7}$	0,042
0.3	$3,9.10^{-3}$	0,040	$6,5.10^{-9}$	0,014
0.5	0,0358	0,040	$2,6.10^{-8}$	$9,5.10^{-3}$
0.7	0,131	0,053	$1.0.10^{-8}$	$9,0.10^{-3}$

TABLE 2.3 – Les p-values conditionnelles sachant $\rho = 0$ et inconditionnelles des sous hypothèses H_{01} et de H_{02} pour différentes valeurs de ϕ_1

2.6 Conclusion

Dans ce chapitre, nous avons développé un test de signification Bayésien pour détecter un changement dans la moyenne et la variance d'une série d'observations issues d'un modèle autorégressif d'ordre connu p , $Ar(p)$. Nous avons obtenu les distributions a posteriori conditionnelles du point de rupture et les autres paramètres du modèle. Nous avons déterminé un test de signification Bayésien Basé sur la construction de région HPD crédibles et le calcul des p-values inconditionnelles. Par une étude de simulation, en utilisant l'algorithme de de l'échantillonneur de Gibbs, nous avons constaté que le test de signification Bayésien basé sur les régions HPD crédibles est sensible à l'autocorrélation dans les observations. Ainsi, les p-values déterminées sous l'hypothèse de l'indépendance sous-estiment les vrais p-values quand le coefficient de corrélation est positif, et les surestiment lorsque le coefficient de corrélation est négatif.

Chapitre 3

Modèle $AR(p)$ avec changement dans les paramètres d'autorégression

3.1 Introduction

Le problème de la détection et l'estimation de point de changement dans les paramètres d'un modèle de série temporelle autorégressif $AR(p)$ a été étudié par plusieurs auteurs. Salazar [98] a donné une analyse Bayésienne d'un changement structurel dans les modèles stationnaires $AR(1)$ et $AR(2)$ avec un point de rupture connu. Davies et al.[31] ont étudié le comportement asymptotique d'un test du rapport de vraisemblance pour tester un changement dans les paramètres d'un modèle $AR(p)$. Huskova et al. [63, 64] ont utilisé une approche basée sur des sommes partielles des résidus pondérés (Méthodes asymptotique et bootstrap). A partir d'une étude de simulation, ils ont montré que pour toutes les statistiques, les tests bootstrap fonctionnent mieux que les tests asymptotiques. Gombay [39] a utilisé le vecteur de score pour détecter un changement dans les paramètres d'une série temporelle autorégressive. Kezim et Abdelli[70] ont proposé une analyse Bayésienne d'un modèle autorégressif d'ordre 1, ($AR(1)$) avec un changement dans le coefficient d'autocorrélation et la variance à un instant inconnu.

Le modèle considéré par Kezim et Abdelli [70] est un modèle autorégressif d'ordre 1 ($AR(1)$) soumis à un changement à la fois dans la variance des erreurs et dans le coefficient

d'autocorrélation à un instant inconnu. Le modèle est donné par :

$$(1) \begin{cases} y_i - \mu = \rho_1(y_{i-1} - \mu) + e_i & i = 1, \dots, m \\ y_i - \mu = \rho_2(y_{i-1} - \mu) + e_i & i = m + 1, \dots, n \end{cases}$$

où ρ_1 et ρ_2 sont les coefficients d'autocorrélation, supposés inconnus. y_i est la i ème observation de la variable dépendante. Les termes d'erreur e_i sont des variables aléatoires indépendantes et suivent une loi normale $N(0; \sigma_1^2)$ pour $i = 1, \dots, m$ et suivent une loi normale $N(0; \sigma_2^2)$ pour $i = m + 1, \dots, n$. m est le point de rupture supposé inconnu et y_0 est une observation initiale, supposé donnée.

Les distributions a priori des paramètres inconnus sont attribuées comme suit : m suit une loi a priori uniforme sur $\{1, \dots, n - 1\}$, la précision $\eta_i = 1/\sigma_i^2$ suit une distribution gamma de paramètres $a_i/2, b_i/2$; $a_i > 0$ et $b_i > 0$; $i = 1, 2$, et la distribution conditionnelle de ρ_i sachant η_i est normale avec une moyenne μ_i et de précision η_i ; $i = 1, 2$, de plus (ρ_1, η_1) et (ρ_2, η_2) sont supposés indépendants.

Ils ont déterminé les distributions a posteriori du point de rupture m , du rapport des variances $\tau = \sigma_1^2/\sigma_2^2$, et de l'amplitude du changement dans les paramètres d'autocorrélation $\delta = \rho_2 - \rho_1$. Par une étude numérique en utilisant l'algorithme de l'échantillonneur de Gibbs, ils ont montré que la détection du point de changement structurel est sensible aux changements à la fois dans la variance et dans l'autocorrélation des observations.

Dans ce chapitre, nous proposons une analyse Bayésienne d'un modèle autorégressif d'ordre connu p ; $Ar(p)$; avec un changement à un instant inconnu m aussi bien dans les paramètres d'autorégression et dans la variance des erreurs. Nous avons présenté un test Bayésien basé sur la construction des régions HPD crédibles et le calcul des p-values pour la détection d'un changement brusque dans les paramètres du modèle $AR(p)$. Dans une partie, ce chapitre présente une généralisation des travaux de Kezim et Abdelli [70] au cas d'un modèle autorégressif d'ordre connu p , $AR(p)$.

Le reste de ce chapitre est organisé comme suit : dans la section 2, nous allons présenté le modèle et les notations utilisés. L'analyse Bayésienne relative à la détection du point de changement dans les paramètres du modèle est donnée dans la section 3. L'étude de simulation est présentée dans la section 4.

3.2 Modèle et notations

Supposons que nous disposons d'une série d'observations $y = (y_1, \dots, y_n)$ considérées comme des réalisations d'un processus stochastique $\{Y_t, t \in \mathbb{Z}\}$. On considère le processus autorégressif d'ordre p ($AR(p)$) :

$$\phi(B)Y_t = \epsilon_t$$

avec ϕ le polynôme défini par $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ et B est l'opérateur retard défini par $B^k Y_t = Y_{t-k}$ et $\{\epsilon_t\}$ est le processus Gaussien de moyenne 0 et de variance σ^2 .

Nous considérons le modèle avec changement dans les paramètres d'autorégression et la variance des erreurs donné par :

$$\begin{cases} y_t - \mu = \sum_{i=1}^p \phi_i (y_{t-i} - \mu) + \epsilon_t & , t = 1 \dots , m \\ y_t - \mu = \sum_{i=1}^p \psi_i (y_{t-i} - \mu) + \epsilon_t & , t = m + 1 \dots , n \end{cases} \quad (3.1)$$

Avec

$$\epsilon_t \xrightarrow{i.i.d} \begin{cases} N(0, \sigma_1^2), & t = 1, \dots, m \\ N(0, \sigma_2^2), & t = m + 1, \dots, n \end{cases}$$

Les paramètres $\phi_i, \psi_i, \sigma_1, \sigma_2$ sont supposés inconnus et m est le point de changement supposé aussi inconnu. Sans restreindre la généralité on suppose que $\mu = 0$. Le but de ce travail est de définir une règle de décision pour détecter l'existence d'un changement dans les paramètres du modèle à partir des observations (y_1, y_2, \dots, y_n) . On suppose que les paramètres $\phi^{(p)} = (\phi_1, \phi_2, \dots, \phi_p)$ sont dans la région de stationnarité définie par $\Phi_1^{(p)} = \{z/1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p = 0 \text{ implique } |z| > 1\}$. De même pour $\psi^{(p)} = (\psi_1, \psi_2, \dots, \psi_p)$ sont dans la région $\Phi_2^{(p)} = \{z/1 - \psi_1 z - \psi_2 z^2 - \dots - \psi_p z^p = 0 \text{ implique } |z| > 1\}$. σ_1^2 et $\sigma_2^2 \in \mathbb{R}_+$. Les quantités initiales $y_{1-p}, \dots, y_{-1}, y_0$ sont supposées données.

L'hypothèse à tester est :

$$H_0 : \delta_i = \psi_i - \phi_i = 0, i = 1, \dots, p \text{ et } \tau = \frac{\sigma_1^2}{\sigma_2^2} = 1$$

$$\text{contre } H_1 : \delta_i = \psi_i - \phi_i \neq 0, \text{ pour certains } i = 1, \dots, p \text{ ou } \tau = \frac{\sigma_1^2}{\sigma_2^2} \neq 1$$

qui pourra être notée :

$$H_0 : \delta = \psi - \phi = 0 \text{ et } \tau = \frac{\sigma_1^2}{\sigma_2^2} = 1 \text{ contre } H_1 : \delta = \psi' - \phi \neq 0 \text{ ou } \tau = \frac{\sigma_1^2}{\sigma_2^2} \neq 1 \quad (3.2)$$

avec $\delta = (\delta_1, \delta_2, \dots, \delta_p)$, $\phi = (\phi_1, \phi_2, \dots, \phi_p)$ et $\psi = (\psi_1, \psi_2, \dots, \psi_p)$

On pose $r_i = \frac{1}{\sigma_i^2}$, $i = 1, 2$ et $\theta = (m, \phi^{(p)}, \psi^{(p)}, r_1, r_2)$, θ est un vecteur de dimension $(2p + 3) \times 1$.

La fonction de vraisemblance conditionnelle du modèle (3.1) basée sur les données (y_1, \dots, y_n) est donnée par :

$$l(\theta | y) \propto r_1^{\frac{m}{2}} r_2^{\frac{n-m}{2}} \exp \left\{ -\frac{r_1}{2} \sum_{t=1}^m (y_t - \sum_{i=1}^p \phi_i y_{t-i})^2 \right\} \times \exp \left\{ -\frac{r_2}{2} \sum_{t=m+1}^n (y_t - \sum_{i=1}^p \psi_i y_{t-i})^2 \right\}, \quad (3.3)$$

La fonction de vraisemblance conditionnelle est basée sur l'hypothèse que les observations initiales $y_0, y_{-1}, \dots, y_{1-p}$ sont aussi disponibles [92]. De plus, si la taille de l'échantillon n est suffisamment grande, la contribution des observations initiales dans la vraisemblance totale est négligeables [49].

3.3 Analyse Bayésienne

3.3.0.1 Loi a priori

En supposant que tous les paramètres sont indépendants, nous choisissons pour m une loi a priori uniforme sur $\{1, \dots, n-2\}$, pour ϕ_i et ψ_i , $i = 1, \dots, p$, une loi uniforme sur $\Phi^{(p)} = \Phi_1^{(p)} \cap \Phi_2^{(p)}$ et pour (r_1, r_2) une loi a priori non informative de Jeffreys, ce qui donne :

$$\pi(\theta) \propto \frac{1}{r_1 r_2}, \quad (3.4)$$

où $m \in \{1, \dots, n-2\}$, $r_1, r_2 \in \mathbb{R}_+^*$ et $\phi_i, \psi_i \in \Phi^{(p)}$ pour $i = 1, \dots, p$

3.3.0.2 Loi a posteriori

La loi de probabilité a posteriori de θ , obtenue par combinaison de (3.3) et (3.4) est donnée par :

$$\begin{aligned} \pi(\theta | y) \propto r_1^{\frac{m}{2}-1} r_2^{\frac{n-m}{2}-1} \exp \left\{ -\frac{r_1}{2} \sum_{t=1}^m (y_t - \sum_{i=1}^p \phi_i y_{t-i})^2 \right\} \\ \times \exp \left\{ -\frac{r_2}{2} \sum_{t=m+1}^n (y_t - \sum_{i=1}^p \psi_i y_{t-i})^2 \right\}, \end{aligned} \quad (3.5)$$

Le Théorème suivant donne les lois a posteriori conditionnelles des δ_j , de ϕ_j , pour $j = 1, \dots, p$, de τ et de m . On note $\delta^{(p)} = (\delta_1, \delta_2, \dots, \delta_p)$, $\delta_{(j)} = (\delta_1, \dots, \delta_{j-1}, \delta_{j+1}, \dots, \delta_p)$, $\phi^{(p)} = (\phi_1, \dots, \phi_p)$ et $\phi_{(j)} = (\phi_1, \dots, \phi_{j-1}, \phi_{j+1}, \dots, \phi_p)$.

Théorème 3.1. 1. Pour $j = 1, \dots, p$, sachant $m, \phi^{(p)}, \delta_{(j)}$, et τ , la distribution a posteriori conditionnelle de δ_j est donnée par :

$$\pi(\delta_j | m, \phi^{(p)}, \delta_{(j)}, \tau, Y) \propto \left\{ 1 + \frac{1}{(n-1)} \left(\frac{\delta_j - \widehat{\delta}_j(m, \phi^{(p)}, \delta_{(j)})}{S_{1j}(m, \phi^{(p)}, \delta_{(j)}, \tau)} \right)^2 \right\}^{-\frac{n}{2}}, \quad (3.6)$$

Avec :

$$\widehat{\delta}_j(m, \phi^{(p)}, \delta_{(j)}) = \frac{\sum_{t=m+1}^n y_{t-j} (y_t - \sum_{i=1}^p \phi_i y_{t-i} - \sum_{i \neq j} \delta_i y_{t-i})}{\sum_{t=m+1}^n y_{t-j}^2},$$

et

$$S_{1j}^2(m, \phi^{(p)}, \delta_{(j)}, \tau) = \frac{\tau SS_1(m, \phi^{(p)}) + SS_{2j}(m, \phi^{(p)}, \delta_{(j)})}{(n-1) \sum_{t=m+1}^n y_{t-j}^2},$$

où

$$SS_1(m, \phi^{(p)}) = \sum_{t=1}^m \left(y_t - \sum_{i=1}^p \phi_i y_{t-i} \right)^2, \quad (3.7)$$

$$\begin{aligned} \text{et } SS_{2j}(m, \phi^{(p)}, \delta_{(j)}) = \sum_{t=m+1}^n (y_t - \sum_{i=1}^p \phi_i y_{t-i} - \sum_{i \neq j} \delta_i y_{t-i})^2 - \\ \frac{\left[\sum_{t=m+1}^n y_{t-j} (y_t - \sum_{i=1}^p \phi_i y_{t-i} - \sum_{i \neq j} \delta_i y_{t-i}) \right]^2}{\sum_{t=m+1}^n y_{t-j}^2}, \end{aligned} \quad (3.8)$$

qui suit une loi de Student de moyenne $\widehat{\delta}_j(m, \phi^{(p)}, \delta_{(j)})$ et de variance $S_{1j}^2(m, \phi^{(p)}, \delta_{(j)}, \tau)$ et de $(n-1)$ degrés de liberté. De plus,

$$T_j(\delta_j) = \frac{\delta_j - \widehat{\delta}_j(m, \phi^{(p)}, \delta_{(j)})}{S_{1j}(m, \phi^{(p)}, \delta_{(j)}, \tau)},$$

est distribuée, conditionnellement à $m, \phi^{(p)}, \delta_{(j)}$ et τ , suivant une loi de Student standard avec $(n - 1)$ degrés de liberté.

2. Pour $j = 1, \dots, p$, Sachant $m, \phi_{(j)}, \delta^{(p)}$ et τ , la distribution a posteriori conditionnelle de ϕ_j est donnée par :

$$\pi(\phi_j | m, \phi_{(j)}, \delta^{(p)}, \tau, Y) \propto \left\{ 1 + \frac{1}{(n-1)} \left(\frac{\phi_j - \widehat{\phi}_j(m, \phi_{(j)}, \delta^{(p)}, \tau)}{S_{2j}(m, \phi_{(j)}, \delta^{(p)}, \tau)} \right)^2 \right\}^{-\frac{n}{2}}, \quad (3.9)$$

Avec

$$\widehat{\phi}_j(m, \phi_{(j)}, \delta^{(p)}, \tau) = \frac{1}{\sum_{t=m+1}^n y_{t-j}^2 + \tau \sum_{t=1}^m y_{t-j}^2} \left[\tau \sum_{t=1}^m y_{t-j} (y_t - \sum_{i \neq j} \phi_i y_{t-i}) + \sum_{t=m+1}^n y_{t-j} (y_t - \sum_{i=1}^p \delta_i y_{t-i} - \sum_{i \neq j} \phi_i y_{t-i}) \right],$$

et

$$S_{2j}^2(m, \phi_{(j)}, \delta^{(p)}, \tau) = \frac{SS_{3j}(m, \delta^{(p)}, \phi_{(j)}, \tau)}{(n-1) SS_{4j}(m, \tau)},$$

où

$$SS_{3j}(m, \delta^{(p)}, \phi_{(j)}, \tau) = \tau \sum_1^m \left(y_t - \sum_{i \neq j} \phi_i y_{t-i} \right)^2 + \sum_{t=m+1}^n \left(y_t - \sum_{i \neq j} \phi_i y_{t-i} - \sum_{i=1}^p \delta_i y_{t-i} \right)^2$$

$$\frac{\left[\tau \sum_{t=1}^m y_{t-j} (y_t - \sum_{i \neq j} \phi_i y_{t-i}) + \sum_{t=m+1}^n y_{t-j} (y_t - \sum_{i \neq j} \phi_i y_{t-i} - \sum_{i=1}^p \delta_i y_{t-i}) \right]^2}{\sum_{t=m+1}^n y_{t-j}^2 + \tau \sum_{t=1}^m y_{t-j}^2}, \quad (3.10)$$

$$SS_{4j}(m, \tau) = \sum_{t=m+1}^n y_{t-j}^2 + \tau \sum_{t=1}^m y_{t-j}^2, \quad (3.11)$$

qui suit une loi de Student t de moyenne $\widehat{a}_j(m, \phi_{(j)}, \delta^{(p)}, \tau)$ et de variance $S_{2j}^2(m, \phi_{(j)}, \delta, \tau)$ et de $(n - 1)$ degrés de liberté.

3. Sachant $m, \phi^{(p)}$ et $\delta_{(j)}$; pour $j = 1, \dots, p$, la distribution a posteriori conditionnelle de τ est donnée par :

$$\pi(\tau | m, \phi^{(p)}, \delta_{(j)}, Y) \propto \tau^{\frac{n}{2}-1} \left\{ \tau SS_1(m, \phi^{(p)}) + SS_{2j}(m, \phi^{(p)}, \delta_{(j)}) \right\}^{-\frac{(n-1)}{2}}, \quad (3.12)$$

pour $j = 1, \dots, p$ ce qui est équivalent à :

$$F(\tau) = \frac{SS_1(m, \phi^{(p)}) | m}{SS_{2j}(m, \phi^{(p)}, \delta_{(j)}) | (n - m - 1)} \tau, \quad (3.13)$$

qui suit, conditionnellement à $m, \phi^{(p)}$ et $\delta_{(j)}$ une loi de Fisher de $(m, n - m - 1)$ degrés de liberté.

4. Sachant $\phi^{(p)}, \delta^{(p)}$ et τ , la distribution a posteriori conditionnelle de m est donnée par :

$$\pi(m | \phi^{(p)}, \delta^{(p)}, \tau, y) = \frac{\pi(m, \phi^{(p)}, \delta^{(p)}, \tau | y)}{\sum_m \pi(m, \phi^{(p)}, \delta^{(p)}, \tau | Y)}, \quad (3.14)$$

où

$$\begin{aligned} \pi(m, \phi^{(p)}, \delta^{(p)}, \tau | y) \propto & \tau^{\frac{m}{2}-1} \left\{ \tau SS_1(m, \phi^{(p)}) + SS_{2j}(m, \phi^{(p)}, \delta_{(j)}) \right. \\ & \left. + \sum_{t=m+1} Y_{t-j}^2 (\delta_j - \widehat{\delta}_j(m, \phi^{(p)}, \delta_{(j)}))^2 \right\}^{-\frac{n}{2}}, \end{aligned} \quad (3.15)$$

Preuve

- Détermination de la densité a posteriori conjointe de $\Phi = (m, \phi^{(p)}, \delta^{(p)}, \tau)$:

Par transformation bijective de l'ensemble des paramètres $\Theta = (m, \phi^{(p)}, \psi^{(p)}, r_1, r_2)$ à $\Phi = (m, \phi^{(p)}, \delta^{(p)}, \tau)$, on peut trouver la densité a posteriori conjointe de Φ comme suit,

$$\begin{aligned} \pi(\Phi | y) &= \int_{r_2} \pi(m, \phi^{(p)}, \delta^{(p)} + \phi^{(p)}, r_2 \tau, r_2 / y | r_2 | dr_2, \\ &\propto \tau^{\frac{m}{2}-1} \left\{ \tau SS_1(m, \phi^{(p)}) + SS_{2j}(m, \phi^{(p)}, \delta_{(j)}) \right. \\ &\quad \left. + \sum_{t=m+1} y_{t-j}^2 (\delta_j - \widehat{\delta}_j(m, \phi^{(p)}, \delta_{(j)}))^2 \right\}^{-\frac{n}{2}}, \\ & \quad j \in \{1, \dots, p\}. \end{aligned} \quad (3.16)$$

i) A partir du théorème de Bayes, la distribution a posteriori de δ_j ($j = 1 \dots, p$) est obtenue comme elle est donnée dans (3.6).

ii) L'intégration de (3.16) par rapport à δ_j donne la distribution a posteriori conjointe de $m, \phi^{(p)}, \delta_{(j)}$ et de τ comme elle est donnée dans (3.15), par application successive du théorème de Bayes, les distributions a posteriori conditionnelles de ϕ_j, τ et de m sont obtenues comme elles sont données dans (3.9), (3.12) et (3.14) respectivement.

La densité a posteriori marginale du point de rupture m est donnée par :

$$\pi(m|y) = \int_{\tau} \left[\int_{\phi^p} \left(\int_{\delta^p} \pi(m, \phi^{(p)}, \delta^{(p)}, \tau | y) d\delta^p \right) d\phi^p \right] d\tau \quad (3.17)$$

pour $p = 2$, Les lois a posteriori inconditionnelles de $T_j(\delta_j)$ ($j = 1, 2$) et de $F(\tau)$ sont données par :

$$\begin{aligned} \pi(T_j(\delta_j) | y) &= \sum_m \int_{\tau} \int_{\phi^p} \int_{\delta_{(j)}} \pi(T_j(\delta_j) | m, \phi^p, \delta_{(j)}, \tau, y) \pi(\delta_{(j)} | m, \phi^p, \tau, y) \\ &\quad \times \pi(\phi^p | m, \tau, y) \pi(\tau | m, y) \pi(m | y) d\delta_{(j)} d\phi^p d\tau \quad j \in \{1, 2\}. \end{aligned} \quad (3.18)$$

$$\begin{aligned} \pi(F(\tau) | y) &= \sum_m \int_{\phi^p} \int_{\delta_{(j)}} \pi(F(\tau) | m, \phi^p, \delta_{(j)}, Y) \pi(\delta_{(j)} | m, \phi^p, \tau, y) \\ &\quad \times \pi(\phi^p | m, \tau, Y) \pi(m | y) d\delta_{(j)} d\phi^p \quad \text{pour } j \in \{1, 2\}. \end{aligned} \quad (3.19)$$

Remarque 3.1. Comme les degrés de liberté m et $n - m - 1$ de la loi F sont supérieures ou égal à 1, ce qui impose à m d'appartenir à $\{1, n - 2\}$.

3.4 Test de signification Bayésien :

L'hypothèse H_0 peut être divisée en deux sous hypothèses $H_{01} : \delta^{(p)} = a^{(p)} - a^{(p)} = 0$ et $H_{02} : \tau = \frac{\sigma_1^2}{\sigma_2^2} = 1$. De même l'hypothèse H_{01} peut être divisée en p sous hypothèses $H_{01j} : \delta_j = 0$, pour $j = 1, \dots, p$ et l'hypothèse H_{01} sera rejetée si l'une des sous hypothèses H_{01j} est rejetée. Pour $\alpha \in]0, 1[$ donné, $j = 1, \dots, p$, conditionnellement à $m, a^{(p)}, \delta_{(j)}$ et τ , le $(1 - \alpha)$ -intervalle H.P.D-crédible de $T_j(\delta_j)$ est donné par : $C_{\delta_j} = \{T_j(\delta_j) / |T_j(\delta_j)| < t_{\alpha/2}(n - 1)\}$ où $t_{\beta}(n)$ est le fractile d'ordre β de la distribution de Student standard avec n degrés de liberté. La sous hypothèse H_{01j} est rejetée, conditionnellement à $m, a^{(p)}, \delta_{(j)}$ et τ , si $T_j(0) \in \overline{C_{\delta_j}}$, où $\overline{C_{\delta_j}}$ est le complémentaire de C_{δ_j} représentant la région de rejet des sous hypothèses $H_{01j}, j = 1, \dots, p$.

Les p-values inconditionnelles des sous hypothèses H_{01j} sont définies par :

$$\begin{aligned} P_{\delta_j=0|y} &= 2 \sum_m \int_{\tau} \int_{\phi^{(p)}} \int_{\delta_{(j)}} \{1 - \mathcal{T}_{n-1}(|t_j(0)|)\} \\ &\quad \pi(m, \delta_{(j)}, \phi^{(p)}, \tau | y) d\delta_{(j)} d\phi^{(p)} d\tau, \\ &= 2 E_m E_{\tau} E_{\phi^{(p)}} E_{\delta_{(j)}} \{1 - \mathcal{T}_{n-1}(|t_j(0)|)\}, \end{aligned} \quad (3.20)$$

où \mathcal{T}_{n-1} est la fonction cumulative de la distribution de Student avec $(n - 1)$ degrés de liberté.

pour $j = 1, \dots, p$, l'hypothèse H_{01j} est rejetée inconditionnellement au niveau de signification α si $P_{\delta_j=0|y} < \alpha$. Ce qui permet de tester d'une manière séparée l'existence d'un changement dans chaque paramètre du modèle.

H_{01} sera rejetée inconditionnellement pour un niveau de signification α si

$$P_{\delta=0|y} := \min_{1 \leq j \leq p} \{P_{\delta_j=0|y}\} < \alpha$$

La p-value inconditionnelle de sous hypothèse H_{02} est définie par :

$$\begin{aligned} P_{\tau=1|y} &= 2 \sum_m \int_{\phi^{(p)}} \int_{\delta_{(j)}} \{1 - \mathcal{F}_{m,n-m-1}[\max(F(1), 1/F(1))]\} \\ &\quad \times \pi(m, \delta_{(j)}, \phi^{(p)} | y) d\delta_{(j)} d\phi^{(p)}, \\ &= 2E_m E_{\phi^{(p)}} E_{\delta_{(j)}} \{1 - \mathcal{F}_{m,n-m-1}[\max(F(1), 1/F(1))]\}, \end{aligned} \quad (3.21)$$

où $\mathcal{F}_{m,n-m-1}$ est la fonction cumulative de la distribution de Fisher avec $(m, n - m - 1)$ degrés de liberté. La sous hypothèse H_{02} sera rejetée inconditionnellement pour un niveau de signification α si $P_{\tau=1|y} < \alpha$.

Les p-values conditionnelles sachant m des sous hypothèses H_{01j} ($j = 1, \dots, p$) et H_{02} sont données respectivement par :

$$\begin{aligned} P_{\delta_j=0|m,y} &= 2 \int_{\tau} \int_{\phi^{(p)}} \int_{\delta_{(j)}} \{1 - \mathcal{T}_{n-1}(|t_j(0)|)\} \pi(m, \delta_{(j)}, \phi^{(p)}, \tau | y) d\delta_{(j)} d\phi^{(p)} d\tau, \\ &= 2E_{\tau} E_{\phi^{(p)}} E_{\delta_{(j)}} \{1 - \mathcal{T}_{n-1}(|t_j(0)|)\}, \end{aligned} \quad (3.22)$$

et

$$\begin{aligned} P_{\tau=1|m,y} &= 2 \int_{\phi^{(p)}} \int_{\delta_{(j)}} \{1 - \mathcal{F}_{m,n-m-1}[\max(F(1), 1/F(1))]\} \pi(m, \delta_{(j)}, \phi^{(p)} | y) d\delta_{(j)} d\phi^{(p)}, \\ &= 2E_{\phi^{(p)}} E_{\delta_{(j)}} \{1 - \mathcal{F}_{m,n-m-1}[\max(F(1), 1/F(1))]\}, \end{aligned} \quad (3.23)$$

Les quantités données dans (3.17), (3.20), (3.21), (3.22) et (3.23) seront évaluées numériquement à l'aide de l'échantillonneur de Gibbs (Gibbs Sampler) en utilisant les lois de probabilités conditionnelles données dans le théorème 3.1.

Notations : $E_{\phi^{(p)}}$ est l'intégrale successive par rapport à ϕ_1, \dots, ϕ_p et $E_{\delta_{(j)}}$ est l'intégrale successive par rapport à $\delta_1, \dots, \delta_{j-1}, \delta_{j+1}, \dots, \delta_p$, E_{τ} est l'intégrale par rapport à τ et E_m est la somme relativement à m .

3.5 Étude de simulation

Dans cette section nous présentons un exemple de simulation pour illustrer les résultats ci-dessus. Nous simulons un échantillon à partir du modèle suivant :

$$y_t = \begin{cases} \mu + 0.2(y_{t-1} - \mu) + 0.3(y_{t-2} - \mu) + \epsilon_t & , t = 1 \cdots , m \\ \mu + 0.4(y_{t-1} - \mu) + 0.6(y_{t-2} - \mu) + \epsilon_t & , t = m + 1, \cdots , 200 \end{cases} \quad (3.24)$$

avec ;

$$\epsilon_t \xrightarrow{i.i.d} \begin{cases} N(0, 0.5^2), & t = 1, \cdots , m \\ N(0, 1^2), & t = m + 1, \cdots , 200 \end{cases}$$

Nous allons considérer deux valeurs pour le point de changement m , a savoir $m = 110$ et $m = 150$. Les valeurs considérées pour les quantités initiales y_{-1} et y_0 sont 0.5 et 0.6 respectivement. Sans perte de généralité, on suppose que $\mu = 0$. Les observations $(y_1, y_2, \cdots, y_{200})$ simulées à partir du modèle (3.24) sont représentées dans les figures (3.1) et (3.2).

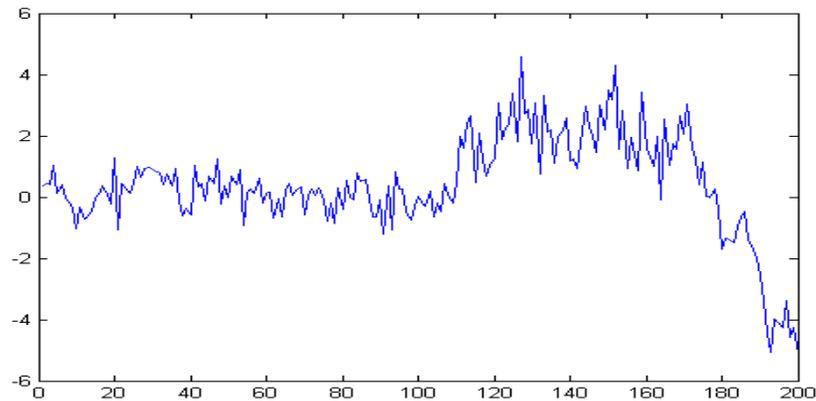


FIGURE 3.1 – Observations simulées y_t avec $m = 110$.

En utilisant ces observations et par l'application de l'algorithme de l'échantillonneur de Gibbs avec 5000 répétitions, nous allons obtenir des approximations des densités a posteriori marginales du point de changement m , des paramètres δ_1 et δ_2 ainsi que les p-values inconditionnelles des sous hypothèses $H_{011} : \delta_1 = 0$, $H_{012} : \delta_2 = 0$ et de $H_{02} : \tau = 1$.

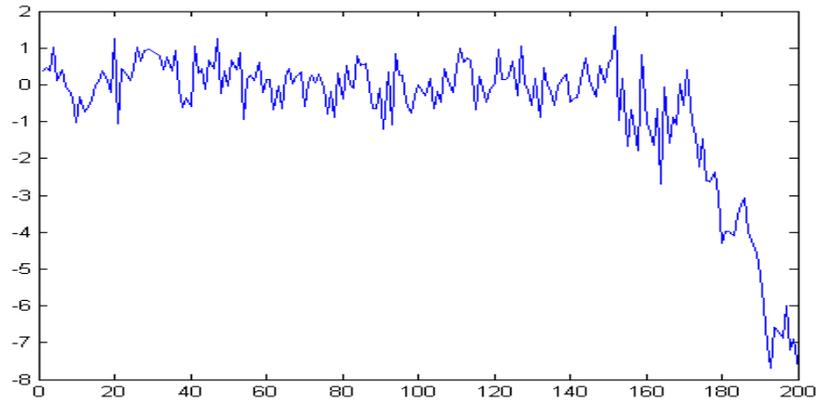


FIGURE 3.2 – Observations simulées y_t avec $m = 150$.

Les résultats sont donnés dans les tables (3.1) et (3.2). Les figures (3.3) et (3.7) illustrent les distributions a posteriori du point de changement m .

3.5.1 Distributions a posteriori des paramètres

Les tables (3.1) et (3.2) donnent les densités a posteriori marginales et conditionnelles du point de changement m , les p-values inconditionnelle et conditionnelle sachant m pour des sous hypothèses H_{011} , H_{012} et H_{02} . Basée sur les p-values inconditionnelles, la stationnarité des paramètres δ_1 , δ_2 et τ est clairement rejetée à un niveau de signification 5%, 10% et 1%, respectivement. Aussi, à partir des tables (3.1) et 3.2 on peut constater facilement que le mode a posteriori est égal à aux vraies valeurs de m , qui sont, 110 et 150 respectivement.

Les graphes (3.4), (refhistd1150), (3.5) et (3.9) montrent que les distributions a posteriori des écarts δ_1 et δ_2 sont bien déplacées de 0, ce qui indique fortement qu'il est peu probable que les valeurs de δ_1 et δ_2 soient égales à 0. De même, à partir des graphes (3.6) et (3.10), on peut voir clairement que la valeur de τ ne peut pas être égale à 1.

Les tables (3.3) et (3.4) résument les estimateurs a posteriori des paramètres m , δ_1 , δ_2 et τ lorsque $m = 110$ et $m = 150$. Ils rassemblent la moyenne et la médiane a posteriori et les écarts types ainsi que les 2.5 et 97.5 centiles de ces paramètres. La table (3.4) donne les estimateurs a posteriori des paramètres du modèle pour les observations simulés quand $m = 150$, la moyenne a posteriori (149.35) et la médiane a posteriori (150) de m estiment

CHAPITRE 3. MODÈLE $AR(P)$ AVEC CHANGEMENT DANS LES PARAMÈTRES D'AUTORÉGRESSION

m	$\pi(m y)$	$P_{\delta_1=0 m,y}$	$P_{\delta_2=0 m,y}$	$P_{\tau=1 m,y}$
105	0.0399	0.0005	0.0780	$1.48 \cdot 10^{-8}$
106	0.0676	0.0009	0.0839	$9.74 \cdot 10^{-9}$
107	0.0879	0.0009	0.0843	$1.03 \cdot 10^{-8}$
108	0.1475	0.0009	0.0749	$7.01 \cdot 10^{-9}$
109	0.2325	0.0022	0.0690	$5.83 \cdot 10^{-9}$
110	0.3338	0.0025	0.0742	$4.88 \cdot 10^{-9}$
111	0.0031	0.0037	0.0550	$3.13 \cdot 10^{-7}$
112	0.0009	0.0104	0.0447	$5.52 \cdot 10^{-7}$
113	0.0001	0.0098	0.0815	$1.19 \cdot 10^{-6}$
114	$3.57 \cdot 10^{-5}$	0.0158	0.0975	$2.62 \cdot 10^{-6}$
115	$2.76 \cdot 10^{-5}$	0.0138	0.0589	$1.70 \cdot 10^{-6}$
Unconditional p-value		0.0240	0.0787	$1.69 \cdot 10^{-8}$

TABLE 3.1 – Densité a posteriori de m , les p-values conditionnelles et inconditionnelles de H_{011} , H_{012} et H_{02} pour $m = 110$.

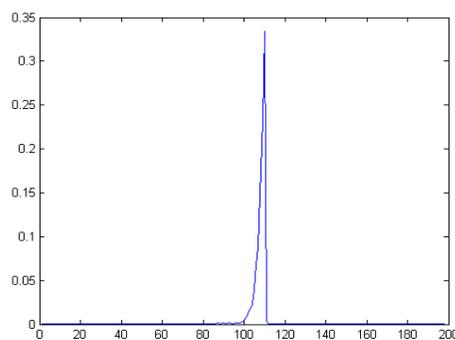


FIGURE 3.3 – Densité a posteriori de m avec $m = 110$.

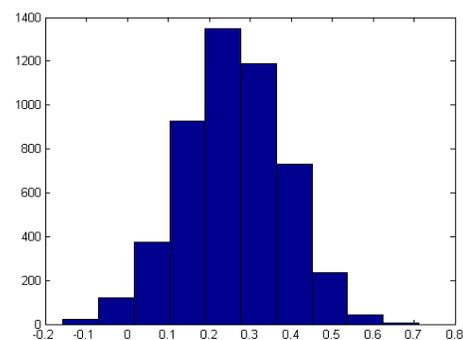


FIGURE 3.4 – Distribution a posteriori de δ_1 avec $m = 110$.

bien la vraie valeur de m égale à 150. Dans l'ensemble, on voit clairement que toutes les 95% régions HPD crédibles des paramètres contiennent les vraies valeurs des paramètres.

CHAPITRE 3. MODÈLE $AR(P)$ AVEC CHANGEMENT DANS LES PARAMÈTRES D'AUTORÉGRESSION

m	$\pi(m y)$	$P_{\delta_1=0 m,y}$	$P_{\delta_2=0 m,y}$	$P_{\tau=1 m,y}$
145	0.0308	0.0001	0.0307	$1.47 \cdot 10^{-5}$
146	0.0512	0.0001	0.0279	$1.40 \cdot 10^{-5}$
147	0.0752	0.0006	0.0276	$1.19 \cdot 10^{-5}$
148	0.0916	0.0005	0.0323	$1.33 \cdot 10^{-5}$
149	0.1594	0.0009	0.0321	$1.02 \cdot 10^{-5}$
150	0.2217	0.0011	0.0340	$8.67 \cdot 10^{-6}$
151	0.1992	0.0015	0.0322	$8.78 \cdot 10^{-6}$
152	0.0249	0.0023	0.0345	$1.84 \cdot 10^{-5}$
153	0.0137	0.0007	0.0381	0.0002
154	0.0217	0.0010	0.0297	0.0001
155	0.0035	0.0012	0.0510	0.0004
Unconditional p-value		0.0103	0.0321	0.0002

TABLE 3.2 – Densité a posteriori de m , les p-values conditionnelles et inconditionnelles de H_{011} , H_{012} et H_{02} pour $m = 150$.

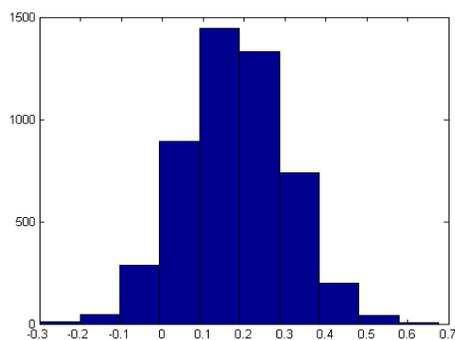


FIGURE 3.5 – Distribution a posteriori de δ_2 avec $m = 110$.

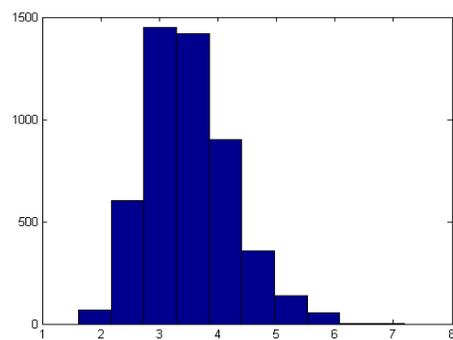


FIGURE 3.6 – Distribution a posteriori de τ avec $m = 110$.

CHAPITRE 3. MODÈLE $AR(P)$ AVEC CHANGEMENT DANS LES PARAMÈTRES D'AUTORÉGRESSION

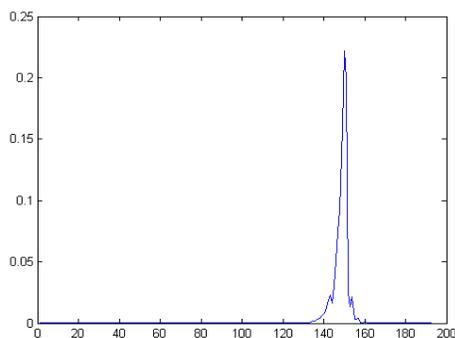


FIGURE 3.7 – Distribution a posteriori de m avec $m = 150$

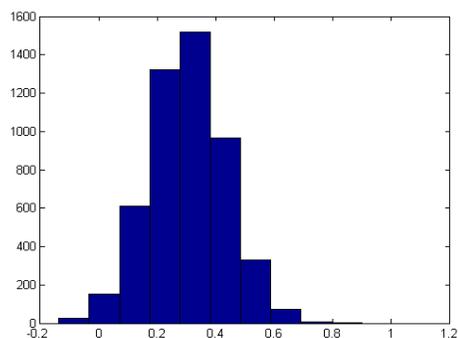


FIGURE 3.8 – Distribution a posteriori de δ_1 avec $m = 150$

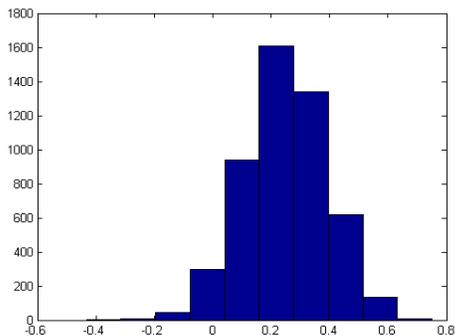


FIGURE 3.9 – Distribution a posteriori de δ_2 avec $m = 150$

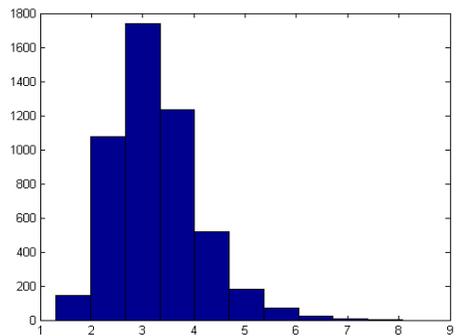


FIGURE 3.10 – Distribution a posteriori de τ avec $m = 150$

Paramètres	vraies valeurs	Médiane	Moyenne	SD	2.5%	97,5%
m	110	109	107.92	3.1529	101	110
δ_1	0.2	0.2590	0.2585	0.1239	0.0070	0.4975
δ_2	0.3	0.1764	0.1780	0.1264	-0.0624	0.4299
τ	4	3.4314	3.5097	0.7344	2.2811	5.2229

TABLE 3.3 – Estimateurs a posteriori de m , δ_1 , δ_2 et de τ lorsau $m = 110$.

Avec la même série simulée de loi normale, nous avons simulée d'autres observations avec différentes valeurs des paramètres, les résultats sont donnés dans la table (3.5).

Paramètres	Vraies valeurs	Médiane	Moyenne	SD	2.5%	97,5%
m	150	150	149.35	2.9249	143	157
δ_1	0.2	0.3022	0.3036	0.1303	0.0544	0.5630
δ_2	0.3	0.2467	0.2491	0.1422	-0.0346	0.5247
τ	4	3.1570	3.2507	0.8222	1.9639	5.2133

TABLE 3.4 – Estimateurs a posteriori de m , δ_1 , δ_2 et de τ lorsque $m = 150$

La table (3.5) donne le mode a posteriori de m et les p-values inconditionnelles de H_{011} , H_{012} et de H_{02} pour différentes valeurs des paramètres m , ϕ_1 , ϕ_2 , ψ_1 , ψ_2 , σ_1 et σ_2 . Les résultats donnés dans la table (3.5) montrent que le mode a posteriori de m estime bien la vraie valeur du point de changement pour les trois ensembles de valeurs des paramètres. Pour le test de signification Bayésien, les p-values inconditionnelles des sous hypothèses H_{012} et H_{02} , calculées à partir du premier ensemble des valeurs des paramètres, indiquent que le non changement dans $\delta_2 = \psi_2 - \phi_2$ et $\tau = \sigma_2^2/\sigma_1^2$ est rejeté au niveau de signification 5% et 1% respectivement. Par contre, le non-changement dans le paramètre δ_1 ne peut pas être rejetée même à 20% niveau de signification, noter que la valeur de $\delta_1 = \psi_1 - \phi_1$ est égale à 0. Pour le second ensemble des valeurs des paramètres, la p-value de H_{02} est égale à 0.4529, donc H_{02} ne peut pas être rejetée. De même pour le troisième ensemble des valeurs des paramètres, la p-value de sous hypothèse H_{012} est 0.2563, ce qui implique que H_{012} ne peut pas être rejeté, alors que les sous hypothèses H_{011} et H_{02} sont rejetées au niveau de signification de 1%. Par conséquent, le test de signification Bayésien est utile pour déterminer quel paramètre du modèle $AR(p)$ est non stationnaire (Présente un changement) à partir l'instant m .

3.5.2 Étude de la performance du test

Pour étudier la performance du test de signification Bayésien pour la détection d'un changement dans les paramètres d'autorégression, on simule 1000 échantillons à partir du

CHAPITRE 3. MODÈLE $AR(P)$ AVEC CHANGEMENT DANS LES PARAMÈTRES D'AUTORÉGRESSION

	$m = 90$	$m = 110$	$m = 120$
Valeurs des paramètres	$\phi_1 = 0.2, \psi_1 = 0.2$ $\phi_2 = 0.2, \psi_2 = 0.6$ $\sigma_1 = 0.5, \sigma_2 = 1$	$\phi_1 = 0.2, \psi_1 = 0.3$ $\phi_2 = 0.4, \psi_1 = 0.6$ $\sigma_1 = 0.5, \sigma_2 = 0.5$	$\phi_1 = 0.2, \psi_1 = 0.5$ $\phi_2 = 0.3, \psi_2 = 0.3$ $\sigma_1 = 0.5, \sigma_2 = 1$
Mode a posteriori de m	90	110	120
$P_{\delta_1=0 y}$	0.3607	0.0277	0.0063
$P_{\delta_2=0 y}$	0.0148	0.0285	0.2563
$P_{\tau=1 y}$	$1.60 \cdot 10^{-5}$	0.4529	$1.21 \cdot 10^{-4}$

TABLE 3.5 – Mode a posteriori de m et la p-value inconditionnelle de H_{011} , H_{012} et de H_{011} pour différentes valeurs de $m, \phi_1, \phi_2, \psi_1, \psi_2, \sigma_1$ et de σ_2 .

modèle (3.1) avec $n = 200, m = 110$ et différentes valeurs de $\phi_1, \phi_2, \psi_1, \psi_2, \sigma_1^2$ et σ_2^2 et on calcul le taux de rejet de l'hypothèse H_0 au niveau de signification de 5%. Les résultats sont donnés dans la Table (3.6).

Valeurs des paramètres	$\delta_1 = 0.2$ $\delta_2 = 0.3$ $\tau = 4$	$\delta_1 = 0.0$ $\delta_2 = 0.3$ $\tau = 4$	$\delta_1 = 0.3$ $\delta_2 = 0.0$ $\tau = 4$	$\delta_1 = 0.2$ $\delta_2 = 0.3$ $\tau = 1.0$
H_{011}	0.581	0.060	0.602	0.618
H_{012}	0.822	0.590	0.071	0.820
H_{02}	1.000	0.999	0.999	0.047

TABLE 3.6 – Taux de rejet de H_{011} , H_{012} et de H_{02} au niveau de signification de 5% pour 1000 échantillons avec différentes valeurs de δ_1, δ_2 et de τ .

La Table (3.6) montre que les taux de rejet des sous hypothèses H_{011} , H_{012} et H_{02} sont plus de 58% au niveau de signification 5% lorsqu'il y a changement dans le paramètre correspondant, et il est de 5% lorsque le paramètre est stable. Pour la sous hypothèse H_{01} ,

le taux de rejet est plus que 58% lorsque il y a un changement dans δ_1 , et il n'est que 6.0% lorsqu'il est stable. Pour la sous hypothèse H_{02} , le taux de rejet de sous hypothèse est plus que 99.9% lorsque il y a un changement dans le paramètre τ et il est de l'ordre de 4.7% lorsqu'on a pas de changement dans ce paramètre. Par conséquent, le test de signification Bayésien détecte bien le paramètre objet de changement dans le modèle autorégressif d'ordre p , $Ar(p)$.

3.5.3 Étude de comparaison :

Une petite étude de simulation a été effectuée pour comparer la puissance du test de signification Bayésien avec celles des tests basés sur les sommes partielles des résidus pondérés introduits par Huskova et al. [64]. Les test sont basés sur les statistiques T_n , $T_n(\varepsilon)$, $T_n^*(q)$ (Huskova et al. [63]).

Les versions Bootstrap des statistiques sont T_{nR}^* , $T_{nR}^*(\varepsilon_0)$, $T_{nR}^*(q_0)$ (Regression bootstrap) et T_{nP}^* , $T_{nP}^*(\varepsilon_0)$, $T_{nP}^*(q_0)$ (Pair bootstrap) (Huskova et al. [64]). Une étude de simulation a montré que, pour toutes les statistiques, les tests bootstrap fonctionnent mieux (plus puissants) que les tests asymptotiques [64].

Nous avons simulé des données à partir du modèle $Ar(1)$ avec $n = 200$, $m = 100$, $\phi_1 = 0,3$, $\sigma_1^2 = \sigma_2^2 = 1$ et avec différentes valeurs de ψ_1 . Les résultats sont donnés dans la table 3.7. Les résultats relatifs aux tests T_{nR}^* , $T_{nR}^*(\varepsilon_0)$, $T_{nR}^*(q_0)$ et T_{nP}^* , $T_{nP}^*(\varepsilon_0)$, $T_{nP}^*(q_0)$ sont donnés à partir du tableau 5 de Huskova et al.[64]. Pour le test de signification Bayésien, les résultats sont obtenus par l'application de l'échantillonneur de Gibbs avec 1000 répétitions.

La table (3.7) résume les puissances des tests T_{nR}^* , $T_{nR}^*(\varepsilon_0)$, $T_{nR}^*(q_0)$, T_{nP}^* , $T_{nP}^*(\varepsilon_0)$, $T_{nP}^*(q_0)$ et le test de signification Bayésien. Pour le test de signification Bayésien, le test pour un changement dans la variance maintient un niveau de signification empirique proche de la valeur nominale $\alpha = 0,05$. A partir des résultats de la table (3.7), on peut observer que le test de signification Bayésien pour un changement dans les paramètres du modèle autorégressif $Ar(p)$ semble plus puissant que les tests basés sur les sommes partielles des résidus pondérés introduits par Huskova et al. [64].

Tests	$\psi_1 = 0.3$	$\psi_1 = 0.6$	$\psi_1 = 0.8$
<i>Regression bootstrap</i>			
$(T_{nR}^*)^{1/2}$	0.049	0.425	0.947
$(T_{nR}^*(\varepsilon_0))^{1/2}$	0.051	0.459	0.954
$(T_{nR}^*(q_0))^{1/2}$	0.058	0.528	0.965
<i>Pair bootstrap</i>			
$(T_{nP}^*)^{1/2}$	0.05	0.419	0.944
$(T_{nP}^*(\varepsilon_0))^{1/2}$	0.058	0.516	0.965
$(T_{nP}^*(q_0))^{1/2}$	0.074	0.559	0.965
<i>Bayesian significance test</i>	0.012	0.600	0.998

TABLE 3.7 – Puissances empiriques des tests T_{nR}^* , $T_{nR}^*(\varepsilon_0)$, $T_{nR}^*(q_0)$, T_{nP}^* , $T_{nP}^*(\varepsilon_0)$, $T_{nP}^*(q_0)$ et le test de signification Bayésien de H_{01} pour un niveau de signification égal à 0.05.

3.6 Conclusion

Dans ce chapitre, nous avons présenté une analyse Bayésienne d'un changement dans les paramètres d'un modèle autorégressif d'ordre connu p , ($AR(p)$). Le modèle est soumis à un changement aussi bien dans les paramètres de régression que dans la variance à un instant inconnu m . Nous avons présenté un test de signification Bayésien inconditionnel pour la détection d'un changement dans les paramètres du modèle $Ar(p)$. Le test est basé sur la construction des régions HPD crédibles et le calcul des p-values inconditionnelles. A partir d'une étude de simulation en utilisant l'algorithme de l'échantillonneur de Gibbs, nous avons déterminé les densités a posteriori marginales du point de rupture, de l'amplitude du changement dans les paramètres d'autorégression, du rapport de la variance ainsi que les p-values inconditionnelles.

Une étude de comparaison a montré que le test de signification Bayésien pour le changement dans les paramètres d'autorégression du modèle $AR(p)$ est plus puissant que

CHAPITRE 3. MODÈLE $AR(P)$ AVEC CHANGEMENT DANS LES PARAMÈTRES D'AUTORÉGRESSION

les tests basés sur les sommes partielles des résidus pondérés proposés dans Huskova et al. [64]. Un autre avantage du test de signification Bayésien sur ces tests est la possibilité de la détection de changement dans les paramètres d'une manière individuelle. Aussi, l'utilisation de l'algorithme de l'échantillonneur de Gibbs nous a permis de calculer des intégrales complexes et de grandes dimensions données par exemple dans les formules (3.20) et (3.21).

Conclusion générale

L'étude des problèmes de changement est assez présent dans la littérature. Ils ont été traités suivant différentes approches statistiques, en utilisant soit l'approche classique basée sur la fonction de vraisemblance soit l'approche Bayésienne ou encore l'approche non paramétrique.

L'approche Bayésienne a été largement utilisée pour la détection et l'estimation des ruptures. Elle est très intéressante dans de nombreux domaines comme la médecine, où l'information a priori est habituellement disponible sous la forme d'études antérieures associées, et le théorème de Bayes permet d'intégrer les informations précédentes (a priori) avec les informations actuelles (contenues dans la vraisemblance). Aussi, le développement des outils de calculs et les méthodes de simulation ont beaucoup favorisé l'intérêt croissant de l'adoption de la méthodologie Bayésienne.

Dans ce travail nous avons présenté une analyse Bayésienne d'un modèle de série temporelle autorégressif avec changement dans les paramètres à un instant inconnu. L'étude de simulation par l'application de l'algorithme de l'échantillonneur de Gibbs, nous a permis de :

1. Montrer que le test de signification Bayésien basé sur la construction des régions HPD crédibles et le calcul des p-values inconditionnelles est sensible à la dépendance (l'autocorrélation) des observations. Ainsi, les p-values déterminées sous l'hypothèse de l'indépendance sous estiment les vrais p-values quand le coefficient de corrélation est positif, et les surestimés quand le coefficient de corrélation est négatif.
2. Montrer que le test de signification Bayésien pour la détection d'un changement dans les paramètres d'autoregression d'un modèle autorégressif d'ordre connu p

CONCLUSION GÉNÉRALE

est plus puissant que les tests basés sur les sommes cumulées des résidus. En plus, l'avantage du test de signification Bayésien est la possibilité de tester le changement dans les paramètres du modèle AR(p) d'une manière individuelle.

Aussi, l'application de l'algorithme de l'échantillonneur de Gibbs nous a permis de simplifier de façon considérable le calcul des intégrales complexes et de grande dimension pour estimer numériquement les paramètres du modèle.

Les perspectives de recherche pour ce travail sont :

1. Considérer un problème d'un changement avec plusieurs points de rupture (Multiple change point models),
2. Considérer un problème d'une rupture épidémique,
3. Étudier la robustesse de la procédure du test de signification Bayésien en considérant des lois a priori conjuguées.
4. Application du test de signification Bayésien pour la détection de rupture dans le taux de service ou l'intensité du trafic des systèmes de files d'attente.

Bibliographie

- [1] Abraham, B., Wei, W. W. S. Inference about the parameters of a time series model with changing variance. *Metrika* 31 :183-194. 1984.
- [2] M. M. Barbieri and C. Conigliani , Bayesian analysis of autoregressive time series with change points. *Journal of the Italian Statistical Society*, Volume 7, Issue 3, pp 243-255. 1998.
- [3] Bai, J. : Estimation of a change point in multiple regressions models. *Rev. Econ. Stat.* 79, 551–563 ,1997.
- [4] J. Bai, Likelihood ratio tests for multiple structural changes, *Journal of Econometrics* , vol. 91, pp. 299–323, 1999.
- [5] M. Basseville, I. Nikiforov, *Detection of Abrupt Changes : Theory and Application*, Prentice Hall Information and System Sciences Series, Prentice Hall, 1993.
- [6] Berger, J. O *Statistical Decision Theory and Bayesian Analysis*, Springer-Verlag, New York, second édition. 1985.
- [7] Bernardo, J. et Smith, A. *Bayesian Theory*. John Wiley, New York. 1994.
- [8] Bhattacharyya, G.K. and R. Johnson . Nonparametric tests for shifts at an unknown time point. *Ann. Math. Statist.* 39, 1731-1743, 1968.
- [9] Bhattacharya, P.K. and D. Frierson, Jr. A nonparametric control chart for detecting small disorders. *Ann. Statist.* 9, 544-554. 1981.
- [10] Box, G. E. P. and Jenkins, G. M. *Time Series Analysis : Forecasting and Control*, revised edn. (San Francisco, Holden-Day), 1976.
- [11] Box, G.E. E. et Tiao, G. C. *Bayesian inference in statistical analysis*. Reading, mass.addison-Wesley, 1973.

BIBLIOGRAPHIE

- [12] P. J. Brockwell et R. A. Davis *Time Series : Theory and Methods*. Springer, 1987.
- [13] Broemeling, L. D., Tsurumi, H. *Econometrics and Structural Change*. New-York : Marcel Dekker Inc. 1987.
- [14] Battaglia, F., Orfei, L., Outlier detection and estimation in nonlinear time series. *Journal of Time Series Analysis* 26, 107-121. 2005.
- [15] Brodsky, B.E., Darkhovsky, B.S. *Nonparametric Methods in Changepoint Problems*. Kluwer, Dordrecht, 1993.
- [16] Brown, R. L., Durbin, J. , Evans, J. M. . Techniques for testing the constancy of regression relationships over time (with discussion). *J. R. Statist. Soc. A* 138, 149-63, 1975.
- [17] Carlin, B. et Louis, T. *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman and Hall, New York, second édition. 2001.
- [18] Carlin, B. P., Gelfand, A. E., Smith, A. F. M. Hierarchical Bayesian analysis of changepoint problems. *Appl. Statist.* 41(2) :389-405. 1992.
- [19] Carlstein, E., Nonparametric changepoints estimation. *Ann. Statist.* 16, 188-197, 1988.
- [20] Casella, G. and E. George , Explaining the Gibbs sampler. *American Statistician*, 46, 167-174, 1992.
- [21] A. Chaturvedi and A. Shrivastava, Bayesian Analysis of Structural Changes in a Linear Regression Model : An Application to Rupee-Dollar Exchange Rate, *J. Quant. Econ.* , Volume 13, Issue 2, pp 185-200 December 2015.
- [22] Chen, C. and Liu, L. M. Forecasting time series with outliers. *Journal of Forecasting*, 12, 13-35. 1993a.
- [23] J. Chen and A. K. Gupta *Parametric Statistical Change Point Analysis, With Applications to Genetics, Medicine and Finance*, Birkhauser, 2012.
- [24] H. Chernoff, S. Zacks. Estimating the current mean distribution which is subjected to change in time, *Ann. Math. Statist*, 35 :999–1018, 1964.
- [25] Chen, C.W.S., Detection of additive outliers in bilinear time series, *Computational Statistics & Data Analysis* 24, 283-294. 1997.

BIBLIOGRAPHIE

- [26] Chib, S., Estimation and Comparison of Multiple Change Point Models, *Journal of Econometrics*, 86, 221-241, 1998.
- [27] Chow G. Test of the equality between two sets of coefficient in two linear models. *Technometrics* 22 pp 72-80. 1960.
- [28] Chin Choy, J. Y., Broemeling, L. D. Some Bayesian inferences for a changing linear model. *Technometrics* 22 :71-78. 1980.
- [29] Csorgo, M., Horvath, L. Nonparametric tests for the changepoint problem, *Journal of Statistical Planning and Inference*, 17, 1-9, 1987.
- [30] Csorgo, M., Horvath, L., *Limit Theorems in Change-point Analysis*. Wiley, New York, 1997.
- [31] R.A. Davies, D. Huang, Y.-C. Yao, Testing for a change in the parameter value and order of an autoregressive model, *Ann. Statist.* 23 282-304, 1995.
- [32] Deutsch, S. J., Richards, J. E. and Swain, J. J. Effects of a single outlier on ARMA identification. *Communications in Statistics - Theory and Methods*, 19, 2207-27, 1990.
- [33] Dumbgen, L., The asymptotic behavior of some nonparametric changepoints estimators. *Ann. Statist.* 19, 1471-1495, 1991.
- [34] H. Fan, K.-C. Chang, and C.-B. Lee, Bayesian estimation of the number of change points in simple linear regression models, *Communications in Statistics, Theory and Methods*, , vol.35, pp. 689-71, 2006
- [35] Ferreira, P. E. A Bayesian analysis of a switching regression model : known number of regimes. *J. Amer. Statist. Assoc.* 70 :370-374. 1975.
- [36] Fisher, R. *Statistical Methods and Scientific Inference*. Oliver and Boyd, Edinburgh. 1956.
- [37] Friede, T., Miller, F., Bishop, W., Kieser, M. A note on change point estimation in dose-response trials. *Comput. Statist. Data Anal.* 37 :219-232. 2001.
- [38] Fuller, W., A. *Introduction to Statistical Time Series*, New York, Wiley, 1976.
- [39] Gombay, E., Change detection in autoregressive time series. *J. Multivariate Anal.* 99, 451-464, 2008.

BIBLIOGRAPHIE

- [40] Gelfand, A.E., S.E. Hills, A. Racine-Poon and A.F.M. Smith, Illustration of Bayesian inference in normal data models using Gibbs sampling, *J. Amer. Statist. Assoc.*, 85, 972-985. 1990.
- [41] Gelfand, A.E. and A.F.M. Smith, Sampling-based approaches to calculate marginal densities, *J. Amer. Statist. Assoc.*, 85 398-409. 1990.
- [42] Gelman, A., Carlin, J., Stern, H. et Rubin, D. *Bayesian Data Analysis*. Chapman and Hall, New York, New York, second édition. 2004.
- [43] Geman, S. et Geman, D. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 6, 721-741. 1984.
- [44] Gilks, W., Richardson, S., et Spiegelhalter, D., *Markov Chain Monte Carlo in Practice*. Chapman and Hall, New York, London. 1996.
- [45] D. Ghorbanzadeh, R. Lounes, Bayesian analysis for detecting a change in exponential family, *Applied Mathematics and Computation*, 124 :1–15, 2001.
- [46] M.A. Girshick and H. Rubin, A Bayes approach to a quality control model, *Ann. Math. Statist.* 23 1 , 114-125. 1952.
- [47] Geurbyenne, H., *Analyse Bayésienne d'un modèle Gaussien à erreurs autocorrélées avec changement dans la variance à un instant inconnu*. Thèse de Magister. Institut de Mathématiques, USTHB, (1992)
- [48] Gurevich, G., Vexler, A. *Change point problems in the model of logistic regression*. Journal of Statistical Planning and Inference 131 :313–331. 2005.
- [49] Hamilton, J.D., *Time Series Analysis*. Princeton University Press, Princeton, 1994.
- [50] Hastings, W. (Monte Carlo sampling methods using Markov chains and their application. *Biometrika*, 57, 97-109. 1970.
- [51] Hawkins, D.M. Testing a sequence of observations for a shift in location, *J. Amer. Statist. Ass.*, 72, 180-186. 1977.
- [52] Henderson, R. Change-point problem with correlated observations, with an application in material accountancy. *Technometrics*, 28, 381-389. 1986.

BIBLIOGRAPHIE

- [53] D.V.Hinkley, Inference about the intersection in two-phase regression, *Biometrika*, 56 3, 495-504. 1969.
- [54] Hinkley, D. V. Inference about the change-point in a sequence of random variables. *Biometrika* 57, 1-17.1970.
- [55] D.V.Hinkley, Inference about the change-point from cumulative sum tests, *Biometrika* 58 3 , 509-523. 1971.
- [56] D.V.Hinkley, Inference in two-phase regression, *J.Amer.Statist.Assoc.* 66336 ,736-743, 1971.
- [57] Holbert, D., Broemeling, L. D. Bayesian inference related to shifting sequences and two-phase regression. *Comm. Statist* A6(3) :265-275. 1977.
- [58] Holbert, D. A Bayesian analysis of a switching linear model. *J. Econometrics* ,19 :77-87.1982.
- [59] Horváth, L. Maximum likelihood method for testing changes in the parameters of normal observations, *Ann. Statist.*, 21, 671-680, 1993.
- [60] Hsu, D. A. (1977). Tests for variance shift at an unknown time point. *JRSS Ser C* 26 :279-284.
- [61] Hsu, D. A. A Bayesian robust detection of shift in the risk structure of stock returns. *JASA* 77 :29-39. 1982.
- [62] S. Hudecova, Structural changes in autoregressive models for binary time series, *Journal of Statistical Planning and Inference* 143, 1744-1752, 2013.
- [63] Huskova, M., Praskova Z., Steinebach, J., On the detection of changes in autoregressive time series. I. Asymptotics. *J. Statist. Plann. Inference*, 137,1243-1259. 2007.
- [64] Huskova, M., Kirch, C., Praskova, Z., Steinebach, J., On the detection of changes in autoregressive time series. II. Resampling procedures. *J.Statist. Plann.Inference*, 138, 1697-1721. 2008.
- [65] Ilmakunnas, P., Tsurumi, H, Testing for parameter shifts in a regression model with two regimes of autocorrelated errors, *Economic studies Quarterly*, 135, 1, 46-56, 1984.

BIBLIOGRAPHIE

- [66] Inclin, C., and Tiao, G. C. Use of cumulative sums of squares for retrospective detection of changes of variance, *Journal of the American Statistical Association*, 89, 913-923. 1994.
- [67] V. Jandhyala, S. Fotopoulos, I. MacNeill, P. Liu, Inference for single and multiple change-points in time series, *Journal of Time Series Analysis* , doi : 10.1111/jtsa12035 (2013).
- [68] Kim, H.-J. Change-point detection for correlated observations, *Statistica Sinica* , 6, 275-287, 1996
- [69] A. Kander, S. Zacks. Test procedure for possible change in parameters of statistical distributions occurring at unknown time point, *Ann. Math. Statist*, 37 : 1196–1210, 1966.
- [70] Kezim, B. , Abdelli, Z., A Bayesian Analysis of a Structural Change in the Parameters of a Time Series, *Comm. Statist. Theory Meth*, 33 :8, 1863-1876, 2004.
- [71] D. Kim. A Bayesian significance test of the stationarity of regression parametres, *Biometrika* : 78 : 667–675, 1991.
- [72] Kim, H. and Siegmund, D. The likelihood ratio test for a change point in simple linear regression. *Biometrika* 76, 409-423. 1989.
- [73] H. Kim, Tests for a change-point in linear regression. Change-point Problems. Institute of Mathematical Statistics, Hayward, California, Lecture Notes and Monograph Series, 23 , 170-176, 1994.
- [74] Lai, T. L. and Xing, H. A simple Bayesian approach to multiple change-points. *Statistica Sinica* ,21, 539-569, 2011.
- [75] G.Lorden, Procedures for reacting to a change in distribution, *Ann. Math. Statist.* 42 6 , 1897-1908. 1971.
- [76] E. Marinari and G. Parisi. Simulated tempering : A new Monte Carlo schemes. *Euro-physics letters* , 19 :451-458, 1992.
- [77] Menzefricke, U. A Bayesian analysis of a change in the precision of a sequence of independent normal random variables at an unknown time point. *Appl. Stat.*, 19, 141-146. 1981.
- [78] Metropolis, N. et Ulam, S.. The Monte Carlo method. *J. American Statist. Assoc.*, 44, 335–341. 1949.

BIBLIOGRAPHIE

- [79] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller , and E. Teller. Equations of state calculations by fast computing machines. *J. Chem. Phys.*, 21 :1087-1091, 1953.
- [80] Ng, V.M., Bayesian analysis of linear models exhibiting changes in mean and precision at an unknown time point. *Comm. Statist. Theory Meth.* 19(1) :111-120. 1990.
- [81] Page, E. S. Continuous inspection schemes. *Biometrika* 41, 100-115. 1954.
- [82] Page, E.S. A test for a change in parameter occurring at an unknown point, *Biometrika*, 42, 523-527. 1955.
- [83] Pena,D. Influential observations in time series. *Journal of Business and Economic Statistics* 8, 235-41, 1990.
- [84] Pettitt, A.N., A non-parametric approach to the change-point problem. *Appl. Statist.* 28, 126-135. 1979.
- [85] D.J. Poirier . *The Econometrics of Structural Change*. North-Holland, Amsterdam, 1976.
- [86] Pollak, M., Optimal detection of a change in distribution. *Ann. Statist.* 13, 206–227. 1985.
- [87] E. Parent, J. Bernier *Le raisonnement bayésien, Modélisation et inférence* , Springer-Verlag France, Paris, 2007.
- [88] Priestley , M. B. *Spectral Analysis and Time Series*, London, Academic Press. 1981.
- [89] Quandt, R. E., The Estimation of the Parameters of a Linear Regression System Obeying Two Separate Regimes, *J. Amer. Statist. Assoc.*, 53 , 873-80, December 1958.
- [90] Quandt, R. E. Tests of the hypothesis that a linear regression system obeys two separate regimes. *J. Amer. Statist. Assoc.*, 55, 324-330, 1960.
- [91] Raiffa, H. et Schlaifer, R. *Applied Statistical decision theory*. Technical report, Division of Research, Graduate School of Business Administration, Harvard Univ. 1961.
- [92] Reinsel, G.C. *Elements of Multivariate Time Series Analysis*. New York : Springer, 1997.
- [93] Roberts, G.O., Smith, A.F.M. Simple Conditions for the Convergence of the Gibbs Sampler and the Metropolis-Hastings Algorithms. *Stochastic processes and their applications*, 49, 207-216. 1993.

BIBLIOGRAPHIE

- [94] C. Robert L'analyse statistique Bayésienne. Economica : Paris, 1992.
- [95] Robert, C. et Casella, G. Monte Carlo Statistical Methods. Springer- Verlag, New York. 2004.
- [96] Robert C., P., *Le choix bayésien, Principes et pratique*, Springer-Verlag France, Paris, 2006.
- [97] Robert C., P., *Méthodes de Monte-Carlo par chaînes de Markov*, ECONOMICA, 1996.
- [98] Salazar, D. Structural changes in time series models. *J. Econometrics* 10 :141-164. 1982.
- [99] Salazar, D., Broemeling, L. D., Chin, A. (1981). Parameter changes in a regression model with autocorrelated errors. *Comm. Statist. Theory Meth.* A10(17) :1751-175
- [100] Shewhart WA *Economic control of quality of manufactured product*. D. Van Nostrand Company, Inc., New York, 1931.
- [101] Sen, A.K. and Srivastava, M.S. On tests for detecting change in mean, *Ann. Statist.*, 3, 98-108, 1975.
- [102] Sen, A. and Srivastava, M. S. Some one-sided tests for change in level. *Technometrics*.17, 61-64. 1975.
- [103] Siegmund, D. Boundary Crossing probabilities and statistical applications. *Ann. Statist.* 14, 361-404, 1986.
- [104] Siegmund, D., Confidence sets in change point problem. *Int. Statist. Rev.* 56, 31-48. 1988.
- [105] Slama, A. A Bayesian significance test of change for correlated observations, *Discusiones Mathematicae, Probability and Statistics*, 34, 51–63, 2014. doi :10.7151/dmps.1169
- [106] Slama, A. A Bayesian significance test of change in the presence of single outlier, *Afrika statistika*, Volume 9(1), pp 685-694, 2014. Doi : <http://dx.doi.org/1016929.as.2014.685.62>
- [107] Slama, A., *Test de signification Bayésien pour la stationnarité des paramètres dans un modèle de rupture à erreurs autocorrélées*, Thèse de Magister, Institut de Mathématiques, USTHB, 1997.
- [108] Slama, A. et Saggou, H., A Bayesian analysis of a change in the parameters of autoregressive time series, *Communications in Statistics - Simulation and Computation*, 2016, <http://dx.doi.org/10.1080/03610918.2016.1222423>.

BIBLIOGRAPHIE

- [109] Tsay, R.S. "Outliers, level shift, and variance changes in time series", *J. Forecasting*, 7, 1-20, 1988.
- [110] Verdinelli, I. et Wasserman, L. , Bayesian analysis of outlier problems using the Gibbs sampler. *Statistics and Computing* 1, 105-117, 1991.
- [111] Yao, Y.C., On the asymptotic behavior of a class of nonparametric tests for a change-point problem. *Statist. Probab. Lett.* 9, 173-177. 1990.
- [112] Venkatesan, D. et Arumugam, P. , Bayesian Analysis of Structural Changes in Autoregressive Models, *American Journal of Mathematical and Management Sciences*, 27, Nos. 1 and 2, 153-162. 2007.
- [113] Von Neumann, J. et Morgenstern, O. *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, seconde édition. 1947.
- [114] Wang, J. et Zivot, E. A Bayesian time series model of multiple structural changes in level, trend and variance. *Journal of Business and Economic Statistics* , 18, 374–86. 2000.
- [115] Wasserman, L. *All of statistics : A concise course in statistical inference* . New York : Springer, (2004).
- [116] Worsley, K.J. On the likelihood ratio tests for a shift in location of normal population, *J. Amer. Statist. Ass.*, 74, 365-367, 1979.
- [117] Worsley, K. J. The power of likelihood ratio and cumulative sum tests for a change in a binomial probability. *Biometrika*, 70, 455-464. 1983.
- [118] Worsley, K.J., Confidence regions and tests for a change-point in a sequence of exponential family random variables. *Biometrika* 73, 91-104. 1986.
- [119] S.Zacks, Survey of classical and Bayesian approaches to the change-point problem, *in Recent advances in statistics*, Academic Press, N.Y., pp. 245-269. 1983.