

RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE  
MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITÉ MOULOUD MAMMERI, TIZI-OUZOU

FACULTÉ DES SCIENCES

DÉPARTEMENT DE MATHÉMATIQUES

**Thèse de Doctorat En-Sciences**

**SPÉCIALITÉ : MATHÉMATIQUES**

**OPTION : Probabilités et Statistique**

**Présentée par AIT MOHAMMED Noura**

Sujet

**Inférence dans un modèle autorégressif avec tendance et à contaminations  
multiples : Approche Bayésienne**

Devant le jury d'examen composé de :

Mr. FELLAG Hocine ;	Professeur ;	UMMTO ;	Président
Mme. GUERBYENNE Hafida ;	Professeur ;	USTHB ;	Directrice de thèse
Mme. ATIL Lynda ;	Maître de Conférence A ;	UMMTO ;	Examinatrice
Mme. BEDOUHENE Fazia ;	Professeur ;	UMMTO ;	Examinatrice
Mme. SADKI Ourida ;	Professeur ;	USTHB ;	Examinatrice
Mr. ZOUGAB Nabil ;	Professeur ;	Univ. Bejaia ;	Examineur

Soutenue : le 21/09/2023

# Remerciements

Je souhaite remercier en premier lieu ma directrice de thèse, Madame GUERBYENNE Hafida, professeur à l'USTHB. Je lui suis reconnaissante pour le temps conséquent qu'elle m'a accordé, ses qualités pédagogiques et scientifiques, sa franchise et sa sympathie. J'ai beaucoup appris à ses côtés et pour tout cela, je lui adresse ma profonde gratitude.

J'adresse de sincères remerciements à Monsieur FELLAG Hocine, professeur à l'UMMTO, pour avoir accepté de présider le jury de cette thèse.

J'exprime également mes remerciements à Madame SADKI Ourida professeur à l'USTHB, Madame BEDOUHENE Fazia professeur à l'UMMTO, Monsieur ZOUGAB Nabil professeur à l'université Abderrahmane Mira de Bejaia et Madame ATIL Lynda M.C.A. à l'UMMTO d'avoir accepté d'évaluer mon travail et d'appartenir au jury de cette thèse.

Je remercie particulièrement Madame BEDOUHENE Fazia et Monsieur ZOUGAB Nabil pour leur aide précieuse et leur disponibilité chaque fois que je les ai sollicités.

Je dédie ce travail à mes parents et toute ma famille et ma belle famille pour leur soutien et leurs encouragements.

Enfin, un grand merci à mon mari Farid, à mes deux enfants Alycia et Aris, témoins de mes joies, de mes fatigues et de mes enthousiasmes. Vous étiez mon soutien quotidien.

# Table des matières

<b>Introduction générale</b>	<b>5</b>
<b>1 Introduction à l'inférence bayésienne et quelques mesures de discrimination</b>	<b>12</b>
1.1 Introduction	12
1.2 Approche statistique classique	13
1.3 Approche bayésienne	13
1.3.1 Modèle bayésien	14
1.3.2 Distributions a priori	15
1.4 Inférence bayésienne	17
1.4.1 Fonctions de coût usuelles	17
1.4.2 Estimateur de Bayes	18
1.4.3 Intervalle de crédibilité bayésien	19
1.5 Test et régions de confiance	19
1.5.1 Région $\alpha$ -crédible Bayésienne et région HPD (highest posterior density)	20
1.5.2 Test d'hypothèses bayésien	20
1.6 Critères de discrimination	21
1.6.1 Le facteur de Bayes	21
1.6.2 Critère d'information de déviance (DIC)	21
1.6.3 Ordonnée prédictive conditionnelle (CPO)	23
1.6.4 Distribution prédictive	23

1.6.5	Les $p$ -values . . . . .	24
1.6.6	Les recommandations pour améliorer l'utilisation des $p$ -values . . . . .	24
1.7	Méthodes d'estimation bayésiennes . . . . .	26
1.7.1	Méthodes d'estimation Monte-Carlo . . . . .	27
1.7.2	Méthodes de Monte-Carlo par chaînes de Markov . . . . .	27
1.7.3	Pratique des méthodes MCMC . . . . .	28
1.7.4	Algorithme de Metropolis-Hastings . . . . .	29
1.7.5	L'échantillonneur de Gibbs . . . . .	30
<b>2</b>	<b>Procédures bayésiennes pour la détection d'observations aberrantes</b>	<b>33</b>
2.1	Introduction . . . . .	33
2.2	Valeurs aberrantes dans des modèles de séries chronologiques . . . . .	34
2.2.1	Les différents types de valeurs aberrantes . . . . .	34
2.2.2	Valeur aberrante additive et innovante dans les modèles autoregressifs d'ordre 1 . . . . .	36
2.2.3	Valeur aberrante additive et innovante dans les modèles de régression linéaire à erreurs AR(1) . . . . .	37
2.3	Détection d'observations aberrantes dans un modèle de régression linéaire univarié . . . . .	38
2.3.1	Procédures bayésiennes de détection utilisant uniquement un modèle de génération . . . . .	38
2.3.2	Procédures bayésiennes de détection utilisant un modèle alternatif de contamination . . . . .	40
2.3.3	Détection de valeurs aberrantes avec l'échantillonnage de Gibbs . . . . .	43
2.4	Test de signification bayésien . . . . .	49
<b>3</b>	<b>Inférence bayésienne dans un modèle autorégressif à contaminations multiples avec tendance</b>	<b>52</b>
3.1	Introduction . . . . .	52

	<b>4</b>
<hr/>	
3.2	Contamination multiple par des valeurs aberrantes additives . . . . . 53
3.2.1	Modèle et notations . . . . . 53
3.2.2	Analyse bayésienne . . . . . 54
3.3	Contamination multiple par des valeurs aberrantes innovantes . . . . . 59
3.3.1	Modèle et notations . . . . . 59
3.3.2	Analyse bayésienne . . . . . 59
3.4	Test de signification bayésien . . . . . 61
<b>4</b>	<b>Application numérique</b> . . . . . <b>64</b>
4.1	Introduction . . . . . 64
4.2	Etude de simulation . . . . . 64
4.2.1	Le modèle affecté par des valeurs aberrantes additives . . . . . 65
4.2.2	Le modèle affecté par des valeurs aberrantes innovantes . . . . . 76
4.3	Application sur des données réelles . . . . . 87
	<b>Conclusion générale</b> . . . . . <b>92</b>
	<b>Bibliographie</b> . . . . . <b>94</b>

# Introduction générale

L'identification des valeurs aberrantes joue un rôle important dans l'analyse statistique. Il est bien connu que les données statistiques, collectées pour être analysées et interprétées dans divers domaines, économie, industrie, finance, etc, contiennent le plus souvent une ou plusieurs observations qui ne ressemblent pas au reste des données. De telles observations sont appelées observations aberrantes. Elles apparaissent naturellement comme des événements rares, en raison d'une erreur humaine dans la collecte des données ou d'une erreur théorique dans la sélection du modèle. Elles peuvent également être dues à des événements qui sortent de l'ordinaire.

Le sujet est important car, en présence de valeurs aberrantes, le processus de modélisation d'une série temporelle donnée est affecté (Galeano et Peña, [28], chapitre 15). Ceci conduit à une conclusion erronée (Tsay [84]). Ces valeurs aberrantes peuvent avoir un impact sur les résultats de la méthodologie standard pour les séries temporelles. Par exemple, la présence de valeurs aberrantes peut entraîner une mauvaise spécification du modèle, des autocorrélations fausses ou des biais dans l'estimation des paramètres (voir, par exemple, Chen [25]; Tsay [84]).

La littérature sur les valeurs aberrantes est florissante; les chercheurs s'intéressent à ce domaine tant du point de vue de l'estimation des paramètres que de la détection des valeurs aberrantes. Il existe des ouvrages sur ce sujet. On peut citer Hawkins [44], Barnett et Lewis [8] et Rousseeuw et Leroy [70] qui se sont intéressés à la relation entre la régression et l'analyse des valeurs aberrantes. Agarwal [6] semble être le livre le plus récent. En adoptant l'approche classique, Fox [27] introduit le contaminant additif (*AO* pour additive outlier) qui affecte

une seule observation et le contaminant innovant (*IO* pour innovation outlier) qui affecte une seule innovation et leurs effets sur un modèle autorégressif d'ordre  $p$  ( $AR(p)$ ). Chen et Liu [24] et Chan [23], distinguent quatre types de valeurs aberrantes dans les données de séries temporelles : la valeur aberrante additive (*AO*), la valeur aberrante innovante (*IO*), le décalage de niveau (*LS*) et le changement temporaire (*TC*).

Supposons qu'un *AO* se présente dans une série temporelle  $\{y_t\}$  au temps  $t = k$  et que  $y_k$  soit l'observation contaminée. L'observation contaminée sera différente des observations initiales selon la règle suivante (voir, par exemple, Tsay [83]) :

$$y_t^* = \begin{cases} y_t & \text{si } t \neq k \\ y_t + \Delta & \text{si } t = k \end{cases}$$

Autrement dit, le choc causé par un *AO* n'affecte que l'observation au temps  $t = k$ , avec une amplitude  $\Delta$ , tandis que le reste n'est pas affecté.

Un *IO* représente un choc extraordinaire au temps  $t = k$  et influence non seulement l'observation  $y_k$ , mais aussi les observations suivantes  $y_{k+1}, y_{k+2}, \dots$  à travers la mémoire d'un système dynamique associé au *IO*. Si le processus est stationnaire, alors les effets des valeurs aberrantes s'éteindront exponentiellement. L'effet d'un *IO* à l'instant  $t = k$  peut être décrit comme suit

$$y_t^* = y_t + \eta_t$$

où  $\eta_t = \varepsilon_t + \Delta\delta_t$  et

$$\delta_t = \begin{cases} 1 & \text{si } t = k \\ 0 & \text{si } t \neq k \end{cases}$$

(voir, par exemple, Battaglia et Orfei [10]; Tsay [84]). Cette approche a été étendue aux modèle *ARMA* par Tsay ([83], [84]). Chang, Tiao et Chen [22] ont proposé une méthode itérative pour la détection d'un *AO* et d'un *IO* et ont estimé les paramètres des séries temporelles dans les modèles *ARIMA* en présence de valeurs aberrantes. Sánchez et Peña [72] ont proposé une méthode de détection multiple basée sur le maximum de vraisemblance pour ces modèles. Berkoun, Fellag et Zieliński [14] ont étudié l'effet d'un *AO* simple sur la robustesse d'un test de corrélation sérielle dans les modèles  $AR(1)$ .

Le problème de l'identification des valeurs aberrantes dans les séries temporelles non linéaires a été étudié par Battaglia et Orfei [10] pour détecter la présence de *AO* ou *IO*.

Si l'ensemble de données qu'on analyse contient plus d'une observation aberrante, le problème de la détection devient plus difficile et conduit à une détection erronée des valeurs aberrantes, à cause de l'effet de masque ("masking effect") et de l'effet d'entraînement ("swamping effect"). L'effet de masque se produit lorsqu'un sous-ensemble d'observations aberrantes n'est pas détecté à cause de sa proximité avec une autre observation aberrante.

L'effet d'entraînement se produit par contre lorsque de "bonnes" observations sont incorrectement identifiées comme aberrantes à cause de la présence d'un autre sous-ensemble d'observations aberrantes (voir Justel, Peña et Tsay [51]). Chen et Liu [24] ont proposé une méthode itérative pour atténuer l'effet de masque. Müller, Peña et Yohai [60] ont abordé le problème de l'existence de blocs de données aberrantes qui entraîne ces effets de masque.

Abd El-Sallam [2] a présenté une étude comparative de simulation de trois méthodes robustes pour l'estimateur des moindres carrés en présence de valeurs aberrantes innovantes et additives dans le cas de modèles autorégressifs.

Pour des données considérées, Moffat et Akpan [59] ont constaté que l'excès d'aplatissement de la loi de probabilité sous-jacente peut être expliqué par la présence de valeurs aberrantes. Des méthodes de détection des valeurs aberrantes ont été proposées. Novika, Siswadi et Bakhtiar [61] ont présenté une méthode de détection des valeurs aberrantes à l'aide de l'analyse biplot et la distance euclidienne. Schwertman, Owens et Adnan [74] ont utilisé le diagramme en boîtes pour identifier les valeurs aberrantes. Pérez, Molina and Peña [64] se sont intéressés à la détection des valeurs aberrantes et à la robustesse de l'estimation lorsque les données sont classées dans des groupes et suivent approximativement un modèle de régression linéaire avec un effet de groupe fixe.

L'un des premiers articles traitant des valeurs aberrantes d'un point de vue bayésien est celui de Box et Tiao [16] qui ont effectué une analyse bayésienne dans le cadre d'un modèle linéaire gaussien. Selon Box et Tiao [16], une bonne observation est normalement distribuée autour de sa moyenne avec une variance  $\sigma^2$  et une mauvaise observation est normale avec la



même moyenne mais une variance plus grande  $k^2\sigma^2$ .

Une approche bayésienne basée sur un modèle a été proposée par Abraham et Box [4] pour les modèles autorégressifs, par Smith et West [79] qui ont considéré des décisions séquentielles avec des systèmes linéaires dynamiques et par McCulloch et Tsay [56], et Barnett, Kohn et Sheather [9] qui ont utilisé des méthodes de Monte Carlo par chaîne de Markov.

Gupta et Singh [39] ont adopté l'approche bayésienne pour estimer les paramètres d'une distribution de Weibull en présence de valeurs aberrantes. Adoptant également la théorie statistique bayésienne, Zhang et Gui [92] ont proposé une méthode de détection simultanée de *AO* et *IO* dans un modèle de série temporelle ARMA.

Silva, Pereira et McCabe [75], ont proposé une approche bayésienne pour estimer la probabilité d'occurrence d'une valeur aberrante additive et son amplitude dans les modèles autorégressifs non gaussiens. Le contexte de travail des chercheurs permet une grande variété de modèles pour les données de comptage et les données positives.

Kim [53] a utilisé des  $p$ -values inconditionnelles, dans un cadre bayésien pour détecter ce qu'il appelle la non-stationnarité des paramètres des équations de régression. Slama et Saggou [78] ont étudié l'approche bayésienne utilisant des régions HPD crédibles et des  $p$ -values pour détecter le point de rupture dans les paramètres d'un modèle  $AR(p)$ . Toutefois, il est connu qu'un point de rupture n'est pas considéré comme une valeur aberrante.

Pour répondre à la question : Dans quelle mesure les résultats favorisent-elles l'hypothèse alternative par rapport à l'hypothèse nulle ? Les  $p$ -values elles-mêmes ne répondent pas directement à cette question (voir, par exemple, Benjamin et al. [11] ; Sellke, Bayarri et Berger, [73] ; Marden, [55]). Held et Ott [46] ont utilisé un facteur de Bayes (BF) (voir, Kass et Raftery [52]) pour montrer que l'évidence contre une hypothèse nulle est donnée par une petite valeur du facteur de Bayes. En supposant qu'une hypothèse alternative  $H_1$  a également été définie, le facteur de Bayes évalue directement si les données ont augmenté ou diminué les chances de  $H_0$ .

Slama [76] a étudié l'impact d'un *AO* sur la performance du test de signification bayésien basé sur les régions HPD crédibles (voir, par exemple, Geweke [35]), où la position et l'amplitude

de la contamination sont supposées être connues. Slama et Fellag [77] ont étendu ces résultats au cas où la position et l'amplitude de la contamination  $AO$  sont inconnues dans une série de données normales indépendantes.

Comme le montrent Justel et Peña [50], les méthodes robustes proposent un modèle pour la génération de toutes les données, y compris les éventuelles valeurs aberrantes. Ensuite, l'estimation est effectuée en utilisant tout l'échantillon, mais de manière à réduire le poids des valeurs aberrantes dans l'estimation. Il existe deux façons de procéder pour obtenir cet effet.

La première consiste à choisir une distribution à queue lourde. Plusieurs distributions à queue lourde ont été proposées. Box et Tiao [17] ont proposé la famille de distributions de puissance exponentielle. West [90] a suggéré d'utiliser des distributions à queue lourde qui peuvent être décomposées en un mélange de lois normales avec différentes échelles.

La seconde consiste à supposer un mélange de distributions : une distribution principale, qui génère les "bons" éléments, et une distribution alternative, qui est responsable des éléments aberrants. Rasmussen [67] a proposé un modèle de mélange gaussien infini qui résout le problème de la recherche du nombre exact de composantes.

Selon Tsay [83], il est nécessaire de développer des méthodes permettant de spécifier un (ou des) modèle(s) approprié(s) pour une série temporelle en présence de valeurs aberrantes.

Dans cette thèse, nous utilisons une méthode de génération des valeurs aberrantes basée sur le modèle. Nous généralisons les deux types de contamination  $AO$  et  $IO$  au cas multiple pour un  $AR(p)$  avec une tendance de régression. Nous adoptons l'approche bayésienne et nous utilisons la méthode du quasi maximum de vraisemblance (GQL) combinée avec une loi a priori non informative ou diffuse. Les lois a posteriori des paramètres du modèle sont obtenues. On utilise la méthode GQL gaussienne dans un cadre bayésien pour estimer nos paramètres. Notre objectif est de montrer la robustesse du GQL lorsque la loi de probabilité sous-jacente des erreurs n'est pas gaussienne.

Une étude de simulation intensive a été menée pour examiner la précision et la robustesse du Gibbs sampler lorsque la distribution du processus d'erreur est gaussienne standard comme

référence, contre, Student avec 5 degrés de liberté ou un mélange de deux lois normales. Une comparaison entre la méthode bayésienne et la méthode classique fondée sur le maximum de vraisemblance a été entreprise.

Un test basé sur la  $p$ -value inconditionnelle est proposé pour détecter la position de la contamination pour des valeurs aberrantes multiples  $AO$  et  $IO$ , quand le nombre de contaminants est connu. Les  $p$ -values sont des outils communément utilisés pour mesurer une évidence contre une hypothèse. Seulement, elles sont souvent considérées, à tort, comme une probabilité de rejet de l'hypothèse. Comme une alternative, le facteur de Bayes a été introduit par Kass et Raftery [52]. Pour plus de détail, voir Benjamin et al [11], Sellke, Bayarri et Berger [73] et Benjamin et Berger [12]. Les  $p$ -values sont utilisées pour calculer la borne supérieure du facteur de Bayes afin d'évaluer les chances de préférer l'hypothèse alternative à l'hypothèse nulle. Pour vérifier dans quelle mesure on favorise l'hypothèse alternative par rapport à l'hypothèse nulle, on a utilisé d'autres critères de discrimination Bayésien comme le critère d'information de déviance (DIC pour Deviance information criterion), l'ordonnée prédictive conditionnelle (CPO pour Conditional predictive ordinate) et le facteur de Bayes (BF pour Bayes factor).

La valeur prédite d'une observation future est donnée et comparée avec la vraie valeur pour les deux cas  $AO$  et  $IO$ . Une application sur des données réelles est faite, les données représentent les relevés horaires de viscosité d'un processus chimique (Box et Jenkins [15], page 529, série D) qui sont ajustées selon un modèle  $AR(1)$ .

Pour étudier l'effet des contaminants  $AO$  et  $IO$  dans cette série de données réelles, on la perturbe avec deux contaminants aux instants  $k_1$  et  $k_2$  d'amplitudes  $\Delta_{k_1}$  et  $\Delta_{k_2}$  respectivement.

Ce travail est structuré en quatre chapitres, que nous décrivons brièvement.

Le premier chapitre intitulé **Introduction à l'inférence bayésienne et quelques mesures de discrimination** est consacré aux rappels des définitions et des notions essentielles de l'inférence bayésienne. Des critères de discrimination sont présentés ainsi que la méthode

du Gibbs sampler.

Dans le chapitre 2 intitulé **Procédures bayésiennes pour la détection d'observations aberrantes**, on introduit d'abord les différents types de valeurs aberrantes notamment les valeurs aberrantes additives (*AO*) et les valeurs aberrantes innovantes (*IO*). Nous étudions aussi leurs impact sur les modèles de régression linéaire à erreur  $AR(1)$ .

Nous donnons ensuite les procédures bayésiennes de détection des valeurs aberrantes en utilisant

- Uniquement un modèle de génération.
- Un modèle alternatif de contamination.

Les chapitres trois et quatre comportent notre contribution.

Dans le chapitre 3 intitulé **Inférence bayésienne dans un modèle autorégressif à contaminations multiples avec tendance**, nous généralisons d'abord les contaminations *AO* et *IO* au cas multiple dans un modèle autorégressif d'ordre  $p$  avec une tendance de régression linéaire multiple. Ensuite, nous donnons les distributions a posteriori conditionnelles complètes des différents paramètres de notre modèle. A la fin, un test bayésien inconditionnel basé sur les  $p$ -values est utilisé pour détecter la position des valeurs aberrantes multiples *AO* et *IO* quand le nombre de contaminant est connu et lorsque le nombre de contaminant est inconnu, un test de détection des valeurs aberrantes est suggéré.

Dans le chapitre 4 intitulé **Application numérique**, nous faisons une étude de simulation intensive, pour comparer les deux estimateurs bayésien et classique, quand le bruit est gaussien comme référence contre une loi de Student à 5 degrés de liberté ou un mélange de deux lois normales. Une application sur des données réelles est faite.

Cette thèse est achevée par une conclusion générale, quelques perspectives de recherche ainsi qu'une bibliographie contenant 92 références.

# Chapitre 1

## Introduction à l'inférence bayésienne et quelques mesures de discrimination

### 1.1 Introduction

L'objet principal de la statistique est de faire, à partir d'observations d'un phénomène aléatoire, une inférence au sujet de la loi générant ces observations en vue d'analyser le phénomène ou de prévoir un événement futur.

Deux approches statistiques sont utilisées dans la littérature : approche non paramétrique et approche paramétrique. Dans la modélisation non paramétrique, on cherche à estimer la distribution de l'ensemble du phénomène, mettant en oeuvre l'estimation des fonctionnelles (densité, fonction de régression, etc.). Par opposition, le modèle paramétrique statistique consiste en l'observation d'une variable aléatoire  $y$ , d'une fonction de densité  $f(y|\theta)$ , où seul le paramètre  $\theta$ , appartenant à un sous-ensemble  $\Theta$  d'un espace vectoriel de dimension finie, n'est pas connu.

D'une manière plus générale, une fois le modèle construit, on cherche à établir une inférence sur  $\theta$ , c'est-à-dire à utiliser les observations  $y$  afin d'évaluer  $\theta$ , en vue d'une décision liée à ces paramètres.

Un aspect important de l'inférence statistique consiste à obtenir des estimations fiables

des caractéristiques d'une population. L'école classique (fréquentiste) et l'école Bayésienne trouvent dans ce domaine un terrain de désaccord, chacune a ses méthodes et arguments spécifiques qui défendent leurs causes.

## 1.2 Approche statistique classique

Dans l'approche fréquentiste, les données observées sont considérées comme des observations de variables aléatoires. Elles servent alors à faire porter l'inférence sur les paramètres ayant dirigé leur mécanisme de génération. Autrement dit, l'information provenant des données observées est l'unique source d'information. Dans cette approche, on écrit

$$l(\theta|y) = f(y|\theta)$$

en considérant  $l(\theta|y)$  comme une fonction de  $\theta$ , appelée fonction de vraisemblance. Par exemple, en estimation, lorsqu'on cherche la valeur  $\hat{\theta}_{MV}$  qui maximise  $l(\theta|y)$ , c'est l'estimation au sens du maximum de vraisemblance :

$$\hat{\theta}_{MV} = \arg \max_{\theta \in \Theta} \{l(\theta|y)\}$$

On utilise  $l(\theta|y)$  comme une fonction du paramètre, sachant les observations  $y$ .

## 1.3 Approche bayésienne

Dans l'approche bayésienne, l'idée de base consiste à traiter le paramètre inconnu  $\theta$  comme une variable aléatoire admettant une distribution de probabilité  $\pi(\theta)$  qui s'appelle distribution a priori.

L'objectif est donc d'utiliser cette information supplémentaire. Sachant que l'information contenue dans les observations  $y$  est contenue dans  $f(y|\theta)$  et l'information a priori sur  $\theta$  dans  $\pi(\theta)$ , on peut utiliser la règle de Bayes pour combiner ces deux types d'informations en définissant la distribution a posteriori par :

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

qui contiendra donc toute l'informations sur  $\theta$  et elle peut s'écrire

$$\pi(\theta|y) \propto l(\theta|y)\pi(\theta)$$

où  $\propto$  désigne le symbole de proportionnalité.

On remarque que l'état de connaissance a priori sur  $\theta$  traduite par la distribution a priori  $\pi(\theta)$  est transformée, après les observations  $y$ , en état de connaissance a posteriori par la distribution a posteriori  $\pi(\theta|y)$ .

### 1.3.1 Modèle bayésien

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

Étant donné la distribution d'échantillonnage,  $f(y|\theta)$ , une distribution a priori sur  $\theta$ ,  $\pi(\theta)$ , on peut construire :

1. La densité jointe de  $(\theta, y)$ ,

$$\Phi(\theta, y) = f(y|\theta)\pi(\theta)$$

2. La densité marginale de  $y$ ,

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

3. La distribution prédictive de  $x$ , où  $x \sim g(x|\theta, y)$ , obtenue par

$$g(x|y) = \int g(x|\theta, y)\pi(\theta|y)d\theta$$

4. La densité a posteriori de  $\theta$ , obtenue par formulation de Bayes,

$$\pi(\theta|y) = \frac{f(y|\theta)\pi(\theta)}{m(y)}$$

Autrement dit, la densité a posteriori représente une actualisation de l'information a priori au vu de l'information apportée par les observations.

### 1.3.2 Distributions a priori

Suivant la conception bayésienne, les distributions a priori permettent d'incorporer toutes les connaissances et opinions a priori sur les paramètres disponibles avant la collecte des données. Cette possibilité constitue à l'évidence un apport potentiel considérable.

Notons que par définition, qu'une classe de distributions a priori forme une famille conjuguée de distribution a priori, si la densité a posteriori,  $\pi(\theta|y)$ , appartient à la même classe que la distribution a priori quelque soit  $y$ . Cependant, la situation qui cause problème est celle où ces distributions a priori sont inconnues et/ou peu fiables. Dans le cadre bayésien, la réponse est apportée par l'utilisation des distributions a priori non-informative, conçues pour exprimer un état d'ignorance sur les paramètres, en ne privilégiant pas a priori des valeurs particulières pour  $\theta$  (Jeffreys [47]). Les distributions a posteriori correspondantes expriment l'apport propre des données. Cette approche de l'inférence bayésienne est maintenant reconnue comme standard. Pour plus de détails voir Lee (1989, chapitre 2).

Mais est-ce que l'utilisation d'une distribution non-informative est toujours possible ?

La réponse est oui dans le cas où l'espace des paramètres est discret et fini, par exemple  $\Theta = \{\theta_1, \dots, \theta_\eta\}$ , la distribution a priori non-informative est

$$\pi(\theta_i) = \frac{1}{\eta}, \quad i = 1, \dots, \eta$$

Dans le cas d'un espace de paramètres continu et borné, comme par exemple  $\Theta = [a, b]$ ,  $-\infty < a < b < +\infty$ , la distribution a priori est uniforme,

$$\pi(\theta) = \frac{1}{b-a}, \quad a < \theta < b,$$

donc il s'agit bien d'une distribution non-informative pour  $\theta$ . De plus, une des solutions proposées est d'utiliser une distribution invariante par toute transformation qui prend le nom de l'a priori de Jeffreys, qui est définie par

$$\pi(\theta) = [I(\theta)]^{\frac{1}{2}}, \tag{1.1}$$



où  $I(\theta)$  représente l'information de Fisher qui a la forme suivante

$$I(\theta) = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2} \log f(y|\theta)\right], \quad (1.2)$$

où  $f(y|\theta)$  est la fonction de vraisemblance. Dans le cas de la distribution a priori de Jeffreys la fonction de vraisemblance joue un rôle important dans la détermination de la distribution a priori (1.1). Pour plus de détails voir Carlin et Louis [20] (chapitre 2).

Box et Tiao [17] ont donné un exemple (pages 14-18) dans lequel la contribution de la distribution a priori à la détermination de la distribution a posteriori d'un paramètre de position  $\theta$  dépend de sa finesse (sharpness) ou de son aplatissement (flatness) par rapport à la finesse ou à l'aplatissement de la fonction de vraisemblance avec laquelle elle est combinée. Les distributions a priori sont influentes sur la distribution a posteriori pour la taille de l'échantillon  $n = 1$ . Mais pour  $n = 100$ , les lois a priori peuvent être relativement plates par rapport à la fonction de vraisemblance. Dans ce cas, les lois a priori ne sont donc pas influentes sur la distribution a posteriori correspondante au paramètre de position  $\theta$ . Pour  $n$  supérieur à 100 observations, les distributions a priori sont dominées par la vraisemblance. On peut également utiliser des lois a priori non informatives approximatives en supposant des lois normales avec de grandes variances pour les paramètres définis par des valeurs réelles ou des lois a priori gamma avec de grandes variances pour les paramètres définis par des valeurs positives.

Des distributions a priori informatives basées sur des avis d'experts peuvent être plus appropriées que des lois a priori vagues (voir, par exemple, Vokó et al. [87]). Albert et al. [7] se sont penchés sur le problème de la combinaison d'opinions de différents experts d'une manière explicitement basée sur un modèle pour construire une loi a priori subjective valide dans une approche statistique bayésienne. Lorsque très peu de données sont disponibles, les auteurs estiment qu'il est nécessaire d'utiliser des opinions d'experts pour compléter les informations fournies par d'autres sous-modèles bien informés.

## 1.4 Inférence bayésienne

Une inférence bayésienne est fondée sur la détermination rigoureuse de trois facteurs :

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

Notons que les critiques fréquentes de l'approche bayésienne se font souvent sur le point 2, alors que, conceptuellement, les points 1 et 3 se trouvent sur le même rang et que le point 3, qui est un point commun entre l'approche classique et l'approche bayésienne, est encore plus subtil. Le choix d'une fonction de coût dépend de l'application concernée. Lorsque le contexte ne conduit pas à la détermination de la fonction d'utilité qui permet d'en déduire la fonction de coût  $C$ , on peut avoir recours à des fonctions de coûts classiques, qui sont à la fois simples et bien étudiées.

### 1.4.1 Fonctions de coût usuelles

**Définition 1.2.** *On appelle coût toute fonction  $C$  de  $\Theta \times D$  dans  $[0, +\infty[$  ( $D$  est l'espace des décisions).*

$$C(\theta, d) = C(\theta, \delta(y))$$

avec  $d \in D$ . La fonction  $C$  évalue la pénalité résultante de l'emploi de la décision  $\delta \in D'$  ( $D'$  est l'espace des règles de décision) quand le paramètre prend la valeur  $\theta$ .

#### Le coût quadratique

**Définition 1.3.** *La fonction de coût quadratique est la fonction définie par*

$$C(\theta, d) = (\theta - d)^2$$

Le coût quadratique est particulièrement intéressant lorsque l'espace des paramètres est borné et le choix d'un coût plus subjectif est impossible. Mais il a donné lieu à de nombreuses

critiques, la plus fréquente étant qu'il pénalise trop fortement les grandes erreurs, à cause de sa convexité stricte.

### Le coût absolu

**Définition 1.4.** *La fonction de coût absolu est la fonction définie par*

$$C(\theta, d) = |\theta - d|$$

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

### 1.4.2 Estimateur de Bayes

L'estimateur de Bayes  $\delta^\pi$  de  $\theta \in \Theta$ , associé à la loi a priori  $\pi(\theta)$  et au coût quadratique  $C(\theta, d) = (\theta - d)^2$ , est la moyenne a posteriori

$$\delta^\pi(y) = \mathbb{E}^\pi[\theta|y] = \frac{\int_{\Theta} \theta f(y|\theta)\pi(\theta)d\theta}{\int_{\Theta} f(y|\theta)\pi(\theta)d\theta} \quad (1.3)$$

En effet, on a

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

qui est minimisé lorsque  $\delta^\pi(y) = \mathbb{E}^\pi[\theta|y]$ .

Un tel estimateur prend aussi la forme de l'extremum de la distribution a posteriori

$$\theta_{MAX} = \arg \max_{\theta \in \Theta} \{\pi(\theta|y)\} \quad (1.4)$$

Dans le cas du coût absolu, l'estimateur de Bayes est la médiane a posteriori.

Le calcul de  $\mathbb{E}^\pi[\theta|y]$  et  $\theta_{MAX}$  nécessitent de mettre en oeuvre des techniques d'intégration et/ou d'optimisation qui sont souvent compliquées et les calculs ne peuvent généralement être menés analytiquement que dans des cas simples, notamment pour des modèles linéaires et des distributions de probabilité standards.

### 1.4.3 Intervalle de crédibilité bayésien

L'intervalle de crédibilité est un reflet réel de la confiance que l'on peut avoir sur la valeur du paramètre concerné. Il met en évidence le point de vue original de l'approche bayésienne qui ne considère pas que le paramètre est une quantité fixe inconnue mais une variable aléatoire dont nous avons une appréciation plus ou moins exacte.

Par définition, un intervalle de crédibilité  $C$ , au niveau  $\alpha$ , est un intervalle tel que la probabilité du paramètre de lui appartenir selon la distribution a posteriori est de  $(1 - \alpha)$ .

$$P(\theta \in C|y) = 1 - \alpha$$

Il y a plusieurs manières de le construire ; par exemple, on peut prendre celui donné par les  $\alpha/2$  et  $(1 - \alpha/2)$  quantiles.

Quand on s'intéresse à plusieurs paramètres simultanément, on parle de région de crédibilité.

## 1.5 Test et régions de confiance

Soit un modèle statistique  $f(y|\theta)$  avec  $\theta \in \Theta$ . Etant donné un sous ensemble d'intérêt de  $\Theta$ ,  $\Theta_0$ , qui se réduit parfois à un singleton  $\{\theta_0\}$ , la question posée est : la vraie valeur du paramètre  $\theta$  appartient-elle à  $\Theta_0$  ? Ceci revient à tester l'hypothèse

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

appelée hypothèse nulle.

Lorsque  $\Theta_0 \neq \Theta_1$ , on définit l'hypothèse alternative pour laquelle on veut tester  $H_1$  par

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

Avec l'hypothèse

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

### 1.5.1 Région $\alpha$ -crédible Bayésienne et région HPD (highest posterior density)

**Définition 1.5.** Pour une loi a priori  $\pi$ , une région  $C$  de  $\Theta$  est dite  $\alpha$ -crédible pour  $\theta$  si

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

**Définition 1.6.** Une région HPD pour  $\theta$  est la région  $C^\pi$  définie par

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

où  $h_\alpha$  est la plus grande 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  $\alpha$ -crédible, par conséquent, peuvent être perçues comme des solutions optimales dans un cadre de décision.

### 1.5.2 Test d'hypothèses bayésien

Supposons que nous voulons choisir l'une des deux hypothèses  $H_0$  et  $H_1$ . Dans le cadre bayésien, nous supposons que nous connaissons les probabilités a priori de  $H_0$  et  $H_1$ . En d'autres termes, nous savons que  $P(H_0) = p_0$  et  $P(H_1) = p_1$  où  $p_0 + p_1 = 1$ . Supposons que les fonctions de vraisemblance  $f(y|H_0)$  et  $f(y|H_1)$  sont connues. En utilisant la règle de Bayes, nous pouvons obtenir les probabilités a posteriori de  $H_0$  et  $H_1$ .

$$P(H_0|y) = f(y|H_0)P(H_0)/f(y) \quad \text{et} \quad P(H_1|y) = f(y|H_1)P(H_1)/f(y)$$

On compare  $P(H_0|y)$  et  $P(H_1|y)$  et on accepte l'hypothèse dont la probabilité a posteriori est la plus élevée. C'est l'idée qui à l'origine du test du maximum a posteriori (MAP). Nous choisissons  $H_0$  si et seulement si  $P(H_0|y) \geq P(H_1|y)$ , c'est à dire si  $f(y|H_0)P(H_0) \geq f(y|H_1)P(H_1)$ .

## 1.6 Critères de discrimination

### 1.6.1 Le facteur de Bayes

**Définition 1.7.** *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  $\Theta_0$  par rapport à celle de l'ensemble  $\Theta_1$  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 souligne le rôle des observations. De fait, il peut être perçu comme un rapport de vraisemblance bayésien, car si  $\pi_0$  est la loi a priori sous  $H_0$  et  $\pi_1$  est la loi a priori sous  $H_1$ , le facteur de Bayes  $B_{01}^{\pi}$  peut s'écrire

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

Ce qui revient donc à remplacer les vraisemblances par des marginales sous les deux hypothèses (Berger [13]).

### 1.6.2 Critère d'information de déviance (DIC)

Une étape très importante à laquelle sont souvent confrontés les statisticiens est le choix des modèles. De nombreux critères d'information basés sur des pénalités ont été proposés pour choisir parmi un ensemble de modèles candidats. Dans le cadre de la statistique fréquentiste, le critère d'information le plus populaire est l'AIC (Critère d'information d'Akaike). L'un

des développements les plus importants pour la sélection de modèles dans la littérature bayésienne au cours des vingt dernières années est sans doute le critère d'information de déviance (DIC) de Spiegelhalter et al. [80].

Le DIC est considéré comme une version bayésienne de l'AIC. Comme l'AIC, il compare une mesure de l'adéquation du modèle à une mesure de la complexité et s'intéresse à la façon dont les données répétées prédisent les données observées. Cependant, contrairement à l'AIC, le DIC prend en compte les informations préalables (a priori).

Spiegelhalter et al. [80] proposent le critère DIC pour la sélection de modèles bayésiens. Le critère est basé sur la déviance

$$D(\underline{\theta}) = -2 \ln f(\underline{y}|\underline{\theta})$$

et se présente sous la forme

$$DIC = \overline{D(\underline{\theta})} + P_D$$

Le premier terme, interprété comme une mesure bayésienne de l'ajustement du modèle, est défini comme l'espérance a posteriori de la déviance, c'est-à-dire,

$$\overline{D(\underline{\theta})} = \mathbb{E}_{\underline{\theta}|\underline{y}} D(\underline{\theta}) = \mathbb{E}_{\underline{\theta}|\underline{y}} [-2 \ln f(\underline{y}|\underline{\theta})]$$

il est estimé par  $\widehat{D}_{avg}(\underline{y}) = \frac{1}{N} \sum_{i=1}^N D(\underline{\theta}^{(i)})$  et  $\underline{\theta}^{(i)}$ ,  $i = 1, \dots, N$  sont les simulations a posteriori. Plus le modèle s'adapte aux données, plus la valeur de la log-vraisemblance est grande, et donc plus la valeur de  $\overline{D(\underline{\theta})}$  est petite. Le second terme, utilisé pour mesurer la complexité du modèle et également connu sous le nom de "nombre effectif de paramètres", est défini comme la différence entre la moyenne a posteriori de la déviance et la déviance évaluée à la moyenne a posteriori des paramètres

$$P_D = \overline{D(\underline{\theta})} - D(\hat{\underline{\theta}}|\underline{y})$$

où  $\hat{\underline{\theta}}|\underline{y}$  est la moyenne a posteriori de  $\underline{\theta}$ .

Le DIC peut être réécrit sous la forme équivalente :

$$DIC = D(\hat{\underline{\theta}}|\underline{y}) + 2P_D$$

Le DIC peut être utilisé pour comparer différents modèles ainsi que différentes méthodes. Les modèles privilégiés ont de faibles valeurs de DIC. Il existe d'autres critères d'information sur la déviance (pour plus de détails, voir Young, Jun et Zeng [91]).

### 1.6.3 Ordonnée prédictive conditionnelle (CPO)

L'ordonnée prédictive conditionnelle (CPO) (Geisser [30], Pettit [65]) est définie comme suit

$$CPO_i = p(y_i|y_{(i)}) = \left[ \int \frac{1}{p(y_i|\theta)} p(\theta|\underline{y}) d\theta \right]^{-1}$$

où  $y_{(i)}$  sont les données omettant la  $i$ -ème observation et  $p(\cdot|y_{(i)})$  est la distribution prédictive d'une nouvelle observation étant donné  $y_{(i)}$ .

Par conséquent, le CPO peut être estimé en prenant l'inverse de la moyenne a posteriori de la fonction de densité inverse de  $y_i$  (la moyenne harmonique de la vraisemblance de  $y_i$ ).

Ainsi,

$$\widehat{CPO}_i = \left[ \frac{1}{S} \sum_{s=1}^S \frac{1}{p(y_i|\theta^{(s)})} \right]^{-1}$$

où les  $\theta^{(s)}$ ,  $s = 1, \dots, S$  sont des simulations a posteriori.

De faibles valeurs de CPO signalent la possibilité d'observations aberrantes, un effet de levier important et/ou observations influentes.

### 1.6.4 Distribution prédictive

La distribution prédictive (posterior predictive distribution) est définie comme la distribution d'une nouvelle observation  $Y_{n+1}$  sachant les observations de l'échantillon. Elle se calcule comme la distribution de  $Y_{n+1}$  sachant  $\underline{y}$ , marginalement par rapport à  $\theta$ . Sa densité  $f_{Y_{n+1}}(y|\underline{y})$  se calcule comme suit :

$$\begin{aligned} f_{Y_{n+1}}(y|\underline{y}) &= \int f_{Y_{n+1}}(y, \theta|\underline{y}) d\theta \\ &= \int f_{Y_{n+1}}(y|\theta, \underline{y}) \pi(\theta|\underline{y}) d\theta \\ &= \int f_{Y_{n+1}}(y|\theta) \pi(\theta|\underline{y}) d\theta \end{aligned}$$



On remarque le lien entre cette formule et celle de la distribution marginale :  $f_Y(y) = \int f_Y(y|\theta)\pi(\theta)d\theta$ , qui peut être vue comme un cas particulier de la distribution prédictive quand il n'y a pas d'information apportée par l'échantillon observé. On note également la différence avec l'approche fréquentiste où l'on estime d'abord  $\theta$  par  $\hat{\theta}$ , puis on remplace  $\theta$  par  $\hat{\theta}$  pour obtenir la distribution prédictive  $f_{Y_{n+1}}(y|\hat{\theta})$ .

### 1.6.5 Les $p$ -values

**Définition 1.8.** *La  $p$ -value associée à un test est le niveau de signification  $\alpha$  le plus petit pour lequel l'hypothèse nulle est rejetée.*

Une définition générale pour les hypothèses nulles ponctuelles (voir Thompson [82]) est qu'une  $p$ -value est une statistique admettant une loi uniforme sous l'hypothèse nulle ; se pose alors le difficile problème du choix de l'une de ces statistiques, comme d'ailleurs pour le test introduit dans la définition ci-dessus. En réalité, si un test de région critique  $R_\alpha$  est disponible pour tout niveau de signification  $\alpha$  et si ces régions sont imbriquées (c'est à dire si  $R_\alpha \subset R_\beta$  pour  $\beta > \alpha$ ), la procédure

$$p(x) = \inf\{\alpha; x \in R_\alpha\}$$

est distribuée selon une loi uniforme si  $\mathbb{E}_{\theta_0}[I_{R_\alpha}(x)] = \alpha$  (voir Goutis et al. [38]).

La  $p$ -value est l'outil le plus fréquemment utilisé pour mesurer les évidences contre une hypothèse ou un modèle posé. Malheureusement, elles sont souvent considérées à tort comme une erreur de probabilité de rejet d'une hypothèse (voir Marden [55]) ou comme une probabilité a posteriori que l'hypothèse soit vraie.

### 1.6.6 Les recommandations pour améliorer l'utilisation des $p$ -values

Les chercheurs utilisent couramment les  $p$ -values pour répondre à la question suivante : Dans quelle mesure les données sont-elles en faveur de l'hypothèse alternative par rapport à l'hypothèse nulle ? Les  $p$ -values ne répondent pas directement à cette question et sont souvent

mal interprétées (Benjamin et Berger [12]), ce qui conduit à surestimer les preuves contre l'hypothèse nulle. Cependant, même quand les  $p$ -values a posteriori ( $post_p$ ), " $post_p < 0,05$ ", il est fort possible que les  $p$ -values continuent à être utilisées pour évaluer la validité de l'hypothèse nulle. Si tel est le cas, le problème des erreurs d'interprétation persistera. Benjamin et Berger [12], recommandent trois pratiques qui aideraient les chercheurs à interpréter les  $p$ -values avec plus de précision. Chacune des trois pratiques recommandées implique l'interprétation des  $p$ -values à l'appui de la "borne du facteur de Bayes" correspondante, qui est la plus grande probabilité en faveur de l'hypothèse alternative par rapport à l'hypothèse nulle qui est en cohérence avec les données observées. La limite du facteur de Bayes indique généralement qu'une  $p$ -value donnée fournit des résultats plus faibles contre l'hypothèse nulle de ce qui est généralement supposé. Ils pensent donc que leurs recommandations peuvent nous éviter certaines interprétations erronées des  $p$ -values (voir, Sellke, Bayarri et Berger [73]).

La question qui se pose donc, est : à partir de quel valeur pourra-t-on dire que l'interprétation n'est plus erronée ?

### Convertir les $p$ -values en facteurs de Bayes

Sellke, Bayarri, et Berger [73], proposent de transformer la  $p$ -value en un facteur de Bayes. Malheureusement, il n'y a pas de correspondance unique entre les  $p$ -values et le facteur de Bayes car, contrairement au calcul de la  $p$ -value, le calcul du facteur de Bayes nécessite la spécification d'une hypothèse alternative (plus précisément, une distribution a priori pour les paramètres sous l'hypothèse alternative).

Heureusement, plusieurs méthodes (Vovk [89]; Johnson [48]; Sellke, Bayarri, et Berger [73]) ont été développées pour utiliser la  $p$ -value afin de calculer une limite supérieure du facteur de Bayes ( $BF$ ), appelée borne supérieure du facteur de Bayes, noté  $\bar{B}$ .

La borne supérieure du facteur de Bayes représente l'argument le plus fort en faveur de l'hypothèse alternative par rapport à l'hypothèse nulle : le  $BF$  le plus élevé possible est compatible avec la  $p$ -value observée. Comme le facteur de Bayes est justifiable à la fois pour

les bayésiens et les fréquentistes, la borne supérieure du facteur de Bayes l'est également. Vovk [89] donnent la formule la plus simple pour calculer une borne supérieure du facteur de Bayes

$$\bar{B} = \frac{1}{-ep \log p} \quad (1.5)$$

La borne supérieure du facteur de Bayes a été déterminée initialement par Vovk [89]. Cette valeur de  $\bar{B}$  peut être directement calculée comme une fonction simple de la p-value,  $p$ , en utilisant seulement le logarithme népérien et sa base, la constante  $e$ . Le calcul de  $\bar{B}$  ne nécessite pas la spécification d'une hypothèse alternative car il s'agit d'une limite supérieure pour une large classe d'hypothèses alternatives raisonnables.

Le tableau suivant donne la valeur de  $\bar{B}$  pour une large gamme de p-values.

p	0.1	0.05	0.01	0.005	0.001	0.0001	0.00001
$\bar{B}$	1.60	2.44	8.13	13.9	52.9	400	3226

Par exemple, une  $p$ -value de 0.05 correspond à un facteur de Bayes de 2.44 :1 au maximum. Cela signifie que les données impliquent des chances en faveur de l'hypothèse alternative par rapport à l'hypothèse nulle de 2.44 sur 1 au maximum.

## 1.7 Méthodes d'estimation bayésiennes

L'approche bayésienne nécessite le calcul d'estimateurs suivant une distribution a posteriori d'intérêt. Les estimateurs de type maximum a posteriori et moyenne a posteriori sont donc calculables par des techniques d'intégration et d'optimisation (voir les expressions (1.3) et (1.4)). La mise en pratique de l'approche bayésienne est toutefois problématique car il est souvent impossible de trouver une expression analytique directement exploitable de ces estimateurs. De plus, les dimensions des variables considérées rendent souvent très difficiles les calculs d'intégration et d'optimisation, même dans le cas où une forme analytique des estimateurs est disponible explicitement.

Des solutions classiques pour résoudre ces problèmes d'implémentation sont basées sur des

méthodes d'approximations asymptotiques et de calculs numériques. Dans ce qui suit, des méthodes de simulation de type Monte-Carlo sont utilisées afin de calculer les estimateurs d'intérêt.

### 1.7.1 Méthodes d'estimation Monte-Carlo

La méthode Monte-Carlo, a été introduite par Metropolis et Ulam [58] et Von Neumann et Morgenstern [88]. Cette méthode permet d'approximer l'intégrale de la forme

$$\int g(\theta)f(y|\theta)\pi(\theta)d\theta \quad (1.6)$$

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  $\pi(\theta)$ , alors la moyenne

$$\frac{1}{m} \sum_{i=1}^m g(\theta_i)f(y|\theta_i) \quad (1.7)$$

converge (presque sûrement) vers l'intégrale de (1.6) lorsque  $m$  tend vers  $+\infty$ , selon la loi des Grands Nombres.

### 1.7.2 Méthodes de Monte-Carlo par chaînes de Markov

Les techniques d'estimation Monte-Carlo permettent d'implémenter le calcul d'estimateurs issus de procédures d'estimation bayésienne. Toutefois, ces approches nécessitent une simulation des séquences d'échantillons suivant les distributions de probabilité d'intérêt. Or, ces distributions sont souvent multivariées, de type non standard. Il est donc nécessaire de considérer des méthodes de simulation statistique élaborées. Les schémas de simulation par chaînes de Markov sont ainsi considérés.

Un moyen de produire des échantillons suivant une loi d'intérêt donnée est d'utiliser une méthode de simulation par chaînes de Markov (voir, par exemple, Gilks, Richardson et Spiegelhalter [37] et Robert [68]). Le principe d'une telle méthode de simulation est de générer une suite d'échantillons  $(\theta_0, \theta_1, \dots, \theta_n, \dots)$  de manière récursive :  $\theta_n \sim q(\theta_n|\theta_{n-1})$  avec une loi

$q(\cdot|\cdot)$  aisément simulable, de telle manière que les échantillons  $\theta_n$  soient asymptotiquement distribués suivant la distribution d'intérêt :  $\theta_n \sim \pi(\theta|y)$  pour  $n$  assez grand. La loi de transition  $q(\cdot|\cdot)$  doit ainsi être choisie en fonction de la distribution d'intérêt  $\pi(\theta|y)$  et certaines conditions sont nécessaires pour qu'un tel schéma de simulation, produisant par définition des échantillons corrélés, puisse converger et générer des variables aléatoires suivant une loi invariante. Ce dernier point est important pour assurer que le calcul des estimateurs (1.3) et (1.4) soit possible. De telles approches, utilisant ces techniques d'estimation et de simulation, sont appelées méthodes de simulation Monte-Carlo par chaîne de Markov (MCMC) et permettent de traiter de nombreux problèmes d'estimation (voir, par exemple, Gilks, Richardson et Spiegelhalter, [37]) même pour des modèles non linéaires et/ou non-Gaussiens.

### 1.7.3 Pratique des méthodes MCMC

L'utilisation intensive des méthodes de simulation MCMC est due à leur adaptabilité pour une vaste classe de problèmes et de modèles mais également à l'amélioration des technologies informatiques. Les algorithmes sont faciles à implémenter et un simple ordinateur de bureau est suffisant pour mettre en oeuvre la réalisation des simulations. La part la plus importante du travail, et sans doute une des plus intéressantes, est l'élaboration du schéma de simulation pour un modèle donné : distribution a priori des paramètres du modèle distribution candidate pour l'algorithme de Metropolis-Hastings et distributions conditionnelles pour l'échantillonnage de Gibbs (Robert [69]) par exemple. Toutefois, un aspect important de l'implémentation de ce type de méthode est qu'il faut utiliser plusieurs réalisations ou simulations sur un ensemble de données pour obtenir des résultats statistiquement significatifs, étant donné la nature aléatoire des algorithmes de simulation. Les chaînes de Markov sont généralement très riches en informations statistiques pour calculer les estimateurs de type (1.3) et (1.4). Deux algorithmes MCMC sont présentés dans la section suivante : l'algorithme de Metropolis-Hastings et l'échantillonneur de Gibbs. Ces méthodes permettent de construire un noyau de transition en fonction d'une distribution d'intérêt donnée telle que les séquences d'échantillons simulés forment une chaîne de Markov convergente vers la

distribution stationnaire voulue.

### 1.7.4 Algorithme de Metropolis-Hastings

La méthode de Metropolis-Hastings est historiquement la première des méthodes MCMC. Elle a été développée par Metropolis et al. [57] au départ pour la physique particulaire, et généralisée par Hastings [43] dans un cadre 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  $(\theta^{(t)})_t$  comme suit :

- Itération 0 : Initialiser avec une valeur arbitraire  $\theta^{(0)}$
- Itération  $t$  : Mettre à jour  $\theta^{(t)}$  par  $\theta^{(t+1)}$  ( $t = 1, 2, \dots$ ), de la façon suivante

a- Générer  $\xi \sim q(\xi|\theta^{(t)}, y)$

b- Poser  $\varrho(\theta^{(t)}, \xi) = \min\left\{\frac{\pi(\xi|y)q(\theta^{(t)}|\xi, y)}{\pi(\theta^{(t)}|y)q(\xi|\theta^{(t)}, y)}, 1\right\}$

c- Prendre  $\theta^{(t+1)} = \begin{cases} \xi & \text{avec probabilité } \varrho(\theta^{(t)}, \xi) \\ \theta^{(t)} & \text{sinon} \end{cases}$

La convergence théorique de l'algorithme de Metropolis-Hastings est prouvée par Robert et Smith [68]. 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 1.1.**  $q$  est quelconque mais elle doit vérifier certaines conditions, telles que :

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

### 1.7.5 L'échantillonneur de Gibbs

L'approche de l'échantillonnage de Gibbs tire son nom des champs aléatoires de Gibbs, où elle a été utilisée pour la première fois par Geman et Geman [33] comme un moyen pour simuler des distributions complexes de grande dimension qui se posent dans le cadre de la restauration d'images et développée plus tard dans le cadre du calcul Bayésien par Gelfand et Smith [32].

L'échantillonneur de Gibbs est une méthode de Monte-carlo par chaîne de Markov (MCMC) permettant d'obtenir des échantillons à partir d'une distribution conjointe par échantillonnage itératif à partir des distribution conditionnelles complètes. Une bonne introduction à l'échantillonneur de Gibbs est donnée par Casella et George [19]. L'échantillonnage de Gibbs tire profit des structures hiérarchiques d'un modèle, i.e, lorsque celui-ci peut s'écrire sous la forme

$$\pi(\theta_1|y) = \int \pi_1(\theta_1|\theta_2, y)\pi_2(\theta_2|y)d\theta_2 \quad (1.8)$$

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

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

- Initialisation : Commencer par une valeur arbitraire  $(\theta_1^{(0)}, \dots, \theta_p^{(0)})$ .

- Itération  $t + 1$  : Pour  $(\theta_1^{(t)}, \dots, \theta_p^{(t)})$  donnée, générer

- $\theta_1^{(t+1)}$  selon la loi  $\pi(\theta_1|\theta_2^{(t)}, \theta_3^{(t)}, \dots, \theta_p^{(t)}, y)$ ,
- $\theta_2^{(t+1)}$  selon la loi  $\pi(\theta_2|\theta_1^{(t+1)}, \theta_3^{(t)}, \dots, \theta_p^{(t)}, y)$
- $\theta_3^{(t+1)}$  selon la loi  $\pi(\theta_3|\theta_1^{(t+1)}, \theta_2^{(t+1)}, \dots, \theta_p^{(t)}, y)$
- ...
- ...

—  $\theta_p^{(t+1)}$  selon la loi  $\pi(\theta_p|\theta_1^{(t+1)}, \theta_2^{(t+1)}, \dots, \theta_{p-1}^{(t+1)}, y)$ .

Après  $m$  itérations, nous obtenons  $\theta^{(m)} = (\theta_1^{(m)}, \theta_2^{(m)}, \dots, \theta_p^{(m)})$ . Sous certaines conditions de régularité sur la loi jointe et sur les lois conditionnelles, le vecteur généré  $\theta^{(m)}$  tend en loi vers le vecteur aléatoire de distribution jointe  $\pi(\theta_1, \dots, \theta_p|y)$  quand  $m$  tend vers l'infini (Geman et Geman [33]). En répétant de façon indépendante ce procédé, avec d'autres valeurs initiales,  $N$  fois, nous obtenons  $N$  répliquats indépendants de  $\theta$ . L'échantillon  $\{\theta^{(m)g}, g = 1, \dots, N\}$  ainsi obtenu constitue un échantillon simulé de la loi jointe a posteriori de  $\theta$  et pour tout  $i = 1, \dots, p$ , la moyenne a posteriori de  $\theta_i$  est calculé comme suit

$$\hat{\theta}_i = \frac{1}{N} \sum_{g=1}^N \theta_i^{(m)g}$$

C'est cette pratique que nous avons employée dans notre étude.

Alternativement, il est possible de générer une seule chaîne de longueur  $m + M$ , d'éliminer les  $m$  premiers tirages et de conserver les  $M$  derniers tirages pour construire l'échantillon  $\{\theta^{(j)}, j = 1, \dots, M\}$  (Gelfand et Smith [32]) et la moyenne a posteriori de  $\theta_i$  est

$$\hat{\theta}_i = \frac{1}{M} \sum_{j=1}^M \theta_i^{(j)}$$

**Théorème 1.9.** (Parent et Bernier. [62])

*La suite de vecteurs  $p$ -dimensionnels simulés  $\{\theta^{(m)g}, g = 1, \dots, N\}$  par l'algorithme précédent est une chaîne de Markov dont la distribution invariante, limite ergodique de la chaîne, est la distribution conjointe :*

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

L'échantillon simulé 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 approchée comme suit (voir Cassela et George [19]).

$$\pi(\theta_i|y) = \frac{1}{N} \sum_{g=1}^N \pi(\theta_i|\theta_1^{(m)g}, \dots, \theta_{i-1}^{(m)g}, \theta_{i+1}^{(m)g}, \dots, \theta_p^{(m)g}, y)$$



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 (Parent et Bernier [\[62\]](#)). Pour plus de détails sur l'application de l'algorithme de l'échantillonneur de Gibbs au calcul Bayésien général, on peut voir Gelfand et Smith [\[32\]](#).

# Chapitre 2

## Procédures bayésiennes pour la détection d'observations aberrantes

### 2.1 Introduction

Recueillir de manière optimale et analyser adéquatement les données sont les deux objectifs fondamentaux de la statistique. Pris dans leur ensemble, l'une de ses tâches les plus importantes est l'épuration de données : vérification, contrôle et traitement des données singulières. Il est en effet essentiel de travailler sur des données fiables. Le statisticien doit disposer d'outils et de pistes permettant de valider la qualité des prélèvements. Il s'agit de vérifier, par des essais statistiques, que les données ne sont pas contaminées, c'est-à-dire que les échantillons ne comportent pas de données douteuses ou erronées, que le système dans lequel on a fait les relevés est stable ou encore que la distribution dans laquelle on a prélevé l'échantillon correspond à une distribution attendue (exemple la loi gaussienne). La collecte des données est une étape importante de la démarche expérimentale et on doit y apporter le plus grand soin. Dans une série de mesures, il est cependant possible que l'on voie apparaître des valeurs s'écartant notablement des autres : ce sont les valeurs singulières. Mais on peut également trouver d'autres valeurs apparues dans l'échantillon à la suite d'une condition irrégulière (erreur de saisie ou d'expérimentation), des erreurs de gestion des données et des

événements inattendus ou inhabituels (par exemple, une catastrophe, une crise politique ou économique soudaine) : ce sont les valeurs erronées qui ne font pas partie de la population et on les appelle valeurs aberrantes.

## 2.2 Valeurs aberrantes dans des modèles de séries chronologiques

Le terme "outliers" fait référence aux observations qui s'écartent manifestement du reste des données. Les valeurs aberrantes sont également dites "observations contaminées", "observations discordantes" ou "valeurs extrêmes".

La présence de telles valeurs dans les séries temporelles, peuvent avoir un effet important sur le processus de construction du modèle pour une série chronologique donnée. D'abord, les observations aberrantes introduisent un biais dans les paramètres du modèle et faussent la puissance des tests statistiques basés sur des estimations biaisées.

Deuxièmement, les valeurs aberrantes peuvent augmenter les intervalles de confiance pour les paramètres du modèle.

Troisièmement, en conséquence des points précédents, les valeurs aberrantes influencent fortement les prédictions.

Ces problèmes ont été étudiés par Abraham et Chuang [5] qui ont découvert qu'un certain nombre de valeurs aberrantes suspectées dans les données de séries temporelles peuvent entraîner des résidus importants, affectant ainsi la spécification du modèle et l'estimation des paramètres.

### 2.2.1 Les différents types de valeurs aberrantes

Nous présentons dans cette section les quatre types de valeurs aberrantes habituellement considérées dans les modèles de séries temporelles (Chen et Liu [24] et Chan [23]) : valeur aberrante additive (*AO*), valeur aberrante innovante (*IO*), le décalage de niveau (*LS*) et le changement temporaire (*TC*).

Le type le plus courant est le  $AO$  qui est de nature déterministe et très probablement causé par un incident isolé, comme une erreur d'enregistrement ou une perturbation soudaine. Parmi les auteurs qui se sont penchés sur l'occurrence de  $AO$  dans leur étude on peut citer Chang et al. [22] et Battaglia et Orfei [10].

Supposons qu'un  $AO$  se produise dans la série temporelle  $y_t$  au temps  $t = k$  et que  $y_k$  soit l'observation contaminée. Suivant Tsay [83], l'observation contaminée différera des observations originales selon la règle suivante :

$$y_t^* = \begin{cases} y_t & si \quad t \neq k \\ y_t + \Delta & si \quad t = k \end{cases}$$

C'est-à-dire que le choc causé par un  $AO$  affecte l'observation au temps  $t = k$  avec une magnitude de  $\Delta$  tandis que le reste est non affecté.

$IO$  est un autre type de valeurs aberrantes que l'on trouve souvent dans les données de séries chronologiques. Il représente un choc extraordinaire au temps  $t = k$  et influence non seulement l'observation  $y_k$ , mais aussi les observations suivantes  $y_{k+1}, y_{k+2}, \dots$  à travers la mémoire du système dynamique associé à  $IO$ . Si le processus est stationnaire, alors les effets aberrants s'éteindront de manière exponentielle. Battaglia et Orfei [10] ont décrit l'effet d'un  $IO$  à l'instant  $t = k$  de la manière suivante

$$y_t^* = y_t + \eta_t$$

où  $\eta_t = \varepsilon_t + \Delta\delta_t$  et

$$\delta_t = \begin{cases} 1 & si \quad t = k \\ 0 & si \quad t \neq k \end{cases}$$

Le LS a un effet permanent sur la série, c'est-à-dire qu'il modifie le niveau de la série de manière permanente. Le TC est un outlier qui agit sur la série comme un LS, cependant l'effet de cet outlier se résorbe et la série retrouve son état normal plus ou moins rapidement, l'effet est donc temporaire sur la série.

## 2.2.2 Valeur aberrante additive et innovante dans les modèles autoregressifs d'ordre 1

Nous allons nous intéresser à des contaminations concernant les observations qui suivent un modèle autoregressif d'ordre 1 ( $AR(1)$ ).

Un modèle  $AR(1)$  sans aberration est donné par  $\mu_t = \phi\mu_{t-1} + \varepsilon_t$  où  $\varepsilon_t$  est un bruit blanc.

Soit  $\mu_t^*$  et  $\varepsilon_t^*$  l'observation et le résidu contaminés au temps  $t$ , lorsqu'une valeur aberrante existe dans l'ensemble des données.

Dans la section suivante, nous formulons les effets de  $AO$  et  $IO$  sur  $\mu_t$  et  $\varepsilon_t$ .

### 1. L'effet d'un $AO$ sur le modèle $AR(1)$

Dans un tel modèle une seule observation est affectée. Donc au lieu d'observer les  $\mu_t$ , nous observons le processus  $(\mu_t^*)_{1 \leq t \leq n}$  tel que :

$$\mu_t^* = \begin{cases} \mu_t & \text{si } t \neq k \\ \mu_t + \Delta & \text{si } t = k \end{cases}$$

$\Delta$  est l'amplitude de la contamination et  $k$  sa position.

L'effet de  $AO$  sur les résidus est résumé comme suit :

$$\varepsilon_t^* = \begin{cases} \varepsilon_t & \text{si } t \neq k, k+1 \\ \varepsilon_k + \Delta & \text{si } t = k \\ \varepsilon_{k+1} - \phi\Delta & \text{si } t = k+1 \end{cases}$$

### 2. L'effet d'un $IO$ sur le modèle $AR(1)$

Ce modèle correspond au cas où une "innovation" prend une valeur extrême et celle-ci n'affecte pas seulement une seule observation mais toutes les observations ultérieures. Dans ce cas on a

$$\varepsilon_t^* = \begin{cases} \varepsilon_t & \text{si } t \neq k \\ \varepsilon_t + \Delta & \text{si } t = k \end{cases}$$

et

$$\mu_t^* = \begin{cases} \mu_t & \text{si } t < k \\ \mu_t + \phi^{t-k}\Delta & \text{si } t = k, k+1, \dots, n \end{cases}$$

### 2.2.3 Valeur aberrante additive et innovante dans les modèles de régression linéaire à erreurs AR(1)

Nous considérons le modèle de régression linéaire à erreur AR(1) sans contamination, tel que

$$y_t = \beta x_t + \mu_t, \quad t = 1, \dots, n.$$

Soit  $y_t^*$  l'observation à l'instant  $t$  contaminée par un *AO* ou un *IO*.

#### 1. Le modèle sous contamination *AO*

$$y_t^* = \begin{cases} y_t & \text{si } t \neq k \\ y_k + \Delta & \text{si } t = k \end{cases} \quad (2.1)$$

Ce modèle peut se réécrire de la manière suivante

$$\begin{cases} y_t - \phi y_{t-1} = \beta(x_t - \phi x_{t-1}) + \varepsilon_t & \text{si } t \neq k, k+1 \\ y_k - \phi y_{k-1} = \beta(x_k - \phi x_{k-1}) + \varepsilon_k + \Delta \\ y_{k+1} - \phi y_k = \beta(x_{k+1} - \phi x_k) + \varepsilon_{k+1} - \phi \Delta \end{cases}$$

#### 2. Le modèle sous contamination *IO*

Le *IO* va affecter les observations  $y_t^*$  à l'instant  $t = k, k+1, k+2, \dots$  comme décrit ci-dessous

$$y_t^* = \begin{cases} y_t & \text{si } t < k \\ y_t + \phi^{t-k}\Delta & \text{si } t = k, k+1, \dots, n \end{cases} \quad (2.2)$$

Ce modèle peut se réécrire de la manière suivante

$$\begin{cases} y_t - \phi y_{t-1} = \beta(x_t - \phi x_{t-1}) + \varepsilon_t & \text{si } t \neq k \\ y_k - \phi y_{k-1} = \beta(x_k - \phi x_{k-1}) + \varepsilon_k + \Delta \end{cases}$$

Dans le chapitre 3, nous généralisons les deux modèles (2.1) et (2.2) au cas d'un modèle autorégressif d'ordre  $p$  avec une tendance de régression multiple sous contamination multiple.

## 2.3 Détection d'observations aberrantes dans un modèle de régression linéaire univarié

Le modèle de régression linéaire univarié est sans doute l'outil statistique le plus souvent mis en oeuvre. La régression linéaire multiple constitue la généralisation naturelle de la régression simple. Le modèle de régression multiple s'écrit matriciellement :

$$Y = X\beta + \varepsilon \quad (2.3)$$

où  $Y$  est un vecteur ( $n \times 1$ ) de variables aléatoires normales appelées variables réponses (ou dépendantes),  $X$  est une matrice ( $n \times p$ ) connue, appelée matrice des variables explicatives (ou indépendantes),  $\beta$  est le vecteur ( $p \times 1$ ) des paramètres et  $\varepsilon$  est le vecteur ( $n \times 1$ ) des erreurs indépendantes et identiquement distribuées selon une loi normale centrée. Ainsi définies, les données sont supposées provenir de l'observation d'un échantillon statistique de taille  $n$  de  $\mathbb{R}^{(p+1)}$  :  $(x_{i1}, \dots, x_{ip}, y_i)$ ,  $i = 1, \dots, n$ .

On distingue deux approches bayésiennes de détection d'observations aberrantes dans les modèles de régression. La première se restreint à postuler un modèle principal de génération de données et chercher alors des méthodes d'identification des observations aberrantes sans toutefois considérer de modèle alternatif (Gambino et Guttman [29] et Chaloner et Brant [21]). En revanche, la deuxième approche prend en compte un modèle alternatif pour caractériser la génération d'un sous-ensemble de contaminants dans l'échantillon de données. Les modèles alternatifs les plus couramment utilisés concernent les glissements de la moyenne (Guttman et al. [40]) et la dilatation de la variance (Box et Tiao [16] et Peña et Tiao [63]).

### 2.3.1 Procédures bayésiennes de détection utilisant uniquement un modèle de génération

Les principales méthodes d'identification d'observations singulières que l'on trouve dans cette catégorie utilisent une distribution prédictive pour la détection, ou encore des probabilités a posteriori de diverses composantes du vecteur des erreurs non observées. Pour la première

méthode, on désigne par  $I$  un ensemble de  $k$  entiers distincts choisis dans l'ensemble  $\{1, \dots, n\}$  de sorte que le vecteur  $Y$  de l'équation (2.3) peut être décomposé en  $Y = (Y'_I, Y'_{(I)})$ , où  $(I)$  signifie : omettre l'ensemble  $I$ . L'idée principale est de calculer la densité prédictive  $p(Y_I|Y_{(I)})$  pour détecter les observations singulières. La procédure de décision est alors la suivante : pour chaque groupe de  $k$  observations calculer  $p(Y_I|Y_{(I)})$ , ensuite déclarer le groupe pour lequel  $p(Y_I|Y_{(I)})$  est la plus petite comme étant composé d'observations singulières à effet de levier (Geisser, [31]).

La deuxième méthode concerne la détection d'observations aberrantes à effet de levier  $h_{ii}$  (avec de grands résidus) (Chaloner et Brant [21]). La technique de détection consiste à calculer, pour chaque résidu non observé  $\varepsilon_i = y_i - x'_i\beta$  de l'équation (2.3), la probabilité a posteriori que  $\varepsilon_i$  soit grand pour une constante  $c$  fixée :

$$p_i = P(|\varepsilon_i| > c\sigma|Y, X)$$

La constante  $c$  peut être choisie telle que le nombre d'observations singulières attendues a priori soit égal à un petit nombre  $\alpha$ .

Si on veut savoir quelles sont les observations aberrantes, on définit la probabilité a posteriori  $p_i$  pour que la  $i^{eme}$  observation soit aberrante, comme suit

$$p_i = Pr(|\varepsilon_i| > c\sigma|Y, X) = p_{i1} + p_{i2}$$

avec

$$p_{i1} = Pr(\varepsilon_i > c\sigma|Y, X)$$

et

$$p_{i2} = Pr(\varepsilon_i < -c\sigma|Y, X)$$



La probabilité a priori est

$$\begin{aligned} Pr(|\varepsilon_i| > c\sigma) &= Pr(\varepsilon_i > c\sigma) + Pr(\varepsilon_i < -c\sigma) \\ &= 2Pr(\varepsilon_i < -c\sigma) \\ &= 2\Phi(-c) \end{aligned}$$

$\Phi$  est la fonction de répartition d'une loi normale centrée réduite.

Les probabilités a posteriori  $p_i$  sont comparées à la probabilité a priori  $2\Phi(-c)$ . Toute observation avec une probabilité a posteriori d'être aberrante supérieure à la probabilité a priori  $2\Phi(-c)$  sera considérée comme suspecte.

La formule de  $p_i$  s'écrit comme suit (voir Chaloner et Brant [21])

$$p_i = 1 - \Phi(u_1) + \Phi(u_2)$$

où

$$u_j = \frac{r_i/\sqrt{h_{ii}} - (-1)^j c/\sqrt{h_{ii}}}{\sqrt{1 + \frac{1}{2}(n-p)^{-1}r_i^2/l_i}} \quad j = 1, 2,$$

Les  $r_i$  étant les résidus studentisés  $\hat{\varepsilon}_i/\sqrt{h_{ii}S^2}$  et  $l_i = \frac{h_{ii}}{1-h_{ii}}$ .

De la même manière, on peut calculer la probabilité a posteriori conjointe pour que la  $i^{eme}$  et la  $j^{eme}$  observations soient aberrantes.

### 2.3.2 Procédures bayésiennes de détection utilisant un modèle alternatif de contamination

Ces méthodes sont robustes, elles proposent un modèle pour la génération de toutes les données, y compris les éventuelles valeurs aberrantes. Ensuite, l'estimation est effectuée en utilisant l'ensemble de l'échantillon, mais de telle manière que le poids des valeurs aberrantes dans l'estimation soit réduit (Justel et Peña [50]).

Il existe deux façons d'obtenir cet effet. La première est de supposer une distribution à queue lourde (voir, par exemple, West [90]). La seconde consiste à supposer un mélange

de distributions : une distribution centrale, qui génère les bons points, et une distribution alternative, qui est responsable des points aberrants (voir, par exemple, Box et Tiao [16] et Abraham et Box [3]).

En pratique, les deux approches conduisent à des résultats similaires car lorsque le modèle central est contaminé par une distribution non spécifiée à queue lourde, cette propriété est transmise à la distribution finale du mélange.

La deuxième méthode consiste à accepter l'hypothèse de normalité pour la plupart des données et de supposer une distribution alternative pour les données aberrantes. Ensuite, le manque d'homogénéité dans l'échantillon est modélisé par un mélange de distributions. Dans cet échantillon, on suppose que les données peuvent provenir d'une distribution centrale avec une forte probabilité,  $(1 - \alpha)$ , et d'une distribution contaminée avec une faible probabilité,  $\alpha$ . Deux principaux outils d'identification des outliers sont utilisés :

1. La distribution a posteriori pour chaque élément issu de la distribution alternative, étant donné un mécanisme de génération particulier pour le reste de l'échantillon.
2. Le facteur de Bayes pour comparer les distributions prédictives avec différents modèles.

Les cas les plus étudiés sont ceux introduits par Tukey [85] de mélanges de lois normales pour la distribution des erreurs.

- Le premier est le modèle de contamination avec inflation de la variance. Box et Tiao [16] ont été les premiers auteurs à proposer une alternative robuste. Ils ont présenté l'idée de la modélisation des erreurs par un mélange de deux lois normales.

$$\varepsilon_i \sim \varphi N(0, k^2 \sigma^2) + (1 - \varphi) N(0, \sigma^2) \quad i = 1, \dots, n$$

où  $0 \leq \varphi \leq 1$  et  $k > 1$ .

- Le second est le modèle normal à décalage de la moyenne, de Guttman [41] et Abraham et Box [3], où les distributions des erreurs sont

$$\varepsilon_i \sim \varphi N(\lambda, \sigma^2) + (1 - \varphi) N(0, \sigma^2) \quad i = 1, \dots, n$$

où  $0 \leq \varphi \leq 1$  et  $\lambda > 0$ .

Introduisant un ensemble de variables de classification non observées  $\delta = (\delta_1, \delta_2, \dots, \delta_n)'$ ,

définies comme

$$\delta_i = \begin{cases} 1 & \text{si l'observation } y_i \text{ est générée par la distribution alternative} \\ 0 & \text{sinon} \end{cases}$$

L'estimation des paramètres du modèle dépend des valeurs de ces variables.

L'observation  $y_i$  sera qualifiée d'aberrante lorsque la probabilité a posteriori  $p_i = p(\delta_i = 1|\underline{y})$  est supérieure à 0.5. Ainsi  $\alpha$  est la probabilité a priori pour qu'une observation soit aberrante. Soit  $A(r)$  l'événement :  $r$  variables particulières sont égales à 1 et  $(n-r)$  restantes sont nulles, la distribution a posteriori de  $\beta$  est

$$p(\beta|\underline{y}) = \sum_r p(A(r)|\underline{y})p(\beta|A(r), \underline{y})$$

où les poids  $p(A(r)|\underline{y})$  sont les probabilités a posteriori de toutes les combinaisons possibles  $A(r)$  (voir Box et Tiao [16] pour le modèle de contamination avec inflation de la variance et Abraham et Box [3] pour le modèle de contamination avec décalage de la moyenne).

Pour identifier les valeurs aberrantes, nous pouvons utiliser les poids  $p(A(r)|\underline{y})$ . Dans le cas particulier d'une seule valeur aberrante dans l'échantillon, les probabilités sont  $p(A_i(1)|\underline{y})$ , où  $A_i(1)$  signifie que  $\delta_i = 1$ , c'est-à-dire que  $y_i$  est l'une des  $r$  observations contaminées (voir Eddy [26]).

Dans le cas général, la probabilité qu'une observation soit aberrante est donnée par  $p_i = \sum_r p(A_i(r)|\underline{y})$ . Cette probabilité nécessite de calculer les probabilités pour toutes les  $2^n$  combinaisons possibles. Nous verrons dans la prochaine sous-section une méthode de calcul de ces probabilités par les méthodes MCMC.

Dans la deuxième méthode, pour identifier les valeurs aberrantes on utilise les facteurs de Bayes. Avec le facteur de Bayes, et en appliquant la règle de Jeffreys (Jeffreys [47]), il est possible de comparer la distribution prédictive d'un modèle ne contenant qu'une seule valeur aberrante avec la distribution prédictive d'un modèle sans valeurs aberrantes. Dans ce cas, le facteur de Bayes peut être exprimé comme suit

$$B_{10}(i) = \frac{p(y|A_i(1))}{p(y|A_i(0))} \quad (2.4)$$

### 2.3.3 Détection de valeurs aberrantes avec l'échantillonnage de Gibbs

L'analyse bayésienne des problèmes de valeurs aberrantes à l'aide de l'échantillonneur de Gibbs a été initialement introduite par Verdinelli et Wasserman [86] pour des données i.i.d. Leur procédure a été généralisée par Justel et Peña [49] au cas des valeurs aberrantes dans les modèles de régression.

#### 1. Détection de valeurs aberrantes dans les modèles de régression linéaires

Justel et Peña [49] ont considéré le modèle de Box et Tiao [16] avec des lois a priori non informatives, en supposant que le paramètre  $\varphi$  est inconnu, de loi a priori  $Beta(\gamma_1, \gamma_2)$ .

$$y = X\beta + \epsilon, \quad (2.5)$$

où  $y$  est le vecteur des observations de dimension  $n$ ,  $X$  une matrice d'éléments fixes de dimension  $n \times p$ ,  $\beta$  le vecteur des paramètres du modèle, de dimension  $p$  et

$$\epsilon_i \sim \varphi N(0, k^2 \sigma^2) + (1 - \varphi) N(0, \sigma^2) \quad i = 1, \dots, n$$

Justel et Peña [49] ont calculé toutes les distributions a posteriori conditionnelles nécessaires et ont montré que la génération à partir de ces distributions est très facile en utilisant des générateurs de nombres aléatoires.

Les distributions a posteriori conditionnelles complètes des paramètres inconnus sont :

1. La distribution a posteriori conditionnelle du vecteur  $\beta$  est une loi normale  $N_p(\tilde{\beta}, \sigma^2(X'V^{-1}X)^{-1})$  où  $\tilde{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}\underline{y}$ ,  $V$  est la matrice diagonale d'éléments  $v_{ii} = k^2$  si  $\delta_i = 1$  et  $v_{ii} = 0$  sinon.
2. La distribution a posteriori conditionnelle de  $\sigma^2$  est une inverse gamma  $IG(\frac{n}{2}, \frac{\sum \mu_i^{*2}}{2})$ , où  $\mu_i^* = \frac{(y_i - x_i' \beta)}{(1 + \delta_i(k-1))}$
3. La distribution a posteriori conditionnelle de  $\alpha$  est une loi  $Beta(\gamma_1 + \sum \delta_i, \gamma_2 + n - \sum \delta_i)$

4. La probabilité a posteriori conditionnelle de  $\delta_i = 1$  est

$$p(\delta_i = 1 | \underline{y}, \beta, \sigma^2, \varphi) = \left(1 + \left(\frac{1 - \varphi}{\alpha}\right) F_{10}(i)\right)^{-1}$$

où  $F_{10}(i) = k \exp\left(\frac{-\mu_i^2}{2\Phi^{-1}\sigma^2}\right)$  est le facteur de Bayes et  $\Phi = 1 - k^2$ .

La probabilité a posteriori conditionnelle,  $p(\delta_i = 1 | \underline{y}, \beta, \sigma^2, \varphi)$ , pour que la  $i$ -ème observation soit aberrante dépend seulement du résidu normalisé  $\frac{\mu_i^2}{\sigma^2}$ . Si le résidu est petit,  $F_{10}$  sera grand et la probabilité  $p(\delta_i = 1 | \underline{y}, \beta, \sigma^2, \varphi)$  sera faible. L'inverse se produit lorsque  $\frac{\mu_i^2}{\sigma^2}$  est grand.

L'échantillonneur de Gibbs permet de calculer facilement la probabilité a posteriori conditionnelle que chaque observation soit aberrante.

## 2. Détection des valeurs aberrantes dans les modèles ARMA

Zhang et Gui [92] proposent une méthode pour détecter simultanément tous les types de valeurs aberrantes, notamment pour détecter  $AO$  et  $IO$ , dans le modèle  $ARMA$  construit sur la base des variables de classification des valeurs aberrantes, et une méthode de détection des valeurs aberrantes est proposée en appliquant le principe du test d'hypothèse bayésien. Soit  $z_t$  le processus qui suit un modèle  $ARMA(p, q)$

$$\phi(B)z_t = \theta(B)\varepsilon_t \quad (2.6)$$

avec  $\varepsilon_t$  i.i.d  $\sim N(0, \sigma^2)$ ,  $\phi(B) = I - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ ,  $\theta(B) = I - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$ ,  $p > q$  et  $B$  est l'opérateur de retard, tel que,  $B^k z_t = z_{t-k}$ ,  $k = 1, 2, \dots$ . Pour s'assurer que le modèle  $ARMA$  est stationnaire et inversible, supposons que tous les zéros de  $\phi(B)$  et  $\theta(B)$  sont à l'extérieur du cercle unitaire.

Sur la base de la définition des  $AO$  et des  $IO$ , l'observation  $y_t$  qui est affectée par un  $AO$  ou un  $IO$  ou par les deux simultanément peut être écrite avec les variables de classification comme suit

$$y_t = z_t + \omega_t^{AO} \delta_t^{AO} + \phi^{-1}(B)\theta(B)\omega_t^{IO} \delta_t^{IO} \quad (2.7)$$

A partir de ce qui précède, un modèle de détection simultanée de  $AO$  et  $IO$  dans  $ARMA(p, q)$  est construit comme suit

$$\begin{cases} y_t = x_t + \omega_t^{AO} \delta_t^{AO} \\ x_t = \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \\ \varepsilon_t = a_t + \omega_t^{IO} \delta_t^{IO} \\ a_t \quad i.i.d \sim N(0, \sigma^2) \end{cases} \quad (2.8)$$

où  $x_t$  est l'observation affectée par le  $IO$ ,  $\delta_t^{AO}$  est la variable de classification du  $AO$  qui suit la loi de Bernoulli. Si  $\delta_t^{AO} = 1$ , l'observation  $y_t$  est un  $AO$  d'amplitude "anormale"  $\omega_t^{AO}$ ; si  $\delta_t^{AO} = 0$ , l'observation  $y_t$  n'est pas un  $AO$ .  $\delta_t^{IO}$  est la variable de classification du  $IO$  qui suit la loi de Bernoulli. Si  $\delta_t^{IO} = 1$ , l'observation  $y_t$  est un  $IO$  d'amplitude  $\omega_t^{IO}$ ; si  $\delta_t^{IO} = 0$ , l'observation  $y_t$  n'est pas un  $IO$ .

Supposons qu'il existe  $n$  observations de la série temporelle  $y_1, \dots, y_n$  et que les  $p$  premières observations  $y_1, \dots, y_p$ ,  $p \ll n$  ne sont pas aberrantes.

Si nous voulons juger si  $y_j$  est un  $AO$  ou un  $IO$  ou non, il est nécessaire de tester une hypothèse multiple :

$$\begin{cases} H_{1,j} : \delta_j^{AO} = 0, \delta_j^{IO} = 0 & H_{2,j} : \delta_j^{AO} = 1, \delta_j^{IO} = 0 \\ H_{3,j} : \delta_j^{AO} = 1, \delta_j^{IO} = 1 & H_{4,j} : \delta_j^{AO} = 0, \delta_j^{IO} = 1 \end{cases} \quad (2.9)$$

Si l'hypothèse  $H_{1,j}$  est accepté, on peut conclure que  $y_j$  n'est pas un  $AO$  ni un  $IO$ . Si  $H_{2,j}$  est accepté,  $y_j$  est un  $AO$  mais n'est pas un  $IO$ . Si  $H_{3,j}$  est accepté signifie que  $y_j$  est à la fois un  $AO$  et un  $IO$ . En outre si  $H_{4,j}$  est accepté,  $y_j$  est un  $IO$  mais n'est pas un  $AO$ .

Selon le principe des tests d'hypothèses bayésiens, nous devons choisir une distribution a priori adéquate pour chaque paramètre inconnu et calculer la probabilité a posteriori de chaque hypothèse de la manière suivante :

$$P(H_{1,j}|Y), P(H_{2,j}|Y), P(H_{3,j}|Y), P(H_{4,j}|Y)$$

où  $Y = (y_{p+1}, \dots, y_n)'$ .

Si  $P(H_{i,j}|Y) = \max \{P(H_{1,j}|Y), P(H_{2,j}|Y), P(H_{3,j}|Y), P(H_{4,j}|Y)\}$ , alors l'hypothèse  $H_{i,j}$

sera accepté, ce qui signifie que  $y_j$  peut être identifiée comme une observation normale ou une sorte de valeur aberrante. A partir du modèle (2.8), on obtient la fonction de densité de probabilité de  $y_t$ , pour  $t = p + 1, \dots, n$ .

$$p(y_t | \Phi, \Theta, \sigma^2, \omega_t^{AO}, \delta_t^{AO}, \omega_t^{IO}, \delta_t^{IO}) = (2\pi)^{-\frac{1}{2}} \sigma^{-1} \exp \left\{ -\frac{1}{2\sigma^2} \left( y_t - \omega_t^{AO} \delta_t^{AO} - \omega_t^{IO} \delta_t^{IO} - \Phi' X_{(-t)} + \Theta' \varepsilon_{(-t)} \right)^2 \right\}, \quad (2.10)$$

avec  $X_{(-t)} = (x_{t-1}, \dots, x_{t-p})'$  et  $\varepsilon_{(-t)} = (\varepsilon_{t-1}, \dots, \varepsilon_{t-p})'$ . Enfin on obtient la fonction de densité de probabilité conjointe de  $Y$  comme suit

$$\begin{aligned} p(Y | x_{-(p+1)}, \varepsilon_{(-t)}, \Phi, \Theta, \sigma^2, \omega_t^{AO}, \delta_t^{AO}, \omega_t^{IO}, \delta_t^{IO}) &= \prod_{t=p+1}^n p(y_t | \Phi, \Theta, \sigma^2, \omega_t^{AO}, \delta_t^{AO}, \omega_t^{IO}, \delta_t^{IO}) \quad (2.11) \\ &= (2\pi)^{-\frac{n-p}{2}} \sigma^{-(n-p)} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{t=p+1}^n \left( y_t - \omega_t^{AO} \delta_t^{AO} - \omega_t^{IO} \delta_t^{IO} - \Phi' X_{(-t)} + \Theta' \varepsilon_{(-t)} \right)^2 \right\} \end{aligned}$$

### Les distributions a posteriori des paramètres inconnus

On suppose qu'il y a une petite probabilité a priori  $\alpha$  pour chaque observation  $y_j$  ( $j = p + 1, \dots, n$ ) d'être un  $AO$  ou un  $IO$ , qui se traduit par  $p(\delta_j^{AO} = 1) = p(\delta_j^{IO} = 1) = \alpha$ . Les probabilités a priori pour les paramètres inconnus sont  $\Phi \sim N_p(\Phi_0, V^{-1})$ ,  $\Theta \sim N_q(\Theta_0, W^{-1})$ ,  $\omega_j^{AO} \sim N(\mu_1, \xi_1^2)$ ,  $\omega_j^{IO} \sim N(\mu_2, \xi_2^2)$ ,  $\delta_j^{AO} \sim b(1, \alpha)$ , tel que  $p(\delta_j^{AO}) = \alpha^{\delta_j^{AO}} (1 - \alpha)^{1 - \delta_j^{AO}}$ ,  $\delta_j^{IO} \sim \text{Bernoulli}(\alpha)$ , tel que  $p(\delta_j^{IO}) = \alpha^{\delta_j^{IO}} (1 - \alpha)^{1 - \delta_j^{IO}}$ .

$\sigma^2 \sim IG(\frac{\nu}{2}, \frac{\nu\lambda}{2})$ , avec  $\Phi_0, V^{-1}, \Theta_0, W^{-1}, \mu_1, \xi_1^2, \mu_2, \xi_2^2, \alpha, \nu$  et  $\lambda$  sont des hyperparamètres connus.

1. La distribution a posteriori de  $\Phi$  est  $\Phi | Y, \Theta, \sigma^2, \omega_t^{AO}, \delta_t^{AO}, \omega_t^{IO}, \delta_t^{IO} \sim N(\hat{\Phi}_0, \hat{V}^{-1})$ , avec  $\hat{V}^{-1} = \left[ \frac{1}{\sigma^2} \sum_{t=p+1}^n X_{(-t)} X_{(-t)}' + V \right]^{-1}$  et  $\hat{\Phi}_0 = \hat{V}^{-1} \left[ \frac{1}{\sigma^2} \sum_{t=p+1}^n X_{(-t)} (y_t - \omega_t^{AO} \delta_t^{AO} - \omega_t^{IO} \delta_t^{IO} - \Phi' X_{(-t)} + \Theta' \varepsilon_{(-t)}) + V \Phi_0 \right]$
2. La distribution a posteriori de  $\sigma^2$  est  $\sigma^2 | Y, \Phi, \Theta, \omega_t^{AO}, \delta_t^{AO}, \omega_t^{IO}, \delta_t^{IO} \sim IG(a, b)$  avec  $a = \frac{n-p+\nu}{2}$  et  $b = \frac{1}{2} \left[ \sum_{t=p+1}^n (y_t - \omega_t^{AO} \delta_t^{AO} - \omega_t^{IO} \delta_t^{IO} - \Phi' X_{(-t)} + \Theta' \varepsilon_{(-t)})^2 + \nu\lambda \right]$
3. La distribution a posteriori de  $\Theta$  est  $\Theta | Y, \Phi, \sigma^2, \omega_t^{AO}, \delta_t^{AO}, \omega_t^{IO}, \delta_t^{IO} \sim N(\hat{\Theta}_0, \hat{W}^{-1})$ , avec  $\hat{W}^{-1} = \left[ \frac{1}{\sigma^2} \sum_{t=p+1}^n \varepsilon_{(-t)} \varepsilon_{(-t)}' + W \right]^{-1}$  et  $\hat{\Theta}_0 = -\hat{W}^{-1} \left[ \frac{1}{\sigma^2} \sum_{t=p+1}^n \varepsilon_{(-t)} (y_t - \omega_t^{AO} \delta_t^{AO} - \omega_t^{IO} \delta_t^{IO} - \Phi' X_{(-t)}) - W \Theta_0 \right]$ .

4. Les distributions a posteriori de  $\delta_j^{AO}$  et  $\delta_j^{IO}$  sont

$$P_{1j} = P(\delta_j^{AO} = 0, \delta_j^{IO} = 0 | Y, \Phi, \Theta, \sigma^2, \omega^{AO}, \delta_{-j}^{AO}, \omega^{IO}, \delta_{-j}^{IO}) = \alpha^2 \exp\left\{-\frac{1}{2\sigma^2} \sum_{t=j}^T (y_t - \omega_t^{AO} \delta_t^{AO} - \omega_t^{IO} \delta_t^{IO} - \Phi' X_{(-t)} + \Theta' \varepsilon_{(-t)})^2\right\}$$

$$P_{2j} = P(\delta_j^{AO} = 1, \delta_j^{IO} = 0 | Y, \Phi, \Theta, \sigma^2, \omega^{AO}, \delta_{-j}^{AO}, \omega^{IO}, \delta_{-j}^{IO}) = \alpha(1 - \alpha) \exp\left\{-\frac{1}{2\sigma^2} \sum_{t=j}^T (y_t - \omega_t^{AO} \delta_t^{AO} - \omega_t^{IO} \delta_t^{IO} - \Phi' X_{(-t)} + \Theta' \varepsilon_{(-t)})^2\right\}$$

$$P_{3j} = P(\delta_j^{AO} = 1, \delta_j^{IO} = 1 | Y, \Phi, \Theta, \sigma^2, \omega^{AO}, \delta_{-j}^{AO}, \omega^{IO}, \delta_{-j}^{IO}) = (1 - \alpha)^2 \exp\left\{-\frac{1}{2\sigma^2} \sum_{t=j}^T (y_t - \omega_t^{AO} \delta_t^{AO} - \omega_t^{IO} \delta_t^{IO} - \Phi' X_{(-t)} + \Theta' \varepsilon_{(-t)})^2\right\}$$

$$P_{4j} = P(\delta_j^{AO} = 0, \delta_j^{IO} = 1 | Y, \Phi, \Theta, \sigma^2, \omega^{AO}, \delta_{-j}^{AO}, \omega^{IO}, \delta_{-j}^{IO}) = \alpha(1 - \alpha) \exp\left\{-\frac{1}{2\sigma^2} \sum_{t=j}^T (y_t - \omega_t^{AO} \delta_t^{AO} - \omega_t^{IO} \delta_t^{IO} - \Phi' X_{(-t)} + \Theta' \varepsilon_{(-t)})^2\right\}$$

où  $T = \min(n, j + p)$ .

5. La distribution a posteriori conditionnelle de  $\omega_j^{AO}$  est

$$\omega_j^{AO} | Y, \Phi, \Theta, \sigma^2, \omega_{-j}^{AO}, \delta^{AO}, \omega^{IO}, \delta^{IO} \sim N(\hat{\omega}_j^{AO}, (\hat{\xi}_j^2)^{AO}) \quad (2.12)$$

où

$$(\hat{\xi}_j^2)^{AO} = \left[ \frac{(\delta_j^{AO})^2}{\sigma^2} (1 + \phi_1^2 + \dots + \phi_{N-j}^2) + \frac{1}{\xi_1^2} \right]^{-1}$$

et

$$\hat{\omega}_j^{AO} = (\hat{\xi}_j^2)^{AO} \left[ \frac{\delta_j^{AO}}{\sigma^2} (y_j - \omega_j^{IO} \delta_j^{IO} - \Phi' X_{(-j)} + \Theta' \varepsilon_{(-j)}) - \sum_{m=j+1}^T \frac{\phi_{m-j}}{\sigma^2} \delta_j^{AO} (y_m - \omega_m^{AO} \delta_m^{AO} - \omega_m^{IO} \delta_m^{IO} - \phi_1 x_{m-1} - \dots - \phi_{m-j} y_j - \dots - \phi_p x_{m-p} + \Theta' \varepsilon_{(-m)} + \frac{u_1}{\xi_1^2}) \right]$$



6. La distribution a posteriori conditionnelle de  $\omega_j^{IO}$  est

$$\omega_j^{IO} | Y, \Phi, \Theta, \sigma^2, \omega_{-j}^{IO}, \delta^{IO}, \omega^{AO}, \delta^{AO} \sim N(\hat{\omega}_j^{IO}, (\hat{\xi}_j^2)^{IO}) \quad (2.13)$$

où

$$(\hat{\xi}_j^2)^{IO} = \left[ \frac{(\delta_j^{IO})^2}{\sigma^2} + \frac{1}{\xi_2^2} \right]^{-1}$$

et

$$\hat{\omega}_j^{IO} = (\hat{\xi}_j^2)^{IO} \left[ \frac{\delta_j^{IO}}{\sigma^2} (y_j - \omega_j^{AO} \delta_j^{AO} - \Phi' X_{(-j)} + \Theta' \varepsilon_{(-j)}) + \frac{u_2}{\xi_2^2} \right]$$

### L'échantillonneur de Gibbs pour la détection des valeurs aberrantes additives et innovantes

Lorsque les ordres  $p$  et  $q$  du modèle *ARMA* sont connus, mais les paramètres sont inconnus, la probabilité a posteriori ne peut être calculée directement par la formule de Bayes. Ainsi, nous pouvons utiliser l'échantillonnage de Gibbs basé sur les distributions a posteriori conditionnelles des paramètres inconnus pour estimer  $P(H_{i,j}|Y)$ .

Une procédure de détection simultanée de *AO* et de *IO* dans une série temporelle *ARMA* est proposée sur la base de l'échantillonnage de Gibbs comme suit

- **Étape 1 :** Choisir les hyperparamètres  $\Phi_0, V^{-1}, \Theta_0, W^{-1}, u_1, \xi_1^2, u_2, \xi_2^2, \alpha, \nu$  et  $\lambda$  pour les distributions a priori inconnues des paramètres du modèle *ARMA*.
- **Étape 2** Choisir les valeurs initiales des paramètres  $\Phi^{(0)}, \Theta^{(0)}, (\sigma^2)^{(0)}, (\omega^{AO})^{(0)}, (\delta^{AO})^{(0)}, (\omega^{IO})^{(0)}$  et  $(\delta^{IO})^{(0)}$  pour l'échantillonneur de Gibbs.
- **Étape 3** Implémenter l'échantillonnage de Gibbs pour obtenir les échantillons de Gibbs. Supposons que le  $(k-1)$ -ième échantillon  $(\Phi^{(k-1)}, \Theta^{(k-1)}, (\sigma^2)^{(k-1)}, (\omega^{AO})^{(k-1)}, (\delta^{AO})^{(k-1)}, (\omega^{IO})^{(k-1)}, (\delta^{IO})^{(k-1)})$  a été obtenu. Alors, le  $k$ -ième échantillon peut être obtenu par la procédure itérative de l'algorithme de Gibbs.
- **Étape 4** Faire l'inférence bayésienne. En supposant que les  $N$  échantillons soient obtenus et que l'échantillonnage de Gibbs soit convergent après l'obtention du  $M$ -ième échantillon, les auteurs utilisent les  $N - M$  derniers échantillons pour faire l'inférence bayésienne suivante.
  - (a) Obtenir les positions et les types de valeurs aberrantes.

Par l'échantillonnage de Gibbs ci-dessus, les probabilités a posteriori des quatre hypothèses

peuvent être calculées approximativement comme suit

$$P(H_{1,j}|Y) = P(\delta_j^{AO} = 0, \delta_j^{IO} = 0) \approx \frac{1}{N - M} \sum_{k=M+1}^N \frac{P_{1j}^{(k)}}{P_{1j}^{(k)} + P_{2j}^{(k)} + P_{3j}^{(k)} + P_{4j}^{(k)}} \quad (2.14)$$

$$P(H_{2,j}|Y) = P(\delta_j^{AO} = 1, \delta_j^{IO} = 0) \approx \frac{1}{N - M} \sum_{k=M+1}^N \frac{P_{2j}^{(k)}}{P_{1j}^{(k)} + P_{2j}^{(k)} + P_{3j}^{(k)} + P_{4j}^{(k)}} \quad (2.15)$$

$$P(H_{3,j}|Y) = P(\delta_j^{AO} = 1, \delta_j^{IO} = 1) \approx \frac{1}{N - M} \sum_{k=M+1}^N \frac{P_{3j}^{(k)}}{P_{1j}^{(k)} + P_{2j}^{(k)} + P_{3j}^{(k)} + P_{4j}^{(k)}} \quad (2.16)$$

$$P(H_{4,j}|Y) = P(\delta_j^{AO} = 0, \delta_j^{IO} = 1) \approx \frac{1}{N - M} \sum_{k=M+1}^N \frac{P_{4j}^{(k)}}{P_{1j}^{(k)} + P_{2j}^{(k)} + P_{3j}^{(k)} + P_{4j}^{(k)}} \quad (2.17)$$

où  $\frac{P_{ij}^{(k)}}{P_{1j}^{(k)} + P_{2j}^{(k)} + P_{3j}^{(k)} + P_{4j}^{(k)}} (i = 1, 2, 3, 4)$  est la probabilité a posteriori conditionnelle de  $H_{i,j}$  à la  $k$ -ème itération de l'échantillonnage de Gibbs.

Si  $P(H_{i,j}|Y) = \max \{P(H_{1,j}|Y), P(H_{2,j}|Y), P(H_{3,j}|Y), P(H_{4,j}|Y)\}$  l'hypothèse  $H_{i,j}$  est acceptée et  $y_j$  peut être identifiée comme une observation normale ou une sorte de valeur aberrante.

**(b) Estimer les amplitudes anormales des valeurs aberrantes**

- Si  $y_j (j = p + 1, \dots, n)$  est identifié comme un *AO*, sa taille anormale peut être estimée par

$$\hat{\omega}_j^{AO} = \frac{\sum_{k=M+1}^N (\omega_j^{AO})^{(k)}}{N - M} \quad (2.18)$$

- Si  $y_j (j = p + 1, \dots, n)$  est identifié comme un *IO*, alors sa taille anormale peut être estimée par

$$\hat{\omega}_j^{IO} = \frac{\sum_{k=M+1}^N (\omega_j^{IO})^{(k)}}{N - M} \quad (2.19)$$

## 2.4 Test de signification bayésien

### Test basé sur la construction des régions HPD crédibles

Kim [53] présente un test de signification bayésien pour la stationnarité d'une équation de régression utilisant les régions HPD crédibles de la densité a posteriori la plus élevée. En outre, une solution au problème de Behrens-Fisher est fournie.

Kim [53] a étudié les modèles de régression à changement univariée

$$Y_1 = X_1\beta_1 + \mu_1, \quad Y_2 = X_2\beta_2 + \mu_2$$

avec  $Y_1 = (y_1, \dots, y_\tau)$ ,  $Y_2 = (y_{\tau+1}, \dots, y_T)$  où  $\tau$  est un instant de repture inconnu. Les matrices  $X_i$  ( $i = 1, 2$ ) sont d'ordre  $n_i \times p$  respectivement, où  $n_1 = \tau$  et  $n_2 = T - \tau$ . Les  $\beta_i \in \mathbb{R}^p$  et les  $\mu_i$  sont des vecteurs d'ordre  $n_i$  normalement distribué de moyenne nulle et de matrice de covariance  $\sigma_i^2 I_{n_i}$  ( $i = 1, 2$ ).

Le test de stationnarité dans les modèles de régression ci-dessus implique l'hypothèse nulle ,

$$H_0 : \quad \delta = \beta_1 - \beta_2 = 0, \quad \rho = \frac{\sigma_2^2}{\sigma_1^2} = 1$$

contre l'hypothèse alternative

$$H_1 : \quad \beta_1 \neq \beta_2 \quad \text{ou} \quad \sigma_1^2 \neq \sigma_2^2$$

Si  $\tau$  est connu, il s'agit du problème de Behrens-Fisher, lorsqu'on teste l'égalité des vecteurs de coefficient de régression dans deux équations de régression avec des variances d'erreur inégales.

L'hypothèse nulle  $H_0$  peut être divisée en deux sous-hypothèses nulles  $H_{01} : \delta = \beta_1 - \beta_2 = 0$  et  $H_{02} : \rho = \frac{\sigma_2^2}{\sigma_1^2} = 1$ , et  $H_0$  sera rejetée si l'une de ces deux sous-hypothèses nulles est rejetée. Sachant que la distribution a priori des paramètres est non informative  $\pi(\tau, \beta_1, \beta_2, r_1, r_2) \propto \frac{1}{r_1 r_2} \pi(\tau)$  avec  $r_i = \frac{1}{\sigma_i^2}$ , ( $i = 1, 2$ ), la distribution a posteriori conditionnelle de  $F(\delta)$  et  $\rho$  sont respectivement

$$\pi(F(\delta)|Y, X) = \sum \left\{ \int_0^\infty \pi(F(\delta)|\tau, \rho, Y, X) \pi(\rho|\tau, Y, X) d\rho \right\} \pi(\tau|Y, X) \quad (2.20)$$

$$\pi(\rho|Y, X) = \sum \pi(\rho|\tau, Y, X) \pi(\tau|Y, X) \quad (2.21)$$

On définit séparément les régions crédibles de densité a posteriori la plus élevée de  $F(\delta)$  et de  $\rho$  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  $\tau$  et  $\rho$ , le  $(1 - \alpha)$ -HPD région crédible de  $F(\delta)$  est définie par

$$C_{F(\delta)} = \{F(\delta) : 0 < F(\delta) < F(p, T - 2p, 1 - \alpha)\}$$

où  $F(p, T - 2p, 1 - \alpha)$  est le quantile d'ordre  $(1 - \alpha)$  de la distribution de Fisher  $F$  avec  $(p, T - 2p)$  degrés de liberté. Donc, étant donné  $\rho$  et  $\tau$ , la règle de décision pour  $H_{01}$ , est de rejeter si  $F(0) \in \overline{C_{F(\delta)}}$ , où  $\overline{C_{F(\delta)}}$  est le complémentaire de  $C_{F(\delta)}$ . La  $p$ -value inconditionnelle de  $H_{01}$  est donc calculée

à partir de (2.20) comme suit

$$\begin{aligned} P_{\delta=0}|Y, X &= 1 - \mathbb{E}_\tau(\mathbb{E}_\rho[\mathfrak{F}_{p, T-2p}\{F(0)\}]) \\ &= \sum_\tau \left( \int_0^\infty [1 - \mathfrak{F}_{p, T-2p}\{F(0)\}] \pi(\rho|\tau, Y, X) d\rho \right) \pi(\tau|Y, X) \end{aligned} \quad (2.22)$$

où  $\mathfrak{F}_{p, T-2p}$  est la fonction cumulative de la distribution de Fisher  $F$  avec  $(p, T - 2p)$  degrés de liberté. Les espérances mathématiques  $\mathbb{E}_\tau$  et  $\mathbb{E}_\rho$  sont pris par rapport à  $\tau$  et  $\rho$  respectivement. Par conséquent, ce test, consiste à rejeter  $H_{01}$ , si  $P_{\delta=0}|Y, X$  est inférieur à  $\alpha$ .

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

$$P_{\rho=1}|Y, X = 2 \sum_\tau [1 - \mathfrak{F}_{n_1-p, n_2-p}\{\max(\rho'_0, \frac{1}{\rho'_0})\}] \pi(\tau|Y, X) \quad (2.23)$$

où  $\rho'_0 = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2}$ . Aussi ce test, consiste à rejeter  $H_{02}$ , si  $P_{\rho=1}|Y, X$  est inférieur à  $\alpha$ .

Les quantités  $P_{\delta=0}|Y, X$  et  $P_{\rho=1}|Y, X$  sont évaluées numériquement par l'algorithme de Gibbs en utilisant les distributions a posteriori conditionnelles complètes des différents paramètres du modèle. Une étude de simulation de Monte Carlo a montré que le test de signification bayésien est plus puissant que les tests de Cusum et de Cusum des carrés (Brown et al. [18]). Le test de signification bayésien est aussi utile pour détecter la non-stationnarité des paramètres de régression individuels voir (Kim [53]).

Dans le chapitre 3, on va utiliser ce test pour détecter les instants de contamination multiple pour les valeurs aberrantes additives et les valeurs aberrantes innovantes.

# Chapitre 3

## Inférence bayésienne dans un modèle autorégressif à contaminations multiples avec tendance

### 3.1 Introduction

Le problème de la détection et l'estimation des valeurs aberrantes  $AO$  et  $IO$  dans les modèles de séries temporelles à été étudié par plusieurs auteurs. Parmi les tests utilisés pour détecter la position et l'amplitude de ces valeurs aberrantes on trouve le test de Kim [53] basé sur la construction des régions HPD crédibles et le calcul des  $p$ -values pour détecter la non-stationnarité des paramètres des équations de régression. Slama et Saggou [78] ont étudié l'approche bayésienne utilisant des régions HPD crédibles et des  $p$ -values pour détecter le point de rupture dans les paramètres d'un modèle  $AR(p)$ . Slama [76] a étudié l'impact d'un  $AO$  sur la performance du test de signification bayésien basé sur les régions HPD crédibles où la position et l'amplitude de la contamination sont supposées être connues. Slama et Fellag [77] ont étendu ces travaux au cas où la position et l'amplitude de la contamination  $AO$  sont inconnues dans une série de données normales indépendantes. Dans ce chapitre, nous généralisons la contamination  $AO$  et  $IO$  au cas multiple dans un modèle  $AR(p)$  avec une tendance de régression linéaire multiple. Nous utilisons des lois a priori non informative pour les paramètres du modèle, nous déterminons les distributions a posteriori conditionnelles

de ces paramètres nécessaires à l'application de l'échantillonnage de Gibbs, dans les deux cas de contamination  $AO$  et  $IO$ . Un test basé sur la  $p$ -value inconditionnelle est proposé pour détecter la position et l'amplitude de la contamination pour les valeurs aberrantes multiples  $AO$  et  $IO$ , quand le nombre de contaminants est connu.

## 3.2 Contamination multiple par des valeurs aberrantes additives

### 3.2.1 Modèle et notations

Dans cette section, on considère un modèle  $AR(p)$  avec une tendance de régression, contaminé par  $m$  valeurs aberrantes additives

$$y_t = \underline{x}'_t \underline{\beta} + \sum_{l=1}^m \delta_{k_l}(t) \Delta_{k_l} + \mu_t, \quad t = 1, \dots, n. \quad (3.1a)$$

et  $\{\mu_t\}$  est la solution causale du modèle  $AR(p)$  suivant,

$$\mu_t = \sum_{i=1}^p \phi_i \mu_{t-i} + \varepsilon_t \quad (3.1b)$$

Ainsi (3.1a) – (3.1b) peuvent se réécrire comme suit :

$$y_t - \underline{x}'_t \underline{\beta} - \sum_{l=1}^m \delta_{k_l}(t) \Delta_{k_l} = \sum_{i=1}^p \phi_i (y_{t-i} - \underline{x}'_{t-i} \underline{\beta} - \sum_{l=1}^m \delta_{k_l}(t-i) \Delta_{k_l}) + \varepsilon_t \quad (3.1c)$$

Avec  $m = 1, 2, \dots, n - 1$  est le nombre de contaminants qui se présentent aux instants  $k_1, k_2, \dots, k_m$ , où  $1 < k_l < n$ ,  $\delta_{k_l}(\cdot)$  est la fonction usuelle de Kronecker,  $l = 1, \dots, m$  et  $k_1 < k_2 < \dots < k_m$ .

On suppose que l'on dispose des  $p$  premières valeurs initiales de  $y$  et  $x$ . Sinon, on fixe  $y_1, \dots, y_p$  et  $\underline{x}_1, \dots, \underline{x}_p$ . Dans nos calculs, on suppose que  $\{\varepsilon_t, t = 1, \dots, n\}$  sont des innovations aléatoires indépendantes de moyenne 0 et de variance  $\sigma^2$ .

Soit  $\underline{\phi} := (\phi_1, \dots, \phi_p)'$ ,  $\underline{\Delta} := (\Delta_{k_1}, \dots, \Delta_{k_m})'$ .

Les paramètres  $\underline{\beta}$ ,  $\underline{\phi}$ ,  $\underline{\Delta}$ ,  $\sigma$  du modèle (3.1c) sont supposés inconnus et  $m$  est connu.  $\underline{\beta} \in \mathbb{R}^K$ ,  $\underline{\Delta} \in \mathbb{R}^m$ ,  $\underline{\phi} \in \mathbb{R}^p$ ,  $X = (x_{t,j}), t = 1, \dots, n$  et  $j = 1, \dots, K$  est une matrice  $(n \times K)$  de rang plein à éléments fixes.

On suppose que les paramètres autorégressifs correspondent au processus causal stationnaire  $\{\mu_t, t \in \mathbb{Z}\}$  de telle sorte que le vecteur des paramètres  $\underline{\phi}$  soit dans la région  $\Phi^{(p)} = \{z \in \mathbb{C} \text{ tel que } 1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p = 0 \text{ implique } |z| > 1\}$ .

Ainsi,  $\underline{\theta} = (\underline{\beta}', \underline{\phi}', \underline{\Delta}', \sigma) \in \mathbb{R}^{(K+p+m)} \times \mathbb{R}_+^*$  est le vecteur des paramètres ( $\mathbb{R}_+^* = (0, +\infty)$ ).

### 3.2.2 Analyse bayésienne

Dans cette section, on calcule les lois a posteriori conditionnelles des paramètres du vecteur  $\underline{\theta}$ .

La fonction de quasi-vraisemblance conditionnelle basée sur les observations  $\underline{y} = (y_1, \dots, y_n)$  est obtenue à partir de (3.1c).

Pour calculer la fonction de vraisemblance conditionnelle, on suppose que les valeurs initiales  $y_0, y_{-1}, \dots, y_{1-p}$  sont disponibles. Sinon, quand la taille de l'échantillon  $n$  est suffisamment grande, les premières observations apportent une contribution négligeable à la vraisemblance [42]. On utilise la quasi vraisemblance gaussienne (GQL) dans un cadre bayésien pour estimer nos paramètres.

Nous supposons que notre connaissance préalable du paramètre  $\sigma$  est vague ou diffuse donnée par  $\pi(\sigma) \propto \frac{1}{\sigma}$  où  $\propto$  désigne le symbole de proportionnalité et les lois a priori des paramètres  $\underline{\beta}$ ,  $\underline{\phi}$  et  $\underline{\Delta}$  sont données par

$$\pi(\underline{\beta}) = \prod_{i=1}^K \pi(\beta_i) \propto c, \quad \pi(\underline{\phi}) = \prod_{i=1}^p \pi(\phi_i) \propto c \quad \text{et} \quad \pi(\Delta_{k_l}) \propto c, \quad l = 1, \dots, m,$$

où  $c$  est une constante positive quelconque. Les paramètres  $\underline{\beta}$ ,  $\underline{\phi}$ ,  $\underline{\Delta}$  et  $\sigma$  sont supposés indépendants. Par conséquent, la loi a priori de  $\underline{\theta}$  est  $\pi(\underline{\theta}) \propto \frac{1}{\sigma}$ .

Avec la loi a priori donnée ci-dessus et la fonction de vraisemblance, le théorème de Bayes permet

d'obtenir la loi a posteriori conjointe suivante

$$\begin{aligned}
 \pi(\theta|X, \underline{y}) &\propto \frac{1}{\sigma^{n+1}} \exp \left\{ -\frac{1}{2\sigma^2} \left[ \sum_{t=1}^{k_1-1} (y_t - \underline{x}'_t \underline{\beta} - \sum_{i=1}^p \phi_i(y_{t-i} - \underline{x}'_{t-i} \underline{\beta}))^2 \right. \right. \\
 &+ \sum_{l=1}^m (y_{k_l} - \underline{x}'_{k_l} \underline{\beta} - \Delta_{k_l} - \sum_{i=1}^p \phi_i(y_{k_l-i} - \underline{x}'_{k_l-i} \underline{\beta} - \sum_{\tau=1}^{l-1} \delta_{k_l-k_\tau}(i) \Delta_{k_\tau}))^2 \\
 &+ \sum_{l=1}^m \sum_{j=1}^{k_{l+1}-k_l-1} (y_{k_l+j} - \underline{x}'_{k_l+j} \underline{\beta} - \sum_{i=1}^p \phi_i(y_{k_l+j-i} - \underline{x}'_{k_l+j-i} \underline{\beta} - \sum_{\tau=1}^{l-1} \delta_{k_l-k_\tau+j}(i) \Delta_{k_\tau} \\
 &\left. \left. - \delta_j(i) \Delta_{k_l})^2 + \sum_{t=k_m+1}^n (y_t - \underline{x}'_t \underline{\beta} - \sum_{i=1}^p \phi_i(y_{t-i} - \underline{x}'_{t-i} \underline{\beta} - \sum_{l=1}^m \delta_{t-k_l}(i) \Delta_{k_l}))^2 \right] \right\}. \tag{3.2}
 \end{aligned}$$

Il convient de noter que, en raison de la non-linéarité découlant de la formulation du modèle, ces calculs a posteriori sont difficiles, si on les envisage directement par le biais du théorème de Bayes, même si les distributions a priori des paramètres sont supposées diffuses. Cependant, le modèle possède une forme conditionnelle commode qui se soumet à la méthode d'échantillonnage de Gibbs. (voir, par exemple, Gelfand et Smith [32]).

En pratique, le bruit blanc ne suit pas toujours une loi gaussienne standard (symétrique, centrée et normalisée). Par exemple, quand le bruit blanc suit une loi de Student  $t(0, \sigma, \nu)$  ou un mélange de lois normales, les lois a posteriori conditionnelles des paramètres ne sont pas de type connu. Dans ce cas, l'algorithme de Gibbs Sampler ne peut pas être appliqué et donc, nous pensons que l'algorithme de Metropolis-Hastings et/ou l'algorithme de Griddy-Gibbs seraient appropriés.

Notre objectif est de montrer la robustesse du GQL lorsque les erreurs sont distribuées selon la loi de Student ou selon une distribution de mélange.

Le théorème suivant donne les lois a posteriori conditionnelles de  $\underline{\beta}$ ,  $\underline{\phi}$  et  $\Delta_{k_l}$ , pour  $l = 1, \dots, m$ .

**Théorème 3.1.** (Ait Mohammed et Guerbyenne [1]) *1. La loi a posteriori conditionnelle du vecteur des paramètres de régression du modèle.*



Soit  $Z(\underline{\phi}, \underline{\Delta}, \underline{y})$  le  $n$ -vecteur dont la  $t$ -ème composante est

$$Z_t(\underline{\phi}, \underline{\Delta}, \underline{y}) = \begin{cases} y_t - \sum_{i=1}^p \phi_i y_{t-i}, & \text{si } t = 1, \dots, k_1 - 1 \\ y_t - \Delta_{k_l} - \sum_{i=1}^p \phi_i y_{t-i} + \sum_{i=1}^p \phi_i \sum_{\tau=1}^{l-1} \delta_{k_l - k_\tau}(i) \Delta_{k_\tau}, & \text{si } t = k_l, l = 1, \dots, m \\ y_t - \sum_{i=1}^p \phi_i y_{t-i} + \sum_{i=1}^p \phi_i \sum_{\tau=1}^{l-1} \delta_{k_l - k_\tau + j}(i) \Delta_{k_\tau} + \sum_{i=1}^p \phi_i \delta_j(i) \Delta_{k_l} & \text{si } t = k_l + j, j = 1, \dots, k_{l+1} - k_l - 1, l = 1, \dots, m - 1 \\ y_t - \sum_{i=1}^p \phi_i y_{t-i} + \sum_{i=1}^p \phi_i \sum_{l=1}^m \delta_{t - k_l}(i) \Delta_{k_l}, & \text{si } t = k_m + 1, \dots, n. \end{cases} \quad (3.3)$$

De plus, soit  $U(\underline{\phi}, X)$  la  $(n \times K)$ -matrice dont la  $t$ -ème ligne est

$$U_t(\underline{\phi}, X) = \underline{x}'_t - \sum_{i=1}^p \phi_i \underline{x}'_{t-i}, \quad t = 1, \dots, n. \quad (3.4)$$

Alors

$$Z(\underline{\phi}, \underline{\Delta}, \underline{y}) = U(\underline{\phi}, X) \underline{\beta} + \underline{\varepsilon} \quad (3.5)$$

où  $\underline{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)'$ .

Pour simplifier les notations, on pose  $U := U(\underline{\phi}, X)$  et  $Z := Z(\underline{\phi}, \underline{\Delta}, \underline{y})$ .

La loi a posteriori conditionnelle de  $\underline{\beta}$  est

$$\pi(\underline{\beta} | \underline{\phi}, \underline{\Delta}, X, \underline{y}) \propto \left\{ 1 + \frac{(\underline{\beta} - \hat{\underline{\beta}})' U' U (\underline{\beta} - \hat{\underline{\beta}})}{(Z - U \hat{\underline{\beta}})' (Z - U \hat{\underline{\beta}})} \right\}^{-\frac{n}{2}} \quad (3.6)$$

qui est une loi  $t_K$  de Student multivariée de paramètre de position  $\hat{\underline{\beta}} = (U'U)^{-1}U'Z$ , de matrice de précision  $\frac{(U'U)(n-K)}{(Z-U\hat{\underline{\beta}})'(Z-U\hat{\underline{\beta}})}$  et  $(n - K)$  degrés de liberté.

## 2. La loi a posteriori conditionnelle du vecteur des paramètres autorégressifs.

Soit  $W(\underline{\beta}, \underline{\Delta}, X, \underline{y})$  la  $(n \times p)$ -matrice dont la  $t$ -ème ligne est

$$W_t(\underline{\beta}, \underline{\Delta}, X, \underline{y}) = \begin{cases} (y_{t-i} - \underline{x}'_{t-i} \underline{\beta}, i = 1, \dots, p), & \text{si } t = 1, \dots, k_1 - 1 \\ (y_{t-i} - \underline{x}'_{t-i} \underline{\beta} - \sum_{\tau=1}^{l-1} \delta_{k_l - k_\tau}(i) \Delta_{k_\tau}, i = 1, \dots, p), & \text{si } t = k_l, l = 1, \dots, m \\ (y_{t-i} - \underline{x}'_{t-i} \underline{\beta} - \sum_{\tau=1}^{l-1} \delta_{k_l - k_\tau + j}(i) \Delta_{k_\tau} - \delta_j(i) \Delta_{k_l}, i = 1, \dots, p), & \text{si } t = k_l + j, j = 1, \dots, k_{l+1} - k_l, l = 1, \dots, m - 1, \\ (y_{t-i} - \underline{x}'_{t-i} \underline{\beta} - \sum_{l=1}^m \delta_{t - k_l}(i) \Delta_{k_l}, i = 1, \dots, p) & \text{si } t = k_m + 1, \dots, n. \end{cases} \quad (3.7)$$

avec  $(x_i, i = 1, \dots, p) = (x_1, \dots, x_p)$ . De plus, soit  $V(\underline{\beta}, \underline{\Delta}, X, \underline{y})$  le  $n$ -vecteur dont la  $t$ -ème composante est

$$V_t(\underline{\beta}, \underline{\Delta}, X, \underline{y}) = \begin{cases} y_t - \underline{x}'_t \underline{\beta}, & \text{si } t \neq k_l, l = 1, \dots, m. \\ y_t - \underline{x}'_t \underline{\beta} - \Delta_{k_l}, & \text{si } t = k_l, l = 1, \dots, m. \end{cases} \quad (3.8)$$

Alors, on a

$$V(\underline{\beta}, \underline{\Delta}, X, \underline{y}) = W(\underline{\beta}, \underline{\Delta}, X, \underline{y})\underline{\phi} + \varepsilon \quad (3.9)$$

Pour simplifier les notations, on pose  $V := V(\underline{\beta}, \underline{\Delta}, X, \underline{y})$  et  $W := W(\underline{\beta}, \underline{\Delta}, X, \underline{y})$ .

Sachant  $\underline{\beta}$ ,  $\underline{\Delta}$ ,  $X$  et  $\underline{y}$ , la loi a posteriori conditionnelle de  $\underline{\phi}$  est

$$\pi(\underline{\phi} | \underline{\beta}, \underline{\Delta}, X, \underline{y}) \propto \left\{ 1 + \frac{(\underline{\phi} - \hat{\underline{\phi}})' W' W (\underline{\phi} - \hat{\underline{\phi}})}{(V - W \hat{\underline{\phi}})' (V - W \hat{\underline{\phi}})} \right\}^{-\frac{n}{2}} \quad (3.10)$$

qui est la loi  $t_p$  de Student multivariée tronquée, de paramètre de position  $\hat{\underline{\phi}} = (W'W)^{-1}W'V$ , de matrice de précision  $\frac{(W'W)(n-p)}{(V-W\hat{\underline{\phi}})'(V-W\hat{\underline{\phi}})}$  et  $(n-p)$  degrés de liberté.

3. La loi a posteriori conditionnelle de  $\Delta_{k_l}$ , sachant  $(\underline{\beta}, \underline{\phi}, \underline{\Delta}^{(-k_l)}, X, \underline{y})$ , où  $\underline{\Delta}^{(-k_l)}$  représente le vecteur  $\Delta_{k_l}$  sans la  $k_l$ -ème composante ( $l = 1, \dots, m$ ).

$$\pi(\Delta_{k_l} | \underline{\beta}, \underline{\phi}, \underline{\Delta}^{(-k_l)}, \sigma, X, \underline{y}) \propto \frac{1}{\sigma^{n+1}} \exp\left\{-\frac{1}{2\sigma^2} I\right\}. \quad (3.11)$$

où  $I$  peut s'écrire comme suit

$$\begin{aligned} I = & [\Delta_{k_l} - \tilde{\Delta}_{k_l}^{(1)}]^2 \\ & + \left[ \sum_{i=1}^p \phi_i \delta_1(i) \Delta_{k_l} + \{y_{k_l+1} - \underline{x}'_{k_l+1} \underline{\beta} - \sum_{i=1}^p \phi_i (y_{k_l+1-i} - \underline{x}'_{k_l+1-i} \underline{\beta} - \sum_{\tau=1}^{l-1} \delta_{k_l-k_\tau+1}(i) \Delta_{k_\tau})\} \right]^2 \\ & + \left[ \sum_{i=1}^p \phi_i \delta_2(i) \Delta_{k_l} + \{y_{k_l+2} - \underline{x}'_{k_l+2} \underline{\beta} - \sum_{i=1}^p \phi_i (y_{k_l+2-i} - \underline{x}'_{k_l+2-i} \underline{\beta} - \sum_{\tau=1}^{l-1} \delta_{k_l-k_\tau+2}(i) \Delta_{k_\tau})\} \right]^2 + \dots \\ & + \sum_{t=k_{l+1}}^{\min(n, p+k_l)} \left[ \sum_{i=1}^p \phi_i \delta_{t-k_l}(i) \Delta_{k_l} + y_t - \underline{x}'_t \underline{\beta} - \sum_{s=l+1}^m \delta_{k_s}(t) \Delta_{k_s} - \sum_{i=1}^p \phi_i (y_{t-i} - \underline{x}'_{t-i} \underline{\beta} - \sum_{\tau=1}^{l-1} \delta_{t-k_\tau}(i) \Delta_{k_\tau}) \right]^2. \end{aligned}$$

avec

$$\tilde{\Delta}_{k_l}^{(1)} = y_{k_l} - \underline{x}'_{k_l} \underline{\beta} - \sum_{i=1}^p \phi_i (y_{k_l-i} - \underline{x}'_{k_l-i} \underline{\beta} - \sum_{\tau=1}^{l-1} \delta_{k_l-k_\tau}(i) \Delta_{k_\tau})$$

En prenant en compte les termes au carré contenant  $\Delta_{k_l}$ , on obtient

$$\pi(\Delta_{k_l} | \underline{\beta}, \underline{\phi}, \underline{\Delta}^{(-k_l)}, \sigma, X, \underline{y}) \propto \frac{1}{\sigma^{n+1}} \exp\left\{-\frac{1}{2\sigma^2} \left[ 1 + \sum_{k=1}^{\min(n, p+k_l)} \left( \sum_{i=1}^p \phi_i \delta_k(i) \right)^2 \right] (\Delta_{k_l} - \tilde{\Delta}_l^{(\lambda)})^2 \right\} \quad (3.12)$$

$\tilde{\Delta}_l^{(\lambda_l)}$  est le paramètre de position de la dernière étape,  $\lambda_l, l = 1, \dots, m$ .

On peut utiliser l'algorithme suivant pour calculer le paramètre de contamination.

### Algorithme de calcul du paramètre de contamination

1. Pour  $j = 1, \dots, \lambda_l = \min(p, k_{l+1} - k_l - 1)$

$$\tilde{\Delta}_{k_l}^{(j+1)} = [1 + \sum_{k=1}^j (\sum_{i=1}^{i=p} \phi_i \delta_k(i))^2]^{-1} [(1 + \sum_{k=1}^{j-1} (\sum_{i=1}^p \phi_i \delta_k(i))^2) \tilde{\Delta}_{k_l}^{(j)} + (\sum_{i=1}^p \phi_i \delta_j(i))^2 \hat{\Delta}_{k_l}^{(j+1)}]$$

où

$$\hat{\Delta}_{k_l}^{(1)} := \tilde{\Delta}_{k_l}^{(1)}$$

et

$$\hat{\Delta}_{k_l}^{(j+1)} = \frac{1}{\sum_{i=1}^p \phi_i \delta_j(i)} [y_{k_l+j} - \underline{x}'_{k_l+j} \underline{\beta} - \sum_{i=1}^p \phi_i (y_{k_l+j-i} - \underline{x}'_{k_l+j-i} \underline{\beta}) - \sum_{i=1}^p (\sum_{\tau=1}^{l-1} \delta_{k_l-k_\tau+j}(i) \Delta_{k_\tau})]$$

de la  $j$ -ème étape avant le contaminant suivant, si  $\sum_{i=1}^p \phi_i \delta_j(i) \neq 0$ , sinon  $\hat{\Delta}_{k_l}^{(j+1)}$  n'existe pas.

2. Pour  $t = k_{l+1}, \dots, \min(n, p + k_l)$

$$\tilde{\Delta}_{k_l}^{(t-k_l+1)} = [1 + \sum_{j=1}^{\min(p, t-k_l)} (\sum_{i=1}^p \phi_i \delta_j(i))^2]^{-1} [(1 + \sum_{s=1}^{t-1} (\sum_{i=1}^p \phi_i \delta_{s-k_l}(i))^2) \tilde{\Delta}_{k_l}^{(t-k_l)} + (\sum_{i=1}^p \phi_i \delta_{t-k_l}(i))^2 \hat{\Delta}_{k_l}^{(t-k_l+1)}]$$

où

$$\hat{\Delta}_{k_l}^{(t-k_l+1)} = \frac{1}{\sum_{i=1}^p \phi_i \delta_{t-k_l}(i)} [y_t - \underline{x}'_t \underline{\beta} - \sum_{s=l+1}^m \delta_{k_s}(t) \Delta_{k_s} - \sum_{i=1}^p \phi_i \{y_{t-i} - \underline{x}'_{t-i} \underline{\beta} - \sum_{\substack{\tau=1 \\ \tau \neq l}}^m \delta_{t-k_l}(i) \Delta_{k_\tau}\}]$$

si  $\sum_{i=1}^p \phi_i \delta_{t-k_l}(i) \neq 0$ , sinon  $\hat{\Delta}_{k_l}^{(t-k_l+1)}$  n'existe pas.

### 3.3 Contamination multiple par des valeurs aberrantes innovantes

#### 3.3.1 Modèle et notations

On considère le modèle  $AR(p)$  avec une tendance de régression, contaminé par  $m$  valeurs aberrantes innovantes

$$y_t = \underline{x}'_t \underline{\beta} + \sum_{j=1}^m \left( \sum_{l=1}^j \gamma_{t-k_l} \Delta_{k_l} \right) \delta_{k_j, \dots, k_{j+1}-1}(t) + \mu_t, \quad t = 1, \dots, n. \quad (3.13a)$$

$$\mu_t = \sum_{i=1}^p \phi_i \mu_{t-i} + \varepsilon_t, \quad (3.13b)$$

avec, par convention  $k_{m+1} = n + 1$ ,

$$\delta_{k_j, \dots, k_{j+1}-1}(t) = \begin{cases} 1 & \text{si } t \in \{k_j, \dots, k_{j+1} - 1\} \\ 0 & \text{sinon.} \end{cases}$$

où

$$\begin{aligned} \gamma_j &= \phi_j \gamma_0 + \phi_{j-1} \gamma_1 + \dots + \phi_1 \gamma_{j-1}, \quad j = 1, 2, \dots \\ \gamma_0 &= 1, \gamma_1 = \phi_1, \gamma_2 = \phi_1^2 + \phi_2, \dots \quad \gamma_j = 0, \quad \text{si } j < 0 \\ \phi_j &= 0 \quad \text{si } j > p \end{aligned}$$

Notre modèle peut se réécrire comme suit

$$y_t = \begin{cases} \underline{x}'_t \underline{\beta} + \sum_{i=1}^p \phi_i (y_{t-i} - \underline{x}'_{t-i} \underline{\beta}) + \varepsilon_t, & \text{si } t \neq k_l, l = 1, \dots, m \\ \underline{x}'_t \underline{\beta} + \sum_{i=1}^p \phi_i (y_{t-i} - \underline{x}'_{t-i} \underline{\beta}) + \Delta_{k_l} + \varepsilon_t, & \text{si } t = k_l, l = 1, \dots, m \end{cases} \quad (3.14)$$

#### 3.3.2 Analyse bayésienne

La densité a posteriori du vecteur des paramètres  $\underline{\theta}$  est donnée par

$$\begin{aligned} \pi(\underline{\theta} | X, \underline{y}) &\propto \sigma^{-(n+1)} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{\substack{t=1 \\ t \notin A_m}}^n (y_t - \underline{x}'_t \underline{\beta} - \sum_{i=1}^p \phi_i (y_{t-i} - \underline{x}'_{t-i} \underline{\beta}))^2 \right. \\ &\quad \left. + \sum_{l=1}^m (y_{k_l} - \underline{x}'_{k_l} \underline{\beta} - \sum_{i=1}^p \phi_i (y_{k_l-i} - \underline{x}'_{k_l-i} \underline{\beta}) - \Delta_{k_l})^2 \right\}. \end{aligned} \quad (3.15)$$

où  $A_m = \{k_1, k_2, \dots, k_m\}$

Le théorème suivant donne les lois a posteriori conditionnelles de  $\underline{\beta}$ ,  $\underline{\phi}$  et  $\Delta_{k_l}$ , pour  $l = 1, \dots, m$ .

**Théorème 3.2.** (Ait Mohammed et Guerbyenne [1]) **1. La loi a posteriori conditionnelle du vecteur des paramètres de régression.**

Soit  $H(\underline{\phi}, \underline{\Delta}, \underline{y})$  le  $n$ -vecteur dont la  $t$ -ème composante est

$$H_t(\underline{\phi}, \underline{\Delta}, \underline{y}) = \begin{cases} y_t - \sum_{i=1}^p \phi_i y_{t-i}, & \text{si } t \neq k_l, l = 1, \dots, m \\ y_t - \Delta_{k_l} - \sum_{i=1}^p \phi_i y_{t-i}, & \text{si } t = k_l, l = 1, \dots, m. \end{cases} \quad (3.16)$$

De plus, soit  $L(\underline{\phi}, X)$  la  $(n \times K)$ -matrice dont la  $t$ -ème ligne est

$$L_t(\underline{\phi}, X) = \underline{x}'_t - \sum_{i=1}^p \phi_i \underline{x}'_{t-i}, \quad t = 1, \dots, n \quad (3.17)$$

Alors

$$H(\underline{\phi}, \underline{\Delta}, \underline{y}) = L(\underline{\phi}, X)\underline{\beta} + \underline{\varepsilon} \quad (3.18)$$

$\underline{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)'$ . Pour simplifier les notations, on pose  $L := L(\underline{\phi}, X)$  et  $H := H(\underline{\phi}, \underline{\Delta}, \underline{y})$ , alors

$$\pi(\underline{\beta} | \underline{\phi}, \underline{\Delta}, X, \underline{y}) \propto \left\{ 1 + \frac{(\underline{\beta} - \hat{\underline{\beta}})' L' L (\underline{\beta} - \hat{\underline{\beta}})}{(H - L \hat{\underline{\beta}})' (H - L \hat{\underline{\beta}})} \right\}^{-\frac{n}{2}} \quad (3.19)$$

qui est une loi de Student  $t_K$ -multivariée, de paramètre de position  $\hat{\underline{\beta}} = (L' L)^{-1} L' H$ , de matrice de précision  $\frac{(L' L)(n-K)}{(H - L \hat{\underline{\beta}})' (H - L \hat{\underline{\beta}})}$  et  $(n - K)$  degrés de liberté.

**2. La loi a posteriori conditionnelle du vecteur des paramètres autorégressifs.**

Soit  $I(\underline{\beta}, \underline{\Delta}, X, \underline{y})$  le  $n$ -vecteur dont la  $t$ -ème composante est

$$I_t(\underline{\beta}, \underline{\Delta}, X, \underline{y}) = \begin{cases} y_t - \underline{x}'_t \underline{\beta} & \text{si } t \neq k_l, l = 1, \dots, m \\ y_t - \underline{x}'_t \underline{\beta} - \Delta_{k_l} & \text{si } t = k_l, l = 1, \dots, m. \end{cases} \quad (3.20)$$

De plus, soit  $J(\underline{\beta}, X, \underline{y})$  la  $(n \times p)$ -matrice dont la  $t$ -ème ligne est

$$J_t(\underline{\beta}, X, \underline{y}) = (y_{t-1} - \underline{x}'_{t-1} \underline{\beta}, \dots, y_{t-p} - \underline{x}'_{t-p} \underline{\beta}) \quad (3.21)$$

Alors, on a

$$I(\underline{\beta}, \underline{\Delta}, X, \underline{y}) = J(\underline{\beta}, X, \underline{y})\underline{\phi} + \underline{\varepsilon} \quad (3.22)$$

Pour simplifier les notations, on pose  $J := J(\underline{\beta}, X, \underline{y})$  et  $I := I(\underline{\beta}, \underline{\Delta}, X, \underline{y})$ , sachant  $\underline{\beta}, \underline{\Delta}, X$  et  $\underline{y}$ , la loi a posteriori conditionnelle de  $\underline{\phi}$  est

$$\pi(\underline{\phi} | \underline{\beta}, \underline{\Delta}, X, \underline{y}) \propto \left\{ 1 + \frac{(\underline{\phi} - \hat{\underline{\phi}})' J' J (\underline{\phi} - \hat{\underline{\phi}})}{(I - J \hat{\underline{\phi}})' (I - J \hat{\underline{\phi}})} \right\}^{-\frac{n}{2}} \quad (3.23)$$

qui est la loi de Student  $t_p$ -multivariée tronquée de paramètre de position  $\hat{\underline{\phi}} = (J'J)^{-1}J'I$ , de matrice de précision  $\frac{(J'J)(n-p)}{(I-J\hat{\underline{\phi}})'(I-J\hat{\underline{\phi}})}$  et  $(n-p)$  degrés de liberté.

3. Pour  $l = 1, \dots, m$  sachant  $\underline{\beta}$ ,  $\underline{\phi}$ ,  $\underline{\Delta}^{(-k_l)}$  la loi a posteriori conditionnelle de  $\Delta_{k_l}$  est

$$\pi(\Delta_{k_l} | \underline{\beta}, \underline{\phi}, \underline{\Delta}^{(-k_l)}, X, \underline{y}) \propto \left\{ 1 + \frac{(\Delta_{k_l} - \tilde{\Delta}_{k_l})^2}{(n-1)S_{k_l}^2} \right\}^{-\frac{n}{2}} \quad (3.24)$$

où

$$\tilde{\Delta}_{k_l} = y_{k_l} - \underline{x}'_{k_l} \underline{\beta} - \sum_{i=1}^p \phi_i (y_{k_l-i} - \underline{x}'_{k_l-i} \underline{\beta}) \quad (3.25)$$

et

$$\begin{aligned} S_{k_l}^2 = & \frac{1}{(n-1)} \left[ \sum_{\substack{t=1 \\ t \notin A_m}}^n \left( y_t - \underline{x}'_t \underline{\beta} - \sum_{i=1}^p \phi_i (y_{t-i} - \underline{x}'_{t-i} \underline{\beta}) \right)^2 \right. \\ & \left. + \sum_{\substack{l'=1 \\ l' \neq l}}^m \left( \Delta_{k_{l'}} - \{ y_{k_{l'}} - \underline{x}'_{k_{l'}} \underline{\beta} - \sum_{i=1}^p \phi_i (y_{k_{l'}-i} - \underline{x}'_{k_{l'}-i} \underline{\beta}) \} \right)^2 \right]. \end{aligned} \quad (3.26)$$

Soit  $\underline{\theta}^{-j}$  le vecteur  $\underline{\theta}$  sans la  $j$ -ème composante.

**Remarque 3.1.** Pour simplifier les calculs, on utilise la matrice de permutation, notée  $R$  pour obtenir  $\beta_j$  comme la première composante de  $R\beta$ . On peut avoir la loi a posteriori conditionnelle de  $\beta_j$  sachant  $\underline{\beta}^{-j}$ ,  $\underline{\phi}$ ,  $\underline{\Delta}$ ,  $X$ ,  $\underline{y}$ ,  $j = 1, \dots, K$  qui est une loi de Student univariée de paramètre de position  $\hat{\beta}_j + \Sigma_{12}\Sigma_{22}(\underline{\beta}^{-j} - \hat{\underline{\beta}}^{-j})$ , de précision  $\Sigma_{11} - \Sigma_{12}(\Sigma_{22})^{-1}\Sigma_{21}$  et  $(n-1)$  degrés de liberté, où

$$\Sigma = R\Sigma_{\beta}R' = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

et  $\Sigma_{\beta} = \left( \frac{U'U(n-K)}{(Z-U\hat{\underline{\beta}})'(Z-U\hat{\underline{\beta}})} \right)^{-1}$ .

**Remarque 3.2.** On peut aussi déterminer la loi d'une composante  $\phi_j$  sachant  $\underline{\phi}^{-j}$ ,  $\underline{\beta}$ ,  $\underline{\Delta}$ ,  $X$ ,  $\underline{y}$ ,  $j = 1, \dots, p$  comme dans la Remarque 3.1.

### 3.4 Test de signification bayésien

Dans les deux types de contamination  $AO$  et  $IO$ , la loi a posteriori conditionnelle de  $\Delta_{k_l}$ , ( $l = 1, \dots, m$ ) est une loi de Student de paramètre de position  $\tilde{\Delta}_{k_l}$ , de paramètre d'échelle  $S_{k_l}$  et à

$(n - 1)$  degrés de liberté.

$$\pi(\Delta_{k_l} | \underline{\beta}, \underline{\phi}, \underline{\Delta}^{(-k_l)}, X, \underline{y}) \propto \left\{ 1 + \frac{(\Delta_{k_l} - \tilde{\Delta}_{k_l})^2}{(n-1)S_{k_l}^2} \right\}^{-\frac{n}{2}} \quad (3.27)$$

De manière équivalente, la quantité,

$$t_l(\Delta_{k_l}) = \frac{(\Delta_{k_l} - \tilde{\Delta}_{k_l})}{S_{k_l}}, \quad l = 1, \dots, m$$

suit une loi a posteriori conditionnelle de Student standard à  $(n - 1)$  degrés de liberté sachant  $\underline{\beta}$ ,  $\underline{\phi}$  et  $\underline{\Delta}^{(-k_l)}$ .

L'hypothèse nulle  $H_0$ , qu'il n'y a pas de contaminant dans le modèle (3.1a) ou (3.13) est

$$H_0 : \Delta_{k_l} = 0, \quad \forall \quad l = 1, \dots, m$$

contre

$$H_1 : \Delta_{k_l} \neq 0, \quad \text{pour certains } l = 1, \dots, m$$

L'hypothèse nulle  $H_0$  peut être décomposée en  $m$  sous-hypothèses nulles

$$H_{0l} : \Delta_{k_l} = 0 \quad \text{pour } l = 1, \dots, m$$

et  $H_0$  peut être rejetée si l'une de ces  $m$  sous-hypothèses est rejetée.

Le test proposé est basé sur la loi a posteriori de  $\Delta_{k_l}$  (voir, Kim [53]).

La loi a posteriori inconditionnelle de  $t_l(\Delta_{k_l})$ ,  $(l = 1, \dots, m)$  est

$$\begin{aligned} \pi(t_l(\Delta_{k_l}) | X, \underline{y}) &= \int_{\Phi^p} \int_{\mathbb{R}^K} \int_{\mathbb{R}^{m-1}} \pi(t_l(\Delta_{k_l}) | \underline{\beta}, \underline{\phi}, \underline{\Delta}^{(-k_l)}, X, \underline{y}) \pi(\underline{\beta} | \underline{\phi}, \underline{\Delta}^{(-k_l)}, X, \underline{y}) \\ &\times \pi(\underline{\phi} | \underline{\Delta}^{(-k_l)}, X, \underline{y}) \pi(\underline{\Delta}^{(-k_l)} | X, \underline{y}) d\underline{\Delta}^{(-k_l)} d\underline{\beta} d\underline{\phi}. \end{aligned} \quad (3.28)$$

On définit la région HPD (highest posterior density) crédible de  $t_l(\Delta_{k_l})$ . La région crédible (voir, e.g., Geweke [35]) sera utilisée pour définir la  $p$ -value inconditionnelle et par conséquent un test inconditionnel.

Pour  $l = 1, \dots, m$ , sachant  $\underline{\beta}$ ,  $\underline{\phi}$  et  $\underline{\Delta}^{(-k_l)}$ , la région  $(1 - \alpha)$  - crédible ( $0 < \alpha < 1$ ) de  $t_l(\Delta_{k_l})$  est définie par

$$C_l = \left\{ t_l(\Delta_{k_l}) : |t_l(\Delta_{k_l})| < t_{\frac{\alpha}{2}}(n-1) \right\}$$

où  $t_{\frac{\alpha}{2}}(n-1)$  est le  $(1 - \frac{\alpha}{2})$ -ème quantile de la loi de Student à  $(n-1)$  degrés de liberté. Ainsi, sachant  $\underline{\beta}$ ,  $\underline{\phi}$  et  $\underline{\Delta}^{(-k_l)}$ , la règle de décision pour  $H_{0l}$  est de rejeter si  $t_l(0) \in \bar{C}_l$ , où  $\bar{C}_l$  est le complémentaire de  $C_l$ .

La  $p$ -value inconditionnelle de la sous-hypothèse  $H_{0l}$  est calculée à partir de (3.28) pour obtenir

$$\begin{aligned} P_{\Delta_{k_l}=0|\underline{y}} &= 2 \int_{\Phi^p} \int_{\mathbb{R}^K} \int_{\mathbb{R}^{m-1}} \{1 - \tau_{n-1}(|t_l(0)|)\} \pi(\underline{\beta}, \underline{\phi}, \underline{\Delta}^{(-k_l)} | X, \underline{y}) d\underline{\Delta}^{(-k_l)} d\underline{\beta} d\underline{\phi} \quad (3.29) \\ &= 2 E_{\underline{\phi}} E_{\underline{\beta}} E_{\underline{\Delta}^{(-k_l)}} \{1 - \tau_{n-1}(|t_l(0)|)\} \end{aligned}$$

où  $\tau_{n-1}$  est la fonction de répartition de la loi de Student standard à  $(n-1)$  degrés de liberté. L'espérance est prise par rapport à  $\underline{\phi}$ ,  $\underline{\beta}$  et  $\underline{\Delta}^{(-k_l)}$ .

La quantité donnée dans la formule (3.29) sera calculée avec l'algorithme de Gibbs Sampler en utilisant les distributions a posteriori conditionnelles donnée dans le théorème 3.1 et théorème 3.2.

1. Si les positions des contaminants sont connues, alors pour  $l = 1, \dots, m$  l'hypothèse  $H_{0l}$  est inconditionnellement rejetée au niveau de signification  $\alpha$  si  $P_{\Delta_{k_l}=0|\underline{y}} < \alpha$ .

Par conséquent,  $H_0$  est inconditionnellement rejetée au niveau de signification  $\alpha$  si

$$P_{\underline{\Delta}=0|\underline{y}} := \min_{1 \leq l \leq m} \{P_{\Delta_{k_l}=0|\underline{y}}\} < \alpha \quad (3.30)$$

2. Si les positions des contaminants sont inconnues, on propose le test basé sur la  $p$ -value inconditionnelle suivant

$$P_{\Delta_{k_l}=0|\underline{y}} = 2 \sum_{\substack{j=1 \\ j \neq l}}^m \sum_{\substack{n-1 \\ k_j = k_{j-1} + 1}} \frac{1}{\prod_{\substack{j=1 \\ j \neq l}}^m (n - k_j - 1)} E_{\underline{\phi}} E_{\underline{\beta}} E_{\underline{\Delta}^{(-k_l)}} \{1 - \tau_{n-1}(|t_l(0)|)\}, \quad k_0 = 1, l = 1, \dots, m \quad (3.31)$$

Ceci nous permet de tester séparément l'existence d'un contaminant dans chaque observation de notre modèle.

**Remarque 3.3.** On peut étendre ce test au cas où le nombre de contaminants  $m$  est inconnu par conditionnement de (3.28) et (3.29) sur différentes valeurs de  $m$ .



# Chapitre 4

## Application numérique

### 4.1 Introduction

Ce chapitre est constitué de deux sections. La première section présente une étude de simulation intensive d'une part, lorsque les contaminations sont additives et d'autre part, lorsque les contaminations sont innovantes. Une comparaison avec des estimations fournies par la méthode du maximum de vraisemblance est faite. Des tests détectant les instants de contaminations fondés sur la  $p$ -value sont mis en oeuvre. L'évaluation des facteurs de Bayes, des CPO, des DIC ainsi qu'un critère empirique de convergence fondés sur les moyennes cumulées est entreprise. La prévision d'une valeur future est faite et comparée à la valeur réelle simulée. Le test de Jarque-Bera (J.B.) est utilisé pour tester la normalité des paramètres pour différentes lois du bruit.

La deuxième section concerne l'étude de la série réelle donnée dans Box et Jenkins [15], page 529, série D. Les données représentent un relevé horaire de viscosité d'un processus chimique.

### 4.2 Etude de simulation

Dans cette partie, on va présenter un exemple de simulation pour étudier l'effet et la détection de la contamination multiple  $AO$  et  $IO$  dans un modèle  $AR(1)$  avec tendance de régression linéaire simple.

### 4.2.1 Le modèle affecté par des valeurs aberrantes additives

La figure suivante représente le graphe d'une série de taille  $n = 200$  simulée à partir du modèle (3.1a) pour  $p = 1$ ,  $\beta = 4$ ,  $\phi = 0.5$ ,  $k_1 = 100$ ,  $k_2 = 120$ ,  $\Delta_{k_1} = 2$ ,  $\Delta_{k_2} = 6$  et  $\varepsilon_t$  i.i.d.  $\sim N(0, 1)$ .

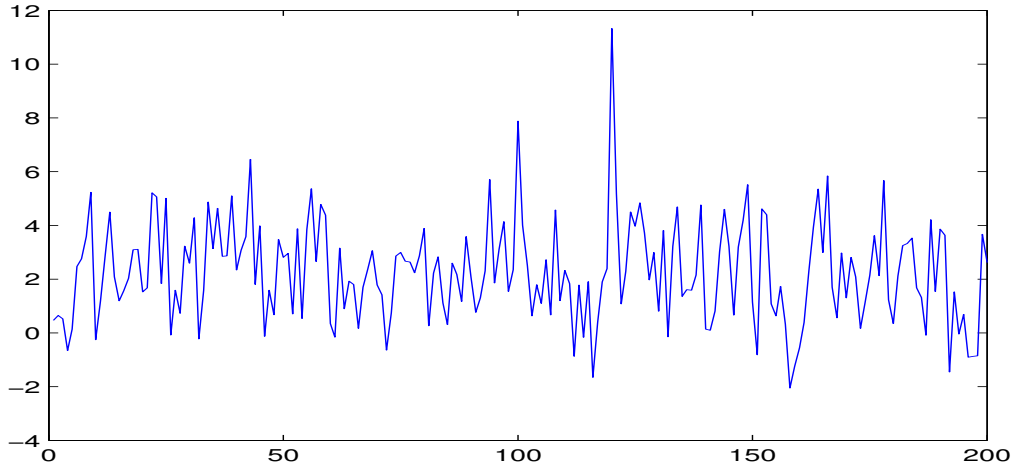


Figure 1.1. Série simulée.

La Figure 1.1 illustre les effets de  $AO$  sur les observations, on peut voir qu'il y a des chocs soudains aux temps  $k_1 = 100$  et  $k_2 = 120$  seulement, correspondants aux amplitudes  $\Delta_{k_1} = 2$  et  $\Delta_{k_2} = 6$  de la contamination  $AO$ . Le reste des observations n'est pas affecté. Maintenant, nous commençons l'étude de simulation.

1. On applique l'algorithme du Gibbs Sampler avec  $N = 1000$  répétitions. Nous approximations la moyenne a posteriori, en utilisant la fonction de coût quadratique, par  $\hat{\theta} = \frac{1}{N} \sum_{i=1}^N \theta_i^{(m)}$ , avec une période de chauffe de  $m - 1$  observations pour éviter la dépendance sur les conditions initiales, pour  $N$  suites de Gibbs de longueur  $m$ . On donne l'écart-type (std), l'erreur quadratique moyenne (rmse) et l'intervalle de crédibilité ( $CI$ ) des paramètres  $\beta$ ,  $\phi$ ,  $\Delta_{k_1}$  et  $\Delta_{k_2}$ . Les estimations des paramètres sont obtenues lorsque les vraies valeurs (T.V) sont :  $\Delta_{k_1} = 2$ ,  $\Delta_{k_2} = 6$ ,  $\phi = 0.5$  et  $\beta = 4$  pour différentes tailles d'échantillons  $n = 50$ ,  $n = 200$  et  $n = 400$ .

Puisqu'en pratique, on ne peut pas garantir que la loi du bruit blanc est gaussienne standard (symétrique, centrée et normalisée), on examine la précision et la robustesse des estimations de Gibbs quand la loi du bruit blanc est gaussienne standard  $\mathbb{N}(0, 1)$  comme

référence contre une loi de Student à 5 degrés de liberté ( $t_5$ ) et un mélange gaussien ( $0.8 \times \mathbb{N}(0, 1) + 0.2 \times \mathbb{N}(0, 2^2)$ ). Pour comparaison, on donne l'estimateur du quasi-maximum de vraisemblance (MLE) de tous les paramètres,  $\hat{\theta} = \arg \max(l(\theta|X, y))$ , où  $l(\theta|X, y)$  est obtenue à partir de (3.1c). Les résultats sont donnés dans les tableaux 4.1 – 4.3.

		$n = 50$		$n = 200$		$n = 400$	
paramètre	T.V.	Gibbs	MLE	Gibbs	MLE	Gibbs	MLE
$\beta_{(rmse)}^{(std)}$	4	4.0184 <sup>(0.3259)</sup> (0.3377) [3.4552, 4.6058]	4.0108 <sup>(0.3947)</sup> (0.4347) [3.2779, 4.7944]	4.0328 <sup>(0.1857)</sup> (0.1908) [3.6658, 4.3855]	3.9953 <sup>(0.1756)</sup> (0.1756) [3.6543, 4.5233]	3.9992 <sup>(0.1242)</sup> (0.1243) [3.6915, 4.2353]	3.9983 <sup>(0.1209)</sup> (0.1208) [3.7653, 4.2393]
$\phi_{(rmse)}^{(std)}$	0.5	0.4975 <sup>(0.1279)</sup> (0.1280) [0.2451, 0.6655]	0.4856 <sup>(0.1962)</sup> (0.2069) [0.1725, 0.8196]	0.4983 <sup>(0.0653)</sup> (0.0765) [0.3547, 0.6220]	0.4979 <sup>(0.0623)</sup> (0.0623) [0.3698, 0.6093]	0.5040 <sup>(0.0402)</sup> (0.0415) [0.4296, 0.5732]	0.4997 <sup>(0.0431)</sup> (0.0431) [0.4123, 0.5813]
$\Delta_{k_1 (rmse)}^{(std)}$	2	2.0129 <sup>(0.8146)</sup> (0.8268) [0.2697, 3.4152]	1.9885 <sup>(0.8909)</sup> (0.9508) [0.1314, 3.8899]	2.0491 <sup>(0.8413)</sup> (0.8430) [0.3010, 3.2097]	1.9645 <sup>(0.8678)</sup> (0.8684) [0.2354, 3.6658]	1.9766 <sup>(0.8308)</sup> (0.8410) [0.3413, 3.5012]	2.0048 <sup>(0.8945)</sup> (0.8940) [0.3894, 3.6011]
$\Delta_{k_2 (rmse)}^{(std)}$	6	6.0230 <sup>(0.8741)</sup> (0.8809) [4.691, 7.8392]	5.9736 <sup>(0.9796)</sup> (0.9998) [4.3047, 7.7796]	6.0692 <sup>(0.8638)</sup> (0.8608) [4.7513, 7.7898]	5.9565 <sup>(0.8826)</sup> (0.8828) [4.5907, 7.5337]	5.9867 <sup>(0.8214)</sup> (0.8222) [4.8205, 7.6824]	5.9914 <sup>(0.9159)</sup> (0.9154) [4.6063, 7.4163]

Table 4.1. Résultats d'estimation quand le bruit blanc est gaussien  $N(0, 1)$ .

		$n = 50$		$n = 200$		$n = 400$	
paramètre	T.V.	Gibbs	MLE	Gibbs	MLE	Gibbs	MLE
$\beta_{(rmse)}^{(std)}$	4	3.9867 <sup>(0.4039)</sup> (0.3411) [3.1950, 4.9167]	3.9883 <sup>(0.4941)</sup> (0.5040) [2.9976, 4.9226]	4.0191 <sup>(0.2156)</sup> (0.2401) [3.6193, 4.5085]	3.9990 <sup>(0.2243)</sup> (0.2242) [3.5499, 4.4714]	4.0173 <sup>(0.1573)</sup> (0.1989) [3.7291, 4.2705]	4.0075 <sup>(0.1627)</sup> (0.1628) [3.6914, 4.3388]
$\phi_{(rmse)}^{(std)}$	0.5	0.4834 <sup>(0.1203)</sup> (0.1355) [0.2096, 0.6948]	0.4849 <sup>(0.1990)</sup> (0.2498) [0.1729, 0.7114]	0.4960 <sup>(0.0623)</sup> (0.0780) [0.3940, 0.6140]	0.4950 <sup>(0.0622)</sup> (0.0624) [0.3660, 0.6114]	0.4908 <sup>(0.0417)</sup> (0.0562) [0.4223, 0.5726]	0.4972 <sup>(0.0440)</sup> (0.0441) [0.4026, 0.5857]
$\Delta_{k_1 (rmse)}^{(std)}$	2	2.0052 <sup>(1.0512)</sup> (0.9401) [-0.1220, 4.3084]	2.0637 <sup>(1.1992)</sup> (1.2391) [-0.1048, 4.6009]	2.0967 <sup>(1.1000)</sup> (0.9311) [0.0215, 4.2002]	1.9854 <sup>(1.1421)</sup> (1.1460) [0.1442, 4.4105]	2.0469 <sup>(1.0971)</sup> (0.9280) [0.2480, 4.1131]	2.0126 <sup>(1.1126)</sup> (1.1022) [0.2186, 4.2008]
$\Delta_{k_2 (rmse)}^{(std)}$	6	5.9908 <sup>(1.1227)</sup> (1.0601) [3.7981, 7.9838]	6.0335 <sup>(1.1831)</sup> (1.1743) [3.6190, 8.2662]	6.0916 <sup>(1.0933)</sup> (1.0100) [3.9765, 7.2901]	5.9732 <sup>(1.0772)</sup> (1.0770) [3.9188, 7.3634]	6.0119 <sup>(0.9267)</sup> (0.9209) [4.1707, 7.2235]	5.9862 <sup>(0.9044)</sup> (0.9161) [3.9878, 7.2831]

Table 4.2. Résultats d'estimation quand le bruit blanc suit une loi de Student  $t_5$ .

		$n = 50$		$n = 200$		$n = 400$	
paramètre	T.V.	Gibbs	MLE	Gibbs	MLE	Gibbs	MLE
$\beta_{(rmse)}^{(std)}$	4	3.9931 <sup>(0.3533)</sup> (0.3731) [3.1374, 4.9836]	4.0223 <sup>(0.4814)</sup> (0.4917) [3.0884, 4.9939]	4.0022 <sup>(0.2242)</sup> (0.2346) [3.6319, 4.4939]	4.0093 <sup>(0.2190)</sup> (0.2193) [3.6043, 4.4625]	4.0030 <sup>(0.1567)</sup> (0.1662) [3.6606, 4.3379]	4.0053 <sup>(0.1511)</sup> (0.1511) [3.7112, 4.3039]
$\phi_{(rmse)}^{(std)}$	0.5	0.4879 <sup>(0.1063)</sup> (0.1066) [0.1975, 0.7004]	0.4809 <sup>(0.2090)</sup> (0.2673) [0.1762, 0.7127]	0.4950 <sup>(0.0741)</sup> (0.0766) [0.3293, 0.6089]	0.4939 <sup>(0.0645)</sup> (0.0647) [0.3649, 0.6108]	0.4976 <sup>(0.0445)</sup> (0.0453) [0.4177, 0.5803]	0.4989 <sup>(0.0432)</sup> (0.0432) [0.4118, 0.5808]
$\Delta_{k_1 (rmse)}^{(std)}$	2	1.9976 <sup>(1.1247)</sup> (1.1303) [-0.1031, 4.0756]	2.0407 <sup>(1.2863)</sup> (1.2968) [-0.2793, 4.2814]	1.998 <sup>(1.0928)</sup> (1.1064) [0.1520, 4.0054]	2.0323 <sup>(1.1597)</sup> (1.1598) [0.0199, 4.1473]	1.9990 <sup>(1.0203)</sup> (1.0246) [0.2202, 3.9783]	1.9907 <sup>(1.1035)</sup> (1.1116) [0.3157, 3.8826]
$\Delta_{k_2 (rmse)}^{(std)}$	6	6.0088 <sup>(1.0088)</sup> (1.0198) [3.8687, 8.6828]	5.9494 <sup>(1.1898)</sup> (1.1799) [3.5398, 8.3619]	6.0017 <sup>(1.0001)</sup> (1.0091) [3.8149, 7.7579]	6.0060 <sup>(1.0960)</sup> (1.0996) [3.7132, 8.3250]	6.0071 <sup>(0.9704)</sup> (0.9806) [3.9456, 7.7356]	5.9852 <sup>(0.9870)</sup> (0.9876) [3.9215, 8.3027]

Table 4.3. Résultats d'estimation quand le bruit blanc est un mélange gaussien

$$0.8 \times \mathbb{N}(0, 1) + 0.2 \times \mathbb{N}(0, 2^2).$$

2. Les tables 4.4 – 4.6 présentent les résultats d'estimation des paramètres quand  $\phi$  est proche

de la région de non-stationnarité ( $\phi = 0.95$ ),  $\Delta_{k_1} = 2$  et  $\Delta_{k_2} = 6$ , pour  $n = 50$ ,  $n = 200$  et  $n = 400$ .

		$n = 50$		$n = 200$		$n = 400$	
paramètre	T.V.	Gibbs	MLE	Gibbs	MLE	Gibbs	MLE
$\beta_{(rmse)}^{(std)}$	4	4.0032 <sup>(0.3044)</sup> (0.3045) [3.2893,4.5510]	4.0104 <sup>(0.3936)</sup> (0.4136) [3.2242,4.7674]	4.0012 <sup>(0.1610)</sup> (0.1709) [3.6650,4.2995]	3.9896 <sup>(0.1766)</sup> (0.1768) [3.6329,4.3185]	4.0016 <sup>(0.1284)</sup> (0.1351) [3.7033,4.2485]	3.9901 <sup>(0.1237)</sup> (0.1241) [3.7482,4.2443]
$\phi_{(rmse)}^{(std)}$	0.95	0.9273 <sup>(0.0613)</sup> (0.0587) [0.7315,0.9881]	0.9179 <sup>(0.1009)</sup> (0.1178) [0.7252,0.9900]	0.9398 <sup>(0.0236)</sup> (0.0320) [0.8934,0.9778]	0.9404 <sup>(0.0251)</sup> (0.0269) [0.8821,0.9765]	0.9448 <sup>(0.0157)</sup> (0.0154) [0.9139,0.9708]	0.9462 <sup>(0.0166)</sup> (0.0170) [0.9072,0.9717]
$\Delta_{k_1}^{(std)}$	2	2.0012 <sup>(0.7182)</sup> (0.7191) [0.7473,3.3851]	2.0033 <sup>(0.7969)</sup> (0.8509) [0.6184,3.4269]	1.9875 <sup>(0.7008)</sup> (0.7049) [0.7884,3.2044]	2.0271 <sup>(0.7012)</sup> (0.7024) [0.7037,3.3244]	2.0363 <sup>(0.6655)</sup> (0.6680) [0.8173,3.1991]	2.0146 <sup>(0.6508)</sup> (0.6519) [0.7766,3.2883]
$\Delta_{k_2}^{(std)}$	6	5.9861 <sup>(0.6560)</sup> (0.6572) [4.6150,7.1798]	5.9752 <sup>(0.7856)</sup> (0.7992) [4.5232,7.4347]	6.0673 <sup>(0.6278)</sup> (0.6289) [4.6982,7.1682]	5.9931 <sup>(0.7148)</sup> (0.7150) [4.5491,7.4041]	6.0618 <sup>(0.6047)</sup> (0.6050) [4.7268,7.1049]	5.9886 <sup>(0.6012)</sup> (0.6015) [4.7002,7.0019]

Table 4.4. Résultats d'estimation quand le bruit blanc est gaussien  $N(0, 1)$ .

		$n = 50$		$n = 200$		$n = 400$	
paramètre	T.V.	Gibbs	MLE	Gibbs	MLE	Gibbs	MLE
$\beta_{(rmse)}^{(std)}$	4	4.0287 <sup>(0.4041)</sup> (0.4106) [3.0949,4.8699]	3.9760 <sup>(0.5674)</sup> (0.5877) [2.9989,4.9124]	4.0488 <sup>(0.2373)</sup> (0.2375) [3.5265,4.4012]	3.9962 <sup>(0.2293)</sup> (0.2293) [3.5332,4.4442]	4.0212 <sup>(0.1493)</sup> (0.1510) [3.8052,4.2885]	4.0010 <sup>(0.1598)</sup> (0.1599) [3.6851,4.3161]
$\phi_{(rmse)}^{(std)}$	0.95	0.9363 <sup>(0.0510)</sup> (0.0512) [0.7602,0.9921]	0.9213 <sup>(0.1150)</sup> (0.1165) [0.7538,0.9900]	0.9341 <sup>(0.0320)</sup> (0.0449) [0.8555,0.9757]	0.9412 <sup>(0.0258)</sup> (0.0271) [0.8747,0.9771]	0.9468 <sup>(0.0181)</sup> (0.0171) [0.9051,0.9724]	0.9454 <sup>(0.0166)</sup> (0.0173) [0.9086,0.9709]
$\Delta_{k_1}^{(std)}$	2	1.9980 <sup>(0.9239)</sup> (0.9241) [0.4280,4.0583]	1.9922 <sup>(0.9894)</sup> (0.9989) [0.2593,3.8790]	2.0026 <sup>(0.8418)</sup> (0.8506) [0.4367,3.5304]	2.0015 <sup>(0.9573)</sup> (0.9568) [0.3540,3.8123]	1.9978 <sup>(0.7904)</sup> (0.7525) [0.5388,3.5212]	2.0082 <sup>(0.9247)</sup> (0.9253) [0.3160,3.8009]
$\Delta_{k_2}^{(std)}$	6	6.0161 <sup>(0.8343)</sup> (0.8501) [4.3832,7.4636]	5.9915 <sup>(0.9273)</sup> (0.9769) [4.1038,7.7173]	6.0507 <sup>(0.8041)</sup> (0.8050) [4.4265,7.3419]	5.9956 <sup>(0.9001)</sup> (0.9002) [4.3172,7.6930]	6.0307 <sup>(0.7762)</sup> (0.7829) [4.6017,7.2120]	5.9921 <sup>(0.8506)</sup> (0.8520) [4.5180,7.7066]

Table 4.5. Résultats d'estimation quand le bruit blanc suit une loi de Student  $t5$ .

		$n = 50$		$n = 200$		$n = 400$	
paramètre	T.V.	Gibbs	MLE	Gibbs	MLE	Gibbs	MLE
$\beta_{(rmse)}^{(std)}$	4	4.0199 <sup>(0.4585)</sup> (0.4569) [3.0837,4.9830]	4.0873 <sup>(0.4886)</sup> (0.4984) [3.0834,5.0069]	4.0207 <sup>(0.2283)</sup> (0.2566) [3.6283,4.5101]	3.9986 <sup>(0.2393)</sup> (0.2391) [3.5349,4.4866]	4.0278 <sup>(0.1504)</sup> (0.1586) [3.7462,4.3283]	3.9999 <sup>(0.1590)</sup> (0.1589) [3.6988,4.3202]
$\phi_{(rmse)}^{(std)}$	0.95	0.9238 <sup>(0.0610)</sup> (0.0615) [0.7898,0.9960]	0.9189 <sup>(0.0771)</sup> (0.0939) [0.7611,0.9900]	0.9395 <sup>(0.0259)</sup> (0.0347) [0.8811,0.9801]	0.9413 <sup>(0.0270)</sup> (0.0284) [0.8703,0.9776]	0.9481 <sup>(0.0175)</sup> (0.0202) [0.9026,0.9696]	0.9461 <sup>(0.0165)</sup> (0.0170) [0.9069,0.9709]
$\Delta_{k_1}^{(std)}$	2	2.0129 <sup>(0.9146)</sup> (0.9150) [0.1089,3.6890]	2.0182 <sup>(1.0751)</sup> (1.0957) [0.1248,3.7517]	2.0180 <sup>(0.9006)</sup> (0.9018) [0.1601,3.6018]	1.9728 <sup>(0.8910)</sup> (0.9091) [0.1566,3.7052]	2.0508 <sup>(0.8806)</sup> (0.8863) [0.3908,3.5041]	2.0127 <sup>(0.8776)</sup> (0.8811) [0.2008,3.6864]
$\Delta_{k_2}^{(std)}$	6	5.9749 <sup>(0.8001)</sup> (0.8109) [4.2645,7.9099]	5.9548 <sup>(0.9785)</sup> (0.9882) [4.0315,7.9332]	5.9707 <sup>(0.7806)</sup> (0.7990) [4.3618,7.7686]	6.0634 <sup>(0.8600)</sup> (0.8628) [4.4161,7.8338]	5.9981 <sup>(0.7445)</sup> (0.7506) [4.4728,7.5506]	6.0166 <sup>(0.8580)</sup> (0.8591) [4.4545,7.6671]

Table 4.6. Résultats d'estimation quand le bruit blanc est un mélange gaussien

$$0.8 \times N(0, 1) + 0.2 \times N(0, 2^2).$$

On note que, pour les trois types de bruits blancs et pour les deux valeurs de  $\phi$ , les deux méthodes d'estimation, Gibbs Sampler et l'estimation du maximum de vraisemblance

(*MLE*), donnent de bons résultats ; quand la taille de l'échantillon augmente, le *rmse* et le *std* des estimations diminuent.

En comparant les deux méthodes, on note que, quand la taille de l'échantillon est petite, la méthode bayésienne est meilleure que la méthode classique sauf pour  $\Delta_{k_1} = 2$ , quand la distribution est à queue lourde pour laquelle les deux méthodes ne donnent pas de résultats satisfaisants quand  $\phi = 0.5$ . Par contre quand  $\phi = 0.95$ , Le Gibbs Sampler semble être meilleur. Quand la taille de l'échantillon est grande, les deux méthodes donnent des résultats presque similaires.

On note aussi que, les *CI* deviennent étroitement concentrés autour des vraies valeurs pour les deux paramètres  $\beta$  et  $\phi$  même dans le cas presque non stationnaire, lorsque la taille de l'échantillon devient plus grande. Par contre pour les amplitudes des contaminants, les *CI* restent plus grands.

Concernant la convergence des chaînes, des tests de diagnostic plus rigoureux ont été développés dans la littérature et peuvent être mis en oeuvre à l'aide du package CODA (Convergence diagnosis and output analysis software for Gibbs Sampler analysis). CODA fournit quatre tests suggérés respectivement par Geweke [36], Gelman et Rubin [34], Raftery et Lewis [66], et Heidelberger et Welch [45].

Quand l'erreur est gaussienne, on donne dans la figure suivante les courbes des moyennes cumulées de  $N$  échantillons, des estimations des paramètres  $\beta$ ,  $\phi$ ,  $\Delta_{k_1}$  et  $\Delta_{k_2}$ .

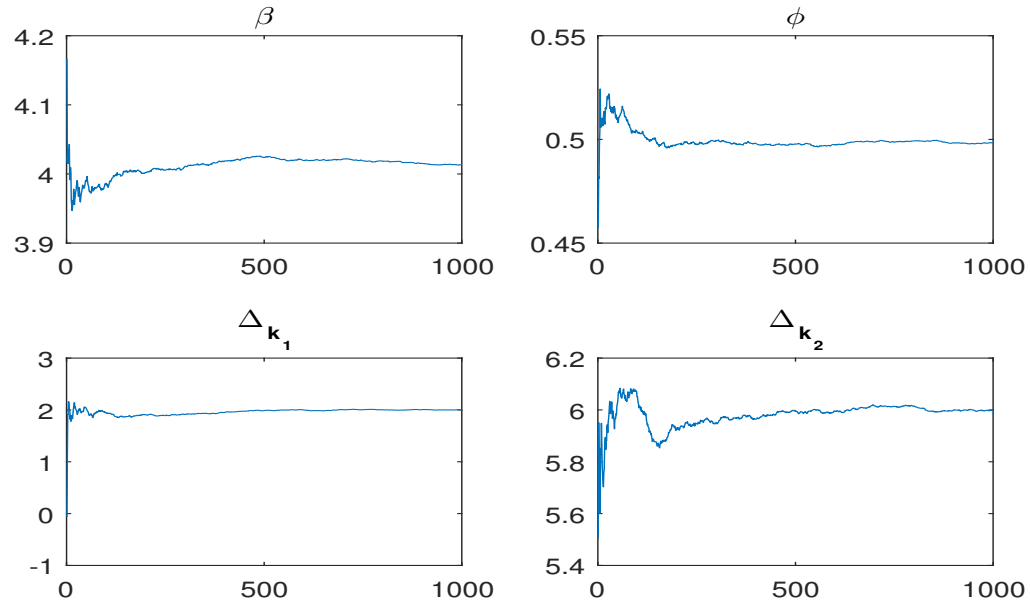


Figure 4.2. Courbes des moyennes cumulées de  $N = 1000$  échantillons, pour  $\beta = 4$ ,  $\phi = 0.5$ ,  $\Delta_{k_1} = 2$ ,  $\Delta_{k_2} = 6$ .

On note que, les estimations se stabilisent aux vraies valeurs, à partir de 500 répétitions. Pour plus de détails théoriques sur les diagnostics de convergence de MCMC, on peut voir, e.g., Roy [71].

3. On applique l'algorithme du Gibbs Sampler avec 1000 répétitions pour  $n = 200$  et  $\phi = 0.5$  pour calculer la  $p$ -value inconditionnelle de l'hypothèse  $H_0 : \Delta_{k_l} = 0$ ,  $l \in \{1, 2\}$  pour différentes valeurs de  $\sigma$  et  $(\Delta_{k_1}, \Delta_{k_2})$  (petites, modérées et grandes valeurs).

Les tests basés sur les  $p$ -values sont critiqués dans la littérature (voir, e.g., Benjamin et al. [11]). Mais des alternatives réalisables aux  $p$ -values ne sont pas spécifiées. Benjamin et Berger [12] suggèrent d'utiliser  $\bar{B}$  ( $\bar{B} = -\frac{1}{ep \ln p}$ , où  $p$  est la  $p$ -value) qui est une borne supérieure, sur tout choix (raisonnable) de la loi a priori pour l'hypothèse alternative, des facteurs de

Bayes (voir, Kass et Raftery [52])  $BF_{10} = \frac{\int \prod_{i=1}^N f_{H_1}(y_i | \theta, \underline{y}_{i-1}, X_i) \pi_{H_1}(\theta) d\theta}{\int \prod_{i=1}^N f_{H_0}(y_i | \theta, \underline{y}_{i-1}, X_i) \pi_{H_0}(\theta) d\theta}$ , ( $BF_{01} = (BF_{10})^{-1}$ ),

où  $\underline{y}_{i-1} = (y_{i-1}, \dots, y_{i-p})'$ ,  $X_i = (\underline{x}'_i, \underline{x}'_{i-1}, \dots, \underline{x}'_{i-p})'$  et  $\underline{y}_0$  est le vecteur des valeurs initiales.

Dans les tables 4.7 – 4.9, on a ajouté la valeur de cette borne supérieure, sous la valeur de la  $p$ -value, quand cette  $p$ -value est inférieure à 0.05, et quand la loi du bruit blanc est gaussienne.

$\sigma = 0.1$	$k_1 - 2$	$k_1 - 1$	$k_1$	$k_1 + 1$	$k_1 + 2$		$k_2 - 2$	$k_2 - 1$	$k_2$	$k_2 + 1$	$k_2 + 2$
$\Delta_{k_1} = 0.5$	0.2408	0.2084	0.0335 (3.2335)	0.6389	0.4151	$\Delta_{k_2} = 1$	0.3301	0.1086	0.00009 (438.8710)	0.7551	0.3891
$\Delta_{k_1} = 2$	0.56797	0.0220 (4.3812)	0.00009 (438.7810)	0.0404 (2.8377)	0.6611	$\Delta_{k_2} = 4$	0.6871	0.0155 (5.6959)	0.00009 (438.7810)	0.0293 (3.5567)	0.6085
$\Delta_{k_1} = 8$	0.8787	0.0062 (11.6728)	0.00009 (438.7810)	0.0097 (8.1814)	0.8864	$\Delta_{k_2} = 15$	0.9805	0.0004 (117.5478)	0.00009 (438.7810)	0.0045 (15.1288)	0.8541

Table 4.7. La  $p$ -value inconditionnelle de  $H_0$ ,  $P_{\Delta_{k_l}=0} | \underline{y}$  pour  $\sigma = 0.1$ .

$\sigma = 0.5$	$k_1 - 2$	$k_1 - 1$	$k_1$	$k_1 + 1$	$k_1 + 2$		$k_2 - 2$	$k_2 - 1$	$k_2$	$k_2 + 1$	$k_2 + 2$
$\Delta_{k_1} = 0.5$	0.9449	0.7812	0.4033	0.5938	0.9360	$\Delta_{k_2} = 1$	0.4402	0.0810	0.00009 (438.7810)	0.0212 (4.5028)	0.6255
$\Delta_{k_1} = 2$	0.1690	0.0808	0.00009 (438.7810)	0.0449 (2.6402)	0.2317	$\Delta_{k_2} = 4$	0.7245	0.0339 (3.2065)	0.00009 (438.7810)	0.0479 (2.5275)	0.1941
$\Delta_{k_1} = 8$	0.8632	0.0098 (8.1158)	0.00009 (438.7810)	0.0211 (4.5186)	0.4190	$\Delta_{k_2} = 15$	0.9508	0.0010 (53.2560)	0.00009 (438.7810)	0.0025 (24.5603)	0.7241

Table 4.8. La  $p$ -value inconditionnelle de  $H_0$ ,  $P_{\Delta_{k_l}=0} | \underline{y}$  pour  $\sigma = 0.5$ .

$\sigma = 1$	$k_1 - 2$	$k_1 - 1$	$k_1$	$k_1 + 1$	$k_1 + 2$		$k_2 - 2$	$k_2 - 1$	$k_2$	$k_2 + 1$	$k_2 + 2$
$\Delta_{k_1} = 0.5$	0.1759	0.4780	0.2042	0.3646	0.3909	$\Delta_{k_2} = 1$	0.1231	0.7433	0.1134	0.2127	0.8529
$\Delta_{k_1} = 2$	0.7493	0.2928	0.0020 (29.5980)	0.1245	0.7861	$\Delta_{k_2} = 4$	0.4361	0.0625	0.0002 (215.9630)	0.0471 (2.5563)	0.7212
$\Delta_{k_1} = 8$	0.8880	0.0495 (2.4725)	0.00009 (438.7810)	0.0297 (3.5223)	0.7625	$\Delta_{k_2} = 15$	0.2665	0.0034 (19.0359)	0.00009 (438.7810)	0.0015 (37.7180)	0.8541

Table 4.9. La  $p$ -value inconditionnelle de  $H_0$ ,  $P_{\Delta_{k_l}=0} | \underline{y}$  pour  $\sigma = 1$ .

On note que, quand la variance du bruit est petite, le test de la  $p$ -value peut détecter même de petites amplitudes de contamination.

On note aussi que, le test de la  $p$ -value détecte les positions  $k_1$  et  $k_2$  des contaminants ; il détecte aussi les positions  $k_1 - 1$ ,  $k_1 + 1$ ,  $k_2 - 1$  et  $k_2 + 1$ . Ceci est due au fait que  $\tilde{\Delta}_{k_l}$  dans le test statistique des additive outliers à l’instant  $k_l$ , dépend seulement des observations  $y_{k_l}$ ,  $y_{k_l-1}$  et  $y_{k_l+1}$ ,  $l = 1, \dots, m$ .

Par exemple, dans la Table 4.7, la  $p$ -value correspondante à  $k_1$ , pour  $\Delta_{k_1} = 2$  est 0.00009 qui correspond à un facteur de Bayes d’au plus 438.8262 :1. C’est-à-dire que les données

impliquent des chances en faveur de l'hypothèse alternative par rapport à l'hypothèse nulle de 438,8262 sur 1 au maximum, qui est interprété comme une décision évidente contre l'hypothèse nulle (Kass et Raftery [52]).

**Remarque 4.1.** *Même pour un échantillon de petite taille, le test basé sur la p-value détecte la présence de valeurs aberrantes.*

4. On donne aussi les résultats de simulations pour d'autres critères de discrimination comme le critère d'information de déviance (DIC) (voir, Spiegelhalter et al. [80], [81]), le facteur de Bayes (BF) (Kass et Raftery [52]) et l'ordonnée prédictive conditionnelle (CPO) (Geisser [30]; voir aussi Pettit [65]).

- On compare le  $DIC_0$  sous l'hypothèse nulle  $H_0$  avec le  $DIC_1$  sous l'hypothèse alternative  $H_1$ , pour différentes valeurs de  $(\Delta_{k_1}, \Delta_{k_2})$  et différents instants de contaminations  $k_1$  et  $k_2$ . L'algorithme de Gibbs Sampler est appliqué avec 1000 répétitions; les résultats sont donnés dans les tableaux suivants pour le modèle  $AO$  ( $n = 200, \beta = 4, \phi = 0.5, \sigma = 1$ ).

	$k_1 = 50, k_2 = 100$			$k_1 = 150, k_2 = 199$		
$(\Delta_{k_1}, \Delta_{k_2})$	(4,8)	(2,6)	(1,2)	(4,8)	(2,6)	(1,2)
$DIC_0$	690.3707	613.4034	577.8949	642.6670	599.4986	565.8701
$DIC_1$	597.4558	579.2586	574.6621	587.7370	569.3011	562.8820

Table 4.10. Le  $DIC_0$  et  $DIC_1$  pour le modèle  $AO$ .

On note que quand l'amplitude de la contamination décroît, les valeurs de  $DIC_1$  et  $DIC_0$  diminuent et  $DIC_1 < DIC_0$  pour n'importe quelle valeur de  $(\Delta_{k_1}, \Delta_{k_2})$ . Ceci montre que le modèle sous l'hypothèse  $H_1$  est le modèle préféré comparé à celui sous l'hypothèse  $H_0$ .

- Le facteur de Bayes  $BF_{10}$  est utilisé pour comparer deux hypothèses  $H_1$  et  $H_0$ . Les résultats sont donnés dans la table 4.11, pour 1000 répétitions de l'échantillonneur de Gibbs et différentes valeurs de  $\Delta_{k_1}$  et  $\Delta_{k_2}$  ( $n = 200, \beta = 4, \phi = 0.5, \sigma = 1$ ).

$(\Delta_{k_1}, \Delta_{k_2})$	(4,8)	(2,6)	(1,2)
$BF_{10}$	387.3273	116.2647	2.6562

Table 4.11. Le facteur de Bayes pour le modèle  $AO$ .



Dans la table 4.11, on note que, quand l'amplitude de la contamination est assez grande, la valeur de  $BF_{10}$  est importante. Quand l'amplitude de la contamination décroît, les valeurs de  $BF_{10}$  diminuent et sont toujours  $> 1$ . Les valeurs ci-dessus du facteur de Bayes sont interprétés comme une évidence décisive contre l'hypothèse nulle pour  $(\Delta_{k_1}, \Delta_{k_2}) = (4, 8)$ , l'évidence est très forte pour  $(\Delta_{k_1}, \Delta_{k_2}) = (2, 6)$  mais on conclut à une faible évidence en faveur de  $H_1$  pour  $(\Delta_{k_1}, \Delta_{k_2}) = (1, 2)$  (voir, par exemple, Held et Ott, [46] et les références infra).

- Les résultats des simulations de CPO sont donnés dans la table 4.12, au voisinage de  $k_1$  et  $k_2$ , pour 1000 répétitions de l'échantillonneur de Gibbs ( $n = 100, \beta = 4, \phi = 0.5, \Delta_{k_1} = 4, \Delta_{k_2} = 8$ ) pour  $\sigma = 0.1$  et  $\sigma = 1$

$i$	$k_1 - 1$	$k_1$	$k_1 + 1$	$k_2 - 1$	$k_2$	$k_2 + 1$
$\sigma = 0.1$	0.4348	0.0000	0.3865	0.4354	0.0000	0.2750
$\sigma = 1$	0.2624	0.0013	0.1995	0.1663	0.0000	0.1167

Table 4.12. CPO pour le modèle  $AO$ .

Les figures suivantes donnent les courbes du CPO pour le modèle  $AO$  avec des amplitudes de contamination  $\Delta_{k_1} = 4, \Delta_{k_2} = 8, \sigma = 0.1$  et  $\sigma = 1$  aux instants  $k_1 = 50$  et  $k_2 = 80$ .

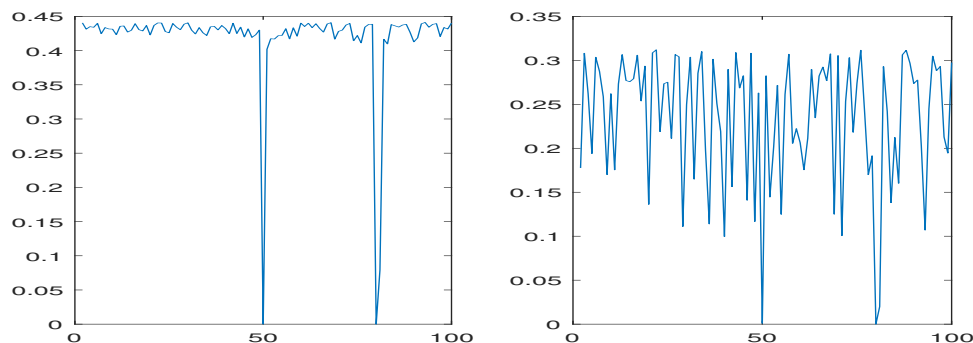


Figure 4.3. Courbe de  $CPO$  ( $k_1 = 50, k_2 = 80$ ) pour  $\sigma = 0.1$  et  $\sigma = 1$  respectivement.

Il est clair que l'instant de contamination est détecté aux instants  $k_1 = 50$  et  $k_2 = 80$ . Ceci est illustré dans les figures par une faible valeur de CPO et on le voit clairement même lorsque  $\sigma = 1$ .

5. Pour  $n = 100$ ,  $\beta = 4$ ,  $\phi = 0.5$ ,  $\Delta_{k_1} = 4$ ,  $\Delta_{k_2} = 8$ ,  $\sigma = 1$  et différents instants  $k_1$  et  $k_2$  de contamination. La valeur prédite  $y_{n+1}$  est estimée comme la moyenne a posteriori ( $y_{n+1}^{Gibbs}(H_i)$ ),  $i = 0, 1$  de la densité prédictive. Les intervalles de prédiction  $PI(H_i)$ ,  $i = 0, 1$  sont donnés en utilisant l'algorithme de Gibbs Sampler pour 1000 répétitions, sous les hypothèses  $H_0$  (absence des valeurs aberrantes) et  $H_1$  (présence des valeurs aberrantes). Les résultats de simulation sont donnés dans la table suivante pour le modèle  $AO$ . Le  $rmse$  et le  $std$  sont calculés et indiqués comme suit :  $y_{n+1}^{Gibbs}(H_0)_{(rmse)}^{(std)}$  et  $y_{n+1}^{Gibbs}(H_1)_{(rmse)}^{(std)}$ . La vraie valeur  $y_T$  de  $y_{n+1}$  est obtenue en simulant  $y_{n+1}$  plusieurs fois à partir du modèle et en prenant la moyenne. On considère  $y_{n+1}$  en utilisant l'algorithme  $MCMC$ , comme un paramètre inconnu.

	$y_T$	$y_{n+1}^{Gibbs}(H_0)$	$y_{n+1}^{Gibbs}(H_1)$	$PI(H_0)$	$PI(H_1)$
$k_1 = 20, k_2 = 30$	2.0042	2.0749 <sub>(1.0581)</sub> <sup>(1.1944)</sup>	2.0324 <sub>(1.0390)</sub> <sup>(1.0860)</sup>	[0.1506, 4.1123]	[0.2941, 3.9368]
$k_1 = 50, k_2 = 70$	1.9601	2.0175 <sub>(1.0169)</sub> <sup>(1.2138)</sup>	1.9656 <sub>(0.9925)</sub> <sup>(1.0922)</sup>	[0.1068, 4.1900]	[0.3470, 3.9701]
$k_1 = 90, k_2 = 99$	1.9931	2.1014 <sub>(1.0730)</sub> <sup>(1.2395)</sup>	2.0004 <sub>(1.0233)</sub> <sup>(1.1501)</sup>	[0.0936, 4.2130]	[0.2930, 3.9843]

Table 4.13. Observation prédite pour le modèle  $AO$  pour  $n = 100$  et différents instants  $k_1$  et  $k_2$  de contamination.

Pour tous les instants de contamination, on note que la valeur estimée  $y_{n+1}^{Gibbs}(H_1)$  de  $y_{n+1}$  par l'algorithme de Gibbs Sampler sous l'hypothèse  $H_1$  est plus proche de  $y_T$  que celle estimée sous l'hypothèse  $H_0$ . On note aussi que les valeurs de  $rmse$  et  $std$  sont plus petites sous l'hypothèse  $H_1$ . La longueur de l'intervalle de prédiction est plus petite sous  $H_1$ . On déduit que le modèle le plus robuste est celui qui prend en compte la présence des valeurs aberrantes.

Quelque soit l'instant de contamination,  $y_T$ ,  $y_{n+1}^{Gibbs}(H_0)$  et  $y_{n+1}^{Gibbs}(H_1)$  ne sont pas affectées.

6. Les figures suivantes montre la courbe de la loi normale (en rouge) et celle de la densité estimée par la méthode du noyau (en bleu) des estimations des paramètres pour  $n = 200$ ,  $\beta = 4$ ,  $\sigma = 1$ ,  $\Delta_{k_1} = 2$  et  $\Delta_{k_2} = 6$  et différentes valeurs de  $\phi$  ( $\phi = 0.5$  et  $\phi = 0.95$ ), pour 1000 répétitions.

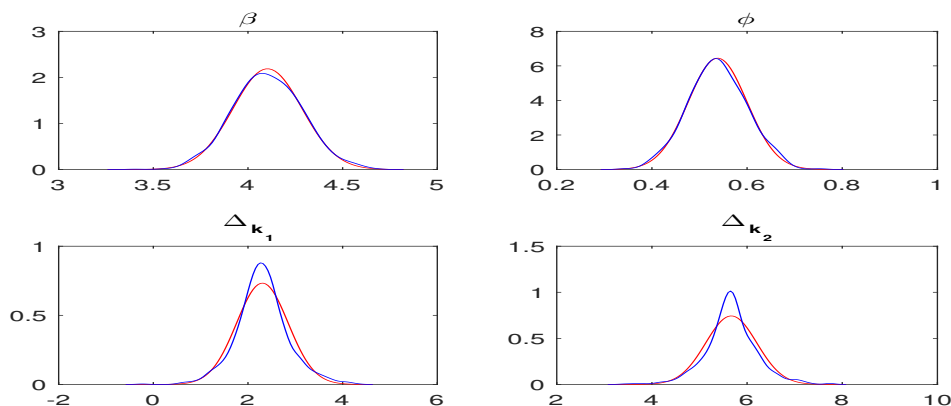


Figure 4.4. Courbe de la loi gaussienne et courbe de la densité estimée par la méthode du noyau pour  $\phi = 0.5$ ,  $\beta = 4$ ,  $\sigma = 1$ ,  $\Delta_{k_1} = 2$ ,  $\Delta_{k_2} = 6$ .

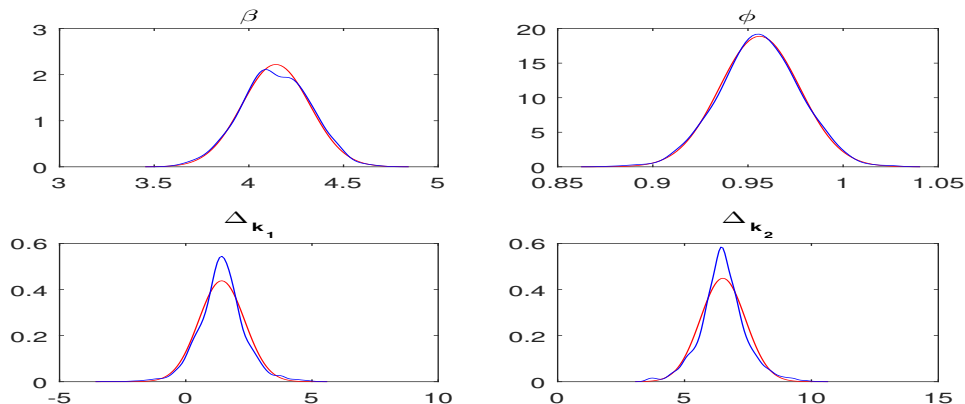


Figure 4.5. Courbe de la loi gaussienne et courbe de la densité par la méthode du noyau pour  $\phi = 0.95$ ,  $\beta = 4$ ,  $\sigma = 1$ ,  $\Delta_{k_1} = 2$ ,  $\Delta_{k_2} = 6$ .

On note que, les distributions des estimations des paramètres sont concentrées autour de la vraie valeur du paramètre pour toutes les valeurs de  $\phi$ .

## 7. Test de Jarque-Bera (J.B.) pour le modèle AO

Pour comparer les résultats des simulations obtenus pour différentes lois du bruit blanc, on représente les estimations des paramètres estimés par un diagramme en boîtes comme le montre la figure 4.6.

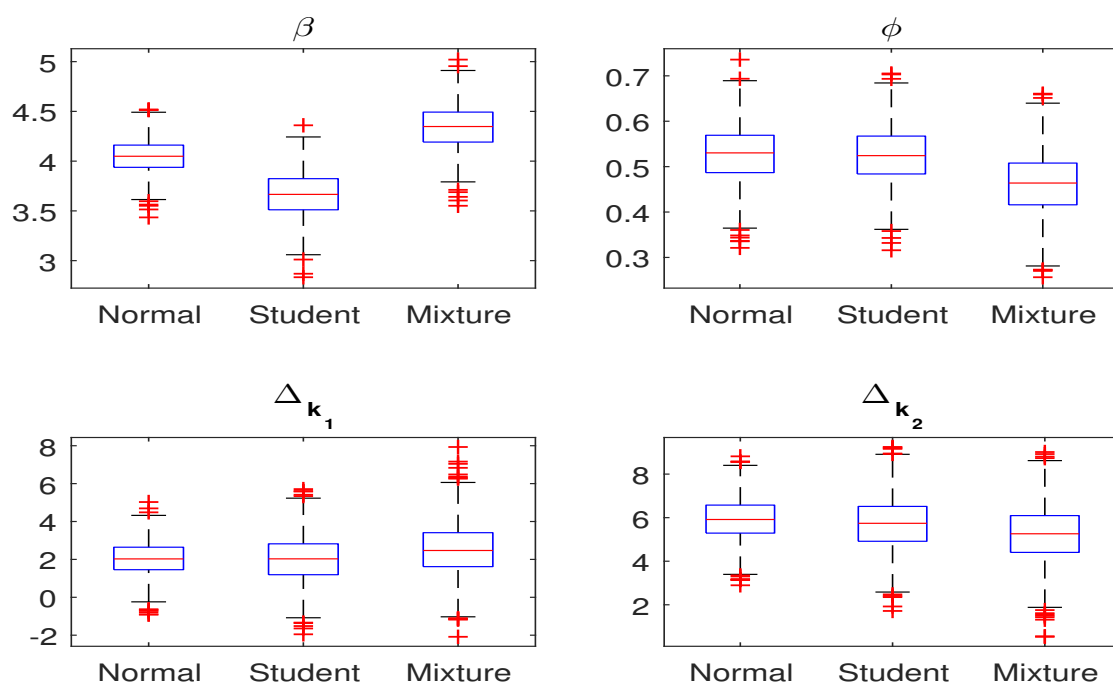


Figure 4.6. Diagrammes en boites pour  $N(0, 1)$ ,  $t_5$  et mélange gaussien  $0.8 \times N(0, 1) + 0.2 \times N(0, 2^2)$ .

Le test de Jarque-Bera (J.B.), nous permet de tester la normalité des paramètres  $\beta$ ,  $\phi$ ,  $\Delta_{k_1}$  et  $\Delta_{k_2}$ . Les résultats obtenus à partir de la figure 4.6, par simulation de la statistique de J.B., les coefficients d'asymétrie  $\gamma_1$  et d'aplatissement  $\gamma_2$  ( $J.B.\gamma_1^2$ ) sont donnés dans la table 4.14.

Pour  $n = 200$  et quand  $\beta = 4$ ,  $\phi = 0.5$ ,  $\sigma = 1$ ,  $\Delta_{k_1} = 2$ ,  $\Delta_{k_2} = 6$ , on compare la statistique de J.B. calculée, pour 1000 répétitions, avec 5.99 qui est la valeur donnée par la table de khi-deux avec 2 degrés de liberté pour  $\alpha = 0.05$ .

Paramètres	Gaussienne	$t_5$	Mixture
$\beta$	1.5200 <sup>(0.0951)</sup> (2.9834)	0.0146 <sup>(0.0067)</sup> (2.9785)	0.4760 <sup>(0.0518)</sup> (3.0264)
$\phi$	3.7770 <sup>(0.1441)</sup> (3.0868)	0.2877 <sup>(-0.0229)</sup> (2.9801)	1.0582 <sup>(0.0766)</sup> (3.0441)
$\Delta_{k_1}$	2.9846 <sup>(0.1192)</sup> (3.1187)	0.2932 <sup>(0.0204)</sup> (3.0861)	5.5994 <sup>(-0.1667)</sup> (3.1528)
$\Delta_{k_2}$	0.3416 <sup>(0.0321)</sup> (3.0639)	1.7432 <sup>(-0.0597)</sup> (3.1010)	3.6629 <sup>(0.0887)</sup> (3.2375)

Table 4.14. Valeurs de la statistique de J.B. pour différentes distributions du bruit blanc.

Dans la table suivante, on donne les pourcentages des valeurs aberrantes des estimations des paramètres.

paramètre	Gaussienne (%)	Student (%)	Mixture (%)
$\beta$	0.50	0.40	0.70
$\phi$	0.60	0.60	0.40
$\Delta_{k_1}$	0.50	0.70	0.90
$\Delta_{k_2}$	0.50	0.60	1.00

Table 4.15. Le pourcentages des valeurs aberrantes dans le diagramme en boîtes.

### 4.2.2 Le modèle affecté par des valeurs aberrantes innovantes

La figure suivante représente le graphe d'une série de taille  $n = 200$  simulée à partir du modèle (3.13) quand  $p = 1$ ,  $\beta = 2$ ,  $\phi = 0.3$ ,  $k_1 = 100$ ,  $k_2 = 120$ ,  $\Delta_{k_1} = 3$ ,  $\Delta_{k_2} = 5$  et  $\varepsilon_t$  i.i.d.  $\sim N(0, 1)$ .

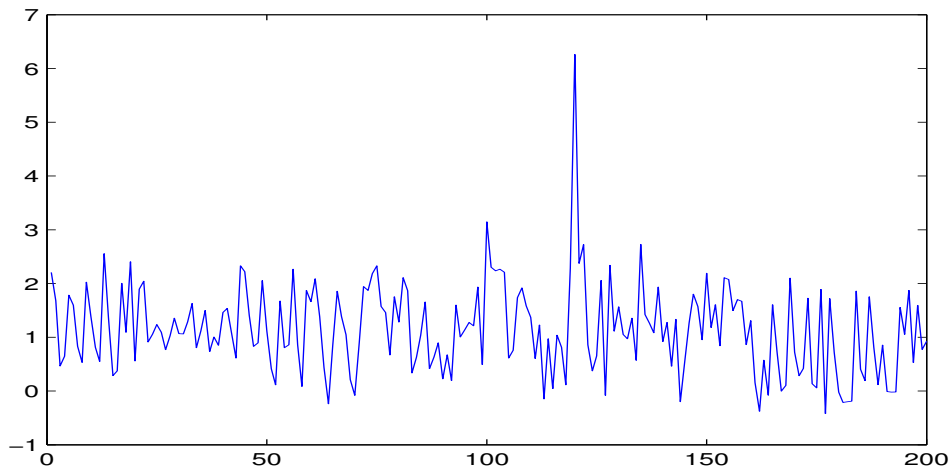


Figure 4.7. Série simulée.

Il est clair que le  $IO$  n'influence pas seulement les observations  $y_{100}$  et  $y_{120}$  mais il influence aussi les observations  $y_{101}, y_{102}, \dots, y_{121}, \dots$ . En outre, les effets de  $IO$  s'éteindront de manière exponentielle. Maintenant, nous commençons l'étude de simulation.

1. On applique l'algorithme du Gibbs Sampler avec  $N = 1000$  répétitions. Nous approximations la moyenne a posteriori en utilisant la fonction de coût quadratique par  $\hat{\theta} = \frac{1}{N} \sum_{i=1}^N \theta_i^{(m)}$ ,

avec une période de chauffe de  $m - 1$  observations pour s'éloigner de la dépendance des conditions initiales, pour  $N$  suites de Gibbs de longueur  $m$ . On donne l'écart-type (std), l'erreur quadratique moyenne (rmse) et l'intervalle de crédibilité ( $CI$ ) des paramètres  $\beta$ ,  $\phi$ ,  $\Delta_{k_1}$  et  $\Delta_{k_2}$  pour les vraies valeurs (T.V.) :  $\Delta_{k_1} = 3$ ,  $\Delta_{k_2} = 5$ ,  $\phi = 0.3$  et  $\beta = 2$  avec différentes tailles d'échantillons  $n = 50$ ,  $n = 200$  et  $n = 400$ .

Puisqu'en pratique, on ne peut pas garantir que la loi du bruit blanc est gaussienne standard, on examine la précision et la robustesse des estimateurs de Gibbs quand la loi du bruit blanc est gaussienne standard  $\mathbb{N}(0, 1)$  comme référence contre une loi de Student à 5 degrés de liberté ( $t5$ ) et un mélange ( $0.8 \times \mathbb{N}(0, 1) + 0.2 \times \mathbb{N}(0, 2^2)$ ). Pour comparaison, on donne l'estimateur du maximum de vraisemblance (MLE) de tous les paramètres,  $\hat{\theta} = \arg \max(l(\theta|X, y))$ , où  $l(\theta|X, y)$  est obtenue à partir de (3.15), avec la même loi du bruit blanc. Les résultats sont donnés dans les tableaux 4.16-4.18.

		$n = 50$		$n = 200$		$n = 400$	
paramètres	T.V.	Gibbs	MLE	Gibbs	MLE	Gibbs	MLE
$\beta_{(rmse)}^{(std)}$	2	2.0029 <sup>(0.3192)</sup> (0.3196) [1.3001,2.5756]	2.01068 <sup>(0.392)</sup> (0.4120) [1.3775,2.6581]	1.9920 <sup>(0.1539)</sup> (0.1540) [1.6968,2.3108]	2.0624 <sup>(0.1533)</sup> (0.1532) [1.6962,2.3011]	1.9970 <sup>(0.1113)</sup> (0.1113) [1.7909,2.2142]	1.9992 <sup>(0.1129)</sup> (0.1129) [1.7725,2.2160]
$\phi_{(rmse)}^{(std)}$	0.3	0.2943 <sup>(0.1067)</sup> (0.1069) [0.1603,0.4924]	0.2824 <sup>(0.1717)</sup> (0.1990) [0.0546,0.4842]	0.2965 <sup>(0.0620)</sup> (0.0623) [0.1668,0.4154]	0.2939 <sup>(0.060)</sup> (0.0611) [0.1671,0.4086]	0.3011 <sup>(0.0444)</sup> (0.0445) [0.2079,0.3851]	0.2967 <sup>(0.0456)</sup> (0.0457) [0.2076,0.3817]
$\Delta_{k_1 (rmse)}^{(std)}$	3	3.0094 <sup>(0.9981)</sup> (0.9988) [1.0574,4.9566]	2.9908 <sup>(1.0496)</sup> (1.0691) [0.9760,4.9309]	2.9965 <sup>(0.9107)</sup> (0.9110) [1.0896,4.8718]	2.9906 <sup>(0.9801)</sup> (0.9810) [1.0102,4.9068]	2.9996 <sup>(0.9078)</sup> (0.9079) [1.1171,4.8963]	3.02148 <sup>(0.9368)</sup> (0.9380) [1.0964,4.7786]
$\Delta_{k_2 (rmse)}^{(std)}$	5	4.9981 <sup>(0.9851)</sup> (0.9842) [3.2428,7.0570]	5.0194 <sup>(1.0758)</sup> (1.0955) [3.0852,7.0142]	5.0123 <sup>(0.9341)</sup> (0.9348) [4.0681,6.8110]	4.9929 <sup>(0.9611)</sup> (0.9620) [4.0160,6.8461]	5.0322 <sup>(0.9006)</sup> (0.9012) [4.3864,6.6906]	4.9871 <sup>(0.9291)</sup> (0.9306) [4.2417,6.8144]

Table 4.16. Résultats d'estimation quand le bruit blanc est gaussien  $\mathbb{N}(0, 1)$ .

		$n = 50$		$n = 200$		$n = 400$	
paramètres	T.V.	Gibbs	MLE	Gibbs	MLE	Gibbs	MLE
$\beta_{(rmse)}^{(std)}$	2	2.0427 <sup>(0.3908)</sup> (0.3916) [1.1985,2.8565]	1.9898 <sup>(0.4654)</sup> (0.4853) [1.2238,2.9173]	2.0157 <sup>(0.1931)</sup> (0.1936) [1.6587,2.3990]	1.9983 <sup>(0.1994)</sup> (0.1993) [1.5913,2.3978]	2.0309 <sup>(0.1399)</sup> (0.1399) [1.7111,2.2664]	2.0004 <sup>(0.1429)</sup> (0.1429) [1.7167,2.2783]
$\phi_{(rmse)}^{(std)}$	0.3	0.2966 <sup>(0.1149)</sup> (0.1149) [0.1613,0.5048]	0.2863 <sup>(0.1971)</sup> (0.2032) [0.0554,0.5037]	0.2963 <sup>(0.0636)</sup> (0.0640) [0.1824,0.4272]	0.2939 <sup>(0.0649)</sup> (0.0651) [0.1667,0.4152]	0.2996 <sup>(0.0481)</sup> (0.0484) [0.2094,0.3880]	0.2949 <sup>(0.0455)</sup> (0.0458) [0.2043,0.3800]
$\Delta_{k_1 (rmse)}^{(std)}$	3	3.0270 <sup>(1.1907)</sup> (1.1961) [0.4047,5.3599]	2.9064 <sup>(1.2657)</sup> (1.2853) [0.0178,5.6135]	3.0241 <sup>(1.1836)</sup> (1.1840) [0.4525,5.2833]	2.9999 <sup>(1.0906)</sup> (1.1006) [0.5557,5.5327]	2.9926 <sup>(1.0906)</sup> (1.1006) [0.5451,5.1561]	3.0168 <sup>(1.0886)</sup> (1.0989) [0.3950,5.4739]
$\Delta_{k_2 (rmse)}^{(std)}$	5	5.0454 <sup>(1.2553)</sup> (1.2609) [2.6552,7.1178]	5.0173 <sup>(1.3405)</sup> (1.3832) [2.5892,7.2048]	5.0681 <sup>(1.2548)</sup> (1.2551) [2.6925,7.0014]	5.0068 <sup>(1.2894)</sup> (1.2888) [2.6298,7.1406]	5.0671 <sup>(1.1416)</sup> (1.1421) [2.7758,7.0001]	5.0038 <sup>(1.1303)</sup> (1.1314) [2.6621,7.0244]

Table 4.17. Résultats d'estimation quand le bruit blanc suit une loi de Student  $t5$ .

		$n = 50$		$n = 200$		$n = 400$	
paramètres	T.V.	Gibbs	MLE	Gibbs	MLE	Gibbs	MLE
$\beta_{(rmse)}^{(std)}$	2	1.9901 <sup>(0.4087)</sup> (0.4084) [1.1434,2.7813]	1.9802 <sup>(0.4940)</sup> (0.4995) [1.0681,2.8137]	2.0051 <sup>(0.1932)</sup> (0.1931) [1.6325,2.3697]	2.0220 <sup>(0.1948)</sup> (0.1947) [1.6229,2.3836]	2.0041 <sup>(0.1410)</sup> (0.1411) [1.7237,2.2685]	1.9982 <sup>(0.1402)</sup> (0.1404) [1.7202,2.2757]
$\phi_{(rmse)}^{(std)}$	0.3	0.3031 <sup>(0.1103)</sup> (0.1102) [0.0779,0.5329]	0.2872 <sup>(0.2199)</sup> (0.2205) [0.0435,0.5437]	0.2964 <sup>(0.0643)</sup> (0.0644) [0.1644,0.4192]	0.2947 <sup>(0.0638)</sup> (0.0640) [0.1629,0.4154]	0.2988 <sup>(0.0485)</sup> (0.0485) [0.1967,0.3896]	0.2989 <sup>(0.0480)</sup> (0.0482) [0.2022,0.3959]
$\Delta_{k_1}^{(std)}$ $\Delta_{k_1}^{(rmse)}$	3	3.0463 <sup>(1.2185)</sup> (1.2187) [0.4263,5.4865]	3.0963 <sup>(1.2980)</sup> (1.30588) [0.3050,5.6728]	2.9860 <sup>(1.0204)</sup> (1.0211) [0.5601,5.3777]	2.9959 <sup>(1.2038)</sup> (1.2068) [0.3755,5.5124]	2.9851 <sup>(1.1639)</sup> (1.1629) [0.6288,5.1952]	2.9491 <sup>(1.1886)</sup> (1.1906) [0.3817,5.2999]
$\Delta_{k_2}^{(std)}$ $\Delta_{k_2}^{(rmse)}$	5	4.9937 <sup>(1.3035)</sup> (1.3038) [1.8536,7.6682]	4.9860 <sup>(1.4730)</sup> (1.4875) [1.8041,7.7544]	4.9901 <sup>(1.2629)</sup> (1.2631) [2.4271,7.4290]	5.0323 <sup>(1.2912)</sup> (1.2920) [2.5522,7.5739]	4.9910 <sup>(1.1707)</sup> (1.1710) [2.5972,7.3313]	5.0126 <sup>(1.1997)</sup> (1.2040) [2.7540,7.4456]

Table 4.18. Résultats d'estimation quand le bruit blanc est un mélange gaussien

$$0.8 \times N(0, 1) + 0.2 \times N(0, 2^2).$$

2. Les tables 4.19-4.21 présentent les résultats d'estimation des paramètres quand  $\phi$  est proche de la zone de non stationnarité ( $\phi = 0.95$ ),  $\Delta_{k_1} = 3$  et  $\Delta_{k_2} = 5$  pour  $n = 50$ ,  $n = 200$  et  $n = 400$ .

		$n = 50$		$n = 200$		$n = 400$	
paramètres	T.V.	Gibbs	MLE	Gibbs	MLE	Gibbs	MLE
$\beta_{(rmse)}^{(std)}$	2	1.9968 <sup>(0.3762)</sup> (0.3758) [1.2328,2.5592]	2.0913 <sup>(0.4629)</sup> (0.4830) [1.2880,2.7221]	1.9991 <sup>(0.1814)</sup> (0.1815) [1.6295,2.3805]	2.0060 <sup>(0.1715)</sup> (0.1714) [1.6694,2.3387]	1.9998 <sup>(0.1265)</sup> (0.1263) [1.7467,2.2585]	2.0008 <sup>(0.1226)</sup> (0.1225) [1.7603,2.2361]
$\phi_{(rmse)}^{(std)}$	0.95	0.9303 <sup>(0.0551)</sup> (0.0584) [0.7871,0.9709]	0.9288 <sup>(0.0981)</sup> (0.0984) [0.7661,0.9900]	0.9415 <sup>(0.0247)</sup> (0.0261) [0.8851,0.9670]	0.9410 <sup>(0.0242)</sup> (0.0258) [0.8837,0.9766]	0.9463 <sup>(0.0176)</sup> (0.0179) [0.9050,0.9622]	0.9458 <sup>(0.0161)</sup> (0.0166) [0.9098,0.9724]
$\Delta_{k_1}^{(std)}$ $\Delta_{k_1}^{(rmse)}$	3	3.0073 <sup>(1.0320)</sup> (1.0310) [1.1863,5.0732]	2.9912 <sup>(1.1201)</sup> (1.1212) [0.9871,5.0622]	3.0618 <sup>(1.0061)</sup> (1.0088) [1.2974,4.9023]	3.0546 <sup>(1.0157)</sup> (1.0166) [1.0741,5.0084]	3.0540 <sup>(0.9949)</sup> (0.9954) [1.3098,4.8075]	2.9782 <sup>(0.9936)</sup> (0.9937) [1.1129,4.8597]
$\Delta_{k_2}^{(std)}$ $\Delta_{k_2}^{(rmse)}$	5	4.9861 <sup>(1.0274)</sup> (1.0264) [2.9998,6.9155]	5.1074 <sup>(1.1523)</sup> (1.1728) [2.9029,6.9979]	5.0387 <sup>(0.9697)</sup> (0.9699) [3.0540,6.7027]	5.0452 <sup>(1.0087)</sup> (1.0093) [2.9850,6.9487]	4.9805 <sup>(0.9556)</sup> (0.9548) [3.2055,6.6104]	5.0107 <sup>(0.9737)</sup> (0.9738) [3.4474,6.9110]

Table 4.19. Résultats d'estimation quand le bruit blanc est gaussien  $N(0, 1)$ .

		$n = 50$		$n = 200$		$n = 400$	
parameter	T.V.	Gibbs	MLE	Gibbs	MLE	Gibbs	MLE
$\beta_{(rmse)}^{(std)}$	2	2.0368 <sup>(0.3962)</sup> (0.3973) [1.0315,3.0095]	1.9721 <sup>(0.4894)</sup> (0.4999) [0.9778,3.0497]	2.0091 <sup>(0.2153)</sup> (0.2152) [1.5830,2.4385]	1.9980 <sup>(0.2319)</sup> (0.2320) [1.5464,2.84383]	2.0135 <sup>(0.1604)</sup> (0.1608) [1.6978,2.3168]	2.0019 <sup>(0.1627)</sup> (0.1629) [1.6815,2.3109]
$\phi_{(rmse)}^{(std)}$	0.95	0.9258 <sup>(0.0504)</sup> (0.0541) [0.7852,0.9980]	0.9241 <sup>(0.0871)</sup> (0.0897) [0.7952,0.9996]	0.9410 <sup>(0.0249)</sup> (0.0265) [0.8789,0.9771]	0.9412 <sup>(0.0258)</sup> (0.0271) [0.8817,0.9773]	0.9463 <sup>(0.0153)</sup> (0.0158) [0.9128,0.9724]	0.9452 <sup>(0.0166)</sup> (0.0173) [0.9086,0.9713]
$\Delta_{k_1}^{(std)}$ $\Delta_{k_1}^{(rmse)}$	3	3.0712 <sup>(1.2822)</sup> (1.2834) [0.8132,5.2343]	3.0920 <sup>(1.3997)</sup> (1.4110) [0.4053,5.6400]	3.0403 <sup>(1.2033)</sup> (1.2041) [0.9953,5.1048]	3.0587 <sup>(1.2558)</sup> (1.2573) [0.9324,5.5011]	3.0315 <sup>(1.1882)</sup> (1.1896) [0.9979,4.9807]	2.9788 <sup>(1.1909)</sup> (1.2002) [0.9289,5.3952]
$\Delta_{k_2}^{(std)}$ $\Delta_{k_2}^{(rmse)}$	5	4.9870 <sup>(1.1924)</sup> (1.2015) [2.7220,7.3231]	5.0841 <sup>(1.2920)</sup> (1.2996) [2.6309,7.4222]	4.9887 <sup>(1.1082)</sup> (1.1109) [2.7761,7.1417]	4.9387 <sup>(1.1806)</sup> (1.1819) [2.7991,7.2933]	5.0792 <sup>(1.0681)</sup> (1.0688) [2.8066,7.0681]	5.0434 <sup>(1.0097)</sup> (1.0168) [2.8002,7.1048]

Table 4.20. Résultats d'estimation quand le bruit blanc suit une loi de Student  $t_5$ .

paramètres	T.V.	$n = 50$		$n = 200$		$n = 400$	
		Gibbs	MLE	Gibbs	MLE	Gibbs	MLE
$\beta_{(rmse)}^{(std)}$	2	2.0285 <sup>(0.3816)</sup> (0.3819) [1.2019, 2.9050]	2.0966 <sup>(0.4688)</sup> (0.4989) [1.1168, 2.9536]	1.9999 <sup>(0.2361)</sup> (0.2359) [1.4895, 2.4050]	1.9945 <sup>(0.2202)</sup> (0.2205) [1.5637, 2.4290]	2.0017 <sup>(0.1570)</sup> (0.1569) [1.6557, 2.2863]	1.9914 <sup>(0.1590)</sup> (0.1593) [1.6867, 2.2915]
$\phi_{(rmse)}^{(std)}$	0.95	0.9373 <sup>(0.0519)</sup> (0.0521) [0.7986, 0.9876]	0.9289 <sup>(0.0944)</sup> (0.0983) [0.7841, 0.9900]	0.9428 <sup>(0.0247)</sup> (0.0257) [0.8800, 0.9755]	0.9424 <sup>(0.0241)</sup> (0.0253) [0.8887, 0.9785]	0.9448 <sup>(0.0175)</sup> (0.0179) [0.9069, 0.9706]	0.9457 <sup>(0.0165)</sup> (0.0170) [0.9094, 0.9717]
$\Delta_{k_1}^{(std)}$ $\Delta_{k_1}^{(rmse)}$	3	3.0419 <sup>(1.3090)</sup> (1.3083) [0.5514, 5.7106]	2.9661 <sup>(1.4923)</sup> (1.4926) [0.0818, 5.5776]	3.0012 <sup>(1.2405)</sup> (1.2410) [0.6606, 5.3814]	3.0561 <sup>(1.2489)</sup> (1.2490) [0.7909, 5.3951]	3.0054 <sup>(1.1106)</sup> (1.1112) [0.7064, 5.1781]	3.0007 <sup>(1.1003)</sup> (1.1120) [0.8198, 5.2608]
$\Delta_{k_2}^{(std)}$ $\Delta_{k_2}^{(rmse)}$	5	4.9749 <sup>(1.2409)</sup> (1.2412) [2.3264, 7.4553]	5.0340 <sup>(1.3565)</sup> (1.3766) [2.2893, 8.1954]	5.0144 <sup>(1.1986)</sup> (1.1981) [2.6074, 7.3849]	4.9996 <sup>(1.1860)</sup> (1.1868) [2.6897, 7.5941]	5.0362 <sup>(1.1007)</sup> (1.1009) [2.7140, 7.2508]	4.9810 <sup>(1.0991)</sup> (1.1008) [2.6902, 7.3379]

Table 4.21. Résultats d'estimation quand le bruit blanc est un mélange gaussien

$$0.8 \times \mathcal{N}(0, 1) + 0.2 \times \mathcal{N}(0, 2^2).$$

On note que, Pour les trois types de bruits blancs, le Gibbs Sampler et le *MLE* donnent de bons résultats pour l'estimation des paramètres de notre modèle et quand la taille de l'échantillon augmente, le *rmse* et le *std* des estimations diminuent.

On note aussi que pour la taille  $n = 50$ , Gibbs Sampler est meilleur que MLE, mais pour des échantillons de grandes tailles ( $n = 200$  et  $n = 400$ ), les deux méthodes donnent des bons résultats, qui sont presque similaires.

On note aussi que, les *CI* deviennent étroitement concentrés autour des vraies valeurs pour les deux paramètres  $\beta$  et  $\phi$  même dans le cas presque non stationnaire, lorsque la taille de l'échantillon devient plus grande. Par contre pour l'amplitude des contaminants, les *CI* restent plus grands.

Quand l'erreur est gaussienne, on donne dans la figure suivante les courbes des moyennes cumulées de  $N$  échantillons, des estimations des paramètres  $\beta$ ,  $\phi$ ,  $\Delta_{k_1}$  et  $\Delta_{k_2}$ .



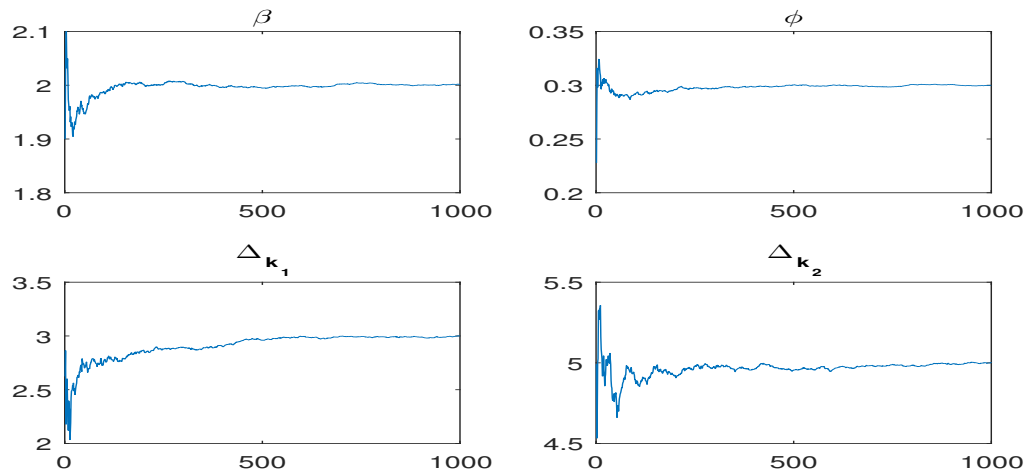


Figure 4.8. Courbes des moyennes cumulées de  $N = 1000$  échantillons, pour  $\beta = 2$ ,  $\phi = 0.3$ ,  $\Delta_{k_1} = 3$ ,  $\Delta_{k_2} = 5$ .

On note que, les estimations se stabilisent aux vraies valeurs, à partir de 500 répétitions.

- On applique l'algorithme du Gibbs Sampler avec 1000 répétitions pour  $n = 200$  et  $\phi = 0.3$  pour calculer la  $p$ -value inconditionnelle de l'hypothèse  $H_0 : \Delta_{k_l} = 0$ ,  $l \in \{1, 2\}$  pour différentes valeurs de  $\sigma$  et  $(\Delta_{k_1}, \Delta_{k_2})$  (petites, modérées et grandes valeurs).

Les tests basés sur les  $p$ -values sont critiqués dans la littérature (voir, par exemple, Benjamin et al. [11]). Mais des alternatives réalisables aux  $p$ -values ne sont pas spécifiées. Benjamin et Berger [12] suggèrent d'utiliser  $\bar{B}$  qui est une borne supérieure de  $BF$ , sur tout choix (raisonnable) de la loi a priori pour l'hypothèse alternative, du facteur de Bayes  $BF_{10}$  (voir, Kass et Raftery [52]).

Dans les tables 4.22-4.24, on a ajouté la valeur de cette borne supérieure, sous la valeur de la  $p$ -value, quand cette  $p$ -value est inférieure à 0.05.

$\sigma = 0.1$	$k_1 - 2$	$k_1 - 1$	$k_1$	$k_1 + 1$	$k_1 + 2$		$k_2 - 2$	$k_2 - 1$	$k_2$	$k_2 + 1$	$k_2 + 2$
$\Delta_{k_1} = 0.5$	0.8410	0.3413	0.0288 (3.6009)	0.7902	0.6516	$\Delta_{k_2} = 1$	0.6188	0.8262	0.00009 (438.7810)	0.4908	0.1846
$\Delta_{k_1} = 3$	0.7754	0.8663	0.00009 (438.7810)	0.6347	0.8958	$\Delta_{k_2} = 5$	0.9065	0.9448	0.00009 (438.7810)	0.4202	0.8012
$\Delta_{k_1} = 7$	0.9069	0.9745	0.00009 (438.7810)	0.7080	0.8982	$\Delta_{k_2} = 15$	0.8010	0.8825	0.00009 (438.7810)	0.3353	0.8446

Table 4.22. La  $p$ -value inconditionnelle de  $H_0$ ,  $P_{\Delta_{k_l}=0}|\underline{y}$  pour  $\sigma = 0.1$ .

$\sigma = 0.5$	$k_1 - 2$	$k_1 - 1$	$k_1$	$k_1 + 1$	$k_1 + 2$		$k_2 - 2$	$k_2 - 1$	$k_2$	$k_2 + 1$	$k_2 + 2$
$\Delta_{k_1} = 0.5$	0.4812	0.6513	0.0279 (3.6840)	0.6012	0.5014	$\Delta_{k_2} = 1$	0.7668	0.6883	0.0202 (4.6672)	0.8006	0.8909
$\Delta_{k_1} = 3$	0.7225	0.4635	0.0010 (53.2560)	0.7905	0.5580	$\Delta_{k_2} = 5$	0.4093	0.5536	0.00009 (438.7810)	0.4994	0.8282
$\Delta_{k_1} = 7$	0.8640	0.2776	0.00009 (438.7810)	0.7452	0.7328	$\Delta_{k_2} = 15$	0.6537	0.7319	0.00009 (438.7810)	0.3464	0.6306

Table 4.23. La  $p$ -value inconditionnelle de  $H_0$ ,  $P_{\Delta_{k_l}=0}|\underline{y}$  pour  $\sigma = 0.5$ .

$\sigma = 1$	$k_1 - 2$	$k_1 - 1$	$k_1$	$k_1 + 1$	$k_1 + 2$		$k_2 - 2$	$k_2 - 1$	$k_2$	$k_2 + 1$	$k_2 + 2$
$\Delta_{k_1} = 0.5$	0.6019	0.3738	0.1988	0.4165	0.8154	$\Delta_{k_2} = 1$	0.7465	0.8385	0.0870	0.2828	0.9611
$\Delta_{k_1} = 3$	0.7319	0.8722	0.0074 (10.1326)	0.6919	0.7598	$\Delta_{k_2} = 5$	0.3818	0.6390	0.00009 (438.7810)	0.6420	0.6430
$\Delta_{k_1} = 7$	0.8930	0.3431	0.0001 (399.4203)	0.5205	0.8774	$\Delta_{k_2} = 15$	0.6004	0.6098	0.0000 (438.7810)	0.4778	0.6577

Table 4.24. La  $p$ -value inconditionnelle de  $H_0$ ,  $P_{\Delta_{k_l}=0}|\underline{y}$  pour  $\sigma = 1$ .

On note que, quand la variance est petite, le test de la  $p$ -value détecte mieux l'amplitude de la contamination.

On note aussi que, le test de la  $p$ -value détecte seulement les positions  $k_1$  et  $k_2$  des contaminants et ceci est dû au fait que la contamination est présente seulement aux instants  $k_1$  et  $k_2$  comme donné dans la formule (3.14).

Par exemple, la  $p$ -value correspondante à  $k_1$ , pour  $\Delta_{k_1} = 3$  est 0.00009 qui correspond à un facteur de Bayes d'au plus 438.7810 :1. C'est-à-dire que les données impliquent des chances en faveur de l'hypothèse alternative par rapport à l'hypothèse nulle de 438,7810 sur 1 au maximum, qui est interprété comme une décision évidente contre l'hypothèse nulle (Kass et Raftery [52]); Benjamin et al. [11]).

- On donne aussi les résultats des simulations pour d'autres critères de discrimination comme le critère d'information de déviance (DIC) (voir, Spiegelhalter et al. [80], [81]), le facteur de Bayes (BF) (Kass et Raftery [52]) et l'ordonnée prédictive conditionnelle (CPO) (Geisser [30]; voir aussi Pettit [65]).

- On compare le  $DIC_0$  sous l'hypothèse nulle  $H_0$  avec le  $DIC_1$  sous l'hypothèse alternative  $H_1$ , pour différentes valeurs de  $(\Delta_{k_1}, \Delta_{k_2})$  et différents instants de contaminations  $k_1$  et  $k_2$ . L'algorithme de Gibbs Sampler est appliqué avec 1000 répétitions, les résultats sont donnés dans le tableau suivant pour le modèle  $IO$  ( $n = 200, \beta = 4, \phi = 0.5, \sigma = 1$ ).

	$k_1 = 50, k_2 = 100$			$k_1 = 150, k_2 = 199$		
$(\Delta_{k_1}, \Delta_{k_2})$	(4,8)	(2,6)	(1,2)	(4,8)	(2,6)	(1,2)
$DIC_0$	653.7866	596.0553	560.0165	611.7714	589.8324	536.0919
$DIC_1$	580.2809	571.8473	556.0319	578.2305	566.5316	534.3317

Table 4.25.  $DIC_0$  et  $DIC_1$  pour le modèle  $IO$ .

On note que, lorsque l'amplitude de la contamination décroît, les valeurs de  $DIC_1$  et  $DIC_0$  diminuent, et  $DIC_1 < DIC_0$  pour n'importe quelle valeur de  $(\Delta_{k_1}, \Delta_{k_2})$ . Ceci montre que le modèle sous l'hypothèse  $H_1$  est le modèle préféré comparé à celui sous l'hypothèse  $H_0$ .

- Le facteur de Bayes  $BF_{10}$  est utilisé pour comparer les deux hypothèses  $H_1$  et  $H_0$ . Les résultats sont donnés dans la Table 4.26, pour 1000 répétitions de l'échantillonneur de Gibbs et différentes valeurs de  $\Delta_{k_1}$  et  $\Delta_{k_2}$  ( $n = 200, \beta = 4, \phi = 0.5, \sigma = 1$ ).

$(\Delta_{k_1}, \Delta_{k_2})$	(4,8)	(2,6)	(1,2)
$BF_{10}$	126.5841	37.4973	1.4327

Table 4.26. Le facteur de Bayes pour le modèle  $IO$ .

Dans la Table 4.26, on note que, quand l'amplitude de la contamination décroît, la valeur de  $BF_{10}$  décroît et elle est toujours  $> 1$ . Les valeurs du facteur de Bayes sont interprétées comme une évidence très forte contre l'hypothèse nulle pour  $(\Delta_{k_1}, \Delta_{k_2}) = (4, 8)$ , évidence forte pour  $(\Delta_{k_1}, \Delta_{k_2}) = (2, 6)$  mais une faible évidence pour  $(\Delta_{k_1}, \Delta_{k_2}) = (1, 2)$  (voir, par exemple, Held et Ott, [46] et les références infra).

- Les résultats des simulations de CPO sont donnés dans la table 4.27, au voisinage de  $k_1$  et  $k_2$ , pour 1000 répétitions de l'échantillonneur de Gibbs ( $n = 100, \beta = 4, \phi = 0.5, \Delta_{k_1} = 4, \Delta_{k_2} = 8$ ) pour  $\sigma = 0.1$  et  $\sigma = 1$

$i$	$k_1 - 1$	$k_1$	$k_1 + 1$	$k_2 - 1$	$k_2$	$k_2 + 1$
$\sigma = 0.1$	0.4230	0.0001	0.3967	0.4321	0.0000	0.2612
$\sigma = 1$	0.2288	0.0005	0.2840	0.2809	0.0000	0.1325

Table 4.27. CPO for  $IO$  model.

Les figures suivantes donnent les courbes du CPO pour le modèle  $IO$  avec des amplitudes de contamination  $\Delta_{k_1} = 4, \Delta_{k_2} = 8, \sigma = 0.1$  et  $\sigma = 1$  aux instants  $k_1 = 50$  et  $k_2 = 80$ .

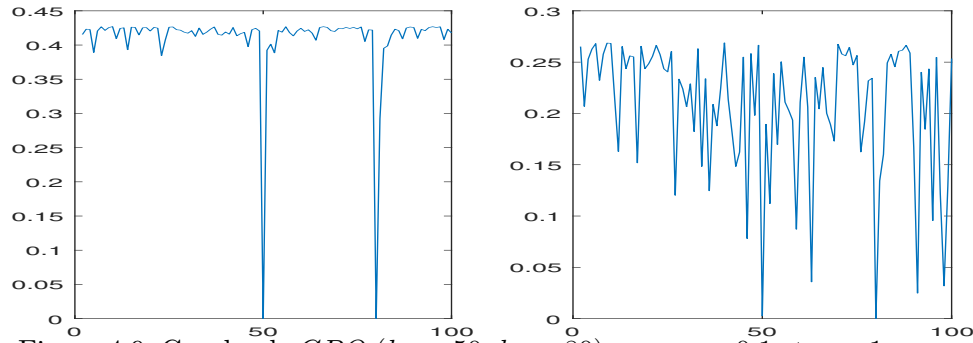


Figure 4.9. Courbe de  $CPO$  ( $k_1 = 50, k_2 = 80$ ) pour  $\sigma = 0.1$  et  $\sigma = 1$  respectivement.

Il est clair que l'instant de contamination est détecté aux instants  $k_1 = 50$  et  $k_2 = 80$ . Ceci est illustré dans les figures par une faible valeur de CPO et on le voit clairement même lorsque  $\sigma = 1$ .

5. En utilisant l'algorithme de Gibbs Sampler avec 1000 répétitions, sous les hypothèses  $H_0$  (absence des valeurs aberrantes) et  $H_1$  (présence des valeurs aberrantes) pour  $n = 100, \beta = 4, \phi = 0.5, \Delta_{k_1} = 4, \Delta_{k_2} = 8, \sigma = 1$  et différents instants de contamination  $k_1$  et  $k_2$ , la valeur prédite  $y_{n+1}$  est estimée comme la moyenne a posteriori ( $y_{n+1}^{Gibbs}(H_i)_{(rmse)}^{(std)}$ ) de la densité prédictive et des intervalles de prédiction  $PI(H_i), i = 0, 1$  sont donnés. Les résultats des simulations sont donnés dans la table suivante pour le modèle  $IO$ . Le  $rmse$  et le  $std$  sont calculés et indiqués comme suit :  $y_{n+1}^{Gibbs}(H_0)_{(rmse)}^{(std)}$  et  $y_{n+1}^{Gibbs}(H_1)_{(rmse)}^{(std)}$ . La vraie valeur  $y_T$  de  $y_{n+1}$  est obtenue en simulant  $y_{n+1}$  plusieurs fois à partir du modèle (3.14) et en prenant la moyenne. On considère  $y_{n+1}$  en utilisant l'algorithme  $MCMC$ , comme un paramètre inconnu.

	$y_T$	$y_{n+1}^{Gibbs}(H_0)$	$y_{n+1}^{Gibbs}(H_1)$	$PI(H_0)$	$PI(H_1)$
$k_1 = 20, k_2 = 30$	2.0398	$2.0967_{(1.1809)}^{(1.2368)}$	$2.0488_{(1.0068)}^{(1.0980)}$	[0.1264, 4.1760]	[0.1912, 3.9658]
$k_1 = 50, k_2 = 70$	2.0623	$2.1006_{(1.0082)}^{(1.1701)}$	$2.0784_{(0.9893)}^{(1.0663)}$	[0.1843, 4.0564]	[0.2034, 3.8602]
$k_1 = 90, k_2 = 99$	4.0531	$4.0996_{(1.1012)}^{(1.1945)}$	$4.0032_{(1.0680)}^{(1.0700)}$	[1.9681, 5.4781]	[1.9921, 5.3360]

Table 4.28. Observation prédite pour le modèle  $IO$  pour  $n=100$  et différents instants de contamination  $k_1$  et  $k_2$ .

Pour tous les instants de contamination, on note que la valeur estimée  $y_{n+1}^{Gibbs}(H_1)$  de  $y_{n+1}$  par l'algorithme de Gibbs Sampler sous l'hypothèse  $H_1$  est plus proche de  $y_T$  que celle estimée sous  $H_0$ . On note aussi que les valeurs de *rmse* et de *std* sont plus petites sous l'hypothèse  $H_1$ . La longueur de l'intervalle de prédiction est plus petite sous  $H_1$ . On déduit que le modèle le plus robuste est celui qui prend en charge la présence des valeurs aberrantes.

Quand l'instant de contamination est proche de la fin de la série d'observations,  $y_T$  est affectée par la contamination et par conséquent  $y_{n+1}^{Gibbs}(H_0)$  et  $y_{n+1}^{Gibbs}(H_1)$  aussi.

6. Les figures 4.10-4.11 représentent la courbe de la loi gaussienne (en rouge) et celle de la densité estimée par la méthode du noyau (en bleu) des estimations des paramètres, pour  $n = 200, \beta = 2, \Delta_{k_1} = 3$  et  $\Delta_{k_2} = 5$ , pour différentes valeurs de  $\phi$  ( $\phi = 0.5$  et  $\phi = 0.95$ ), pour 1000 répétitions.

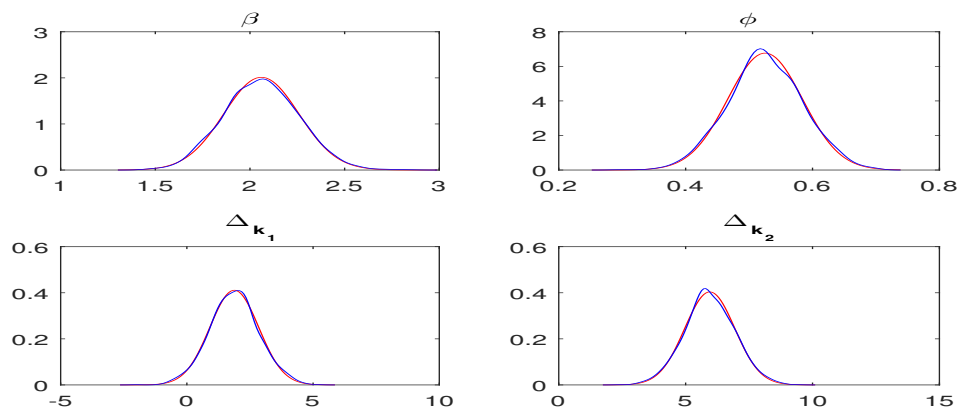


Figure 4.10. Courbe de la loi gaussienne et courbe de densité estimée par la méthode du noyau pour  $\phi = 0.5, \beta = 2, \Delta_{k_1} = 3, \Delta_{k_2} = 5$ .

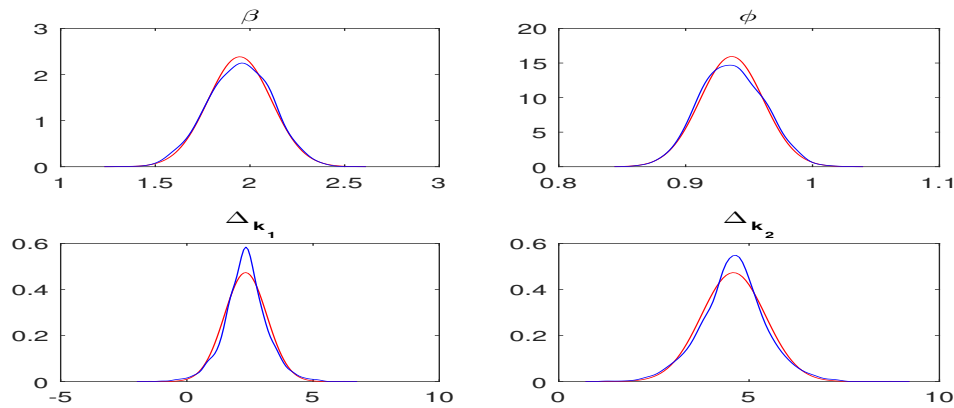


Figure 4.11. Courbe de la loi gaussienne et courbe de la densité estimée par la méthode du noyau pour  $\phi = 0.95$ ,  $\beta = 2$ ,  $\Delta_{k_1} = 3$ ,  $\Delta_{k_2} = 5$ .

On note que, les distributions des estimations des paramètres sont concentrées autour de la vraie valeur du paramètre pour toutes les valeurs de  $\phi$ .

### 7. Test de Jarque-Bera pour le modèle IO

Les résultats obtenus par simulation de la statistique de J.B., le coefficient d'asymétrie  $\gamma_1$  et le coefficient d'aplatissement  $\gamma_2$  ( $J.B. \gamma_2^1$ ) sont donnés dans la Table 4.29, pour 1000 répétitions ( $n = 200$ ,  $\beta = 2$ ,  $\phi = 0.3$ ,  $\sigma = 1$ ,  $\Delta_{k_1} = 3$ ,  $\Delta_{k_2} = 5$ ).

Paramètres	Gaussienne	t5	Mélange
$\beta$	5.1386 <sup>(0.1443)</sup> (2.7998)	0.3698 <sup>(0.0387)</sup> (3.0536)	0.3389 <sup>(0.0413)</sup> (3.0342)
$\phi$	4.2380 <sup>(0.0830)</sup> (2.7277)	3.0992 <sup>(0.1353)</sup> (3.0336)	0.8093 <sup>(0.0618)</sup> (3.0645)
$\Delta_{k_1}$	0.1431 <sup>(0.0234)</sup> (3.0354)	4.2150 <sup>(0.1363)</sup> (3.1743)	0.5093 <sup>(-0.0074)</sup> (2.8904)
$\Delta_{k_2}$	3.6878 <sup>(0.1444)</sup> (3.0716)	0.5983 <sup>(-0.0373)</sup> (3.0761)	1.2687 <sup>(-0.0861)</sup> (3.0277)

Table 4.29. Valeurs calculées de la statistique de J.B. pour différentes distributions du bruit blanc.

Pour comparer les résultats des simulations obtenus pour différentes distributions du bruit blanc, on représente les paramètres estimés par un diagramme en boîte.

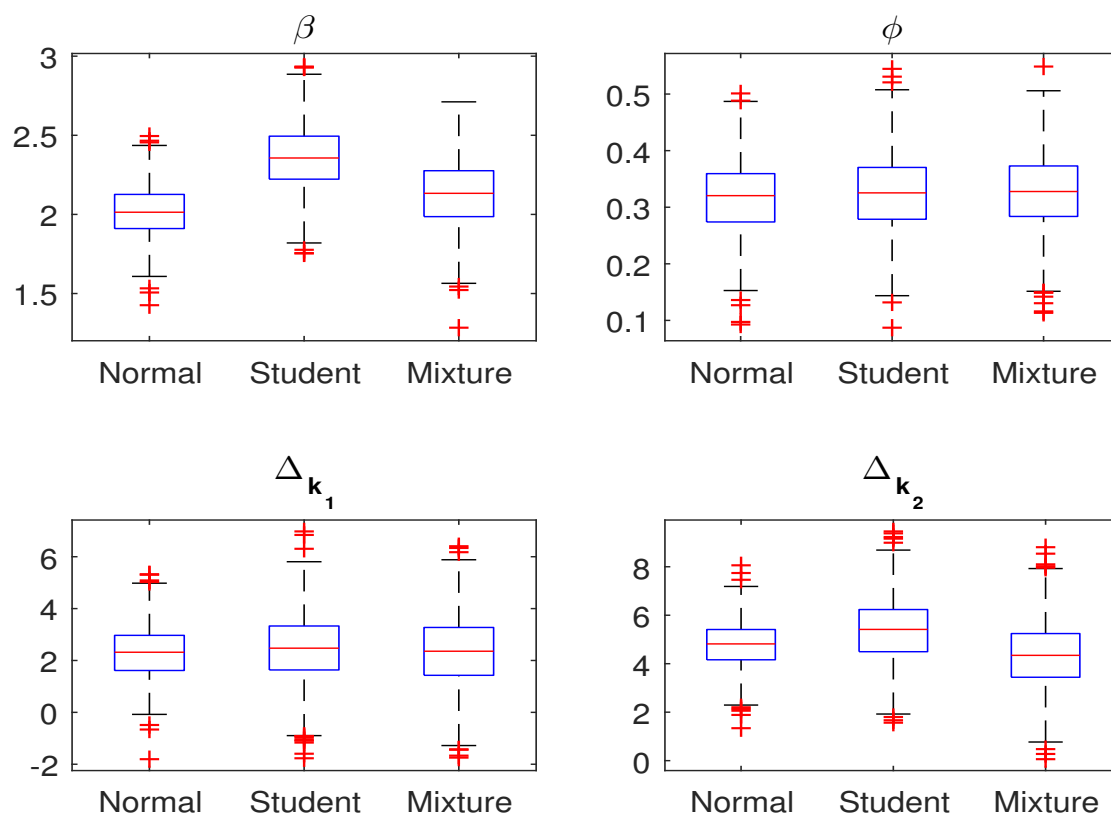


Figure 4.12. Diagrammes en boites pour  $N(0, 1)$ ,  $t_5$  et un mélange gaussien  $0.8 \times N(0, 1) + 0.2 \times N(0, 2^2)$ .

Dans cette table, on donne les pourcentages des valeurs aberrantes des estimations des paramètres donné dans la figure 4.12.

paramètres	Gaussienne (%)	Student (%)	Mélange (%)
$\beta$	0.60	0.40	0.30
$\phi$	0.50	0.50	0.60
$\Delta_{k_1}$	0.50	0.60	0.40
$\Delta_{k_2}$	0.60	0.60	0.60

Table 4.30. Le pourcentage des valeurs aberrantes dans le diagramme en boites.

### 4.3 Application sur des données réelles

Les données utilisées dans cette partie sont extraites du livre de Box et Jenkins [15], page 529, series

D. Les données représentent les relevés de viscosité d'un processus chimique : toutes les heures.

Le modèle ajusté suit un modèle  $AR(1)$  avec constante.

$$y_t - \underset{(0.123357)}{9.158383} = \underset{(0.028331)}{0.861469}(y_{t-1} - 9.158383) + \varepsilon_t, \quad t = 2, \dots, 310.$$

Les écart-types sont donnés sous les coefficients estimés.

Pour étudier l'effet des contaminants  $AO$  et  $IO$  dans cette série de données réelles, on la perturbe avec deux contaminants aux instants  $k_1 = 100$  et  $k_2 = 120$  d'amplitudes  $\Delta_{k_1} = 3$  et  $\Delta_{k_2} = 6$  respectivement.

Les résultats d'estimation sont donnés dans la table suivante, avec 1000 répétitions, pour les deux cas de contaminants  $AO$  et  $IO$ .

paramètres	Le modèle $AO$		Le modèle $IO$	
	Gibbs	MLE	Gibbs	MLE
$\beta$ <sup>(std)</sup> <sub>(rmse)</sub>	9.1523 <sup>(0.1308)</sup> <sub>(0.1308)</sub>	9.1995 <sup>(0.1382)</sup> <sub>(0.1461)</sub>	9.1508 <sup>(0.1415)</sup> <sub>(0.1416)</sub>	9.1823 <sup>(0.1723)</sup> <sub>(0.1726)</sub>
$\phi$ <sup>(std)</sup> <sub>(rmse)</sub>	0.8612 <sup>(0.0295)</sup> <sub>(0.0205)</sub>	0.8509 <sup>(0.0281)</sup> <sub>(0.0319)</sub>	0.8602 <sup>(0.0180)</sup> <sub>(0.0197)</sub>	0.8748 <sup>(0.0672)</sup> <sub>(0.0666)</sub>
$\Delta_{k_1}$ <sup>(std)</sup> <sub>(rmse)</sub>	2.8192 <sup>(0.8465)</sup> <sub>(0.8653)</sub>	2.7974 <sup>(0.8016)</sup> <sub>(0.8104)</sub>	2.8804 <sup>(0.8100)</sup> <sub>(0.8352)</sub>	2.7772 <sup>(0.8812)</sup> <sub>(0.8814)</sub>
$\Delta_{k_2}$ <sup>(std)</sup> <sub>(rmse)</sub>	5.8775 <sup>(0.8131)</sup> <sub>(0.8365)</sub>	5.7013 <sup>(0.8040)</sup> <sub>(0.8475)</sub>	5.7837 <sup>(0.8085)</sup> <sub>(0.8797)</sub>	5.9772 <sup>(0.8303)</sup> <sub>(0.8838)</sub>

Table 4.31. Résultats d'estimation pour la contamination  $AO$  et  $IO$ .

On note que, les deux méthodes Gibbs et  $MLE$  semblent être similaires.

On donne, dans les figures suivantes, les courbes de la densité gaussienne (en rouge) et la densité estimée par la méthode du noyau (en bleu) des estimations des paramètres pour les deux types de contamination.



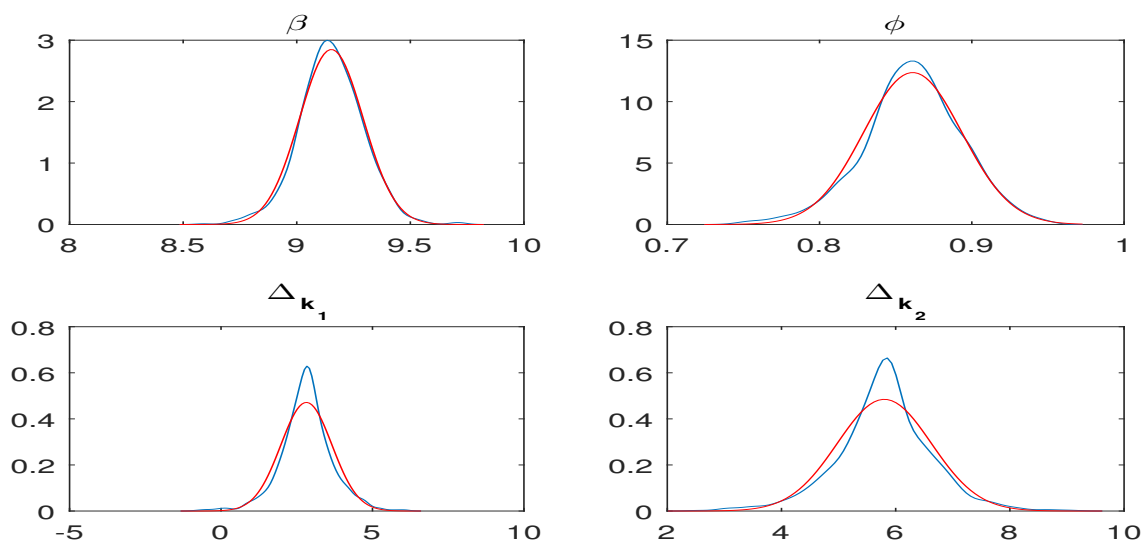


Figure 4.13. Courbe de densité gaussienne et courbe de densité estimée par la méthode du noyau des estimations des paramètres pour le modèle  $AO$ .

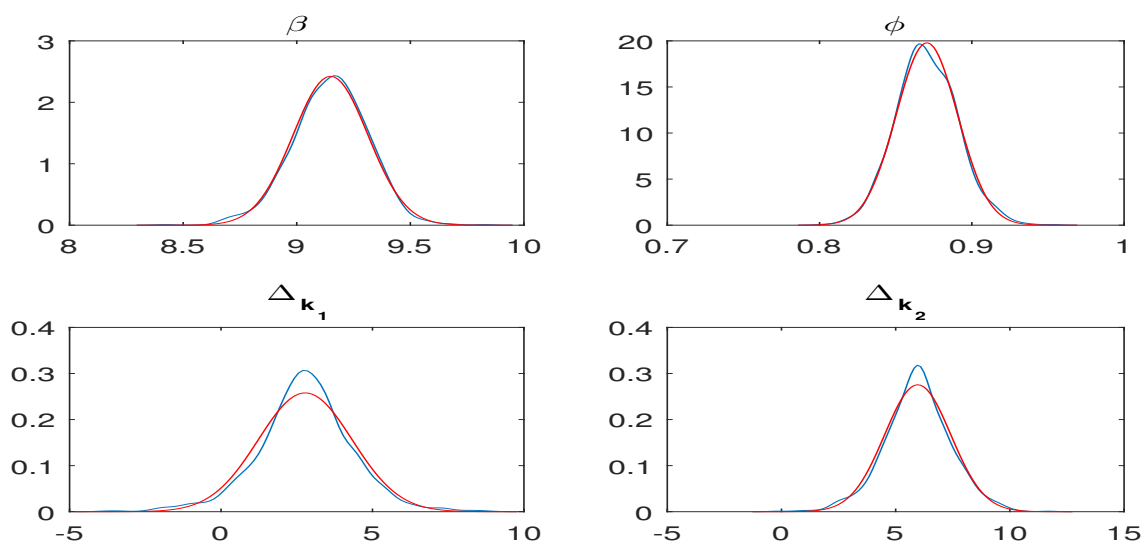


Figure 4.14. Courbe de densité gaussienne et courbe de densité estimée par la méthode du noyau des estimations des paramètres pour le modèle  $IO$ .

On note que, dans les deux types de contamination  $AO$  et  $IO$ , les courbes de la densité gaussienne et celle estimée par la méthode du noyau des estimations des paramètres sont concentrées

autour de la vraie valeur du paramètre.

On donne dans la table 4.32 et la table 4.33 les  $p$ -values pour les contaminations  $AO$  et  $IO$ . On ajoute la borne supérieure du facteur de Bayes, au dessous des  $p$ -values, quand les  $p$ -values sont inférieures à 0.05.

	$k_1 - 2$	$k_1 - 1$	$k_1$	$k_1 + 1$	$k_1 + 2$		$k_2 - 2$	$k_2 - 1$	$k_2$	$k_2 + 1$	$k_2 + 2$
$\Delta_{k_1} = 3$	0.6721	0.0053 (13.2463)	0.0003 (151.1719)	0.0097 (8.1814)	0.9402	$\Delta_{k_2} = 6$	0.6604	0.0006 (82.6483)	0.00009 (438.7810)	0.00009 (438.7810)	0.9448

Table 4.32. Les  $p$ -values inconditionnelles de  $H_0, P_{\Delta_{k_l}=0}|\underline{y}$  pour le modèle  $AO$ .

	$k_1 - 2$	$k_1 - 1$	$k_1$	$k_1 + 1$	$k_1 + 2$		$k_2 - 2$	$k_2 - 1$	$k_2$	$k_2 + 1$	$k_2 + 2$
$\Delta_{k_1} = 3$	0.6844	0.9558	0.0529 (2.3659)	0.9091	0.7634	$\Delta_{k_2} = 6$	0.6101	0.9099	0.0027 (23.0369)	0.7837	0.8552

Table 4.33. Les  $p$ -values inconditionnelles de  $H_0, P_{\Delta_{k_l}=0}|\underline{y}$  pour le modèle  $IO$ .

Pour le modèle  $AO$ , le test de la  $p$ -value détecte les positions de contaminations  $k_1$  et  $k_2$  ; il détecte aussi les positions  $k_1 - 1, k_1 + 1, k_2 - 1$  et  $k_2 + 1$  et ceci est du au fait que  $\tilde{\Delta}_{k_l}$  dans la statistique de test d'un additive outlier à l'instant  $k_l, l = 1, \dots, m$  dépend des observations  $y_{k_l}, y_{k_l-1}$  et  $y_{k_l+1}$ . Pour le modèle  $IO$ , le test de la  $p$ -value détecte seulement les positions de contamination  $k_1$  et  $k_2$  et ceci est du au fait que la contamination est présente seulement aux instants  $k_1$  et  $k_2$  comme donné dans (3.14).

Dans la table suivante, on donne les valeurs de CPO pour les modèles  $AO$  et  $IO$ .

$i$	$k_1 - 1$	$k_1$	$k_1 + 1$	$k_2 - 1$	$k_2$	$k_2 + 1$
$AO$	0.6519	0.0001	0.3189	0.7229	0.0000	0.3508
$IO$	0.8308	0.0000	0.8053	0.8185	0.0000	0.7019

Table 4.34. CPO pour les modèles  $AO$  et  $IO$ .

Les valeurs de  $DIC_0, DIC_1$  et  $BF_{10}$  sont données dans la table suivante.

	Le modèle $AO$	Le modèle $IO$
$DIC_0$	508.8622	425.0048
$DIC_1$	140.1539	138.9610
$BF_{10}$	145.3183	17.4179

Table 4.35.  $DIC_0$ ,  $DIC_1$  et le facteur de Bayes  $BF_{10}$  pour les modèles  $AO$  et  $IO$ .

Pour les contaminations  $AO$  et  $IO$ , on note que  $DIC_1 < DIC_0$  pour  $\Delta_{k_1} = 3$  et  $\Delta_{k_2} = 6$ . Ceci montre que le modèle sous l'hypothèse  $H_1$  est le modèle préféré par rapport à celui sous  $H_0$ .

Les valeurs du facteur de Bayes  $BF_{10}$  sont interprétées comme une très forte évidence en faveur de  $H_1$  pour la contamination  $AO$  et comme évidence importante pour la contamination  $IO$ .

Pour prédire une observation future, on considère que la taille de notre échantillon est 309 et on va prédire la valeur  $y_{310} = 9.10$ . Les résultats de simulation sont donnés dans les tables suivantes pour les modèles  $AO$  et  $IO$  respectivement.

	$y_n$	$y_n^{Gibbs}(H_0)$	$y_n^{Gibbs}(H_1)$	$PI(H_0)$	$PI(H_1)$
$k_1 = 100, k_2 = 200$	9.1000	9.1240 <sup>(0.0537)</sup> <sub>(0.0588)</sub>	9.1088 <sup>(0.0308)</sup> <sub>(0.0320)</sub>	[9.0183, 9.2233]	[9.0467, 9.1709]
$k_1 = 300, k_2 = 308$	9.1000	9.1270 <sup>(0.0570)</sup> <sub>(0.0631)</sub>	9.1064 <sup>(0.0301)</sup> <sub>(0.0307)</sub>	[9.0152, 9.2416]	[9.0495, 9.1657]

Table 4.36. Observation prédite pour le modèle  $AO$  pour  $n = 309$  et différents instants de contamination  $k_1$  et  $k_2$ .

	$y_n$	$y_n^{Gibbs}(H_0)$	$y_n^{Gibbs}(H_1)$	$PI(H_0)$	$PI(H_1)$
$k_1 = 100, k_2 = 200$	9.1000	9.1313 <sup>(0.2091)</sup> <sub>(0.2113)</sub>	9.1054 <sup>(0.1330)</sup> <sub>(0.1330)</sub>	[8.7148, 9.5480]	[8.8318, 9.3716]
$k_1 = 300, k_2 = 308$	14.2281	14.4397 <sup>(0.2526)</sup> <sub>(0.3294)</sub>	14.2081 <sup>(0.1469)</sup> <sub>(0.1482)</sub>	[13.9234, 14.9791]	[13.9318, 14.4953]

Table 4.37. Observation prédite pour le modèle  $IO$  pour  $n = 309$  et différents instants de contamination  $k_1$  et  $k_2$ .

Pour tous les instants de contamination, on note que la valeur estimée  $y_n^{Gibbs}(H_1)$  de  $y_n$  par l'algorithme de Gibbs sampler sous hypothèse  $H_1$  est plus proche de  $y_n$  que celle estimée sous l'hypothèse

$H_0$ . On note aussi que les valeurs du *rmse* et du *std* sont plus petites sous l'hypothèse  $H_1$ . L'étendue des intervalles de prédiction est plus étroite sous  $H_1$ . On déduit que le modèle le plus robuste est celui qui prend en charge la présence des valeurs abberantes.

Quand l'instant de contamination est proche de la fin de la série d'observations,  $y_n$  est affectée par la contamination dans le cas *IO* et par conséquent  $y_n^{Gibbs}(H_0)$  et  $y_n^{Gibbs}(H_1)$ .

# Conclusion générale

Dans cette thèse, une inférence bayésienne a été réalisée pour un modèle de série chronologique autorégressif d'ordre  $p$  avec une tendance de régression, quand il est contaminé premièrement, par des valeurs aberrantes additives multiples et ensuite, par des valeurs aberrantes innovantes multiples. En utilisant la méthode du quasi maximum de vraisemblance combinée avec une loi a priori non informative ou diffuse, des distributions a posteriori conditionnelles des paramètres du modèle sont obtenues.

Une étude de simulation intensive a été menée pour examiner la précision et la robustesse des estimateurs de Gibbs lorsque la distribution du processus d'erreur est gaussienne standard comme référence contre, Student avec 5 degrés de liberté ou un mélange de deux lois gaussiennes. On a montré que les paramètres du modèle sont bien estimés dans les deux cas de contamination  $AO$  et  $IO$ . Une comparaison avec l'approche MLE a montré que la méthode bayésienne est meilleure que la méthode classique quand la taille de l'échantillon est petite.

Un test basé sur la  $p$ -value inconditionnelle a été proposé pour détecter la position et l'amplitude des contaminations pour des valeurs aberrantes multiples  $AO$  et  $IO$ , quand le nombre de contaminants est connu. Les  $p$ -values sont utilisées pour calculer la borne supérieure du facteur de Bayes afin d'évaluer les chances de préférer l'hypothèse alternative à l'hypothèse nulle. Pour vérifier dans quelle mesure on favorise l'hypothèse alternative par rapport à l'hypothèse nulle, on a utilisé d'autres critères de discrimination bayésien comme le critère d'information de déviance (DIC), l'ordonnée prédictive conditionnelle (CPO) et le facteur de Bayes (BF).

La valeur prédite d'une observation future est donnée et comparée avec la vraie valeur pour les deux cas  $AO$  et  $IO$ . Nos résultats ont montré que le modèle le plus robuste est celui qui prend en charge la présence des valeurs aberrantes. Une application sur des données réelles a montré l'efficacité de notre méthode dans l'estimation des paramètres et la détection des valeurs aberrantes  $AO$  et  $IO$ .

Les perspectives de recherche pour ce travail sont diverses, on peut citer par exemple :

1. Etude de la robustesse de la procédure en considérant des lois a priori conjuguées.
2. Généralisation de la contamination  $AO$  et  $IO$  dans les modèles  $GARCH$ .
3. Extension de nos travaux au cas d'un modèle  $ARMA$  avec une tendance de régression.
4. Extension de la méthode au cas d'un modèle  $VAR(p)$  avec tendance.
5. Etude de modèles à coefficients aléatoires en présence de contamination.

# Bibliographie

- [1] Ait Mohammed N, Guerbyenne H. Bayesian inference in a multiple contaminated autoregressive model with trend. *Journal of Statistical Computation and Simulation*. 2023; DOI : 10.1080/00949655.2023.2172171
- [2] Abd El-Salam MEF. Methods of estimation for autoregressive models with outliers. *Asian Journal of Mathematics & Statistics*. 2013;6(2) :57.
- [3] Abraham B, Box GEP. Linear models and spurious observations. *Journal of the Royal Statistical Society Series C : Applied Statistics*. 1978;27(2) :131-138.
- [4] Abraham B, Box GEP. Bayesian analysis of some outlier problems in time series analysis. *Biometrika*. 1979;66 :229-236.
- [5] Abraham B, Chuang A. Outlier detection and time series modeling. *Technometrics*. 1989;31(2) :241-248.
- [6] Aggarwal CC. *Outlier analysis*. Springer, Cham; 2017.
- [7] Albert I, et al. Combining expert opinions in prior elicitation. *Bayesian Analysis*. 2012;7(3) :503-532.
- [8] Barnett V, Lewis T. *Outliers in statistical data*. New York : John Wiley and Sons; 1994.
- [9] Barnett G, Kohn R, Sheather S. Bayesian estimation of an autoregressive model using Markov chain Monte Carlo. *Journal of Econometrics*. Elsevier. 1996;74(2) :237-254.
- [10] Bataglia F, Orfei L. Outlier detection and estimation in nonlinear time series. *Journal of Time Series Analysis*. 2002;26(1) :108-120.
- [11] Benjamin D, et al. Redefine Statistical Significance. *Nature Human Behaviour*. 2017;2 :6-10.

- [12] Benjamin D, Berger J. Three recommendations for improving the use of  $p$ -values. *The American Statistician*. 2019;73 :186-191.
- [13] Berger JO. *Statistical Decision Theory and Bayesian Analysis*. 2nd ed. New York : Springer-Verlag; 1985.
- [14] Berkoun Y, Fellag H, Zieliński R. Robust testing serial correlation in  $AR(1)$  processes in the presence of a single additive outlier. *Communications in Statistics-Theory and Methods*. 2003; 32(8) :1527-1540.
- [15] Box GEP, Jenkins GM. *Time series analysis forecasting and control*. California : Revised Edition, Holden-Day Oakland; 1976.
- [16] Box GEP, Tiao GC. A bayesian approach to some outlier problems. *Biometrika*. 1968;55(1) :119-129.
- [17] Box GEP, Tiao GC. *Bayesian inference in statistical analysis*. Addison-Wesley, Reading, Mass; 1973.
- [18] Brown RL, Durbin J, Evans JM. Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society : Series B*. 1975;37(2) :149-163.
- [19] Casella G, George EI. Explaining the Gibbs sampler. *The American Statistician*. 1992;46(3) :167-174.
- [20] Carlin BP, Louis TA. Empirical Bayes : Past, present and future. *Journal of the American Statistical Association*. 2000;95(452) :1286-1289.
- [21] Chaloner K, Brant R. A Bayesian approach to outlier detection and residual analysis. *Biometrika*. 1988;75(4) :651-659.
- [22] Chang I, Tiao GC, Chen C. Estimation of time series parameters in the presence of outliers. *Technometrics*. 1988;30(2) :193-204.
- [23] Chan WS. Understanding the effect of time series outliers on sample autocorrelation. *Test* 4. 1995;179-186.
- [24] Chen C, Liu L. Joint estimation of model parameters and outlier effects in time series. *Journal of the American Statistical Association*. 1993;88 :284-297.
- [25] Chen CW. Detection of additive outliers in bilinear time series. *Computational Statistics and Data Analysis*. 1997;24(3) :283-294.



- [26] Eddy WF. The distribution of the convex hull of a Gaussian sample. *Journal of Applied Probability*. 1980 ;17(3) :686-695.
- [27] Fox AJ. Outliers in time series. *Journal of the Royal Statistical Society*. 1972 ;B,34 :350-363.
- [28] Galeano P, Peña D. Finding outliers in linear and nonlinear time series. In *Robustness and complex Data Structures*. (pp. 243-260). Springer, Berlin, Heidelberg ; 2013.
- [29] Gambino J, Guttman I. A Bayesian approach to prediction in the presence of spurious observations for several models. *Communication in Statistics Theory and Methods*. 1984 ;13(7) :791–812.
- [30] Geisser S. Discussion on Sampling and Bayes' inference in scientific modelling and robustness (by G.E.P. Box). *J. R. Statist. Soc. A*. 1980 ;143 :416-417.
- [31] Geisser S. Interval prediction for Pareto and exponential observables. *Journal of Econometrics*. 1985 ;29(1-2) :173-185.
- [32] Gelfand AE, Smith AFM. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*. 1990 ;85 :398-409.
- [33] Geman S, Geman D. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*. 1984 ;6 :721-741.
- [34] Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. *Statistical science*. 1992 ;457-472.
- [35] Geweke J. *Contemporary bayesian econometrics and statistics*. A John Wiley and Sons, INC. ; 2005.
- [36] Geweke J. Evaluating the accuracy of sampling-based approaches to the calculations of posterior moments. *Bayesian statistics*.1992 ;4 :641-649.
- [37] Gilks WR, Richardson S, Spiegelhalter D. *Markov chain Monte Carlo in practice*. Chapman and Hall. New york ;1995.
- [38] Goutis C, Casella G, Wells MT. Assessing evidence in multiple hypotheses. *Journal of the American Statistical Association*. 1996 ;91(435) :1268-1277.
- [39] Gupta PK, Singh AK. Classical and Bayesian estimation of Weibull distribution in presence of outliers. *Cogent Mathematics*. 2017 ;4(1) :1300975.

- [40] Guttman I, Dutter R, Freeman, PR. Care and handling of univariate outliers in the general linear model to detect spuriousity—A Bayesian approach. *Technometrics*. 1978 ;20(2) :187-193.
- [41] Guttman I. Care and handling of univariate or multivariate outliers in detecting spuriousity—a Bayesian approach. *Technometrics*. 1973 ;15(4) :723-738.
- [42] Hamilton JD. *Time series analysis*. Princeton university press. Princeton ; 1994.
- [43] Hastings W. Monte Carlo sampling methods using Markov chains and their application. *Biometrika*. 1970 ;57 :97–109.
- [44] Hawkins DM. *Identification of outliers*. Chapman and Hall. Cambridge ; 1980.
- [45] Heidelberger P, Welch PD. Simulation run length control in the presence of an initial transient. *Operations Research*. 1983 ;31(6) :1109-1144.
- [46] Held L, Ott M. On  $p$ -values and Bayes factor. *Annual Review of Statistics and Its Application*. 2018 ;5 :393-419.
- [47] Jeffreys H. *Theory of Probability* (3rd edition). Oxford University Press. Oxford ;1961.
- [48] Johnson V. Revised Standards for Statistical Evidence. *Proceedings of the National Academy of Sciences*. 2013 ;110 :19313–19317. DOI : 10.1073/pnas.1313476110.
- [49] Justel A, Peña D. Gibbs sampling will fail in outlier problems with strong masking. *Journal of Computational and Graphical Statistics*. 1996 ;5(2) :176-189.
- [50] Justel A, Peña D. Heterogeneity and model uncertainty in Bayesian regression models. *Rev.R.Acad.Cienc.Exact.Fis.Nat*. 1999 ;9(3) :357-366.
- [51] Justel A, Peña D, Tsay RS. Detection of outlier patches in autoregressive time series. *Statistica Sinica* 2001 :651-673.
- [52] Kass RE, Raftery AE. Bayes Factors. *Journal of the American Statistical Association*. 1995 ;90 :773-795.
- [53] Kim D. A Bayesian significance test of the stationarity of regression parameters. *Biometrika*. 1991 ;78(3) :667-675.
- [54] Lee PM. *Bayesian statistics*. Oxford University Press. London ; 1989.
- [55] Marden JJ. Hypothesis testing : from  $p$ -values to Bayes factors. *Journal of the American Statistical Association*. 2000 ;95(452) :1316-1320.

- [56] McCulloch RE, Tsay RS. Bayesian analysis of autoregressive time series via the Gibbs sampler. *Journal of Time series Analysis*. 1994;15 :235-250.
- [57] Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculations by fast computing machines. *The journal of chemical physics*. 1953;21(6) :1087-1092.
- [58] Metropolis N, Ulam U. The Monte Carlo method. *Journal of American Statistical Association*. 1949;44 :335-341.
- [59] Moffat IM, Akpan EA. Identification and modeling of outliers in a discrete-time stochastic series. *American Journal of Theoretical and Applied Statistics*. 2017;6(4) :191-197.
- [60] Müller N, Peña D, Yohai VJ. Robust estimation for ARMA models. *The Annals of Statistics*. 2009;37(2) :816-840.
- [61] Novika F, Siswadi, Bakhtiar T. The Use of biplot analysis and euclidean distance with procrustes measure for outliers detection. *International Journal of Engineering and Management Research*. 2018;8 :194-200.
- [62] Parent E, Bernier J. *Le raisonnement bayésien : modélisation et inférence*. Springer-verglas. France ;2007.
- [63] Pena D, Tiao GC. Bayesian robustness functions for linear models. *Bayesian Statistics 4* (Edited by JM Bernardo, JO Berger, AP Dawid and AFM Smith). 1992 ;365-388.
- [64] Pérez B, Molina I, Peña D. Outlier detection and robust estimation in linear regression models with fixed group effects. *Journal of Statistical Computation and Simulation*. 2014;84(12) :2652-2669.
- [65] Pettit LI. The conditional predictive ordinate for the normal distribution. *Journal of the Royal Statistical Society : Series B (Methodological)*. 1990 ;52(1) :175-184.
- [66] Raftery AE, Lewis SM. [Practical Markov Chain Monte Carlo] : comment : one long run with diagnostics : implementation strategies for Markov Chain Monte Carlo. *Statistical science*. 1992;7(4) :493-497.
- [67] Rasmussen CE. The infinite Gaussian mixture model. *Advances in neural information processing systems*. 1999;12 :554-560.
- [68] Roberts GO, Smith AFM. Sample conditions for the convergence of the Gibbs Sampler and the Metropolis-Hastings. *Stochastic processes and their applications*. 1993;49 :207-216.

- [69] Robert CP. Le choix bayésien, Principes et pratique. Springer-Verlag. France ;2006.
- [70] Rousseeuw P, Leroy A. Robust regression and outlier detection. New York : Wiley ; 2003.
- [71] Roy V. Convergence diagnostics for Markov Chain Monte Carlo. Annual review of statistics and its application. 2020 ;7 :387-420.
- [72] Sánchez MJ, Peña D. The identification of multiple outliers in ARIMA models. Communications in Statistics. 2003 ;32 :1265-1287.
- [73] Sellke T, Bayarri MJ, and Berger JO. Calibration of p-values for testing precise null hypotheses. The American Statistician. 2001 ;55(1) :62-71.
- [74] Schwertman NC, Owens MA, Adnan R. A simple more general boxplot method for identifying outliers. Computational Statistics and Data Analysis. 2004 ; 47(1) :165-174.
- [75] Silva ME, Pereira I, and McCabe B. Bayesian outlier detection in non-Gaussian autoregressive time series. Journal of Time Series Analysis. 2019 ;40(5) :631-648.
- [76] Slama A. A Bayesian significance test of change in the presence of single outlier. Africa statistica. 2014 ;9(1) :685-694.
- [77] Slama A, Fellag H. A Bayesian analysis of a change in the mean of independent normal sequence with contaminated observation. Africa Statistica. 2018 ;13(2) :1779-1794.
- [78] Slama A, Saggou H. A Bayesian analysis of a change in the parameter of autoregressive time series. Communications in Statistics-Simulation and Computation. 2017 ;46(9) :7008-7021.
- [79] Smith AFM, West M. Monitoring renal transplants : an application of the multiprocess Kalman filter. Biometrics. 1983 ;867-878.
- [80] Spiegelhalter DJ, Best NG, Carlin BP, Van der Linde A. Bayesian measures of model complexity and fit. Journal of the Royal Statistical : Society B. 2002 ;64(4) :583-639.
- [81] Spiegelhalter DJ, Best NG, Carlin BP, Van der Linde A. The deviance information criteria : 12 years on. Journal of the Royal Statistical : Society B. 2014 ;76(3) :485-493.
- [82] Thompson KR, Page FH. Detecting synchrony of recruitment using short, autocorrelated time series. Canadian Journal of Fisheries and Aquatic Sciences. 1989 ; 46(10) :1831-1838.
- [83] Tsay RS. Time series model specification in the presence of outliers. Journal of the American Statistical Association. 1986 ;81 :132-141.

- [84] Tsay RS. Outliers level shifts and variance change in time series. *Journal of Forecasting*. 1988;7 :1-20.
- [85] Tukey JW. A survey of sampling from contaminated distributions. *Contributions to probability and statistics*. 1960 :448-485.
- [86] Verdinelli I, Wasserman L. Bayesian analysis of outliers problems using the Gibbs sampler. *Statistics and Computing*. 1992;1 :105–117.
- [87] Vokó et al. Using informative prior based on expert opinion in Bayesian estimation of the transition probability matrix in Markov modelling—an example from the cost-effectiveness analysis of the treatment of patients with predominantly negative symptoms of schizophrenia with cariprazine. *Cost Effectiveness and Resource Allocation*. 2020;18(1) :1-9.
- [88] Von Neumann J, Morgensterne O. *Theory of games and economic behavior*. Princeton university press. Priceton. Second edition. 1947
- [89] Vovk VG. A logic of probability, with application to the foundations of statistics. *Journal of the Royal statistical society. series B (Methodological)*. 1993 ;55(2) :317-341.
- [90] West M. Outlier models and prior distributions in Bayesian linear regression. *Journal of the Royal Statistical Society : Series B (Methodological)*. 1984 ;46(3) :431-439.
- [91] Young LI, Jun YU, Zeng T. Deviance information criterion for Bayesian model selection : Justification and variation. *Economic Statistics Working Paper*. 2017 ;10 :1-25.
- [92] Zhang G, Gui Q. A Bayesian approach for detecting outliers in ARMA time series. *WSEAS Transactions on Mathematics*, 2017 ;16 :103-112.

## Abstract

Two types of outliers that may occur in data are considered in this thesis : additive outliers (*AO*) and innovational outliers (*IO*). We have generalized the two types of contaminations *AO* and *IO* to the multiple case for an autoregressive model of order  $p$  with a regression trend. We adopt the Bayesian approach combined with Gibbs sampling to jointly estimate the model parameters and the outliers on the first hand, and on the other hand we use a test based on  $p$ -values and other discrimination Bayesian criteria to detect the location and the magnitude of the two types of outliers. An intensive simulation study is presented for illustrating the performance of the method relative to maximum likelihood estimation, mainly for small sample sizes. The prediction of a future value is given. Our method is applied to a real data set.

**keywords** : Gibbs sampling ; multiple additive outliers ; multiple innovational outliers ; autoregressive model ; bayesian analysis ; outliers detection.

## Résumé

Deux types de contaminations pouvant affecter des données d'observation sont considérées dans cette thèse : les contaminants additifs (*AO*) et les contaminants innovants (*IO*). Nous avons généralisé les deux types de contaminations *AO* et *IO* au cas multiple pour un modèle autorégressif d'ordre  $p$  avec une tendance de régression. Nous adoptons l'approche bayésienne et l'échantillonnage de Gibbs pour estimer d'abord les paramètres du modèle et les amplitudes des contaminants, ensuite nous utilisons un test basé sur les  $p$ -values et d'autres critères de discrimination bayésiens pour détecter la position des deux types de contaminants. Une étude de simulation intensive est présentée pour illustrer la performance de la méthode par rapport à l'estimation par maximum de vraisemblance, principalement pour des échantillons de petite taille. La prévision d'une valeur future est donnée. Notre méthode est appliquée à un ensemble de données réelles.

**Mots clés** : Échantillonnage de Gibbs ; valeurs aberrantes additives multiples ; valeurs aberrantes innovantes multiples ; modèle autorégressif ; analyse bayésienne ; détection des valeurs aberrantes.