

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE
UNIVERSITE MOULOUD MAMMERI, TIZI-OUZOU.

FACULTE : DES SCIENCES
DEPARTEMENT : MATHEMATIQUES

THESE DE DOCTORAT

SPECIALITE : MATHEMATIQUES
OPTION : PROBABILITES ET STATISTIQUES

Présentée par :
M^{me} BELKACEM Cherifa ép. BOULAHIA

Sujet :

Stabilité des modèles de rupture sous contamination

Devant le jury d'examen composé de :

HAMADOUCHE Djamel	; Professeur	; UMMTO	;Président
FELLAG Hocine	; Professeur	; UMMTO	;Rapporteur
NECIR Abdelhakim	; Professeur	; U. Biskra	;Examineur
YOUSFATE Abderahmane	; Professeur	; U. S.Belabbes	;Examineur
MERAGHNI Djamel	; Maître de conférence A	; U. Biskra	;Examineur
BOUDIBA Mohand Arezki	; Maître de conférence A	; UMMTO	;Examineur

Soutenue le : 30 /05 /2013

Table des matières

Introduction générale	3
1 L'analyse statistique Bayésienne	8
1.1 Introduction	8
1.2 Notions de base	9
1.2.1 Espace de la théorie de la décision statistique	9
1.2.2 Théorème de Bayes	10
1.3 Les lois a priori	12
1.3.1 Le choix d'une loi a priori	12
1.3.2 Lois a priori non informatives	13
1.3.3 Lois a priori d'entropie maximale	17
1.3.4 Lois a priori conjuguées	18
1.3.5 Lois a priori subjectives	20
1.3.6 Lois a priori impropres	20
1.4 Les bases de la théorie de la décision	21
1.4.1 Fonction de perte et risque	21
1.4.2 Estimateurs de Bayes	22
1.4.3 Fonctions de coût usuelles	23
1.4.4 Admissibilité et minimaxité	26
1.4.5 Estimateur du maximum a posteriori MAP	30
1.4.6 Tests et intervalles de crédibilité	30
1.5 Méthodes de calcul Bayésien	32
1.5.1 Méthodes classiques d'approximation	32
1.5.2 Méthodes de Monte Carlo par chaînes de Markov (MCMC)	34
1.6 Conclusion	36

2	Les problèmes de rupture	37
2.1	Introduction	37
2.2	Modèle Gaussien	38
2.2.1	Rupture dans la moyenne	38
2.2.2	Rupture dans la variance	43
2.3	Modèle de régression lineaire multiple	49
2.3.1	Présentation du modèle	50
2.3.2	Cas où les erreurs sont indépendantes	50
2.3.3	Cas où les erreurs sont autocorrélées	52
2.4	Processus autoregréssif d'ordre (1)	57
2.4.1	Présentation du modèle	57
2.4.2	La fonction de vraisemblance	58
2.4.3	L'analyse a posteriori	58
2.5	Tests d'hypothèses	60
2.5.1	Statistique du rapport de vraisemblance	61
2.5.2	Statistique de Pettitt, 1980	62
2.5.3	Statistique de Brown et Al, 1975	62
2.5.4	Statistique de Chernoff et Zacks, 1964	62
2.6	Détection de point de rupture en utilisant les tests d'hypothèses séquentiels	63
2.6.1	Tests d'hypothèses séquentiels :	63
2.6.2	Problème de point de rupture :	67
3	Analyse Bayésienne d'un modèle de rupture sous contamination	69
3.1	Introduction	69
3.2	Les densités a priori	70
3.3	contamination avant rupture	71
3.3.1	Présentation du modèle	71
3.3.2	La fonction de vraisemblance des différents paramètres	72
3.3.3	L'analyse a posteriori	72
3.4	contamination après rupture	78
3.4.1	Présentation du modèle	78
3.4.2	La fonction de vraisemblance des différents paramètres	78
3.4.3	L'analyse a posteriori	79
3.5	Application numérique	84

Table des matières **3**

3.5.1	Position et amplitude du contaminant	84
3.5.2	Position du point de rupture	87
3.5.3	La taille de l'échantillon	87
3.6	conclusion	90
Conclusion générale		97
Bibliographie		97

Introduction générale

Les problèmes de rupture surgissent dans beaucoup de domaines, tel que, le traitement du signal, les systèmes de contrôle industriel, l'économie, la médecine, l'agronomie, la météorologie, etc... Chronologiquement ordonnées, les données sont collectées dans un intervalle de temps dont on suspecte qu'un changement s'est produit dans le processus qui génère ces données. Typiquement, ce problème est connu comme un problème de rupture.

La détection de rupture est la recherche d'un changement soudain dans la distribution des données. Ce problème a suscité un intérêt croissant et a fait l'objet de plusieurs études chez les économètres et les statisticiens. En économétrie où des prévisions sont à envisager, ce phénomène risque de s'aggraver et entraîner des conséquences inattendues. Il est donc impératif de s'assurer de la stabilité ou non du modèle étudié.

Comme exemples pratiques des problèmes de rupture, citons, Hsu (1977,1979(a)), en analysant une série du taux de revenu de l'indice industriel moyen de Dow-Jons "Dow-Jons Industrial Average" entre le 1^{er} Juillet 1971 et le 2 Août 1974, il a trouver qu'un changement brusque s'est produit au cours du mois de mars 1973. Incidemment, il note que la période identifiée coïncide avec le début de l'affaire de Watergate.

Dans le monde de l'industrie pharmaceutique Japonaise, une question a été posée : est ce que oui ou non le mouvement de protection des consommateurs contre les produits pharmaceutiques (drogues,...), lancé à la fin de 1960, a apporté un changement significatif dans le comportement des consommateurs, en ce qui concerne la consommation des vitamines et autres nutriments supplémentaires ?. Tsurumi(1977) a étudié des données sur une population entre 1960 et 1974 concernée par la consommation de ces produits. Il a constaté qu'un changement dans les paramètres a eu lieu en 1971, et que ce changement est, sans doute, causé par le mouvement de protection contre ces produits pharmaceutiques.

Dans une production en chaîne, un dérèglement de machine peut provoquer la perte d'une partie des articles produits; le dérèglement peut être détecté par l'existence d'un point de rupture. La modélisation des cours d'un produit stratégique (pétrole, blé, café, sucre,...) sur une longue période permet de déceler, et estimer l'amplitude des changements brutaux (à la hausse comme à la baisse) de ces cours. Ces changements sont en général non périodiques et provoqués par des événements politiques, sociaux ou climatiques. Beaucoup d'autres exemples peuvent être cités en agronomie, en météorologie, en médecine, en biochimie, en hydrogéologie, en sociologie etc....

Parmi les chercheurs qui se sont intéressés au problème de rupture de modèle, Page (1954,1955,1957) est considéré comme le pionnier dans ce domaine. Durant les années cinquante, en s'intéressant au contrôle de qualité, il a mis au point un test non paramétrique basé sur les sommes cumulatives, appelées techniques de CUSUM, pour détecter un changement dans les distributions d'une suite de variables aléatoires indépendantes.

Chernoff et Zacks (1964) et Kander et Zacks (1966) ont étudié une suite de variables aléatoires gaussiennes et indépendantes et ont proposé un test (paramétrique) Bayésien pour mettre en évidence un changement dans la moyenne. Broemeling (1985) et Broemeling et Tsurumi (1986) se sont intéressés à une analyse Bayésienne de quelques modèles linéaires.

Manzefricke (1981) a étudié une suite de variables aléatoires indépendantes et gaussiennes, où les densités de probabilité a posteriori du point de rupture et du rapport des variances ont été déterminées dans les trois cas suivants :

- Les moyennes sont distinctes et connues.
- Les moyennes sont égales et inconnues.
- Les moyennes sont distinctes et inconnues.

Hsu (1982) a déterminé les distributions a posteriori pour la différence des paramètres de deux régimes conditionnellement à l'estimateur du point de rupture, et ce pour un modèle de loi exponentielle. Abdelli (1993) a considéré le problème de rupture dans les séries chronologiques.

Andrews et al (1996), ont testé la constance d'un modèle de régression linéaire gaussien standard contre une alternative où la régression précédente présente $(m + 1)$ phases distinctes en m points inconnus.

Dans ce cadre, Ilmakunnas et Tsurumi (1984) ont étudié un modèle de régression à erreurs autocorrélées où le changement concerne le coefficient d'autocorrélation. Ils se sont intéressés au paramètre de translation pour lequel il détermine un intervalle de confiance conditionnel.

Récemment, de nouveaux résultats théoriques ont vu le jour, comme ceux de Fotopoulos et al (2010). Ceux de Fearnhead (2006) et de Rigai et al (2012) qui se sont intéressés au problème de ruptures multiples.

Plusieurs approches sont utilisées dans les problèmes de rupture. Soit l'approche classique de l'inférence fondée sur la fonction de vraisemblance, où encore les méthodes non paramétriques basées sur les sommes cumulatives. Mais l'approche Bayésienne a été largement sollicitée. Cette méthodologie consiste à assigner des lois a priori (propres ou non informatives) aux paramètres inconnus du modèle, afin de déterminer les lois de probabilité a posteriori des paramètres d'intérêt et cela en utilisant le théorème de Bayes qui stipule que la loi de probabilité a posteriori d'un paramètre θ est proportionnelle au produit de la fonction de vraisemblance et de sa loi a priori :

$$\pi(\theta/S_n) \propto L(\theta/S_n).\pi(\theta)$$

où :

θ est un paramètre vectoriel.

$S_n = (X_1, X_2, \dots, X_n)$ échantillon des observations.

$\pi(\theta)$ densité a priori du paramètre θ .

$\pi(\theta/S_n)$ densité a posteriori.

Par des intégrations successives, nous obtenons les différentes densités a posteriori marginales et conditionnelles des différents paramètres d'intérêt, qui nous permettent de déterminer le mode, la moyenne ou la médiane de ces paramètres. Le mode de la densité a posteriori marginale $\pi(m/S_n)$ nous donne le point où a eu lieu le changement.

Le document comporte trois chapitres :

Le premier chapitre comprend quelques notions de base de l'analyse Bayésienne, qui apparaît comme une approche attrayante devant les difficultés théoriques des approches paramétriques et non paramétriques classiques.

Au chapitre deux, Nous abordons les problèmes de rupture. Nous citons quelques modèles qui ont subi un changement de structure à un instant inconnu dans un contexte Bayésien. Et nous introduisons quelques tests d'hypothèses qui nous permettent de dire s'il y a rupture ou pas.

Dans le troisième chapitre, nous proposons d'étudier l'effet d'un contaminant sur l'estimation du point de rupture de certains modèles ayant subi un changement de structure à un instant inconnu en adoptant l'approche Bayésienne. Plus précisément nous nous intéressons à l'étude d'une suite de variables aléatoires gaussiennes et indépendantes, avec rupture dans la moyenne. Nous donnons l'expression de la masse de probabilité a posteriori du point de rupture en présence d'une valeur aberrante (outlier).

En guise de conclusion, nous citons quelques problèmes susceptibles de constituer des pistes de recherches futures.

Chapitre 1

L'analyse statistique Bayésienne

1.1 Introduction

Apprendre en observant, est l'objet principal de l'analyse statistique : il s'agit de développer les outils et le cadre permettant de mieux comprendre un phénomène observé, en vue d'aider à une prise de décision ou parfois, plus simplement en vue de déceler des structures complexes du processus qui engendre les données.

La Statistique doit être considérée comme l'interprétation d'un phénomène naturel, plutôt que son explication. En effet, l'inférence statistique s'accompagne d'une modélisation probabiliste du phénomène observé et implique nécessairement une étape de formalisation réductrice. Sans cette base probabiliste, aucune conclusion utile ne pourra être tirée.

L'objet principal de la Statistique est de faire, grâce à l'observation d'un phénomène aléatoire, une inférence sur la distribution probabiliste à l'origine de ce phénomène, c'est-à-dire faire une analyse afin de donner une description d'un phénomène passé, ou faire une prédiction d'un phénomène arrivant avec une nature similaire.

Depuis un siècle, la statistique s'est considérablement développée, initiant une révolution dans les modes de pensée, car elle porte un langage de représentation du monde et de ses incertitudes. C'est aujourd'hui une science mathématique dont l'objectif est de décrire ce qui s'est produit et de faire des projections quant à ce qu'il peut advenir dans le futur. Parfois, la situation peut être simplement décrite par quelques représentations graphiques d'analyse élémentaire des données. Bien souvent, le problème est beaucoup plus compliqué car de multiples facteurs d'influence doivent être pris en compte.

Le paradigme Bayésien offre un cadre de raisonnement bien adapté à l'intégration des opinions et des faits de toutes provenances qui interviennent dans la gestion des risques et la prise de décision en contexte d'incertitude. De la collecte de données à la prévision, l'analyse statistique pose plusieurs défis. L'élaboration du modèle représente sans doute la phase la plus délicate de l'exercice, car elle doit répondre à un double impératif de réalisme et de parcimonie. Hormis quelques cas de figure, une démarche Bayésienne n'est envisageable qu'à charge de disposer d'outils efficaces pour la quantification et la mise à jour de l'information.

1.2 Notions de base

Nous considérons l'approche paramétrique où nous supposons que les observations x_1, \dots, x_n , sur lesquelles l'analyse statistique se fonde, proviennent de lois de probabilité paramétriques, donc $x_i (i = 1 : n)$ a une distribution de densité $f_i(x_i | \theta_i, x_1, \dots, x_{i-1})$ sur \mathbb{R}^p , telle que le paramètre θ_i soit inconnu et la fonction f_i soit connue. Ce modèle peut être représenté par $x \sim f(x | \theta)$, où x est le vecteur des observations et θ l'ensemble des paramètres. Cette représentation est unificatrice dans le sens où elle aborde de manière similaire une observation isolée, et des observations distribuées de façon indépendantes et identiquement distribuées (*iid*) x_1, \dots, x_n de même loi, $f(x_1 | \theta)$, on a $x = (x_1, \dots, x_n)$ et

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

Une fois le modèle statistique identifié, l'objectif principal de l'analyse statistique est de nous conduire à une inférence sur le paramètre θ . Nous utilisons l'observation de x pour améliorer notre connaissance du paramètre θ , afin de pouvoir prendre une décision concernant le paramètre, c'est à dire d'estimer une fonction de θ ou un futur événement dont la distribution dépend de θ .

1.2.1 Espace de la théorie de la décision statistique

Un système est caractérisé par un certain nombre de variables définissant son état. Les valeurs de ces variables, les mesurandes, sont obtenues par le biais d'un système de mesure. Les résultats de la mesure, que nous appellerons aussi observations, ne permettent qu'une estimation de l'état, car interviennent dans le processus des phénomènes de nature

aléatoire non maîtrisés par l'observateur. On obtient une correspondance entre l'état et l'observation qui est de nature statistique, associant à un état fixé une répartition des observations (gaussienne, poissonnienne ou autre). Le but de la mesure est alors d'inverser cette relation, en ce sens que l'observateur ayant obtenu un résultat, doit en inférer l'état.

La théorie bayésienne de la décision statistique distingue trois espaces.

Le premier espace est celui des états, c'est-à-dire les valeurs des variables ou mesurandes relatives au système. Les états sont notés θ et leur espace Θ .

Le deuxième espace est celui des observations ou des résultats de mesure. Ces résultats seront notés x et leur espace \mathcal{X} .

Le troisième espace est l'espace des décisions noté D , car l'observateur est souvent amené à prendre des décisions. Ces décisions sont envisagées d'une manière très générale. Par exemple, un test d'hypothèses conduisant au choix d'une alternative est associé à un espace discret de décisions, qui n'est autre que celui de l'ensemble des différentes hypothèses envisagées. S'il existe une règle s associant une décision d à une observation x , la fonction $d = s(x)$ est appelée une stratégie.

La relation donnant la répartition des observations pour un état donné se nomme *le modèle*. Elle est objective, extérieure à l'observateur, car elle n'est dépendante que de l'objet mesuré et de l'instrument utilisé. En répétant l'expérience en face du même état, l'observateur verra ses résultats se distribuer selon le modèle.

1.2.2 Théorème de Bayes

Soient A et B deux événements aléatoires tels que $P[B] \neq 0$. La probabilité de A , conditionnellement à la réalisation de B , est par définition exprimée par la relation suivante :

$$P[A|B] = \frac{P[A, B]}{P[B]}$$

où $P[A, B]$ est la probabilité que les deux événements A et B aient lieu simultanément.

Puisque $P[B, A] = P[A, B]$, alors les deux probabilités conditionnelles $P[A|B]$ et $P[B|A]$ sont reliées par :

$$P[A|B] = \frac{P[A].P[B|A]}{P[B]}$$

Cette équation est une conséquence triviale de la définition de la probabilité conditionnelle, est appelée *Formule de Bayes* (ou Théorème de Bayes) en l'honneur du Révérend Thomas Bayes (1702 – 1761) .

Une version continue de ces résultats, nous donne la distribution conditionnelle de y sachant x définie par

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

Avec x et y deux variables aléatoires de distributions conditionnelle $f(x | y)$ et marginale $g(y)$.

L'incertitude sur le paramètre θ d'un modèle peut être décrite par une distribution de probabilité π sur Θ appelée distribution a priori. L'appellation *a priori* exprime le fait qu'elle a été établie préalablement à l'observation des données x . L'inférence est alors fondée sur la distribution de θ conditionnelle à x , $\pi(\theta | x)$, appelée distribution a posteriori définie par

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

Avec $f(x|\theta)$ la probabilité des observations conditionnellement à la valeur θ du paramètre du modèle statistique qu'on utilise pour leur description. Il s'agit de la vraisemblance des données, sous le modèle paramétré par θ .

Comme le dénominateur est indépendant de θ , c'est uniquement une constante de normalisation, la formule de Bayes peut s'écrire comme suit :

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

Le passage de la distribution a priori à la distribution a posteriori des paramètres du modèle statistique, exprimé par la formule de Bayes, peut être alors interprété comme une mise à jour de la connaissance, sur la base des observations.

1.3 Les lois a priori

La loi a priori est le point le plus signifiant de l'inférence Bayésienne, et sa détermination est donc l'étape la plus importante dans la mise en oeuvre de cette inférence.

La loi a priori est une transformation sous une distribution de probabilité du savoir de l'expert, qu'on appelle encore l'expertise déjà connu sur le problème en main, en dehors des informations apportées par les résultats expérimentaux.

Comment passer des informations a priori à des lois a priori ? est la question fondamentale de l'analyse Bayésienne.

1.3.1 Le choix d'une loi a priori

Le point le plus critiqué de l'analyse Bayésienne est le choix de la loi a priori. Car une fois que cette loi est connue, l'inférence peut être conduite d'une façon quasimécanique en minimisant le coût a posteriori, en calculant les régions de plus fortes densités a posteriori ou en intégrant les paramètres pour obtenir la distribution prédictive.

La loi a priori est la clé de voute de l'inférence Bayésienne. Sa détermination, dans une certaine mesure, est l'étape la plus difficile. Evidemment, dans la pratique, il est rare que l'information a priori soit suffisamment précise pour conduire à une détermination exacte de la loi a priori, au sens où plusieurs lois de probabilité peuvent être compatibles avec cette information. Cela revient au fait que le décideur ou le statisticien n'a pas forcément le temps ou les ressources de chercher à construire une loi a priori exacte. Il doit compléter l'information partielle qu'il a rassemblé à l'aide de données subjectives afin d'obtenir une loi a priori.

Même dans le cas où l'information a priori est disponible, il est rare de pouvoir proposer une détermination exacte de lois a priori à partir de cette information.

Il est donc nécessaire le plus souvent de faire un choix arbitraire de loi a priori, ce qui peut avoir un impact considérable sur l'inférence qui en découle. En particulier, l'utilisation systématique de lois usuelles (normale, gamma, bêta, etc.) n'est pas toujours justifiée, car la détermination subjective de la loi a priori qui en résulte se fait au prix d'un traitement analytique plus fruste du problème. Certaines situations requièrent cependant une détermination partiellement automatisée de la loi a priori comme dans le cas extrême où l'information a priori est complètement absente.

Dans la plus part des cas, une certaine imprécision sur la loi a priori employée dans une inférence Bayésienne demeure toujours.

Toutes ces critiques contre l'approche Bayésienne ont une certaine validité au sens où elles attirent l'attention sur le fait qu'il n'y a pas une façon unique de choisir une loi a priori, et que ce choix a un impact sur l'inférence résultante. Cet impact peut être négligeable, modéré ou énorme, puisqu'il est toujours possible d'obtenir la loi a priori qui donnera la réponse qu'on souhaite obtenir. Mais le point essentiel est que, premièrement, les lois a priori non fondées fournissent des inférences a posteriori non justifiées et, deuxièmement, le concept d'une loi a priori unique n'as pas de sens, sauf dans des cas très particuliers.

Nous considérons deux techniques usuelles : une approche qui nécessite une quantité limitée d'informations, et l'approche non informative qui est obtenue à partir de la distribution de l'échantillon.

1.3.2 Lois a priori non informatives

Les lois a priori non informatives représentent une ignorance sur le problème considéré, mais ne signifient pas que l'on sache absolument rien sur la distribution statistique du paramètre.

Lorsque aucune information a priori n'est disponible, le choix de la loi a priori est analytique, puisqu'elles donnent des expressions exactes pour quelques quantités a posteriori. Dans de telles situations, il est impossible de justifier le choix d'une loi a priori sur des bases subjectives. Ces lois a priori particulières doivent être construites à partir de

la distribution d'échantillonnages, puisque c'est la seule information disponible. Pour des raisons évidentes, de telles lois sont dites *non informatives*.

Les lois a priori non informatives peuvent être considérées comme des lois de références, auxquelles chacun pourrait avoir recours quand toute information a priori est absente. Certaines de ces lois sont plus utiles ou plus efficaces que d'autres, mais ne peuvent être perçues comme moins informatives que d'autres.

1.3.2.1 Les lois a priori invariantes

Le fait de formaliser l'absence d'information a priori par une propriété d'invariance est naturelle au sens où seuls les paramètres de la distribution de θ changent lorsqu'on effectue une transformation de θ .

Cette technique de construction des lois non informatives n'est que partiellement satisfaisante, car elle implique la référence à une structure d'invariance, qui peut être parfois choisie de plusieurs manières, ne pas exister, ou être sans intérêt pour le décideur.

1.3.2.2 Les lois a priori de Laplace

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

Trois critiques ont été plus tard avancées sur ce choix. Premièrement, les lois résultantes sont impropres quand l'espace des paramètres n'est pas compact et certains statisticiens se refusent à utiliser de telles lois, car elles mènent à des difficultés comme le paradoxe de marginalisation.

Deuxièmement, le principe des événements équiprobables de Laplace n'est pas cohérent en termes de partitionnement : si $\Theta = \{\theta_1, \theta_2\}$, la règle de Laplace donne $\pi(\theta_1) = \pi(\theta_2) = 1/2$ mais, si la définition de Θ est plus détaillée, avec $\Theta = \{\theta_1, \theta_2, \theta_3\}$, la règle de Laplace mène à $\pi(\theta_1) = 1/3$, ce qui évidemment n'est pas cohérent avec la première formulation, cette cohérence n'est pas un problème important : il peut être évacué en argumentant que le niveau de partitionnement doit être fixé à un certain stade de l'analyse et que l'introduction

d'un degré plus fin dans le partitionnement modifie le problème d'inférence.

La troisième critique est plus fondamentale, car elle concerne le problème de l'invariance par reparamétrisation. Si on passe de $\theta \in \Theta$ à $\eta = g(\theta)$ par une transformation bijective g , l'information a priori reste totalement inexistante et ne devrait pas être modifiée. Cependant, si $\pi(\theta) = 1$, la loi a priori sur η est :

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

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

1.3.2.3 Loi a priori de Jeffreys

les lois a priori non informatives de Jeffreys sont fondées sur l'information de Fisher, donnée par

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

dans le cas unidimensionnel.

Sous certaines conditions, cette information est aussi égale à

$$I(\theta) = -E_{\theta} \left[\frac{\partial^2 \log f(x|\theta)}{\partial \theta^2} \right] \quad (1.1)$$

La loi a priori de Jeffreys est

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

La loi a priori de Jeffreys est invariante par reparamétrisation, puisque pour une transformation bijective donnée h qui transforme le paramètre θ en $h(\theta)$, nous avons la transformation Jacobienne

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

Le choix d'une loi a priori dépendant de l'information de Fisher se justifie par le fait que $I(\theta)$ est largement accepté comme un indicateur de la quantité d'information apportée par le modèle ou l'observation sur θ . (voir Fisher, 1956)

Dans le cas où le paramètre θ est multidimensionnel, on définit la matrice d'information de Fisher par généralisation de l'équation (1.1). Pour $\theta \in \mathbb{R}^k$, $I(\theta)$ a les éléments suivant :

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

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

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

L'approche de Jeffreys fournit une des meilleures techniques pour obtenir une loi a priori non informative, de plus elle nous permet de retrouver les estimateurs classiques. Mais elle a été critiquée par certains Bayésiens comme étant un outil sans justification subjective en termes d'information a priori.

1.3.2.3 Loi a priori de référence

Cette approche est une modification de l'approche de Jeffreys, proposée par Bernardo (1979). Une différence majeure est que cette méthode fait la distinction entre paramètres d'intérêt et paramètres de nuisance. Par conséquent, la loi résultante ne dépend pas seulement de la loi d'échantillonnage, mais aussi du problème inférentiel considéré.

Quand $x \sim f(x/\theta)$ et $\theta = (\theta_1, \theta_2)$, où θ_1 est le paramètre d'intérêt, la loi de référence est obtenue en définissant d'abord $\pi(\theta_2/\theta_1)$ comme la loi de Jeffreys associée à $f(x/\theta)$ pour θ_1 fixé, puis en calculant la loi marginale

$$\tilde{f}(x/\theta_1) = \int f(x/\theta_1, \theta_2) \pi(\theta_2/\theta_1) d\theta_2$$

et la loi de Jeffreys $\pi(\theta_1)$ associée à $\tilde{f}(x/\theta_1)$.

Cette stratégie peut se généraliser si $\theta = (\theta_1, \dots, \theta_n)$, et si l'on ordonne les θ_i par intérêt croissant. Cependant, il est clair que ce raisonnement n'est pas purement objectif parce que donner plus d'importance à un paramètre qu'à un autre relève une fois encore d'un choix.

1.3.2.4 Lois a priori coïncidentes (matching priors)

Une approche particulière de la modélisation non informative, est de s'intéresser aux propriétés fréquentistes de la loi a priori, c'est à dire en moyenne sur x plutôt que conditionnellement à x . En effet, il existe des lois a priori donnant des estimateurs optimaux selon des critères fréquentistes comme la minimaxité ou l'admissibilité, et on peut souhaiter restreindre le choix de la loi a priori à ces distributions optimales.

Une approche plus standard, est d'imposer que certaines probabilités a posteriori coïncident, jusqu'à un certain degré d'approximation, avec la couverture fréquentiste correspondantes ; d'où l'appellation de *lois a priori coïncidentes* (matching priors). (Voir Severeni (1991), Liseo(1993)).

1.3.3 Lois a priori d'entropie maximale

Si l'on dispose de certaines caractéristiques de la loi a priori de type :

$$E^\pi[g_k(\theta)] = \mu_k$$

où pour chaque $k = 1, \dots, n$, g_k est une fonction donnée, la méthode d'entropie maximale développée par Jaynes durant les années (1980,1983), nous permet de déterminer une loi a priori qui satisfait ces contraintes.

Pour $\theta \in \{1, \dots, n\}$ et $\pi(\theta) = (\pi_1, \dots, \pi_n)$ tel que $\pi_i > 0$ et $\sum_{i=1}^n \pi_i = 1$, l'entropie de la loi est définie par :

$$Ent(\pi) = - \sum_{i=1}^n \pi_i \log(\pi_i) \leq - \sum_{i=1}^n \frac{1}{n} \log\left(\frac{1}{n}\right) = \log n \quad (1.2)$$

Ce dernier terme correspond à une répartition uniforme. Pour la masse de Dirac $\delta(j)$ (telle que $\pi_j = 1$ et $\forall i \neq j, \pi_i = 0$), $Ent(\delta(j)) = 0$ ce qui correspond à l'intuition puisqu'il n'y a plus d'incertitude et l'information est totale. Une entropie petite s'interprète comme une loi concentrée et informative. La maximisation de l'entropie sous les contraintes permet de chercher la loi qui apporte le moins d'information. Le principe à la base de cette méthode est donc de chercher à calculer :

$$\arg \max_{\pi} Ent(\pi) \quad \text{sous la contrainte} \quad E^\pi[g_k(\theta)] = \mu_k$$

La solution de ce problème est alors donnée par :

$$\pi^* \propto e^{\sum_{k=1}^n \lambda_k g_k(\theta)}$$

où les λ_k sont les multiplicateurs de Lagrange associés. Dans la pratique, on détermine ces valeurs λ à partir des contraintes.

1.3.4 Lois a priori conjuguées

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

Définition 1.1. Famille conjuguée

Une famille F de distributions de probabilité sur Θ est dite conjuguée (ou fermée par échantillonnage) par une fonction de vraisemblance $f(x|\theta)$ si, pour tout $\pi \in F$, la distribution a posteriori $\pi(\cdot|x)$ appartient également à F .

Un exemple trivial d'une famille conjuguée est l'ensemble F_0 de toutes les lois de probabilité sur Θ . L'avantage des familles conjuguées est avant tout de simplifier les calculs. Avant l'essor du calcul numérique, ces familles étaient pratiquement les seules qui permettaient de faire aboutir des calculs.

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

Définition 1.2. Familles exponentielles

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

$$f(x|\theta) = C(\theta)h(x) \exp\{R(\theta).T(x)\} \quad (1.3)$$

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

$$f(x|\theta) = C(\theta)h(x) \exp\{\theta.x\} \quad (1.4)$$

la famille est dite *naturelle*.

D'un point de vue analytique, les familles exponentielles ont certaines caractéristiques intéressantes, pour tout échantillon de (1.3), en particulier, il existe une statistique exhaustive de dimension constante, en effet, si $x_1, \dots, x_n \sim f(x|\theta)$, avec f satisfaisant (1, 4),

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \in \mathbb{R}^k$$

est exhaustive pour tout n . La réciproque de ce résultat a été aussi établie par Koopman en 1936 et Pitman en 1936.

Théorème 1.1. (*Lemme de Pitman-Koopman*) (*Christian P. Robert 2006*)

Si une famille de lois $f(\cdot|x)$ à support constant est telle que, à partir d'une taille d'échantillon suffisamment grande, il existe une statistique exhaustive de taille fixe, la famille est exponentielle.

Définition 1.3. Soit $f(x|\theta) = C(\theta)h(x) \exp\{\theta \cdot x\}$, une famille exponentielle naturelle. L'espace naturel des paramètres est

$$N = \{\theta; \int_{\mathcal{X}} e^{\theta \cdot x} h(x) d\mu(x) < +\infty\}$$

La famille est dite régulière si N est un ensemble ouvert et minimale si $\dim(N) = \dim(K) = k$, où K est la clôture de l'enveloppe convexe du support de μ .

Remarque 1.1. Les familles exponentielles naturelles peuvent aussi être réécrites sous la forme

$$f(x|\theta) = h(x)e^{\theta \cdot x - \psi(\theta)} \quad (1.5)$$

et $\psi(\theta)$ est dite *fonction cumulante des moments*.

Lois conjuguées des familles exponentielles

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

Proposition 1.1. (*Christian P. Robert 2006*)

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

$$\pi(\theta|\mu, \lambda) = K(\mu, \lambda)e^{\theta \cdot \mu - \lambda\psi(\theta)} \quad (1.6)$$

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

Le tableau ci-dessus représente quelques lois a priori conjuguées naturelles pour quelques familles exponentielles usuelles.

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

Tab 1.2-Lois a priori conjuguées usuelles.

1.3.5 Lois a priori subjectives

Précisons tout d'abord que cette démarche n'est pas forcément facile dans la pratique. L'idée est d'utiliser les données antérieures. Par exemple dans un cadre paramétrique, cela revient à présenter des valeurs ponctuelles de θ à l'expert et pour chacune d'entre elles, de lui demander les chances qu'il lui accorde.

Ces distributions sont dites subjectives parce qu'elles sont propres à l'expert. Elles doivent être interprétées comme un pari de l'expert.

1.3.6 Lois a priori impropres

Une loi impropre ou généralisée est une mesure σ -finie sur l'espace des paramètres Θ , c'est une mesure π telle que

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

Ces lois sont obtenues lorsqu'on dispose des critères subjectifs ou théoriques sur la distribution a priori du paramètre, qui conduisent à une mesure σ -finie sur Θ plutôt qu'à

une mesure de probabilité.

1.4 Les bases de la théorie de la décision

L'objectif général de la plus part des études inférentielles est de fournir une décision au statisticien.

Un problème de décision est en général fondé sur les trois éléments suivants :

- Un ensemble des décisions D
- Un espace des paramètres Θ
- Une fonction de coût (de perte) $L(\theta, \delta)$ qui décrit la perte de prendre la décision δ lorsque le paramètre est θ

1.4.1 Fonction de perte et risque

Soit $\delta \in D$ une règle de décision. La fonction de perte (de coût) est une fonction mesurable de $(\Theta \times D)$ à valeurs réelles positives notée $L(\theta, \delta)$

$$L : \Theta \times D \rightarrow \mathbb{R}_+$$

Elle est définie selon le problème étudié et constitue l'armature du problème statistique.

Définition 1.4. Risque fréquentiste

Pour une fonction de perte donnée $L(\theta, \delta)$, le risque fréquentiste est défini par :

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

où $\delta(x)$ est la règle de décision.

C'est une fonction de θ et ne définit pas un ordre total sur D et ne permet pas de comparer toutes décisions et estimateurs. Il n'existe donc pas de meilleur estimateur dans un sens absolu. Ainsi, l'approche fréquentiste restreint l'espace d'estimation en préférant la classe des estimateurs sans biais dans laquelle il existe des estimateurs de risque uniformément minimal ; l'école Bayésienne ne perd pas en généralité en définissant un risque a posteriori. L'idée est d'intégrer sur l'espace des paramètres pour pallier cette difficulté.

Définition 1.5. Risque a posteriori

Une fois données la loi a priori sur le paramètre et la fonction de perte, le risque a posteriori est défini par :

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

Ainsi, le problème change selon les données ; ceci est dû à la non existence d'un ordre total sur les estimateurs.

Définition 1.6. Risque intégré

Pour une fonction de perte donnée, le risque intégré est défini par :

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

Une fois la loi a posteriori sur le paramètre est disponible, le problème de l'estimation Bayésienne ponctuelle peut être exprimé comme un problème de décision.

1.4.2 Estimateurs de Bayes

Soit une fonction de coût $L(\theta, \delta)$, et une loi de probabilité a priori (ou une loi impropre) π , pour trouver l'estimateur de Bayes $\delta^\pi(x)$, on applique la règle suivante :

$$\delta^\pi(x) = \min_{\delta} E^\pi[L(\theta, \delta)/x]$$

L'estimateur $\delta^\pi(x)$ sera déterminé analytiquement ou numériquement, ceci dépendra de la fonction de perte, de sa nature et complexité.

Généralement, les solutions associées à des coûts classiques sont formellement connues et correspondent aux caractéristiques usuelles d'une distribution (moyenne, médiane, fractiles, etc.).

L'estimateur de Bayes δ^π est un estimateur minimisant $r(\pi, \delta)$. C'est à dire vérifiant :

$$r(\pi, \delta^\pi) = \inf_{\delta \in D} r(\pi, \delta) < \infty$$

La valeur $r(\pi, \delta^\pi)$ est alors appelée risque de Bayes.

Pour obtenir la valeur de l'infimum du risque intégré il faut donc en théorie minimiser une intégrale double de δ . L'introduction du risque intégré se justifie par le théorème suivant. Il suffira de minimiser une grandeur qui ne dépend plus que des données, ceci permet donc d'arriver à des estimateurs satisfaisants.

Théorème 1.2. *Méthode de calcul (Judith Rousseau (2009))*

Si $\exists \delta \in D, r(\pi, \delta) < \infty$ et $\forall X \in \chi, \delta^\pi(X) = \arg \min_{\delta} \rho(\pi, \delta|X)$, alors $\delta^\pi(X)$ est un estimateur Bayésien.

1.4.3 Fonctions de coût usuelles

1.4.3.1 Perte quadratique

Introduit par Légendre (1805) et Gauss (1810), ce coût est sans conteste le critère d'évaluation le plus commun. Fondant sa validité sur l'ambiguïté de la notion d'erreur dans un contexte statistique (soit erreur de mesure, soit variation aléatoire), il a aussi donné lieu à de nombreuses critiques, la plus fréquente étant sans doute le fait que le coût quadratique

$$L(\theta, \delta) = (\theta - \delta)^2 \tag{1.7}$$

pénalise trop fortement les grandes erreurs.

les estimateurs de Bayes associés au coût quadratique sont les moyennes a posteriori. Cependant, notons que le coût quadratique n'est pas le seul coût à avoir cette caractéristique. Les fonctions de coût conduisant à la moyenne a posteriori comme estimateur de Bayes sont appelées *fonctions de coût propres* et ont été identifiées par Lindley 1985 et Hwang et Pemantle(1994).

Proposition 1.2. *(Christian P. Robert 2006)*

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

$$\delta^\pi(x) = E^\pi[\theta|x] = \frac{\int_{\theta} \theta f(x|\theta)\pi(\theta)d\theta}{\int_{\theta} f(x|\theta)\pi(\theta)d\theta} \tag{1.8}$$

Corollaire 1.1. (Christian P. Robert 2006)

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

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

est la moyenne a posteriori, $\delta^\pi(x) = E^\pi[\theta|x]$, pour toute matrice Q ($p \times p$) symétrique définie positive.

Le coût quadratique est particulièrement intéressant lorsque l'espace des paramètres est borné et le choix d'un coût plus subjectif est impossible.

Le tableau ci-dessus représente quelques estimateurs de Bayes du paramètre θ sous coût quadratique pour les lois a priori conjuguées des familles exponentielles usuelles.

Loi de x	Loi conjuguée	Moyenne a posteriori
Normale $\mathcal{N}(\theta, \sigma^2)$	Normale $\mathcal{N}(\mu, \tau^2)$	$\frac{\mu\sigma^2 + \tau^2x}{\sigma^2 + \tau^2}$
Poisson $P(\theta)$	Gamma $\mathcal{G}(\alpha, \beta)$	$\frac{\alpha + x}{\beta + 1}$
Gamma $\mathcal{G}(\nu, \theta)$	Gamma $\mathcal{G}(\alpha, \beta)$	$\frac{\alpha + \nu}{\beta + x}$
Binomiale $B(n, \theta)$	Beta $Be(\alpha, \beta)$	$\frac{\alpha + x}{\alpha + \beta + n}$
Binomiale Négative $Neg(m, \theta)$	Bêta $Be(\alpha, \beta)$	$\frac{\alpha + n}{\alpha + \beta + x + n}$
Multinomiale $M_k(\theta_1, \dots, \theta_k)$	Dirichlet $D(\alpha_1, \dots, \alpha_k)$	$\frac{\alpha_i + x_i}{(\sum_j \alpha_j) + n}$
Normale $\mathcal{N}(\mu, 1/\theta)$	Gamma $\mathcal{G}(\alpha, \beta)$	$\frac{\alpha + 1}{\beta + (\mu - x)}^2$

Tab 1.1-Quelques estimateurs de Bayes usuels.

1.4.3.2 L'erreur de coût absolu

Une solution alternative au coût quadratique en dimension un est d'utiliser le coût absolu,

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

déjà considéré par Laplace en 1773 ou, plus généralement, une fonction linéaire par morceaux

$$L_{k_1, k_2}(\theta, d) = \begin{cases} k_2(\theta - d) & \text{si } \theta > d \\ k_1(d - \theta) & \text{sinon} \end{cases}$$

Proposition 1.3. (*Christian P. Robert 2006*)

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

En particulier, si $k_1 = k_2$, dans le cas du coût absolu, l'estimateur de Bayes est la médiane a posteriori, qui est l'estimateur obtenu par Laplace

1.4.3.3 Le coût 0 – 1

Ce coût est surtout utilisé dans l'approche classique des tests d'hypothèse, proposée par Neyman et Pearson. Plus généralement, c'est un exemple typique d'un coût non quantitatif. En effet, pour ce coût, la pénalité associée à un estimateur δ est 0 si la réponse est correcte et 1 sinon.

Exemple 1.1. *Soit le test de $H_0 : \theta \in \Theta_0$ contre $H_1 : \text{sinon}$. Alors $D = \{0, 1\}$, où 1 représente l'acceptation de H_0 et 0 son rejet, pour la fonction de coût 0 – 1, qui vaut*

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

le risque associé est

$$\begin{aligned} R(\theta, \delta) &= E_\theta[L(\theta, \delta(x))] \\ &= \begin{cases} P_\theta(\delta(x) = 0) & \text{si } \theta \in \Theta_0 \\ P_\theta(\delta(x) = 1) & \text{sinon} \end{cases} \end{aligned}$$

ce qui donne les erreurs de première et deuxième espèce qui sous-tendent la Théorie de Neyman-Pearson

Proposition 1.4. (*Christian P. Robert 2006*)

L'estimateur de Bayes associé à π et au coût $0 - 1$ est

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

donc $\delta^\pi(x)$ vaut 1 si et seulement si $P(\theta \in \Theta_0|x) > 1/2$.

1.4.4 Admissibilité et minimaxité

Définition 1.7. Estimateur randomisé

Pour un ensemble de décisions D , on définit D^* comme l'ensemble des probabilités sur D . L'estimateur $\delta^* \in D^*$ est appelé estimateur randomisé.

Cette extension est nécessaire au traitement des notions d'admissibilité et de minimaxité. L'ensemble D^* apparaît comme complétion topologique de D . Cependant cette modification de l'espace de décision ne modifie pas les réponses Bayésiennes, comme le montre le résultat suivant

Théorème 1.3. (*Christian P. Robert 2006*)

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

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

1.4.4.1 Minimaxité

Le critère de minimaxité apparaît comme une assurance contre le pire, car il vise à minimiser le coût moyen dans le cas le moins favorable. Il représente aussi un effort fréquentiste pour éviter de recourir au paradigme Bayésien, tout en engendrant un ordre (faible) sur D^* .

Définition 1.8. On appelle risque minimax associé à la fonction de coût L la valeur :

$$\bar{R} = \inf_{\delta \in D^*} \sup_{\theta} R(\theta, \delta) = \inf_{\delta \in D^*} \sup_{\theta} \mathbb{E}_{\theta}[L(\theta, \delta(x))] \quad (1.9)$$

et estimateur minimax tout estimateur δ_0 tel que

$$\bar{R} = \sup_{\theta} R(\theta, \delta_0) \quad (1.10)$$

L'estimateur minimax correspond au point de vue de faire le mieux dans le pire des cas, c'est-à-dire à s'assurer contre le pire. Il est utile dans des cadres complexes mais trop conservateur dans certains cas où le pire est très peu probable. Il peut être judicieux de voir l'estimation comme un jeu entre le statisticien (choix de δ) et la Nature (choix de θ), l'estimation minimax rejoint alors celle de la Théorie des Jeux.

règle minimax et stratégie maximin

Une difficulté importante liée à la notion de minimaxité est que les estimateurs minimax n'existent pas nécessairement. En particulier, il existe une stratégie minimax quand Θ est fini et la fonction de coût est continue. Plus généralement, Brown (1976) (voir aussi Le Cam, 1986, et Strasser, 1985) considère l'espace de décision D comme plongé dans un autre espace de manière telle que l'ensemble des fonctions de risque sur D est compact dans ce grand espace. Dans cette perspective et sous des hypothèses supplémentaires, il est alors possible de construire des estimateurs minimax lorsque la fonction de coût est continue.

Théorème 1.4. (*Blackwel et Girshick 1954*)

Si $D \subset \mathbb{R}^k$ est convexe et compact et si $L(\theta, d)$ est continue et convexe en tant que fonction de d , pour chaque $\theta \in \Theta$, alors, il existe un estimateur minimax non randomisé.

Lemme 1.1. (*Christian P. Robert 2006*)

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

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

La première valeur est dite *risque maximin* et une distribution π^* telle que $r(\pi^*) = \underline{R}$ est appelée *distribution a priori la moins favorable*, quand de telles distributions existent. En général, la borne supérieure $r(\pi^*)$ est atteinte plutôt par une distribution impropre pouvant s'exprimer comme une limite de distributions a priori propres π_n . Mais ce phénomène n'empêche pas nécessairement la construction d'estimateurs minimax. Dans le cas où elles

existent, les distributions les moins favorables sont celles qui ont le risque de Bayes le plus grand, donc aussi les moins intéressantes en terme de coût lorsqu'elles ne sont pas suggérées par l'information a priori disponible. Le résultat ci-dessus est assez logique au sens où l'information a priori ne peut qu'améliorer l'erreur d'estimation, même dans le pire des cas.

Un cas particulier intéressant correspond à la définition suivante :

Définition 1.9. Un problème d'estimation est dit *admettre une valeur* si $\underline{R} = \overline{R}$, c'est-à-dire quand

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

Quand le problème admet une valeur, certains estimateurs minimax sont des estimateurs de Bayes correspondant aux lois a priori les moins favorables. Cependant, ils peuvent être randomisés. Par conséquent le principe minimax ne fournit pas toujours des estimateurs acceptables.

D'un point de vue pratique, le lemme suivant fournit des conditions suffisantes de minimaxité.

Lemme 1.2. (*Christian P. Robert 2006*)

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

1.4.4.2 Admissibilité

Le critère d'admissibilité induit un ordre partiel sur D^* en comparant les risque fréquentistes des estimateurs $R(\theta, \delta)$.

Définition 1.10. Estimateur admissible

Soit un modèle paramétrique et une fonction de perte L sur $\Theta \times D$ où D est l'ensemble des décisions. On dit que $\delta \in \Theta$ est inadmissible s'il existe un estimateur δ_0 qui domine δ , c'est -à- dire tel que pour tout θ ,

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

et pour au moins une valeur θ_0 du paramètre

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

Dans le cas contraire, δ est admissible.

Proposition 1.5. (Christian P. Robert 2006)

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

Notons que la réciproque de ce résultat est fautive, car il peut exister plusieurs estimateurs minimax admissibles. Par exemple, dans le cas $N_p(\theta, I_p)$, il existe des estimateurs de Bayes réguliers minimax pour $p \geq 5$. Quand la fonction de coût L est absolument convexe (en d), la caractérisation suivante est aussi possible.

Proposition 1.6. (Christian P. Robert 2006)

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

Proposition 1.7. (Christian P. Robert 2006)

Si l'estimateur Bayésien δ^π associé à une fonction de perte L et une loi a priori π est unique, alors il est admissible

Proposition 1.8. (Christian P. Robert 2006)

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

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

soit fini, δ^π est admissible.

Définition 1.11. π -admissibilité

Un estimateur δ_0 est π -admissible si et seulement si

$$\forall(\delta, \theta), R(\theta, \delta) \leq R(\theta, \delta_0) \Rightarrow \pi(\{\theta \in \Theta, R(\theta, \delta) < R(\theta, \delta_0)\}) = 0$$

Propriété 1.1. (Christian P. Robert 2006)

Tout estimateur Bayésien tel que $r(\pi) < \infty$ est π -admissible

Théorème 1.5. *Continuité et π -admissibilité, (Judith Rousseau (2009))*

Si $\pi > 0$ sur Θ , $r(\pi) < \infty$ pour une fonction de perte L donnée, si δ^π estimateur Bayésien correspondant existe et si $\theta \mapsto R(\theta, \delta)$ est continu, alors δ^π est admissible.

1.4.5 Estimateur du maximum a posteriori MAP

Définition 1.12. On appelle estimateur MAP tout estimateur $\delta^\pi(x)$ qui maximise l'information sur θ représentée par sa loi a posteriori, c'est-à-dire tout estimateur $\delta^\pi(x)$ tel que

$$\delta^\pi(x) \in \arg \max_{\theta} \pi(\theta|x)$$

Cet estimateur naturel peut s'exprimer comme un estimateur du maximum de vraisemblance pénalisée au sens classique, il a le grand avantage de ne pas dépendre d'une fonction de perte et est utile pour les approches théoriques. Ses inconvénients sont les mêmes que l'estimateur du maximum de vraisemblance : non unicité, instabilité (dus aux calculs d'optimisation) et dépendance vis-à-vis de la mesure de référence (dominant Θ). En outre, il ne vérifie pas la non invariance par reparamétrisation qui peut apparaître importante intuitivement.

1.4.6 Tests et intervalles de crédibilité

D'un point de vue statistique, un test soit au sens Bayésien ou classique, peut être considéré comme une des deux approches :

-Soit comme un procédé statistique, c'est-à-dire une fonction définie sur l'espace des observations à valeurs dans un espace à deux points que l'on appelle "accepter" et "rejeter" une hypothèse. Dans ce cas, on peut envisager un problème de test comme un problème de décision avec deux actions possibles.

-Sinon, comme une façon pour le statisticien de gérer ses doutes relatifs à son modèle statistique.

Comme dans le cas de l'estimation, un test Bayésien se fait après avoir calculé la loi a posteriori.

1.4.6.1 Intervalles de crédibilité

Définition 1.13. Région α -crédible

Pour $0 < \alpha < 1$, une région α -crédible de $100(1 - \alpha)\%$ pour θ est un sous-ensemble $C \in \Theta$ tel que $P^\pi\{\theta \in C \mid X = x\} = 1 - \alpha$ habituellement C est un intervalle.

Il existe une infinité de région α -crédibles, il est logique de s'intéresser donc à celle qui a un volume minimal. Pour cela, nous allons introduire la notion d'une région HPD (Highest Posterior density).

Définition 1.14. Région HPD

Une région HPD est la région C_α^π définie par

$$C_\alpha^\pi = \{\theta, \pi(\theta \mid X = x) \geq h_\alpha\}$$

où $h_\alpha = \sup\{h; P^\pi(\{\theta, \pi(\theta \mid x) \geq h\} \mid x) \geq 1 - \alpha\}$

Les régions HPD peuvent être calculées numériquement ou approximativement comme elles peuvent être calculées par des méthodes de simulation.

1.4.6.2 Le facteur de Bayes

Définition 1.15. 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 :

$$B_F = \frac{P(\theta \in \Theta_0 \mid x)}{P(\theta \in \Theta_1 \mid x)} / \frac{\pi(\theta \in \Theta_0)}{\pi(\theta \in \Theta_1)}$$

Ce rapport est une analogue Bayésienne du rapport de vraisemblances des tests classiques, il évalue la modification de la vraisemblance de l'ensemble Θ_0 par rapport à celle de l'ensemble Θ_1 due à l'observation et peut se comparer naturellement à 1.

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 π_0 est la loi a priori sous H_0 , et π_1 la loi a priori sous H_1 , B_F peut s'écrire :

$$B_F = \frac{\int_{\Theta_0} f(x | \theta_0) \pi_0(\theta) d\theta}{\int_{\Theta_1} f(x | \theta_1) \pi_1(\theta) d\theta}$$

1.5 Méthodes de calcul Bayésien

La simplicité ultime de l'approche Bayésienne est que, pour une fonction de coût et une loi a priori données, l'estimation Bayésienne associée à une observation x est la décision d minimisant le coût a posteriori, dans la pratique, minimiser un tel coût peut être difficile car le calcul explicite de la loi a posteriori peut être impossible des fois.

Une réponse simple à ces difficultés de calcul est de n'utiliser que des modèles d'échantillonnage des lois a priori et des coûts qui mènent à des solutions explicites pour minimiser les coûts.

Cette approche restrictive est justifiée lorsque les outils de calculs ne sont pas applicables en termes subjectifs, car la fonction de coût et la loi a priori devraient être construites en fonction du problème de décision et non pas parce qu'elles fournissent des réponses analytiques.

Pour éviter le recours systématique à des lois a priori et à des coûts simples, différentes méthodes d'approximation peuvent être utilisées lorsque la loi a posteriori ou un estimateur donné n'admettent pas d'expression analytique.

1.5.1 Méthodes classiques d'approximation

L'intégration numérique est l'une des techniques classiques utilisées dans les calculs Bayésiens. En effet, à partir de la simple méthode de Simpson, plusieurs approches ont été conçues en mathématiques appliquées pour l'approximation numérique d'intégrales. Par exemple, la quadrature polynomiale est censée approcher les intégrales liées à des

distributions proches de la loi normale. L'approximation de base est donnée par

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

où

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

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

D'autres approximations d'intégrales reliées à cette méthode sont disponibles, qui reposent sur différentes bases orthogonales classiques (voir Abramowitz et Stegun, 1964) mais ces méthodes requièrent généralement des hypothèses de régularité sur la fonction f , ainsi que des études préliminaires pour déterminer quelle base est la plus adéquate et à quel point cette approximation est précise. Par exemple, des transformations du modèle peuvent être nécessaires pour mettre en pratique l'approximation d'Hermite (voir Naylor et Smith, 1982, et Hills et Smith, 1992).

Quelle que soit la méthode d'intégration numérique utilisée, sa précision diminue dramatiquement lorsque la dimension de Θ augmente. De façon plus spécifique, l'erreur associée aux méthodes numériques se comporte comme une puissance de la dimension de Θ . En pratique, une règle empirique est que la plupart des méthodes standard ne devraient pas être utilisées pour l'intégration en dimension supérieure à 4. En effet, la taille de la partie de l'espace non pertinente pour le calcul d'une intégrale donnée augmente considérablement avec la dimension de l'espace. Ce problème est appelé fléau de la dimension.

Dans un problème statistique, l'approximation de l'intégrale

$$\int_{\Theta} g(\theta) f(x|\theta) \pi(\theta) d\theta \tag{1.11}$$

doit tirer avantage de la nature particulière, à savoir le fait que π soit une densité de probabilité ou plutôt, que $f(x|\theta)\pi(\theta)$ soit proportionnel à une densité. Une conséquence naturelle de cette perspective est d'utiliser la méthode de Monte Carlo, introduite par Metropolis et Ulam (1949) et Von Neumann (1951).

Une discussion plus détaillée sur les méthodes de Monte Carlo et leurs applications à la statistique Bayésienne peuvent être trouvées dans Robert et Casella (2004) et Chen et al

(2000).

Par comparaison aux méthodes d'intégration numérique, les méthodes de Monte Carlo présentent en effet l'avantage que, une fois l'échantillon $\theta_1, \dots, \theta_n$ produit, celui-ci peut être utilisé à plusieurs reprises pour tous les objectifs inférentiels, incluant l'obtention des règles de Bayes à partir du coût a posteriori approché.

Lorsque la fonction à intégrer est assez régulière, il existe une solution alternative analytique mais asymptotique aux simulations de Monte Carlo. Cette méthode a été introduite par Laplace et est par conséquent appelée *approximation de Laplace*. (Voir Tierney et al (1989), Kass et Steffey(1989))

1.5.2 Méthodes de Monte Carlo par chaînes de Markov (MCMC)

La méthode de Monte Carlo par chaînes de Markov (abrégée en MCMC) est une méthode de Monte Carlo plus générale, permettant d'approcher la génération de variables aléatoires d'une loi a posteriori $\pi(\theta/x)$, lorsque cette loi ne peut pas être simulée directement.

L'avantage de cette méthode sur les méthodes de Monte Carlo classiques est qu'elle ne demande pas de connaître la constante de normalisation, ce qui est en pratique le cas de la distribution a posteriori de la statistique Bayésienne. Les méthodes de Monte-Carlo par chaîne de Markov (MCMC) génèrent une suite de variables aléatoires $(\theta^1, \dots, \theta^n, \dots)$ et, hormis la première à laquelle on donne une valeur arbitraire, chacune d'entre elles dépend uniquement de celle qui la précède. Les calculs sont ensuite poursuivis en appliquant à cette séquence une loi des grands nombres pour les chaînes markoviennes érgodiques de forme identique.

Les techniques les plus importantes conçues pour créer des chaînes de Markov de loi stationnaire donnée, sont les algorithmes de Metropolis-Hastings et l'échantillonnage de Gibbs.

1.5.2.1 Algorithme de Metropolis-Hastings

La technique de Metropolis-Hastings a été développée par Metropolis et al. (1953), au départ pour la physique particulaire, et généralisée par Hastings (1970) dans un cadre plus statistique.

Pour une densité donnée $\pi(\theta)$, et une densité conditionnelle $q(\theta'/\theta)$, l'algorithme génère la chaîne $(\theta^{(t)})_t$ comme suit :

Pour $\theta^{(0)}$ choisie arbitrairement comme une valeur initiale, on définit par récurrence les valeurs de $\theta^{(t)}$.

A l'étape t , à partir de $\theta^{(t-1)}$, $\theta^{(t)}$ est construit en tirant un θ' à l'aide d'une distribution de probabilité instrumentale : $\theta' \sim q(\cdot|\theta^{(t-1)})$. $\theta^{(t)}$ est alors donné par :

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

où

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

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

Notons qu'il est possible suivant cette construction de rester au même endroit après une itération. On peut alors montrer en écrivant la condition de balance, que pour ce choix de α , on obtient une chaîne de Markov de loi stationnaire π .

Cette chaîne de Markov est ergodique si et seulement si $(\theta^{(t)})_t$ est irréductible et apériodique.

1.5.2.2 L'échantillonnage de Gibbs

Une autre méthodes de Monte Carlo par chaînes de Markov (MCMC) est l'échantillonnage de Gibbs. Cette méthode tire son nom des champs aléatoires de Gibbs où elle a été utilisée pour la première fois par Geman et Geman (1984). L'approche de l'échantillonnage de Gibbs repose sur une perspective différente de l'algorithme de Metropolis-Hastings, elle ne demande pas de mettre en place une fonction de proposition, elle est fondée sur la loi cible et essentiellement basée sur des distributions conditionnelles complètes. De plus

l'algorithme de Gibbs pour l'estimation et construction de modèle par conditionnement probabiliste donne souvent de meilleurs résultats.

Pour $\theta = (\theta_1, \dots, \theta_p)$, on veut simuler $\pi(\theta)$ à partir de $\pi_i(\theta_i|\theta_{(-i)}) = \pi_i(\theta_i|\theta_j, j \neq i)$ pour tout i . On initialise avec $\theta^{(0)}$ et à l'instant t , on écrit :

$$\begin{aligned}(\theta_1^{(t)}|\theta^{(t-1)}) &\sim \pi_1(\theta_1^{(t)}|\theta_{(-1)}^{(t-1)}) \\(\theta_2^{(t)}|\theta^{(t-1)}, \theta_1^{(t)}) &\sim \pi_2(\theta_2^{(t)}|\theta_1^{(t)}, \theta_3^{(t-1)}, \dots, \theta_p^{(t-1)}) \\(\theta_p^{(t)}|\theta^{(t-1)}, \theta_{(-p)}^{(t)}) &\sim \pi_p(\theta_p^{(t)}|\theta_{(-p)}^{(t)})\end{aligned}$$

Dans le cas où une telle loi π existe, $\theta^{(t)}$ issu de cet algorithme est une chaîne de Markov ergodique de loi stationnaire π .

1.6 Conclusion

La pertinence et l'efficacité de l'approche Bayésienne, comme guide du raisonnement scientifique face à l'incertitude, sont reconnues depuis longtemps (de Finetti, 1937 ; Savage, 1954 ; etc.). La mise en oeuvre des principes Bayésiens, en dehors de cas d'école, s'est longtemps heurtée aux difficultés pratiques de calcul. Les moyens informatiques, dont on disposait avant les années 1990, n'étaient pas suffisamment puissants. Les problèmes réels, avec leurs dimensions et leurs complexités importantes, faisaient alors la part belle aux méthodes statistiques classiques. La situation, depuis lors, a subi une véritable révolution (Brooks, 2003). Maintenant on peut affirmer qu'il n'existe, au moins a priori, aucun contre-argument justifie à l'emploi des méthodes Bayésiennes quelle que soit la complexité du cas envisagé. On doit ce nouveau paysage scientifique au développement de nouveaux outils de calcul : les méthodes MCMC (simulations Monte Carlo par Chaînes de Markov) et à l'amélioration des anciens (échantillonnage pondéré ou importance sampling et méthodes des particules). Les fondements conceptuels de ces méthodes de calcul sont solidaires des modes de raisonnements conditionnels de la modélisation Bayésienne et le paradigme Bayésien apparaît comme une démarche rationnelle, efficace et solidement intégrée du programme complet : (modélisation, calcul et décision).

Chapitre 2

Les problèmes de rupture

2.1 Introduction

L'étude de rupture de modèle est devenue l'un des passages incontournables pour les économètres et les statisticiens. En effet, la stabilité d'un modèle si elle venait à faire défaut doit être décelée sinon elle risque de conduire à des résultats erronés et entraîner des interprétations fausses. Il est donc impératif de s'assurer de la stabilité ou non du modèle étudié. Pour cela, le statisticien est appelé à faire des tests pour détecter un éventuel changement des paramètres de son modèle, estimer s'il y a lieu le ou les points de rupture, identifier les paramètres sujets à un changement et enfin estimer leur amplitude.

Quand un modèle paramétrique change de structure, il est important de connaître le point où a eu lieu le changement ainsi que les valeurs des paramètres du modèle avant et après la rupture.

On dira qu'il y a rupture dans une suite de variables aléatoires X_1, X_2, \dots, X_n s'il existe un entier m ; ($1 < m \leq n - 1$) ; tel que les variables aléatoires X_1, X_2, \dots, X_m suivent une loi de probabilité de fonction de répartition $F_1(x/\theta_1)$ et les variables aléatoires X_{m+1}, \dots, X_n suivent une autre loi de probabilité de fonction de répartition $F_2(x/\theta_2)$, où θ_1 et θ_2 sont des paramètres inconnus réels ou vectoriels ($\theta_1 \neq \theta_2$). Le point m est appelé point de rupture ou point de changement. Dans la littérature anglo-saxonne il est connu sous le nom de shift point ou change point.

Dans ce chapitre, nous citons quelques résultats d'estimation d'un point de rupture de certains modèles, ayant subi un changement de structure, par la méthodologie Bayésienne.

La nature précise du changement du modèle est déterminée par le problème considéré. En analyse des échantillons aléatoires, le changement peut porter sur la moyenne, la variance ou sur d'autres paramètres inconnus de la distribution. Dans l'étude de régression il concerne les paramètres de régression ou bien la distribution des erreurs. Dans la deuxième partie, nous introduisons les tests d'hypothèses.

2.2 Modèle Gaussien

2.2.1 Rupture dans la moyenne

Nous nous intéressons dans cette section à une analyse Bayésienne d'une suite de variables aléatoires Gaussiennes, ayant subi un changement dans la moyenne à un instant inconnu. Soient X_1, X_2, \dots, X_n cette suite de moyennes ϕ_i , $i = 1, \dots, n$ et de variance σ^2 .

2.2.1.1 Présentation du modèle

Considérons le modèle suivant :

$$\begin{cases} X_i = \phi_0 + \epsilon_i; & i = 1, 2, \dots, m \\ X_i = \phi_1 + \epsilon_i; & i = m + 1, \dots, n \end{cases} \quad (2.1)$$

Où :

ϕ_0 et ϕ_1 sont des constantes réelles, inconnues et différentes. Elles représentent les moyennes des variables aléatoires X_i avant et après le changement respectivement.

Les ϵ_i représentent les erreurs aléatoires, qui peuvent être indépendantes ou autocorrélées. Le paramètre m représente le point de rupture. C'est un entier qui varie entre 1 et $n - 1$, n étant la taille de l'échantillon.

2.2.1.2 Cas où les erreurs sont indépendantes

Nous considérons le modèle (2.1) et nous supposons que les erreurs aléatoires ϵ_i ; $i = 1, \dots, n$ sont gaussiennes, indépendantes et identiquement distribuées de moyenne nulle et de variance constante et inconnue σ^2 ; ($\epsilon_i \sim N(0, \sigma^2)$).

La fonction de vraisemblance

Comme $X_i \sim N(\phi_0, \sigma^2)$ pour $i = 1, \dots, m$,
 et $X_i \sim N(\phi_1, \sigma^2)$ pour $i = m + 1, \dots, n$.

Alors la fonction de vraisemblance L associée aux variables X_1, X_2, \dots, X_n est donnée par :

$$L(m, \sigma^2, \phi_0, \phi_1/X) \propto (\sigma^2)^{-n/2} \exp \left\{ -1/(2\sigma^2) \left[\sum_{i=1}^m (X_i - \phi_0)^2 + \sum_{i=m+1}^n (X_i - \phi_1)^2 \right] \right\}$$

L'analyse a posteriori

La densité conjointe a posteriori des paramètres $(m, \sigma^2, \phi_0, \phi_1)$ est donnée en vertu du théorème de Bayes par :

$$\pi_1(m, \sigma^2, \phi_0, \phi_1) \propto L(m, \sigma^2, \phi_0, \phi_1/X) \cdot \pi_0(m, \sigma^2, \phi_0, \phi_1)$$

Avec $\pi_0(m, \sigma^2, \phi_0, \phi_1)$ la densité conjointe a priori des différents paramètres du modèle.

Premier cas :

En assignant des lois a priori impropres (ou non informatives) aux paramètres σ^2 et (ϕ_0, ϕ_1) et une loi uniforme sur $1, 2, \dots, n - 1$ pour le point de rupture m , le théorème suivant nous détermine la masse de probabilité marginale a posteriori du point de rupture m .

Théorème 2.1. (Broemeling L.D et Holbert D, 1977)

Etant donné le modèle (2.1)

$$\begin{cases} X_i = \phi_0 + \epsilon_i; & i = 1, 2, \dots, m \\ X_i = \phi_1 + \epsilon_i; & i = m + 1, \dots, n \end{cases}$$

Où : $\phi_0 \neq \phi_1$ et $\epsilon_i \sim N(0, \sigma^2)$; ($\sigma > 0$) avec ϕ_0, ϕ_1 et σ sont des paramètres inconnus.

Si les densités a priori des paramètres $\sigma^2, (\phi_0, \phi_1)$ et du point de rupture m sont données respectivement par :

$$\pi_0(\sigma^2) \propto 1/\sigma^2$$

$$\pi_0(\phi_0, \phi_1) \propto \text{constante sur } \mathbb{R}^2$$

$$\pi_0(m) \propto 1/(n-1) \text{ pour } m = 1, 2, \dots, n-1$$

et $\sigma^2, (\phi_0, \phi_1)$ et m sont indépendants.

Alors la masse de probabilité a posteriori du point de rupture m est donnée par :

$$\pi_1(m) = \pi(m/X) \propto [m(n-m)]^{-1/2} [S_1^m + S_{m+1}^n]^{-(n-2)/2}$$

$$\text{où : } S_j^k = \sum_{i=j}^k (X_i - \bar{X}_j^k)^2$$

$$\text{et } \bar{X}_j^k = \frac{1}{k-j+1} \sum_{i=j}^k X_i, \text{ pour } k = 1, \dots, n.$$

Pour déterminer les densités a posteriori marginales des paramètres ϕ_0, ϕ_1 et σ^2 nous avons le corollaire suivant :

Corollaire 2.1. (Broemeling L.D et Holbert D, 1977)

Sous les hypothèses du théorème précédent, les densités a posteriori marginales des paramètres ϕ_0, ϕ_1 et σ^2 sont données respectivement par :

$$i) \pi(\phi_0/X) \propto \sum_{m=1}^{n-1} \pi_1(m) t(n-2, y) \cdot \sqrt{m} (S_1^m + S_{m+1}^n)^{-1/2}.$$

où $t(n-2, y)$ est la densité d'une loi de Student à $(n-2)$ degrés de liberté de variable y . avec $y = \sqrt{m(n-2)} ((\bar{X}_1^m - \phi_0) / \sqrt{(S_1^m + S_{m+1}^n)})$.

$$ii) \pi(\phi_1/X) \propto \sum_{m=1}^{n-1} \pi_1(m) t(n-2, z) \cdot \sqrt{n-m} (S_1^m + S_{m+1}^n)^{-1/2}.$$

où $t(n-2, z)$ est la densité d'une loi de Student à $(n-2)$ degrés de liberté de variable z . avec $z = \sqrt{n-m}(\bar{X}_{m+1}^n - \phi_1) / \sqrt{(S_1^m + S_{m+1}^n)}$.

$$iii) \pi(\sigma^2/X) \propto \sum_{m=1}^{n-1} \pi_1(m) \cdot g(r, a^*, S_1^m + S_{m+1}^n).$$

où : $r = \sigma^{-2}$ et g est la densité d'une loi gamma de paramètres $a^* = n/2 + 1$ et $(S_1^m + S_{m+1}^n)$.

Deuxième cas :

Après avoir assigné une densité a priori non informative à la variance σ^2 , intéressons-nous au cas où nous assignons à la précision des erreurs ($R = 1/\sigma^2$) une densité a priori propre, par exemple une loi gamma de paramètres a, b strictement positifs. D'où le théorème suivant :

Théorème 2.2. (Broemeling L.D. et Holbert D, 1977)

Reprenons le même modèle (2.1) en gardant les mêmes hypothèses pour les paramètres m, ϕ_0, ϕ_1 et assignons à la précision des erreurs ϵ_i ($R = 1/\sigma^2$) une loi gamma de paramètres a, b strictement positifs comme densité a priori, ie $\pi_0(r) \propto b^a r^{a-1} e^{-br}$.

Alors la masse de probabilité a posteriori du point de rupture m est donnée par :

$$\pi_1(m) \propto [m(n-m)]^{-1/2} [2b + S_1^m + S_{m+1}^n]^{-(a^*-1)}$$

où $a^* = a + n/2$

2.2.1.3 Cas où les erreurs sont autocorrélées

Supposons que les moyennes ϕ_0 et ϕ_1 du modèle (2.1) sont différentes et inconnues. Les $\epsilon_i; i = 1, \dots, n$ suivent un processus autorégressif d'ordre un.
 $\epsilon_i = \rho\epsilon_{i-1} + e_i; e_i \sim N(0, \sigma_i^2); i = 1, 2, \dots, n; \rho$ et σ_i^2 sont inconnus.

La dépendance des variables $X_i; i = 1, \dots, n$ intervient dans la corrélation des erreurs $\epsilon_i; i = 1, \dots, n$

La fonction de vraisemblance

Comme les $e_i \sim N(0, \sigma^2)$ alors :

$$L(\theta/X) \propto (\sigma^2)^{-(n/2+1)} \exp\left(-\frac{1}{2\sigma^2} A(\theta)\right)$$

Avec :

$$\begin{aligned} A(\theta) &= \sum_{i=1}^m (X_i - \rho X_{i-1} - \phi_0(1 - \rho))^2 + (X_{m+1} - \rho X_m - \phi_1 + \rho\phi_0)^2 \\ &+ \sum_{i=m+2}^n (X_i - \rho X_{i-1} - \phi_1(1 - \rho))^2 \end{aligned}$$

L'analyse a posteriori

En assignant une densité conjointe a priori impropre aux paramètres m, ϕ_0, ϕ_1, ρ et σ^2 nous déterminons la masse de probabilité marginale a posteriori du point de rupture m .

Soit le théorème suivant :

Théorème 2.3. (Broemeling L.D et Holbert D, 1977)

Considérons le modèle (2.1) :

$$\begin{cases} X_i = \phi_0 + \epsilon_i; & i = 1, 2, \dots, m \\ X_i = \phi_1 + \epsilon_i; & i = m + 1, \dots, n \end{cases}$$

où :

$\epsilon_i = \rho\epsilon_{i-1} + e_i$; $e_i \sim N(0, \sigma^2)$; $i = 1, 2, \dots, n$.

ρ et σ^2 sont inconnus.

Si on suppose que la densité conjointe a priori des paramètres m, ϕ_0, ϕ_1, ρ et σ^2 est :

$$\pi_0(m, \phi_0, \phi_1, \rho, \sigma^2) \propto 1/((n-1)\sigma^2)$$

alors la masse de probabilité a posteriori du point de rupture m est donnée par :

$$\pi_1(m) = \int_R \frac{1}{\sqrt{\alpha_1 a - \beta_1^2}} \cdot \left[\alpha_3 - \frac{\beta_2^2}{a} - \left((\alpha_2 a + \beta_1 \beta_2) / \sqrt{\alpha_1 a^2 - \beta_1^2 a} \right)^2 \right]^{-(n/2-1)} d\rho$$

où :

$$a = m(1 - \rho)^2 + \rho^2,$$

$$\alpha_1 = (n - m - 1)(1 - \rho)^2 + 1,$$

$$\alpha_2 = (1 - \rho) \sum_{i=m+2}^n (X_i - \rho X_{i-1}) + (X_{m+1} - \rho X_m),$$

$$\alpha_3 = \sum_{i=2}^n (X_i - \rho X_{i-1})^2,$$

$$\beta_1 = \rho,$$

$$\beta_2 = (1 - \rho) \sum_{i=2}^m (X_i - \rho X_{i-1}) + \rho (X_{m+1} - \rho X_m).$$

2.2.2 Rupture dans la variance

Dans cette section, nous considérons le modèle précédent et nous nous intéressons au cas où une rupture dans la variance apparaît à un instant inconnu m

Il s'agit donc d'un modèle du type :

$$\begin{cases} X_i = \phi_0 + \epsilon_i; & i = 1, 2, \dots, m \\ X_i = \phi_1 + \epsilon_i; & i = m + 1, \dots, n \end{cases}$$

Le point de rupture m est supposé inconnu.

ϕ_0 et ϕ_1 représentent la moyenne des variables aléatoires X_i avant et après le changement respectivement.

Les ϵ_i sont des variables aléatoires gaussiennes et indépendantes de moyenne nulle et de variances σ_1^2 si $i = 1, 2, \dots, m$ et σ_2^2 si $i = m + 1, \dots, n$ avec $\sigma_1^2 \neq \sigma_2^2$.

$$ie \quad \epsilon_i \sim \begin{cases} N(0, \sigma_1^2); & i = 1, 2, \dots, m \\ N(0, \sigma_2^2); & i = m + 1, \dots, n \end{cases}$$

$\sigma_j^2 = \frac{1}{t_j}$ pour $j = 1, 2$ où t_j , $j = 1, 2$ désigne la précision des erreurs.

En assignant des densités de probabilité a priori aux paramètres m, ϕ_0, ϕ_1, t_1 et t_2 , Menzefricke en 1981 a déterminé les lois de probabilité a posteriori du point de rupture m et du rapport des précisions $\tau = \frac{t_2}{t_1}$ dans les trois cas suivants :

- 1) ϕ_0 et ϕ_1 sont connues.
- 2) $\phi_0 = \phi_1 = \phi$ inconnue.
- 3) ϕ_0 et ϕ_1 sont distincts et inconnues.

Supposons que la densité de probabilité conjointe a priori des paramètres m, t_1, t_2, ϕ_0 et ϕ_1 de ce modèle est donnée par :

$$\pi_0(m, t_1, t_2, \phi_0, \phi_1) = \pi_0(m) \cdot \pi_0(t_1) \cdot \pi_0(t_2) \cdot \pi_0(\phi_0/t_1) \cdot \pi_0(\phi_1/t_2)$$

où :

$$\pi_0(m) \quad \text{est toute masse de probabilité sur } \{1, 2, \dots, n\} \quad (2.2)$$

$$\pi_0(\phi_0/t_1) = N\left(v_1, \frac{1}{\gamma_1 t_1}\right) \quad (2.3)$$

$$\pi_0(\phi_1/t_2) = N\left(v_2, \frac{1}{\gamma_2 t_2}\right) \quad (2.4)$$

$$\pi_0(t_j) = \Gamma\left(\frac{\alpha_j}{2}, \frac{\beta_j}{2}\right) \quad j = 1, 2. \quad (2.5)$$

Où : v_j, γ_j, α_j et β_j $j = 1, 2$ sont des constantes connues, avec : $\alpha_j > 0$ et $\beta_j > 0$

Premier cas : ϕ_0 et ϕ_1 sont connues.

Les lois de probabilité a priori sont $\pi_0(m)$, $\pi_0(t_1)$ et $\pi_0(t_2)$ telle que définies dans (2.2) et (2.5).

Alors :

1) La masse de probabilité a posteriori du point de rupture m est donnée par :

$$\pi(m/\phi_0, \phi_1, X) \propto \pi_0(m) \cdot \beta \left[\frac{\alpha_1 + m}{2}, \frac{\alpha_2 + n - m}{m} \right] \cdot B_{1m}^{-\frac{\alpha_1 + m}{2}} \cdot B_{2m}^{-\frac{\alpha_2 + n - m}{2}}$$

2) La densité de probabilité a posteriori marginale de $\tau = t_2/t_1$ est :

$$\pi(\tau/\phi_0, \phi_1, X) \propto \sum_{m=1}^n \pi_0(m) \frac{\tau^{\frac{\alpha_2 + n - m}{2} - 1}}{[B_{1m} + \tau B_{2m}]^{\frac{\alpha_1 + \alpha_2 + n}{2}}}$$

3) La densité de probabilité a posteriori conditionnelle de τ est :

$$\pi(\tau/m, \phi_0, \phi_1, X) \propto \frac{1/[B_{1m}]^{\frac{\alpha_1 + m}{2}} \cdot [\tau B_{2m}]^{\frac{\alpha_2 + n - m}{2} - 1}}{\beta \left[\frac{\alpha_1 + m}{2}, \frac{\alpha_2 + n - m}{2} \right] \cdot [B_{1m} + \tau B_{2m}]^{\frac{\alpha_1 + \alpha_2 + n}{2}}}$$

où :

$$B_{1m} = m \cdot s_{1m}^2 + \beta_1 \quad , \quad B_{2m} = (n - m) \cdot s_{2m}^2 + \beta_2$$

Avec :

$$s_{1m}^2 = \frac{\sum_{i=1}^m (X_i - \phi_0)^2}{m} \quad , \quad s_{2m}^2 = \frac{\sum_{i=m+1}^n (X_i - \phi_1)^2}{n - m}$$

Deuxième cas : $\phi_0 = \phi_1 = \phi$ inconnue.

Les densités de probabilité a priori utilisées sont $\pi_0(m), \pi_0(\phi_0/t_1)$ et $\pi_0(t_j), j = 1, 2$ telles que définies dans (2.2),(2.3) et (2.5)

avec $v_1 = v_2 = v$ et $\gamma_1 = \gamma_2 = \gamma$.

Alors :

1) La densité de probabilité conditionnelle de $\tau = t_2/t_1$ sachant m est :

$$\pi(\tau/m, X) \propto \frac{\tau^{\frac{\alpha_2 + n - m}{2}} [\gamma + m + \tau(n - m)]^{\frac{\alpha_1 + \alpha_2 + n - 1}{2}}}{[a_m \tau^2 + b_m \tau + c_m]^{\frac{\alpha_1 + \alpha_2 + n}{2}}}$$

Où :

$$a_m = (n - m) [(n - m) s_{2m}^2 + \beta_2]$$

$$b_m = (\gamma + m) [(n - m) s_{2m}^2 + \beta^2] + (n - m) \left[m s_{1m}^2 + \beta_1 + \gamma (\bar{X}_{2m} - v)^2 + m (\bar{X}_{1m} - \bar{X}_{2m})^2 \right]$$

$$c_m = (\gamma + m) (m s_{1m}^2 + \beta - 1) + \gamma_m (\bar{X}_{1m} - v)^2$$

Avec :

$$s_{1m}^2 = \frac{\sum_{i=1}^m (X_i - \bar{X}_{1m})^2}{m}, \quad s_{2m}^2 = \frac{\sum_{i=m+1}^n (X_i - \bar{X}_{2m})^2}{n - m}$$

$$\bar{X}_{1m} = \frac{\sum_{i=1}^m (X_i)}{m}, \quad \bar{X}_{2m} = \frac{\sum_{i=m+1}^n (X_i)}{n - m}$$

2) La masse de probabilité a posteriori du point de rupture m est :

$$\pi(m/X) = \pi_0(m) \int_0^{+\infty} \pi(\tau/m, X) d\tau$$

Troisième cas : ϕ_0 et ϕ_1 sont distinctes et inconnues.

La loi de probabilité a priori conjointe des paramètres m, ϕ_0, ϕ_1, t_1 et t_2 est :

$$\pi_0(m, \phi_0, \phi_1, t_1, t_2) = \pi_0(m) \cdot \pi_0(t_1) \cdot \pi_0(t_2) \cdot \pi_0(\phi_0/t_1) \cdot \pi_0(\phi_1/t_2)$$

Alors :

1) La masse de probabilité a posteriori du point de rupture m est :

$$\pi(m/X) \propto \frac{\pi_0(m) \cdot \beta \left[\frac{\alpha_1 + m}{2}, \frac{\alpha_2 + n - m}{2} \right]}{[\gamma_1 + m]^{\frac{1}{2}} [\gamma_2 + n - m]^{\frac{1}{2}} \left[\acute{B}_{1m} \right]^{\frac{\alpha_1 + m}{2}} \left[\acute{B}_{2m} \right]^{\frac{\alpha_2 + n - m}{2}}}$$

2) La densité de probabilité a posteriori conditionnelle de τ sachant m est :

$$\pi(\tau/m, X) = \frac{1/\tau \left[\acute{B}_{1m} \right]^{\frac{\alpha_1 + m}{2}} \left[\acute{B}_{2m}\tau \right]^{\frac{\alpha_2 + n - m}{2}}}{\beta \left[\frac{\alpha_1 + m}{2}, \frac{\alpha_2 + n - m}{2} \right] \left[\acute{B}_{1m} + \tau \acute{B}_{2m} \right]^{\frac{\alpha_1 + \alpha_2 + n}{2}}}$$

Où :

$$\acute{B}_{1m} = \frac{m\gamma_1}{\gamma_1 + m} (v_1 - \bar{X}_{1m})^2 + ms_{1m}^2 + \beta_1$$

$$\acute{B}_{2m} = \frac{(n-m)\gamma_2}{\gamma_2 + n - m} (v_2 - \bar{X}_{2m})^2 + (n-m)s_{2m}^2 + \beta_2$$

$s_{1m}^2, s_{2m}^2, \bar{X}_{1m}$ et \bar{X}_{2m} ont été définis précédemment.

Nous allons successivement étudier les trois cas précédents en utilisant des lois a priori impropres .

Premier cas :

Si les lois a priori impropres suivantes sont utilisées :

$$\pi_0(m, t_1, t_2) \propto \pi_0(m) \cdot \frac{1}{t_1 t_2}, \quad (m, t_1, t_2) \in \{1, 2, \dots, n\} \times (\mathbb{R}_*^+)^2$$

Alors :

$$\pi(m/\phi_0, \phi_1, X) \propto \pi_0(m) \beta \left[\frac{m}{2}, \frac{n-m}{2} \right] [B_{1m}]^{-\frac{m}{2}} [B_{2m}]^{-\frac{n-m}{2}}$$

$$\pi(\tau/\phi_0, \phi_1, X) \propto \sum_{m=1}^n \pi_0(m) \frac{\tau^{\frac{n-m}{2} - 1}}{[B_{1m} + \tau B_{2m}]^{\frac{n}{2}}}$$

$$\pi(\tau/m, \phi_0, \phi_1, X) \propto \frac{1/\tau [B_{1m}]^{\frac{m}{2}} [\tau B_{2m}]^{\frac{n-m}{2}}}{\beta \left[\frac{m}{2}, \frac{n-m}{2} \right] [B_{1m} + \tau B_{2m}]^{\frac{n}{2}}}$$

Où :

$$B_{1m} = m s_{1m}^2$$

$$\text{et } B_{2m} = (n-m) s_{2m}^2.$$

Deuxième cas :

En utilisant les lois a priori impropres suivantes :

$$\pi_0(m, \phi, t_1, t_2) \propto \pi_0(m) \frac{1}{t_1 t_2}, \quad (m, \phi, t_1, t_2) \in \{1, 2, \dots, n\} \times \mathbb{R} \times (\mathbb{R}_*^+)^2$$

nous obtenons :

$$\pi(\tau/m, X) \propto \frac{\tau^{\frac{n-m}{2} - 1} [m + \tau(n-m)]^{\frac{n-2}{2}}}{\left\{ (n-m)^2 s_{2m}^2 \tau^2 + m(n-m) \left[s_{1m}^2 + s_{2m}^2 + (\bar{X}_{1m} - \bar{X}_{2m})^2 \right] \tau + m^2 s_{1m}^2 \right\}^{\frac{n-1}{2}}}$$

Et

$$\pi(m/X) = \pi_0(m) \int_0^{+\infty} \pi(\tau/m, X) d\tau$$

Troisième cas :

Si les lois a priori impropres suivantes sont utilisées :

$$\pi_0(m, \phi_0, \phi_1, t_1, t_2) \propto \pi_0(m) \frac{1}{t_1 t_2}; \quad (m, \phi_0, \phi_1) \in \{1, 2, \dots, n\} \times \mathbb{R}^2, \quad (t_1, t_2) \in (\mathbb{R}_*^+)^2$$

Alors :

$$\pi(\tau/m, X) \propto \frac{1/\tau [B_{1m}]^{\frac{m-1}{2}} [\tau B_{2m}]^{\frac{n-m-1}{2}}}{\beta \left[\frac{m-1}{2}, \frac{n-m-1}{2} \right] [B_{1m} + \tau B_{2m}]^{\frac{n-2}{2}}}$$

Et

$$\pi(m/X) \propto \frac{\pi_0(m) \beta \left[\frac{m-1}{2}, \frac{n-m-1}{2} \right]}{\frac{1}{[m]^2} \frac{1}{[n-m]^2} \frac{1}{[B_{1m}]^{\frac{m-1}{2}}} \frac{1}{[B_{2m}]^{\frac{n-m-1}{2}}}}$$

Où :

$$B_{1m} = m s_{1m}^2 \quad \text{et} \quad B_{2m} = (n-m) s_{2m}^2$$

2.3 Modèle de régression lineaire multiple

A la fin des années soixante et début des années soixante dix, beaucoup de travaux ont été initiés pour des modèles de régression où la méthodologie Bayésienne a été largement sollicitée ; on peut citer entre autres les travaux de Quandt (1960), ceux de Ferreira (1975), Holbert et Broemeling (1977), Holbert (1982) et ceux de Lyle Broemeling (1985). Dans cette section, nous nous intéressons à un modèle de régression linéaire multiple avec un changement dans les paramètres de régression.

2.3.1 Présentation du modèle

Le modèle considéré est le suivant :

$$\begin{cases} Y_i = X_i\beta_1 + \epsilon_i, & i = 1, 2, \dots, m \\ Y_i = X_i\beta_2 + \epsilon_i, & i = m + 1, \dots, n \end{cases} \quad (2.6)$$

où :

X_i ; $i = 1, \dots, n$ est un vecteur à $(1 \times p)$ observations constitué de p variables indépendantes.

Y_i est la i ème réalisation de la variable dépendante.

β_1 et β_2 deux vecteurs $(p \times 1)$, ils représentent les paramètres de la régression, avec $(\beta_1 \neq \beta_2)$.

ϵ_i ; $i = 1, \dots, n$ sont les erreurs du modèle, elles sont supposées d'une part indépendantes, identiquement distribuées suivant une loi normale $N(0, \sigma^2)$, et d'autre part autocorrélées, plus précisément, elles sont régies par un processus autorégressif d'ordre p , $p \geq 1$, $(AR(p))$.

m ; point de rupture, c'est un paramètre aléatoire prenant les valeurs $1, 2, \dots, n - 1$.

Le changement concerne uniquement les paramètres de régression.

2.3.2 Cas où les erreurs sont indépendantes

Intéressons-nous au cas où les erreurs ϵ_i sont indépendantes et gaussiennes, de moyenne nulle et de variance σ^2 .

$$\epsilon_i \sim iid \quad N(0, \sigma^2)$$

Ce modèle a été considéré par J.H. Chin choy et L.D.Broemeling en 1980. Le théorème suivant nous donne les densités a posteriori des différents paramètres de ce modèle.

Théorème 2.4. (Chin choy et Broemeling,1980)

Etant donné le modèle (2.6).

m, β, σ^2 sont des paramètres inconnus, avec $\beta = (\beta'_1, \beta'_2)$.
 m est uniformément distribué sur l'ensemble $\{1, 2, \dots, n - 1\}$
 La loi conjointe a priori de β et de $R = 1/\sigma^2$ est :

$$\pi_0(\beta/R = r) \equiv 2p - N(\beta_\mu, r\tau)$$

tel que $\beta_\mu \in \mathbb{R}^{2p}$, et τ est une matrice $(2p \times 2p)$ donnée, symétrique, définie positive.
 Une telle loi est appelée une loi normale-Gamma.
 R suit une distribution Gamma de paramètres a, b ($a > 0, b > 0$).
 m est indépendant de β et de R .

Alors :

i) La masse de probabilité a posteriori du point de rupture m est :

$$\pi(m/y) \propto \begin{cases} D(m)^{-a^*} |x'(m)x(m) + \tau|^{-1/2}; & m = 1, 2, \dots, n - 1 \\ 0 & \text{sinon} \end{cases}$$

où $a^* = a + n/2$.

$$\begin{aligned} D(m)^{-a^*} &= b + 1/2 \{y'y + \beta'_\mu \tau \beta_\mu - \beta^{*'}(m) [x'(m)x(m) + \tau] \beta^*(m)\} \\ &= b + 1/2 \{[y - x(m)\beta^*(m)]' \cdot y + [\beta_\mu - \beta^*(m)]' \tau \beta_\mu\} \end{aligned}$$

ii) La densité de probabilité a posteriori (d.d.p) du paramètre de régression $\beta = (\beta'_1, \beta'_2)$ est :

$$\pi(\beta/y) = \sum_{m=1}^{n-1} \pi(m/y) t[\beta, 2p, 2a^*, \beta^*(m), p(m)]$$

où $p(m) = a^*/D(m) (x'(m)x(m) + \tau)$
 $t[\beta, 2p, 2a^*, \beta^*(m), p(m)]$ est une densité de Student $2p$ - dimensionnelle du vecteur aléatoire β à $2a^*$ degré de liberté, de vecteur de position $\beta^*(m)$, et de matrice de précision $p(m)$.

Si on pose :

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, \beta^*(m) = \begin{pmatrix} \beta_1^*(m) \\ \beta_2^*(m) \end{pmatrix}, p(m) = \begin{pmatrix} p_{11}(m) & p_{12}(m) \\ p_{21}(m) & p_{22}(m) \end{pmatrix}$$

où β_i et $\beta_i^*(m)$, $i = 1, 2$ sont $p \times 1$ et $p_{ij}(m)$, $i, j = 1, 2$ sont des matrices $p \times p$
Alors :

iii) La densité de probabilité a posteriori de β_1 est :

$$\pi(\beta_1/y) = \sum_{m=1}^{n-1} \pi(m/y) t(\beta, p, 2a^*, \alpha_1^*(m), p_1^*(m))$$

où $p_1^*(m) = p_{11}(m) - p_{12}(m)p_{22}^{-1}(m)p_{21}(m)$

iv) La d.d.p a posteriori de β_2 est :

$$\pi(\beta_2/y) = \sum_{m=1}^{n-1} \pi(m/y) t(\beta, p, 2a^*, \alpha_2^*(m), p_2^*(m))$$

où $p_2^*(m) = p_{22}(m) - p_{21}(m)p_{11}^{-1}(m)p_{12}(m)$

v) La d.d.p de R est :

$$\pi(r/y) = \sum_{m=1}^{n-1} \pi(m/Y) g(r, a^*, D(m))$$

où :

$g(r, a^*, D(m))$ est la distribution Gamma de paramètres a^* et $D(m)$

2.3.3 Cas où les erreurs sont autocorrélées

Après avoir considéré le cas d'un modèle de régression linéaire où les erreurs ϵ_i sont gaussiennes et indépendantes, nous considérons le cas où elles sont autocorrélées.

Le cas AR(1).

Dans cette section, Salazar et al (1981) ont considéré le même modèle de régression avec des erreurs régies par un processus autorégressif d'ordre un (AR(1)).

Le théorème suivant nous détermine la masse de probabilité a posteriori du point de rupture m , du paramètre de régression β et celle de R .

Théorème 2.5. (Salazar, Broemeling et Chi, 1981) *Considérons le modèle suivant :*

$$\begin{cases} Y_i = X_i\beta_1 + \mu_i; & i = 1, 2, \dots, m \\ Y_i = X_i\beta_2 + \mu_i; & i = m + 1, \dots, n \end{cases}$$

$$\mu_i = \rho\mu_{i-1} + e_i; \quad i = 1, \dots, n$$

où $e_i \sim iid \ N(0, \sigma^2)$; $0 < \rho = \sigma^{-2}$

$m, \beta_1, \beta_2, \rho$, et $r = \sigma^{-2}$ inconnus.

Y_i est la i ème observation de la variable dépendante.

X_i est la i ème observation de K variables indépendantes.

$\rho \in \mathbb{R}$, il est muni d'une distribution non informative sur \mathbb{R} ($\rho \propto cste$ sur \mathbb{R}).

m : point de rupture, uniformément distribuée sur $I_{n-1} = \{1, 2, \dots, n-2\}$

$\beta = (\beta'_1, \beta'_2) \in \mathbb{R}^{2K}$: coefficient de régression. ($\beta_1 \neq \beta_2$).

La distribution conjointe a priori pour $\beta_i (i = 1, 2)$ et R est :

$\pi_0(\beta/R = r) \equiv N(\mu, rP)$ c'est la loi normale $2K$ -dimensionnelle.

$\mu = (\mu'_1, \mu'_2)' \in \mathbb{R}^{2K}$, vecteur moyen.

P une matrice $2K \times 2K$ symétrique, définie positive.

La densité a priori de R est une Gamma de paramètres a et b strictement positifs.

$\pi_0(r) \propto r^{a-1} \exp(-br)$

On suppose de plus $m, (\beta, R), \rho$ indépendants.

Alors :

i) la masse de probabilité a posteriori du point de rupture m est donnée par :

$$\pi(m/Y) = \begin{cases} \int_{\mathbb{R}} |H(\rho)|^{-1/2} C(\rho)^{-(a+n/2)}; & m = 1, 2, \dots, n-2 \\ 0 & \text{ailleurs} \end{cases}$$

Où :

$$C(\rho) = 2b + \mu' P \mu + \sum_{t=1}^n (Y_t - \rho Y_t - 1)^2 \tilde{\beta}' H(\rho) \tilde{\beta}$$

$$Z_1 = ((X_1 - \rho X_0)', (X_2 - \rho X_1)', \dots, (X_m - \rho X_{m-1})')$$

$$Z_2 = ((X_{m+2} - \rho X_{m+1})', \dots, (X_n - \rho X_{n-1})')$$

$$V_1 = ((Y_1 - \rho Y_0)', (Y_2 - \rho Y_1)', \dots, (Y_m - \rho Y_{m-1})')$$

$$V_2 = ((Y_{m+2} - \rho Y_{m+1})', \dots, (Y_n - \rho Y_{n-1})')$$

$$\begin{aligned} H(\rho) &= X'X + P = \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix} \\ &= \begin{pmatrix} P_{11} + Z_1'Z_1 + \rho^2 X_m'X_m & P_{12} - \rho X_m'X_m \\ P_{21} - \rho X_{m+1}'X_m & P_{22} + Z_2'Z_2 + X_{m+1}'X_m \end{pmatrix} \end{aligned}$$

$$P = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix}, \quad \tilde{\beta} = (\tilde{\beta}_1', \tilde{\beta}_2')$$

$$\tilde{\beta}_1 = H_{11.2}^{-1}(\rho)(\alpha_1 - H_{12}(\rho)H_{22}^{-1}(\rho).\alpha_2)$$

$$\tilde{\beta}_2 = H_{22.1}^{-1}(\rho)(\alpha_2 - H_{21}(\rho)H_{11}^{-1}(\rho).\alpha_1)$$

$$\begin{aligned} P\mu + X'y &= \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} Z_1' & -\rho X_m' & 0 \\ 0 & X_{m+1}' & Z_2' \end{pmatrix} \\ &= \begin{pmatrix} P_{11}.\mu_1 + P_{12}.\mu_2 + Z_1'V_1 - \rho X_m'(Y_{m+1} - \rho Y_m) \\ P_{21}.\mu_1 + P_{22}.\mu_2 + Z_2'V_2 + X_{m+1}'(Y_{m+1} - \rho Y_m) \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \end{aligned}$$

ii) La densité a posteriori marginale de $\beta = (\beta_1', \beta_2')$ est :

$$\pi(\beta/Y) \propto \sum_{m=1}^{n-1} \int_{\mathbb{R}} |H(\rho)|^{-1/2} C(\rho)^{-(a+n/2)} f(\beta) d\rho, \quad \beta \in \mathbb{R}^{2K}$$

Où $f(\beta)$ est une densité de Student $2K$ dimensionnelle à $(2a + n)$ degrés de liberté, de vecteur de position $\tilde{\beta}$ et de matrice de précision $(2a + n)H(\rho)/C(\rho)$.

iii) La densité a posteriori marginale de R est :

$$\pi(r/Y) \propto \sum_{m=1}^{n-1} \int_{\mathbb{R}} r^{a+n/2-1} \exp(-rC(\rho)/2) |H(\rho)|^{-1/2} d\rho, \quad r > 0$$

Le cas AR $(p), (p \geq 2)$.

Salazar et al (1981) ont étendu les travaux précédents au cas où les erreurs sont régies par un processus autorégressif d'ordre p , $(p \geq 2)$, plus précisément à des modèles du type :

$$\begin{cases} Y_t = X_t\beta_1 + \mu_t; & t = 1, 2, \dots, m \\ Y_t = X_t\beta_2 + \mu_t; & t = m + 1, \dots, n \end{cases}$$

$m \in \{p + 1, \dots, n - p - 1\}$; $p < m < n - p$

où $\mu_t \sim AR(p)$

$$\mu_t = \sum_{k=1}^p \rho_k \mu_{t-k} + e_t; \quad e_t \sim iid \quad N(0, \sigma^2) \quad (t = 1, \dots, n)$$

$X_0, \mu_0, \dots, Y_0, \dots, X_{1-p}, \mu_{1-p}, \dots, Y_{1-p}$ sont des valeurs initiales fixées.

Ce modèle peut s'écrire de la manière suivante :

$$Y_t - \sum_{k=1}^p \rho_k Y_{t-k} = (X_t - \sum_{k=1}^p \rho_k X_{t-k})\beta_1 + e_t, \quad t = 1, \dots, m$$

$$Y_t - \sum_{k=1}^p \rho_k Y_{t-k} = [X_t - \sum_{k=1}^{t-m-1} \rho_k X_{t-k}]\beta_2 - [\sum_{k=t-m}^p \rho_k X_{t-k}]\beta_1 + e_t, \\ t = m + 1, \dots, m + p$$

$$Y_t - \sum_{k=1}^p \rho_k Y_{t-k} = (X_t - \sum_{k=1}^p \rho_k X_{t-k})\beta_2 + e_t, \quad t = m + p + 1, \dots, n$$

On pose :

$$\begin{cases} y_t = Y_t - \sum_{k=1}^p \rho_k Y_{t-k} \\ x_t = X_t - \sum_{k=1}^p \rho_k X_{t-k} \end{cases} \quad \text{pour } t = 1, \dots, m \text{ et } t = m + p + 1, \dots, n$$

Ce qui donne :

$$y_t = x_t\beta_1 + e_t; \quad t = 1, 2, \dots, m$$

$$y_t = x_t\beta_2 + e_t; \quad t = m + p + 1, \dots, n$$

En posant :

$$V_1 = (y_1, \dots, y_m)'; \quad Z_1 = (x'_1, \dots, x'_m)'; \quad E_1 = (e_1, \dots, e_m)'$$

$$V_2 = (y_{m+p+1}, \dots, y_n)'; \quad Z_2 = (x'_{m+p+1}, \dots, x'_n)'; \quad E_2 = (e_{m+p+1}, \dots, e_n)'$$

$$M_1 = \begin{pmatrix} -\sum_{k=1}^p \rho_k X_{m+1-k} \\ \vdots \\ -\sum_{k=t-m}^p \rho_k X_{t-k} \\ \vdots \\ \rho_p X_m \end{pmatrix}; \quad M_2 = \begin{pmatrix} X_{m+1} \\ X_{m+2} - \rho_1 X_{m+1} \\ \vdots \\ X_t - \sum_{k=1}^{t-m-1} \rho_k X_{t-k} \\ \vdots \\ X_{m+1} - \sum_{k=1}^{p-1} \rho_k X_{m+p-k} \end{pmatrix}$$

on a :

$$Y = \begin{pmatrix} V_1 \\ y_{m+1} \\ \cdot \\ \cdot \\ y_{m+p} \\ V_2 \end{pmatrix} = \begin{pmatrix} Z_1 & 0 \\ M_1 & M_2 \\ 0 & Z_2 \end{pmatrix} \cdot \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} E_1 \\ e_{m+1} \\ \cdot \\ \cdot \\ \cdot e_{m+p} \\ E_2 \end{pmatrix}$$

Donc le modèle peut s'écrire sous une forme matricielle :

$$Y = X\beta + E$$

$\rho = (\rho_1, \dots, \rho_p)'$ est muni d'une probabilité a priori non informative sur \mathbb{R}^p . donc $\pi_0(\rho) \propto cste$ sur \mathbb{R}^p

En vertu du théorème de Bayes on a :

$$\pi(\beta, r, \rho, m/Y) \propto r^{a^*+k-1} \cdot \exp[-r(b + 1/2(\beta - \mu)'P(\beta - \mu) + 1/2(Y - X\beta)'(Y - X\beta))]$$

Une intégration par rapport à r nous donne :

$$\pi(\beta, \rho, m) \propto A_\beta^{-(a^*+k)}$$

Après intégration par rapport à β on trouve :

$$\pi(m, \rho/Y) \propto C(\rho)^{-a^*} |H(\rho)|^{-1/2}$$

Avec $H(\rho) = X'X + P$ où $P = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix}$
 $C(\rho) = 2b + \mu'P\mu + Y'Y - \beta'^*H(\rho)\beta^*$
 où $\beta^* = H(\rho)^{-1}(P \cdot \mu + X'Y)$

Ce qui nous donne finalement :

$$\pi(m/Y) \propto \int_{\mathbb{R}^p} |H(\rho)|^{-1/2} C(\rho)^{-a^*} d\rho$$

où $\rho = (\rho_1, \dots, \rho_p)'$

2.4 Processus autoregressif d'ordre (1)

La littérature sur la détection de rupture dans les séries chronologiques est vaste, nous pouvons citer les travaux de Chu (1995), ceux de Kokoszka et Leipus (1999,2000) et Tsay (1988) sur les modèles ARCH ainsi ceux de Berkes, Gombay, Horvath et Kokoszka (2004) sur les modèles GARCH.

Dans cette section, nous nous intéressons aux travaux de Mayuri Pandya, Krishnam Bhatt et Hardik Pandya chetan Thakar (octobre 2012). En adoptant l'approche Bayésienne, ils se sont intéressés à la détection de rupture dans un processus autorégressif d'ordre 1 (AR(1)).

2.4.1 Présentation du modèle

Considérons un processus autorégressif d'ordre 1 (AR(1)), avec rupture dans le coefficient d'autocorrélation et la variance des erreurs à un instant inconnu m . Le modèle est donc donné par :

$$\begin{cases} X_i = \beta_1 X_{i-1} + \epsilon_i; & i = 1, 2, \dots, m \\ X_i = \beta_2 X_{i-1} + \epsilon_i; & i = m + 1, \dots, n \end{cases}$$

β_1 et β_2 représentent les coefficients d'autocorrélation, supposés inconnus.

X_i est la i ème observation de la variable dépendante.

Le paramètre m est le point de rupture. C'est un entier qui varie entre 1 et $n - 1$, n étant la taille de l'échantillon.

Les ϵ_i sont des erreurs aléatoires gaussiennes, indépendantes, de moyenne nulle et de variances connues σ_1^2 et σ_2^2 avant et après le changement respectivement.

2.4.2 La fonction de vraisemblance

La fonction de vraisemblance associée aux paramètres β_1, β_2 et m est donnée par :

$$L(\beta_1, \beta_2, m | X) = k_1 \exp \left[-\frac{1}{2} \beta_1^2 (S_{m1}/\sigma_1^2) + \beta_1 (S_{m2}/\sigma_1^2) - S_{m3}/2\sigma_1^2 \right] \\ \sigma_1^{-m} \sigma_2^{-(n-m)} \exp \left[-\frac{1}{2} \beta_2^2 (S_{n1} - S_{m1}/\sigma_2^2) + \beta_2 (S_{n2} - S_{m2}/\sigma_2^2) - (S_{n3} - S_{m3}/2\sigma_2^2) \right]$$

Avec :

$$S_{k1} = \sum_{i=1}^k X_{i-1}^2 \\ S_{k2} = \sum_{i=1}^k X_i X_{i-1}^2 \\ S_{k3} = \sum_{i=1}^k X_i^2 \\ k_1 = (2\pi)^{-\frac{n}{2}}$$

2.4.3 L'analyse a posteriori

Premier cas

Dans ce premier cas, les auteurs ont considéré des lois a priori informatives pour les paramètres du modèle. Une loi normale de moyenne μ_i et de variance $a_i, i = 1, 2$ pour les coefficients d'autocorrélation β_1 et β_2 respectivement. Une distribution uniforme sur $[1, n - 1]$ pour le point m . Les paramètres m, β_1 et β_2 sont supposés indépendants.

En appliquant le théorème de Bayes la densité a posteriori de m est donnée par :

$$\pi_1(m) = T_1(m) / \sum_{m=1}^{n-1} T_1(m)$$

Avec :

$$T_1(m) = I_1(m) I_2(m) \exp \left[-\frac{1}{2} \{ S_{m3} (\sigma_1^{-2} \sigma_2^{-2}) \} \sigma_1^{-m} \sigma_2^{-(n-m)} \right]$$

où :

$$I_i(m) = \left[\frac{\sqrt{2\pi} \exp \frac{B_{im}^2}{2A_{im}}}{\sqrt{A_{im}}} \right], i = 1, 2.$$

$$A_{1m} = \frac{S_{m1}}{\sigma_1^2} + \frac{1}{a_1} > 0, B_{1m} = \frac{S_{m2}}{\sigma_1^2} + \frac{\mu_1}{a_1}$$

$$A_{2m} = \frac{S_{n1} - S_{m1}}{\sigma_2^2} + \frac{1}{a_2} > 0, B_{2m} = \frac{S_{n2} - S_{m2}}{\sigma_2^2} + \frac{\mu_2}{a_2}$$

S_{k1}, S_{k2} et S_{k3} sont définis précédemment.

Deuxième cas

Intéressons nous à présent, à des densités a priori non informatives concernant les coefficients d'autocorrélation. Pandya et al (2012), ont considéré pour les coefficients β_1 et β_2 une loi normale $N(0, 1)$.

Dans ce cas, la densité a posteriori de m est :

$$\pi_1(m) = T_2(m) / \sum_{m=1}^{n-1} T_2(m)$$

Avec :

$$T_1(m) = I_3(m)I_4(m)\exp\left[-\frac{1}{2}\{S_{m3}(\sigma_1^{-2}\sigma_2^{-2})\}\sigma_1^{-m}\sigma_2^{-(n-m)}\right]$$

où :

$$I_i(m) = \left[\frac{\sqrt{2\pi} \exp \frac{B_{im}^2}{2A_{im}}}{\sqrt{A_{im}}} \right], i = 3, 4.$$

$$A_{3m} = \frac{S_{m1}}{\sigma_1^2} + 1 > 0, B_{3m} = \frac{S_{m2}}{\sigma_1^2}$$

$$A_{4m} = \frac{S_{n1} - S_{m1}}{\sigma_2^2} + 1 > 0, B_{4m} = \frac{S_{n2} - S_{m2}}{\sigma_2^2}$$

S_{k1}, S_{k2} et S_{k3} sont définis précédemment.

2.5 Tests d'hypothèses

Le plus souvent, les procédures considérées supposent que le point de rupture correspond à l'un des instants d'observations et tentent de répondre à la question suivante : ya-t-il une rupture dans le modèle considéré ou non?. Ces procédures sont en général basées sur les tests d'hypothèses. Dans le cas où l'hypothèse nulle est rejetée, certains auteurs s'intéressent à l'estimation de ce point de rupture.

Dans la littérature, de nombreuses statistiques de tests de detection de ruptures ont été étudiées, telles que les statistiques basées sur la moyenne (Sen et Srivastava, 1975), celles basées sur le maximum de vraisemblance (Sen et Srivastava, 1975), (Cobb,1978) ou encore (Csörgo et Horváth 1988) qui ont proposé des méthodes basées sur les rangs de la fonction de répartition empirique. Pour déterminer la loi de la statistique de tests, Deshayes et Picard (1986) ont proposé des méthodes asymptotiques.

Dans cette section, nous allons nous intéresser à tester l'apparition d'un changement dans les moyennes d'une suite de variables aléatoires indépendantes et Gaussiennes de variance constante connue ou inconnue. Dans le but de savoir s'il y a rupture ou non, nous allons introduire quelques statistiques de tests qui nous permettent de détecter ce changement.

Considérons donc le modèle de rupture suivant :

Soient x_1, x_2, \dots, x_n une suite de variables aléatoires. Le but est de tester l'hypothèse H_0 où les x_i avec $i \in \{1, 2, \dots, n\}$ sont toutes identiquement distribuées contre l'hypothèse alternative H_1 où un seul point de rupture apparaît. C'est à dire les x_1, x_2, \dots, x_m sont toutes identiquement distribuées et les x_{m+1}, \dots, x_n sont aussi identiquement distribuées mais de distribution différente de celle des m premières observations.

Le paramètre m représente le point de rupture, c'est un entier qui varie entre 1 et $n - 1$, n étant la taille de l'échantillon.

On considère le cas spécial où les x_i sont gaussiennes de moyenne μ_i et de variance 1.
 $x_i \sim N(\mu_i, 1)$

Donc les hypothèses du test peuvent s'écrire comme suit :

$$\begin{aligned}
 H_0 : & \quad \mu_1 = \mu_2 = \dots = \mu_n \\
 H_1 : & \quad \mu_1 = \mu_2 = \dots = \mu_m \neq \mu_{m+1} = \mu_{m+2} = \dots = \mu_n
 \end{aligned}$$

avec $1 \leq m < n$

Plusieurs statistiques ont été proposées pour ce test, et parmi ces statistiques on retrouve celle du test du rapport de vraisemblance, celles suggérées par Chernoff et Zack en 1964, Brown et Al en 1975 et Pettitt en 1980.

2.5.1 Statistique du rapport de vraisemblance

Posons : $S_i = x_1 + x_2 + \dots + x_i$. pour tout $i = 1, 2, \dots, n$.

La racine carrée du log de la statistique du rapport de vraisemblance pour ce test est donnée par :

$$\max_{1 \leq K < n} [|KS_n/n - S_K| / \{K(1 - K/n)\}^{1/2}]$$

L'expression $[|KS_n/n - S_K| / \{K(1 - K/n)\}^{1/2}]$ représente la différence entre les moyennes des K premières observations et les $n - K$ dernières, c'est à dire un test d'égalité de deux moyennes de deux échantillons gaussiens.

Le maximum est recherché pour retrouver le point qui sépare l'échantillon en deux sous-échantillons de moyennes différentes.

Ce test est invariant si on change l'emplacement des observations, on peut aussi réduire le nombre d'observations et dans ce cas le test ne dépend pas directement des x_i mais des différences $y_i = x_i - x_1$, $i = 2, 3, \dots, n$

Pour des valeurs données de m et de $\delta = \mu_i - \mu_1$, le rapport de vraisemblance de y sous H_1 et H_0 est :

$$\exp\{\delta(mS_n/n - S_m) - 1/2m(1 - m/n)\delta^2\}$$

En s'intéressant au test unilatéral, on suppose que le signe de δ est connu, voir $\delta > 0$, dans ce cas la statistique du rapport de vraisemblance peut s'écrire sans valeur absolue et donc

la généralisation suivante peut être considérée comme une statistique de ce test :

$$\max_{n_0 \leq K < n_1} [KS_n/n - S_K] / \{K(1 - K/n)\}^{1/2}$$

avec $1 \leq n_0 < n_1 < n$

2.5.2 Statistique de Pettitt, 1980

Reprenons la statistique précédente :

$$\exp\{\delta(mS_n/n - S_m) - 1/2m(1 - m/n)\delta^2\}$$

En dérivant le log de cette dernière expression, avec $\delta = 0$ on obtient la statistique suggérée par Pettitt en 1980 qui est :

$$\max_{1 \leq K \leq n} (KS_n/n - S_K)$$

2.5.3 Statistique de Brown et Al, 1975

Une autre statistique intéressante pour ce test est celle proposée par Brown et Al en 1975 qui est défini comme suit :

$$\max_{1 < K \leq n} (\tilde{S}_{n-1} - \tilde{S}_{n-K}) / (k-1)^{1/2}$$

avec :

$$\tilde{S}_i = z_1 + z_2 + \dots + z_i$$

où : $z_i = \{i/(i+1)\}^{1/2}(x_{i+1} - \bar{x}_i)$, on remarque que sous H_0 , les z_n sont des variables aléatoires indépendantes de loi normale centrée et réduite.

2.5.4 Statistique de Chernoff et Zacks, 1964

En supposant que m possède une loi a priori qui est la loi uniforme sur $\{1, 2, \dots, n\}$ et que δ tend vers zero, Chernoff et Zacks en 1964 ont proposé une statistique pour ce test

dite "statistique quasi-bayésienne" qui est :

$$\sum_{i=1}^{n-1} \{i(i+1)\}^{1/2} . z_i$$

2.6 Détection de point de rupture en utilisant les tests d'hypothèses séquentiels

On considère une suite d'observations indépendantes X_1, X_2, \dots d'un certain processus. Initialement les X ont tous la même distribution f , dans ce cas le processus est dit "en contrôle". A un certain temps inconnu v , les X prennent une autre distribution g différente de f et dans ce cas le processus devient "hors contrôle". Le problème est de détecter la présence d'un tel changement et pour cela on utilise les tests d'hypothèses séquentiels.

2.6.1 Tests d'hypothèses séquentiels :

Dans les tests statistiques usuels, la taille de l'échantillon est fixée, elle est indépendante des autres données du test par contre dans les tests séquentiels c'est une variable aléatoire qui dépend de ses données.

Les critères de choix dans les tests séquentiels sont :

- La fonction caractéristique d'opération (operating characteristic OC).
- Les fonctions du nombre de l'échantillon moyen (average sample number ASN).

Supposons que X_1, X_2, \dots sont des variables aléatoires séquentiellement observées, indépendantes et identiquement distribuées et P leur fonction de densité.

Les sous-ensembles $\{f_\theta : \theta \in \Theta\}$ et $\{g_\lambda : \lambda \in \Lambda\}$ sont deux sous-ensembles disjoints de densités de probabilité.

On veut tester l'hypothèse nulle $H_0 : P = f_\theta$ pour un certain $\theta \in \Theta$ contre l'hypothèse alternative $H_1 : P = g_\lambda$ pour un certain $\lambda \in \Lambda$.

On note par P_θ et P_λ les densités de probabilité lorsque X_1, X_2, \dots sont indépendantes et identiquement distribuées de densité f_θ et g_λ respectivement, et par E_θ et E_λ les espérances correspondantes.

La fonction caractéristique d'opération (OC) du test séquentiel est déterminée par :

i) La probabilité d'erreur de type I, $P_\theta(\text{rejeter } H_0)$;

et

ii) La probabilité d'erreur de type II, $P_\lambda(\text{accepter } H_0)$.

Dans les tests séquentiels, un bon test est celui qui prend les plus petites valeurs possibles des probabilités d'erreurs.

Donc on doit prendre :

$P_\theta(\text{rejeter } H_0) \leq \alpha$ pour tout $\theta \in \Theta$,

et $P_\lambda(\text{accepter } H_0) \leq \beta$ pour tout $\lambda \in \Lambda$

où α et β sont les limites des probabilités d'erreurs de type I et II respectivement.

Le nombre d'observations N exigé par le test séquentiel est aléatoire, donc sa distribution est bien considérée, elle est généralement caractérisée par une valeur espérée lorsque f_θ ou g_λ sont les densités des observations.

Donc on peut définir les fonctions du nombre de l'échantillon moyen (ASN) par :

i) $E_\theta N$;

et

ii) $E_\lambda N$.

Une autre propriété importante pour ce test est d'avoir les plus petites valeurs de $E_\theta N$; et $E_\lambda N$, ie plus petites que α et β respectivement.

Notons que N peut être considéré comme le temps où le test ne prend plus d'observation.

Puisque la décision de ne plus prendre d'observations en un temps N est basée seulement sur les N premières alors N est appelé temps d'arrêt, "stopping time".

Définition 2.1. (Temps d'arrêt)

Une variable aléatoire N correspondant à la suite X_1, X_2, \dots , qui prend des valeurs dans \mathbb{N} (qui peut valoir ∞) est dite temps d'arrêt si pour tout $1 \leq n < \infty$, l'évènement $\{N = n\}$ dépend seulement des n premières observations.

Considérons le cas le plus simple où on suppose que X_1, X_2, \dots sont indépendantes et identiquement distribuées de fonction de densité P , et nous voudrions tester l'hypothèse simple $H_0 : P = f_\theta$ contre l'hypothèse alternative simple $H_1 : P = g_\lambda$ où θ et λ sont donnés.

La meilleure procédure est le test du rapport de probabilité séquentiel (sequential probability ratio test SPRT), qui est développé par Abraham Wald durant la deuxième guerre mondiale.

On prend deux constantes A et B tel que $0 < A < B < \infty$, et on définit la fonction du rapport de vraisemblance L_n par :

$$L_n = \prod_{i=1}^n \frac{g_\lambda(X_i)}{f_\theta(X_i)}$$

On définit le temps d'arrêt N par la première valeur de $n \geq 1$ tel que $L_n \notin (A, B)$, en d'autres termes :

$$N = \inf\{n \geq 1 : L_n \notin (A, B)\}$$

Donc le SPRT arrête son échantillon en temps n et :

$$\begin{aligned} & \text{rejeter } H_0 \quad \text{si } L_N \geq B, \\ & \text{accepter } H_0 \quad \text{si } L_N \leq A \end{aligned}$$

Théorème 2.6. (Wald et Wolfowitz, 1948)

Parmi tous les tests (séquentiel ou pas) pour lesquels :

$$P_{\theta}(\text{Rejeter } H_0) \geq \alpha, \text{ et } P_{\lambda}(\text{Accepter } H_0) \leq \beta$$

et pour lesquels $E_{\theta}N$ et $E_{\lambda}N$ sont finis, le test du rapport de probabilité séquentiel (SPRT) d'erreurs de probabilité α et β minimise $E_{\theta}N$ et $E_{\lambda}N$ simultanément.

Dans le but de déterminer la fonction caractéristique d'opération (OC) et les fonctions du nombre de l'échantillon moyen (ASN) du test (SPRT), Wald a développé les deux propositions suivantes :

Proposition 2.1. (Equation de Wald) :

Soient X_1, X_2, \dots des variables aléatoires indépendantes et identiquement distribuées de moyenne $\mu = EX_1$

Pour tout temps d'arrêt N avec $EN < \infty$ on a :

$$E\left(\sum_{i=1}^N X_i\right) = \mu EN$$

Proposition 2.2. (L'identité du rapport de vraisemblance de Wald) :

Supposons que X_1, X_2, \dots des variables aléatoires indépendantes et identiquement distribuées de densité f et g sous P_f et P_g respectivement. On définit la suite du rapport de vraisemblance

$$L_n = \prod_{i=1}^n \frac{g(X_i)}{f(X_i)}$$

Pour tout temps d'arrêt N :

$$P_f(N < \infty) = E_g(L_N^{-1}; N < \infty).$$

2.6.2 Problème de point de rupture :

Intéressons-nous maintenant aux problèmes de rupture. Considérons une suite d'observations indépendantes d'un certain processus. Initialement le processus est "en contrôle" et la vraie distribution des X est f_θ avec $\theta \in \Theta$. A un certain temps inconnu v , le processus devient "hors contrôle" dans le sens où la distribution de X_v, X_{v+1}, \dots est g_λ avec $\lambda \in \Lambda$. Ce problème est connu comme un problème de rupture où un problème de détection d'un changement.

On note par $P_{\theta,\lambda}^v$ la probabilité de mesure pour laquelle f_θ est la distribution avant le changement des X_1, X_2, \dots, X_{v-1} et g_λ celle des variables X_v, X_{v+1}, \dots c'est à dire après le changement.

On note aussi par P_θ la probabilité de mesure pour laquelle toutes les variables sont indépendantes et identiquement distribuées de densité f_θ (elle correspond à $v = \infty$).

Pour les problèmes de rupture, l'apparition d'un changement est définie par le temps d'arrêt N .

L'interprétation du temps d'arrêt est comme suit :

Lorsque $N = n$, on arrête et on déclare que le changement apparaît dans les n premières observations.

Le problème est de trouver le temps d'arrêt N pour lequel on arrête l'échantillon et dans le cas où le changement n'apparaît pas, on continue et on prend le plus grand échantillon possible.

La performance du temps d'arrêt est évaluée par deux critères :

- Le délai de détection (detection delay).
- La fréquence de la fausse alerte (frequency of false alarms).

Le délai de détection est défini par :

$$\bar{E}_\lambda N = \sup_{v \geq 1} (ess \sup E_{\theta,\lambda}^v [(N - v + 1)^+ / X_1, \dots, X_{v-1}])$$

Notons que la définition de $\bar{E}_\lambda N$ ne dépend pas de la distribution avant le changement f_θ . Le délai de détection est aussi appelé "short average run length" (ARL)

Si $\bar{E}_\lambda N$ est fini, alors $P_\theta(N < \infty) = 1$, ce qui montre qu'on a une fausse alerte avec une probabilité 1 même lorsqu'il n'y a pas de changement.

La fausse alerte est généralement mesurée par $1/E_\theta N$ où $E_\theta N$ est souvent appelée "long average run length".

Le problème de trouver une valeur optimale de N est résolue dans le cas simple où Θ et Λ sont simples ie : les deux distributions f_θ et g_λ avant et après le changement sont données.

La meilleure procédure est d'avoir $E_\theta N$ grand pour tout $\theta \in \Theta$ lorsque E_λ est petit pour tout $\lambda \in \Lambda$.

Les procédures optimales (exactes) sont données par la méthode CUSUM de Page où il utilise comme temps d'arrêt :

$$T_a = \inf\{n \geq 1 : \max_{1 \leq K \leq n} \sum_{i=K}^n \log \frac{g_\lambda(X_i)}{f_\theta(X_i)} \geq a\}$$

Théorème 2.7. (Moustakides, 1986)

Pour tout $a \geq 0$, la procédure T_a donnée par la méthode CUSUM de Page minimise le délai de détection $E_\lambda N$ parmi tous les temps d'arrêts N qui satisfont $E_\theta N \geq E_\theta T_a$.

Chapitre 3

Analyse Bayésienne d'un modèle de rupture sous contamination

3.1 Introduction

Dans ce chapitre, nous proposons une analyse Bayésienne d'un modèle de rupture sous contamination. Nous considérons un modèle Gaussien ayant subi un changement de structure à un instant inconnu en présence d'une valeur aberrante.

Soient X_1, X_2, \dots, X_n une suite de variables aléatoires gaussiennes et indépendantes, avec :

$$X_1, X_2, \dots, X_m \sim N(\phi_0, \sigma^2)$$

et

$$X_{m+1}, \dots, X_n \sim N(\phi_1, \sigma^2)$$

Où :

ϕ_0 et ϕ_1 sont des constantes réelles et inconnues . Elles représentent la moyenne des variables aléatoires X_i avant et après le changement respectivement.

σ^2 représente la variance des variables, elle est supposée inconnue.

Le paramètre m est le point de rupture. C'est un entier qui varie entre 1 et $n - 1$, n étant la taille de l'échantillon.

Le modèle peut donc s'écrire comme suit :

$$\begin{cases} X_i = \phi_0 + \epsilon_i; & i = 1, 2, \dots, m \\ X_i = \phi_1 + \epsilon_i; & i = m + 1, \dots, n \end{cases}$$

Les ϵ_i sont des erreurs aléatoires gaussiennes, indépendantes, de moyenne nulle et de variance inconnue σ^2 .

En adoptant l'approche Bayésienne, nous nous proposons d'étudier l'effet d'un contaminant sur l'estimation de ce point de rupture. nous donnons l'expression de la masse de probabilité a posteriori de ce dernier en présence d'une valeur aberrante (outlier). Nous considérons le cas où une seule observation est contaminée.

Supposons donc qu'il existe un k tel que $k \in \{1, \dots, n\}$ avec :

$$\begin{cases} Y_k = X_k + \Delta; \\ Y_i = X_i; \end{cases} \quad \forall i \in \{1, \dots, n\} \quad \text{avec } i \neq k \quad (3.1)$$

Où : Δ est l'amplitude de la contamination et k sa position.

Dans une première étape, nous supposons que l'observation affectée est avant la rupture puis nous considérons le cas où elle l'est après.

1) $k \in \{1, \dots, m\}$

2) $k \in \{m + 1, \dots, n\}$

3.2 Les densités a priori

Les paramètres d'intérêts du modèle considéré sont : la variance σ^2 , les moyennes (ϕ_0, ϕ_1) et le point de rupture m .

Supposons qu'ils sont indépendants et leurs densités a priori sont données respectivement par :

$$\begin{cases} \pi_0(\sigma^2) & \propto 1/\sigma^2 \\ \pi_0(\phi_0, \phi_1) & \propto \text{constante sur } \mathbb{R}^2 \\ \pi_0(m) & \propto 1/(n-1) \text{ pour } m = 1, 2, \dots, n-1 \end{cases} \quad (3.2)$$

La densité conjointe a priori des différents paramètres

On note par $\pi_0(m, \sigma^2, \phi_0, \phi_1)$, la densité conjointe a priori des paramètres $(m, \sigma^2, \phi_0, \phi_1)$.

En utilisant l'indépendance des paramètres m, σ^2 et (ϕ_0, ϕ_1) , on déduit :

$$\pi_0(m, \sigma^2, \phi_0, \phi_1) \propto \pi_0(m) \cdot \pi_0(\sigma^2) \cdot \pi_0(\phi_0, \phi_1)$$

Ce qui nous donne :

$$\pi_0(m, \sigma^2, \phi_0, \phi_1) \propto \frac{1}{\sigma^2} \cdot \frac{1}{n-1}$$

3.3 contamination avant rupture

3.3.1 Présentation du modèle

Dans le cas où la valeur contaminée X_k se situe avant la rupture, c'est-à-dire $k \in \{1, 2, \dots, m\}$, le modèle considéré s'écrit comme suit :

$$\begin{cases} Y_i = X_i = \phi_0 + \epsilon_i & i = 1, \dots, k-1 \quad \text{et} \quad i = k+1, \dots, m. \\ Y_k = X_k + \Delta = \phi_0 + \epsilon_k + \Delta \\ Y_i = X_i = \phi_1 + \epsilon_i & i = m+1, \dots, n. \end{cases}$$

3.3.2 La fonction de vraisemblance des différents paramètres

Comme :

$$\begin{aligned} Y_i &\longrightarrow N(\phi_0, \sigma^2) & i = 1, \dots, k-1 \quad \text{et} \quad i = k+1, \dots, m. \\ Y_k &\longrightarrow N(\phi_0 + \Delta, \sigma^2) \\ Y_i &\longrightarrow N(\phi_1, \sigma^2) & i = m+1, \dots, n. \end{aligned}$$

Alors la fonction de vraisemblance des paramètres m, σ^2, ϕ_0 et ϕ_1 notée $L(m, \sigma^2, \phi_0, \phi_1/Y)$, est donnée par :

$$\begin{aligned} L(m, \sigma^2, \phi_0, \phi_1/Y) \propto (\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} \left[\sum_{i=1}^{k-1} (Y_i - \phi_0)^2 + (Y_k - (\phi_0 + \Delta))^2 \right. \right. \\ \left. \left. + \sum_{i=k+1}^m (Y_i - \phi_0)^2 + \sum_{m+1}^n (Y_i - \phi_1)^2 \right] \right\} \end{aligned}$$

3.3.3 L'analyse a posteriori

D'après le théorème de Bayes, la densité conjointe a posteriori des paramètres $(m, \sigma^2, \phi_0, \phi_1)$ est donnée par :

$$\pi_1(m, \sigma^2, \phi_0, \phi_1) \propto L(m, \sigma^2, \phi_0, \phi_1/Y) \cdot \pi_0(m, \sigma^2, \phi_0, \phi_1)$$

En remplaçant $L(m, \sigma^2, \phi_0, \phi_1/Y)$ et $\pi_0(m, \sigma^2, \phi_0, \phi_1)$ par leurs valeurs, nous aurons le résultat suivant :

$$\begin{aligned} \pi_1(m, \sigma^2, \phi_0, \phi_1) \propto (\sigma^2)^{-n/2-1} \exp \left\{ -\frac{1}{2\sigma^2} \left[\sum_{i=1}^{k-1} (Y_i - \phi_0)^2 + (Y_k - (\phi_0 + \Delta))^2 \right. \right. \\ \left. \left. + \sum_{i=k+1}^m (Y_i - \phi_0)^2 + \sum_{m+1}^n (Y_i - \phi_1)^2 \right] \right\} \end{aligned}$$

Et la densité a posteriori du point de rupture m est :

$$\pi_1(m) \propto \int_0^{+\infty} \int_R \int_R \pi_1(m, \sigma^2, \phi_0, \phi_1) d\phi_0 d\phi_1 d\sigma^2$$

En intégrant $\pi_1(m, \sigma^2, \phi_0, \phi_1)$ par rapport à σ^2 , on retrouve la densité a posteriori des paramètres m, ϕ_0, ϕ_1 :

$$\pi_1(m, \phi_0, \phi_1) = \int_0^{+\infty} \pi_1(m, \sigma^2, \phi_0, \phi_1) d\sigma^2$$

Pour calculer cette intégrale posons :

$$\sigma^2 = \frac{1}{r} \Rightarrow d\sigma^2 = -\frac{1}{r^2} dr$$

Donc :

$$\begin{aligned} \pi_1(m, \phi_0, \phi_1) \propto \int_0^{+\infty} r^{\frac{n}{2}-1} \cdot \exp \left\{ -\frac{r}{2} \left[\sum_{i=1}^{k-1} (Y_i - \phi_0)^2 + (Y_k - (\Delta + \phi_0))^2 \right. \right. \\ \left. \left. + \sum_{i=k+1}^m (Y_i - \phi_0)^2 + \sum_{i=m+1}^n (Y_i - \phi_1)^2 \right] \right\} dr \end{aligned}$$

En utilisant les propriétés de la densité d'une loi Gamma ($\int_0^{+\infty} (b^a / \Gamma(a)) r^{a-1} e^{-br} dr = 1$), on obtient :

$$\pi_1(m, \phi_0, \phi_1) \propto \left[\sum_{i=1}^{k-1} (Y_i - \phi_0)^2 + (Y_k - (\Delta + \phi_0))^2 + \sum_{i=k+1}^m (Y_i - \phi_0)^2 + \sum_{i=m+1}^n (Y_i - \phi_1)^2 \right]^{-\frac{n}{2}}$$

Pour retrouver $\pi_1(m, \phi_0)$ on intègre $\pi_1(m, \phi_0, \phi_1)$ par rapport à ϕ_1

$$\pi_1(m, \phi_0) = \int_R \pi_1(m, \phi_0, \phi_1) d\phi_1$$

Pour le calcul de cette intégrale posons :

$$I = \sum_{i=1}^{k-1} (Y_i - \phi_0)^2 + (Y_k - (\Delta + \phi_0))^2 + \sum_{i=k+1}^m (Y_i - \phi_0)^2 + \sum_{i=m+1}^n (Y_i - \phi_1)^2$$

Cette expression peut être réécrite comme suit :

$$\begin{aligned} I &= \sum_{i=1}^{k-1} \left(Y_i - \bar{Y}_1^{k-1} + \bar{Y}_1^{k-1} - \phi_0 \right)^2 + (Y_k - \Delta - \phi_0)^2 \\ &\quad + \sum_{i=k+1}^m \left(Y_i - \bar{Y}_{k+1}^m + \bar{Y}_{k+1}^m - \phi_0 \right)^2 + \sum_{i=m+1}^n \left(Y_i - \bar{Y}_{m+1}^n + \bar{Y}_{m+1}^n - \phi_1 \right)^2 \end{aligned}$$

Avec : $\bar{Y}_j^k = \frac{1}{k-j+1} \sum_{i=j}^k Y_i$

On a :

$$\begin{aligned} 2 \sum_i^{k-1} \left[(Y_i - \bar{Y}_1^{k-1}) (\bar{Y}_1^{k-1} - \phi_0) \right] &= 2 \sum_{k+1}^m \left[(Y_i - \bar{Y}_{k+1}^m) (\bar{Y}_{k+1}^m - \phi_0) \right] \\ &= 2 \sum_{m+1}^n \left[(Y_i - \bar{Y}_{m+1}^n) (\bar{Y}_{m+1}^n - \phi_1) \right] = 0 \end{aligned}$$

Ce qui nous donne :

$$\begin{aligned} I &= (k-1) \left(\bar{Y}_1^{k-1} - \phi_0 \right)^2 + (m-k) \left(\bar{Y}_{k+1}^m - \phi_0 \right)^2 + (n-m) \left(\bar{Y}_{m+1}^n - \phi_1 \right)^2 \\ &\quad + S_1^{k-1} + S_{k+1}^m + S_{m+1}^n + (Y_k - \phi_0 - \Delta)^2 \\ &= Q_1(\Delta) + S_{m+1}^n + (n-m) \left(\bar{Y}_{m+1}^n - \phi_1 \right)^2 \end{aligned}$$

$$\text{Avec : } S_j^k = \sum_{i=j}^k \left(Y_i - \bar{Y}_j^k \right)$$

Et :

$$\begin{aligned} Q_1(\Delta) &= (k-1) \left(\bar{Y}_1^{k-1} - \phi_0 \right)^2 + (m-k) \left(\bar{Y}_{k+1}^m - \phi_0 \right)^2 \\ &\quad + (Y_k - \phi_0 - \Delta)^2 + S_1^{k-1} + S_{k+1}^m \end{aligned}$$

Donc :

$$\begin{aligned} \pi_1(m, \phi_0) &\propto \int_R \left[Q_1(\Delta) + S_{m+1}^n + (n-m) \left(\bar{Y}_{m+1}^n - \phi_1 \right)^2 \right]^{-\frac{n}{2}} d\phi_1 \\ &= \left(Q_1(\Delta) + S_{m+1}^n \right)^{-\frac{n}{2}} \int_R \left[1 + \left(\frac{\bar{Y}_{m+1}^n - \phi_1}{\sqrt{(Q_1(\Delta) + S_{m+1}^n) / (n-m)}} \right)^2 \right]^{-\frac{n}{2}} d\phi_1 \end{aligned}$$

En posant

$$\frac{z}{\sqrt{n-1}} = \frac{\bar{Y}_{m+1}^n - \phi_1}{\sqrt{(Q_1(\Delta) + S_{m+1}^n) / (n-m)}}$$

et en tenant compte des propriétés d'une loi de Student à n degrés de liberté on obtient :

$$\begin{aligned} \pi_1(m, \phi_0) &\propto \left(Q_1(\Delta) + S_{m+1}^n \right)^{-\frac{n-1}{2}} (n-1)^{-\frac{1}{2}} (n-m)^{-\frac{1}{2}} \int_R \left(1 + \frac{z^2}{n-1} \right)^{-\frac{n}{2}} dz \\ &\propto \left(Q_1(\Delta) + S_{m+1}^n \right)^{-\frac{n-1}{2}} \cdot [(n-m)(n-1)]^{-\frac{1}{2}} (n-1)^{\frac{1}{2}} \end{aligned}$$

Ce qui nous donne :

$$\pi_1(m, \phi_0) \propto (Q_1(\Delta) + S_{m+1}^n)^{-\frac{n-1}{2}} (n-m)^{-\frac{1}{2}}$$

Une dernière intégration de $\pi_1(m, \phi_0)$ par rapport à ϕ_0 nous donne la densité a posteriori du point de rupture m

$$\pi_1(m) \propto \int_{\mathcal{R}} (Q_1(\Delta) + S_{m+1}^n)^{-\frac{n-1}{2}} (n-m)^{-\frac{1}{2}} d\phi_0$$

Nous avons :

$$\begin{aligned} Q_1(\Delta) + S_{m+1}^n &= (k-1) \left(\bar{Y}_1^{k-1} - \phi_0 \right)^2 + (m-k) \left(\bar{Y}_{k+1}^m - \phi_0 \right)^2 \\ &\quad + (Y_k - \phi_0 - \Delta)^2 + S_1^{k-1} + S_{k+1}^m + S_{m+1}^n \\ &= m\phi_0^2 - 2\phi_0 \left(\sum_{i=1}^m Y_i - \Delta \right) + \frac{(\sum_{i=1}^{k-1} Y_i)^2}{k-1} + \frac{(\sum_{i=k+1}^m Y_i)^2}{m-k} \\ &\quad + (Y_k - \Delta)^2 + S_1^{k-1} + S_{k+1}^m + S_{m+1}^n \\ &= m \left[\phi_0 - \frac{\sum_{i=1}^m Y_i - \Delta}{m} \right]^2 - \frac{(\sum_{i=1}^m Y_i - \Delta)^2}{m} + \frac{(\sum_{i=1}^{k-1} Y_i)^2}{k-1} + \frac{(\sum_{i=k+1}^m Y_i)^2}{m-k} \\ &\quad + (Y_k - \Delta)^2 + S_1^{k-1} + S_{k+1}^m + S_{m+1}^n \end{aligned}$$

D'où :

$$Q_1(\Delta) + S_{m+1}^n = m \left[\phi_0 - \frac{\sum_{i=1}^m Y_i - \Delta}{m} \right]^2 + Q_2(\Delta)$$

Avec :

$$Q_2(\Delta) = -\frac{(\sum_{i=1}^m Y_i - \Delta)^2}{m} + \frac{(\sum_{i=1}^{k-1} Y_i)^2}{k-1} + \frac{(\sum_{i=k+1}^m Y_i)^2}{m-k} + (Y_k - \Delta)^2 + S_1^{k-1} + S_{k+1}^m + S_{m+1}^n$$

Donc :

$$\pi_1(m) \propto \int_{\mathcal{R}} (n-m)^{-\frac{1}{2}} \left[m \left(\phi_0 - \frac{\sum_{i=1}^m Y_i - \Delta}{m} \right)^2 + Q_2(\Delta) \right]^{-\frac{n-1}{2}} d\phi_0$$

Que nous pouvons écrire :

$$\pi_1(m) \propto (n-m)^{-\frac{1}{2}} \cdot Q_2(\Delta)^{-\frac{n-1}{2}} \int_{\mathcal{R}} \left[1 + \left(\frac{\phi_0 - \frac{\sum_{i=1}^m Y_i - \Delta}{m}}{\sqrt{Q_2(\Delta)/m}} \right)^2 \right]^{-\frac{n-1}{2}} d\phi_0$$

Posons :

$$\frac{z}{\sqrt{n-2}} = \left(\phi_0 - \frac{\sum_{i=1}^m Y_i - \Delta}{m} \right) / \sqrt{Q_2(\Delta)/m}$$

Donc :

$$\pi_1(m) \propto (n-m)^{-\frac{1}{2}} \cdot Q_2(\Delta)^{-\frac{n-1}{2}} \cdot \frac{\sqrt{Q_2(\Delta)/m}}{\sqrt{n-2}} \int_{\mathcal{R}} \left(1 + \frac{z^2}{n-2} \right)^{-\frac{n-1}{2}} dz$$

Et comme : $\int_{\mathcal{R}} \left(1 + \frac{z^2}{n-2} \right)^{-\frac{n-1}{2}} dz \propto (n-2)^{1/2}$ (loi de student à $(n-1)$ degrés de liberté)

Alors la densité a posteriori du point de rupture m est :

$$\pi_1(m) \propto [m(n-m)]^{-\frac{1}{2}} \cdot Q_2(\Delta)^{-\frac{n-2}{2}}$$

Avec :

$$Q_2(\Delta) = -\frac{(\sum_{i=1}^m Y_i - \Delta)^2}{m} + \frac{(\sum_{i=1}^{k-1} Y_i)^2}{k-1} + \frac{(\sum_{i=k+1}^m Y_i)^2}{m-k} + (Y_k - \Delta)^2 + S_1^{k-1} + S_{k+1}^m + S_{m+1}^n$$

3.4 contamination après rupture

3.4.1 Présentation du modèle

Dans ce cas $k \in \{m+1, \dots, n\}$. Le modèle s'écrit alors :

$$\begin{cases} Y_i = X_i = \phi_0 + \epsilon_i & i = 1, \dots, m. \\ Y_i = X_i = \phi_1 + \epsilon_i & i = m+1, \dots, k-1 \quad \text{et} \quad i = k+1, \dots, n. \\ Y_k = X_k + \Delta = \phi_1 + \epsilon_k + \Delta \end{cases}$$

3.4.2 La fonction de vraisemblance des différents paramètres

Comme :

$$\begin{aligned} Y_i &\longrightarrow N(\phi_0, \sigma^2) & i = 1, \dots, m. \\ Y_i &\longrightarrow N(\phi_1, \sigma^2) & i = m+1, \dots, k-1 \quad \text{et} \quad i = k+1, \dots, n. \\ Y_k &\longrightarrow N(\phi_1 + \Delta, \sigma^2) \end{aligned}$$

Alors la fonction de vraisemblance L des paramètres m, σ^2, ϕ_0 et ϕ_1 est donnée par :

$$L(m, \sigma^2, \phi_0, \phi_1 / Y) \propto (\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} \left[\sum_{i=1}^m (Y_i - \phi_0)^2 + (Y_k - (\phi_1 + \Delta))^2 + \sum_{i=m+1}^{k-1} (Y_i - \phi_1)^2 + \sum_{k+1}^n (Y_i - \phi_1)^2 \right] \right\}$$

3.4.3 L'analyse a posteriori

La densité conjointe a posteriori des paramètres $(m, \sigma^2, \phi_0, \phi_1)$ est donnée en vertu du théorème de Bayes par :

$$\pi_1(m, \sigma^2, \phi_0, \phi_1) \propto L(m, \sigma^2, \phi_0, \phi_1/Y) \cdot \pi_0(m, \sigma^2, \phi_0, \phi_1)$$

Ce qui est équivalent à :

$$\begin{aligned} \pi_1(m, \sigma^2, \phi_0, \phi_1) \propto & (\sigma^2)^{-n/2-1} \exp \left\{ -\frac{1}{2\sigma^2} \left[\sum_{i=1}^{m+1} (Y_i - \phi_0)^2 + (Y_k - (\phi_1 + \Delta))^2 \right. \right. \\ & \left. \left. + \sum_{i=m+1}^{k-1} (Y_i - \phi_0)^2 + \sum_{k+1}^n (Y_i - \phi_1)^2 \right] \right\} \end{aligned}$$

Et la densité a posteriori du point de rupture m est :

$$\pi_1(m) \propto \int_0^{+\infty} \int_R \int_R \pi_1(m, \sigma^2, \phi_0, \phi_1) d\phi_0 d\phi_1 d\sigma^2$$

En intégrant $\pi_1(m, \sigma^2, \phi_0, \phi_1)$ par rapport à σ^2 , on retrouve la densité a posteriori des paramètres m, ϕ_0, ϕ_1 :

$$\pi_1(m, \phi_0, \phi_1) = \int_0^{+\infty} \pi_1(m, \sigma^2, \phi_0, \phi_1) d\sigma^2$$

Posons :

$$\sigma^2 = \frac{1}{r} \quad \Rightarrow \quad d\sigma^2 = -\frac{1}{r^2} dr$$

Donc :

$$\begin{aligned} \pi_1(m, \phi_0, \phi_1) \propto & \int_0^{+\infty} r^{\frac{n}{2}-1} \cdot \exp \left\{ -\frac{r}{2} \left[\sum_{i=1}^m (Y_i - \phi_0)^2 + (Y_k - (\Delta + \phi_1))^2 \right. \right. \\ & \left. \left. + \sum_{i=m+1}^{k-1} (Y_i - \phi_1)^2 + \sum_{i=k+1}^n (Y_i - \phi_1)^2 \right] \right\} dr \end{aligned}$$

Chapitre 3. Analyse Bayésienne d'un modèle de rupture sous contamination 80

En utilisant les propriétés de la densité d'une loi Gamma ($\int_0^\infty (b^a/\Gamma(a)) r^{a-1} e^{-br} dr = 1$), on obtient :

$$\pi_1(m, \phi_0, \phi_1) \propto \left[\sum_{i=1}^m (Y_i - \phi_0)^2 + (Y_k - (\Delta + \phi_1))^2 + \sum_{i=m+1}^{k-1} (Y_i - \phi_1)^2 + \sum_{i=k+1}^n (Y_i - \phi_1)^2 \right]^{-\frac{n}{2}}$$

Dans ce deuxième cas, pour retrouver la densité a posteriori de m , nous intégrons $\pi_1(m, \phi_0, \phi_1)$ d'abord par rapport à ϕ_0 puis par rapport à ϕ_1

$$\pi_1(m, \phi_1) = \int_R \pi_1(m, \phi_0, \phi_1) d\phi_0$$

Pour le calcul de cette intégrale posons :

$$I = \sum_{i=1}^m (Y_i - \phi_0)^2 + (Y_k - (\Delta + \phi_1))^2 + \sum_{i=m+1}^{k-1} (Y_i - \phi_1)^2 + \sum_{i=k+1}^n (Y_i - \phi_1)^2$$

que l'on peut aussi écrire comme suit :

$$\begin{aligned} I &= \sum_{i=1}^m (Y_i - \bar{Y}_1^m + \bar{Y}_1^m - \phi_0)^2 + (Y_k - \Delta - \phi_1)^2 \\ &+ \sum_{i=m+1}^{k-1} (Y_i - \bar{Y}_{m+1}^{k-1} + \bar{Y}_{m+1}^{k-1} - \phi_1)^2 + \sum_{i=k+1}^n (Y_i - \bar{Y}_{k+1}^n + \bar{Y}_{k+1}^n - \phi_1)^2 \end{aligned}$$

$$\text{Avec : } \bar{Y}_j^k = \frac{1}{k-j+1} \sum_{i=j}^k Y_i$$

Comme :

$$\begin{aligned} 2 \sum_i^m [(Y_i - \bar{Y}_1^m) (\bar{Y}_1^m - \phi_0)] &= 2 \sum_{m+1}^{k-1} [(Y_i - \bar{Y}_{m+1}^{k-1}) (\bar{Y}_{m+1}^{k-1} - \phi_1)] \\ &= 2 \sum_{k+1}^n [(Y_i - \bar{Y}_{k+1}^n) (\bar{Y}_{k+1}^n - \phi_1)] = 0 \end{aligned}$$

Alors :

$$\begin{aligned} I &= (m) (\bar{Y}_1^m - \phi_0)^2 + (k - m - 1) (\bar{Y}_{m+1}^{k-1} - \phi_1)^2 + (n - k) (\bar{Y}_{k+1}^n - \phi_1)^2 \\ &\quad + S_1^m + S_{m+1}^{k-1} + S_{k+1}^n + (Y_k - \phi_1 - \Delta)^2 \\ &= Q_1(\Delta) + S_1^m + (m) (\bar{Y}_1^m - \phi_0)^2 \end{aligned}$$

Avec : $S_j^k = \sum_{i=j}^k (Y_i - \bar{Y}_j^k)$

Et :

$$\begin{aligned} Q_1(\Delta) &= (k - m - 1) (\bar{Y}_{m+1}^{k-1} - \phi_1)^2 + (n - k) (\bar{Y}_{k+1}^n - \phi_1)^2 \\ &\quad + (Y_k - \phi_1 - \Delta)^2 + S_{m+1}^{k-1} + S_{k+1}^n \end{aligned}$$

Ce qui nous donne :

$$\begin{aligned} \pi_1(m, \phi_1) &\propto \int_R [Q_1(\Delta) + S_1^m + (m) (\bar{Y}_1^m - \phi_0)^2]^{-\frac{n}{2}} d\phi_0 \\ &= (Q_1(\Delta) + S_1^m)^{-\frac{n}{2}} \int_R \left[1 + \left(\frac{\bar{Y}_1^m - \phi_0}{\sqrt{(Q_1(\Delta) + S_1^m) / (m)}} \right)^2 \right]^{-\frac{n}{2}} d\phi_0 \end{aligned}$$

posons :

$$\frac{z}{\sqrt{n-1}} = \frac{\bar{Y}_1^m - \phi_0}{\sqrt{(Q_1(\Delta) + S_1^m) / m}}$$

$$\pi_1(m, \phi_1) \propto (Q_1(\Delta) + S_1^m)^{-\frac{n-1}{2}} (n-1)^{-\frac{1}{2}} \sqrt{(Q_1(\Delta) + S_1^m)/m} \int_R \left(1 + \frac{z^2}{n-1}\right)^{-\frac{n}{2}} dz$$

Puisque; $\int_R \left(1 + \frac{z^2}{n-1}\right)^{-\frac{n}{2}} dz \propto (n-1)^{1/2}$

Alors :

$$\pi_1(m, \phi_1) \propto (Q_1(\Delta) + S_1^m)^{-\frac{n-1}{2}} \cdot (m)^{-\frac{1}{2}}$$

Une dernière intégration de $\pi_1(m, \phi_1)$ par rapport à ϕ_1 nous donne la densité a posteriori du point de rupture m

$$\pi_1(m) \propto \int_R (Q_1(\Delta) + S_1^m)^{-\frac{n-1}{2}} (m)^{-\frac{1}{2}} d\phi_1$$

Nous avons :

$$\begin{aligned} Q_1(\Delta) + S_1^m &= (k-m-1) \left(\bar{Y}_{m+1}^{k-1} - \phi_1\right)^2 + (n-k) \left(\bar{Y}_{k+1}^n - \phi_1\right)^2 \\ &\quad + (Y_k - \phi_1 - \Delta)^2 + S_1^m + S_{m+1}^{k-1} + S_{k+1}^n \\ &= (n-m)\phi_1^2 - 2\phi_1 \left(\sum_{i=m+1}^n Y_i - \Delta\right) + \frac{(\sum_{i=m+1}^{k-1} Y_i)^2}{k-m-1} + \frac{(\sum_{i=k+1}^n Y_i)^2}{n-k} \\ &\quad + (Y_k - \Delta)^2 + S_1^m + S_{m+1}^{k-1} + S_{k+1}^n \\ &= (n-m) \left[\phi_1 - \frac{\sum_{i=m+1}^n Y_i - \Delta}{n-m}\right]^2 - \frac{(\sum_{i=m+1}^n Y_i - \Delta)^2}{n-m} + \frac{(\sum_{i=m+1}^{k-1} Y_i)^2}{k-m-1} \\ &\quad + \frac{(\sum_{i=k+1}^n Y_i)^2}{n-k} + (Y_k - \Delta)^2 + S_1^m + S_{m+1}^{k-1} + S_{k+1}^n \end{aligned}$$

D'où :

$$Q_1(\Delta) + S_1^m = (n-m) \left[\phi_1 - \frac{\sum_{i=m+1}^n Y_i - \Delta}{n-m}\right]^2 + Q_2(\Delta)$$

Avec :

$$Q_2(\Delta) = -\frac{(\sum_{i=m+1}^n Y_i - \Delta)^2}{n-m} + \frac{(\sum_{i=m+1}^{k-1} Y_i)^2}{k-m-1} + \frac{(\sum_{i=k+1}^n Y_i)^2}{n-k} + (Y_k - \Delta)^2 + S_1^m + S_{m+1}^{k-1} + S_{k+1}^n$$

Donc :

$$\pi_1(m) \propto \int_R m^{-\frac{1}{2}} \left[(n-m) \left(\phi_1 - \frac{\sum_{i=m+1}^n Y_i - \Delta}{n-m} \right)^2 + Q_2(\Delta) \right]^{-\frac{n-1}{2}} d\phi_1$$

$$\pi_1(m) \propto m^{-\frac{1}{2}} \cdot Q_2(\Delta)^{-\frac{n-1}{2}} \int_R \left[1 + \left(\frac{\phi_1 - \frac{\sum_{i=m+1}^n Y_i - \Delta}{n-m}}{\sqrt{Q_2(\Delta)/(n-m)}} \right)^2 \right]^{-\frac{n-1}{2}} d\phi_1$$

Posons :

$$\frac{z}{\sqrt{n-2}} = \left(\phi_1 - \frac{\sum_{i=m+1}^n Y_i - \Delta}{n-m} \right) / \sqrt{Q_2(\Delta)/(n-m)}$$

Donc :

$$\pi_1(m) \propto m^{-\frac{1}{2}} \cdot Q_2(\Delta)^{-\frac{n-1}{2}} \cdot \frac{\sqrt{Q_2(\Delta)/(n-m)}}{\sqrt{n-2}} \int_R \left(1 + \frac{z^2}{n-2} \right)^{-\frac{n-1}{2}} dz$$

Et comme : $\int_R \left(1 + \frac{z^2}{n-2} \right)^{-\frac{n-1}{2}} dz \propto (n-2)^{1/2}$ (loi de student à $(n-1)$ degrés de liberté)

Alors la densité a posteriori du point de rupture m est :

$$\pi_1(m) \propto [m(n-m)]^{-\frac{1}{2}} \cdot [Q_2(\Delta)]^{-\frac{n-2}{2}}$$

Avec :

$$Q_2(\Delta) = -\frac{(\sum_{i=m+1}^n Y_i - \Delta)^2}{n-m} + \frac{(\sum_{i=m+1}^{k-1} Y_i)^2}{k-m-1} + \frac{(\sum_{i=k+1}^n Y_i)^2}{n-k} + (Y_k - \Delta)^2 + S_1^m + S_{m+1}^{k-1} + S_{k+1}^n$$

3.5 Application numérique

Dans cette section, nous allons présenter quelques résultats de simulation de l'application de la méthode d'estimation du point de rupture. Notre but est de voir si le mode a posteriori calculé estime bien la vraie valeur du paramètre d'intérêt en présence de valeurs aberrantes.

Pour cela, nous avons simulé 100 échantillons à partir du modèle (3.1), en considérant $\phi_0 = 0$, $\phi_1 = 1$ et $\sigma^2 = 0.5$.

Dans le but d'étudier l'influence de la taille de l'échantillon, de la position du contaminant ainsi que la position du point de rupture, sur la performance de cette méthode d'estimation, nous avons fait varier les valeurs de n, k, m et Δ .

3.5.1 Position et amplitude du contaminant

Afin d'étudier l'influence de la position et de l'amplitude du contaminant sur la performance de la méthode, nous avons simulé 100 échantillon de taille $n = 50$ à partir du modèle (3.1) pour différentes valeurs de Δ , en considérant $m = 20$, et nous avons fait varier la position du contaminant, pour ce dernier nous avons considéré les valeurs suivantes : $k = 10$ (contamination avant la rupture) et $k = 30$ (contamination après la rupture).

Les résultats obtenus sont donnés par les figures (*FIG.3.1*), (*FIG.3.2*) et (*FIG.3.3*) :

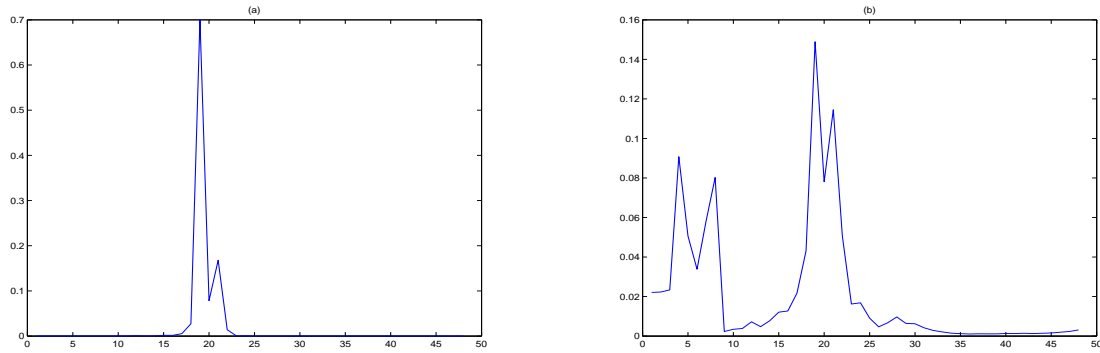


FIGURE 3.1 – Graphe d'une série simulée à partir du modèle (3,1), pour $n=50$, $k=10$ et $m=20$ (a) : $\Delta = 0$, (b) $\Delta = 4$

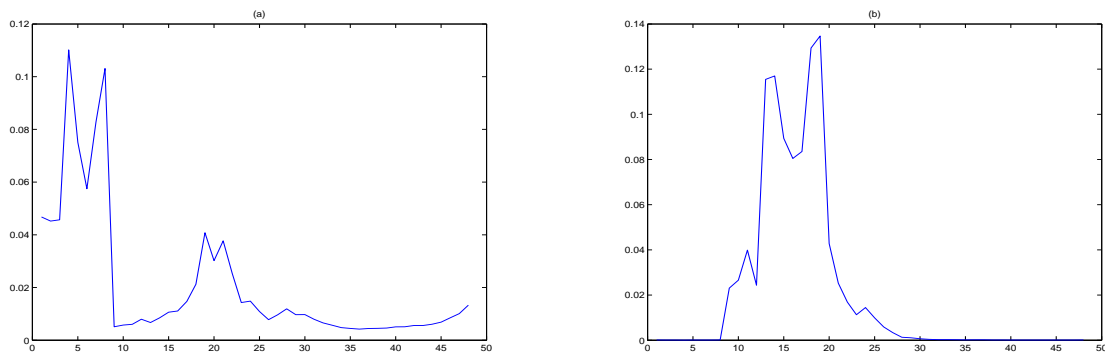


FIGURE 3.2 – Graphe d'une série simulée à partir du modèle (3,1), pour $n=50$, $k=10$ et $m=20$ (a) : $\Delta = 6$, (b) $\Delta = -6$

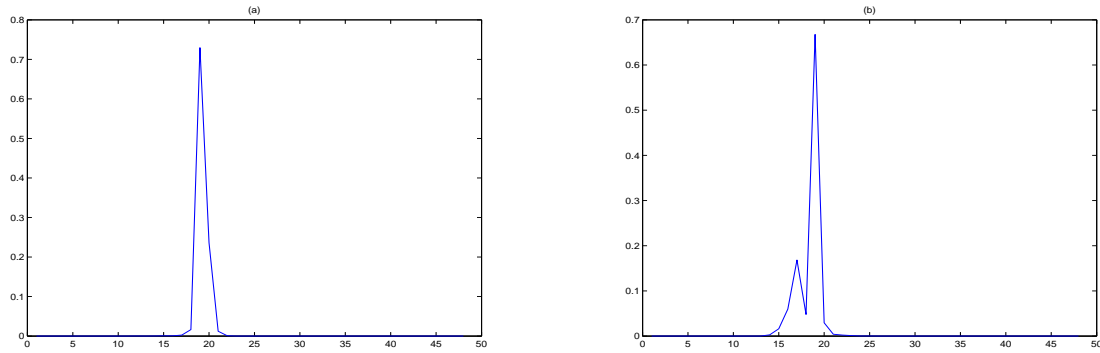


FIGURE 3.3 – Graphe d'une série simulée à partir du modèle (3,1), pour $n=50$, $k=30$ et $m=20$ (a) : $\Delta = 0$, (b) $\Delta = 4$

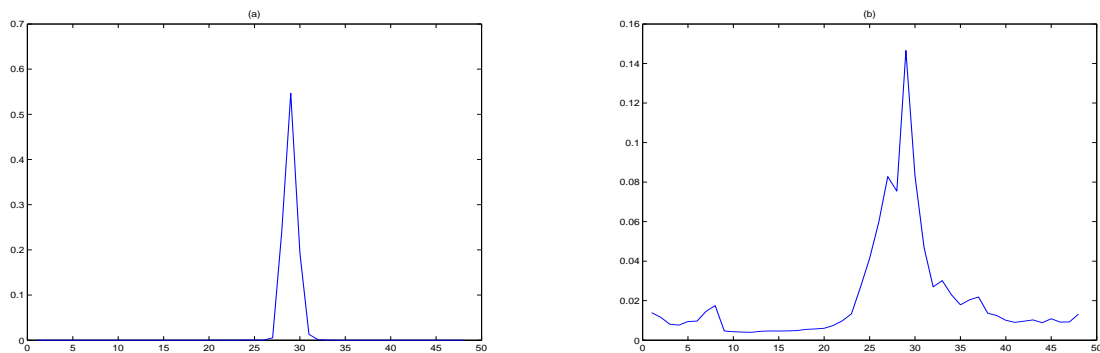


FIGURE 3.4 – Graphe d'une série simulée à partir du modèle (3,1), pour $n=50$, $k=10$ et $m=30$ (a) : $\Delta = 0$, (b) $\Delta = 6$

Au vu des résultats, nous remarquons que cette méthode d'estimation du point de rupture est insensible aux petites valeurs absolues du contaminant Δ , ($|\Delta| < 6$). Ce qu'on peut voir dans les figures ci-dessus (*FIG.3.1*). Mais lorsque la valeur absolue du contaminant augmente ($|\Delta| \geq 6$), la méthode perd son efficacité. (voir figure (*FIG.3.2*))

Dans le cas où $k = 30$, les résultats sont donnés par les figures (*FIG.3.3*). Nous remarquons que ces résultats sont quasiment similaires à ceux retrouvés pour $k = 10$, donc la position du contaminant n'a pas d'influence sur la variation de la méthode en fonction de Δ .

3.5.2 Position du point de rupture

Pour vérifier l'influence de la position du point de rupture sur la performance de la méthode d'estimation, nous avons simulé 100 échantillons pour $m = 30$ en gardant les autres paramètres inchangés. Les résultats sont donnés par la figure *FIG.3.4*.

Ces résultats montrent que la performance de la méthode n'est pas influencée par l'emplacement du point de rupture m , (voir (*FIG.3.4*)). D'autre part, les résultats donnés par les figures (*FIG.3.2*), (*FIG.3.3*) et (*FIG.3.4*) montrent que la variation de la méthode en fonction de Δ n'est pas influencée par l'emplacement de m .

3.5.3 La taille de l'échantillon

Nous avons conduit une autre étude de simulation pour investiguer l'influence de la taille de l'échantillon sur la performance de la méthode. Pour cela, nous avons simulé 100 échantillons de taille $N = 100$ en considérant $m = 30$ et $k = 10$ et 100 échantillons de taille $N = 25$ en considérant $m = 15$ et $k = 5$.

Les résultats obtenus sont donnés par les figures (*FIG.3.5*), (*FIG.3.6*) et (*FIG.3.7*) :

Ces résultats montrent que la performance de la méthode s'améliore lorsque la taille de l'échantillon augmente, ce qu'on peut voir dans les figures (*FIG.3.5.a*), et (*FIG.3.7.a*). (cas sans contamination, $\Delta = 0$)

Nous remarquons aussi que la taille de l'échantillon a une influence importante sur la variation de la méthode en fonction de l'amplitude du contaminant. En effet, lorsque

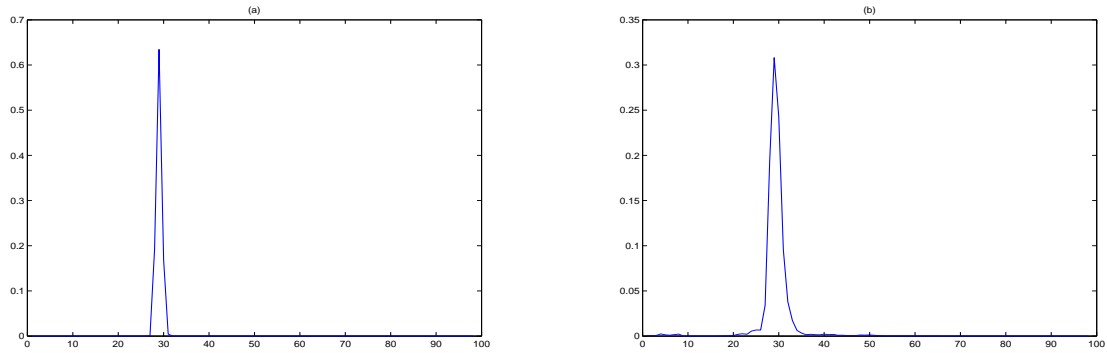


FIGURE 3.5 – Graphe d'une série simulée à partir du modèle (3,1), pour $n=100$, $k=10$ et $m=30$ (a) : $\Delta = 0$, (b) $\Delta = 6$

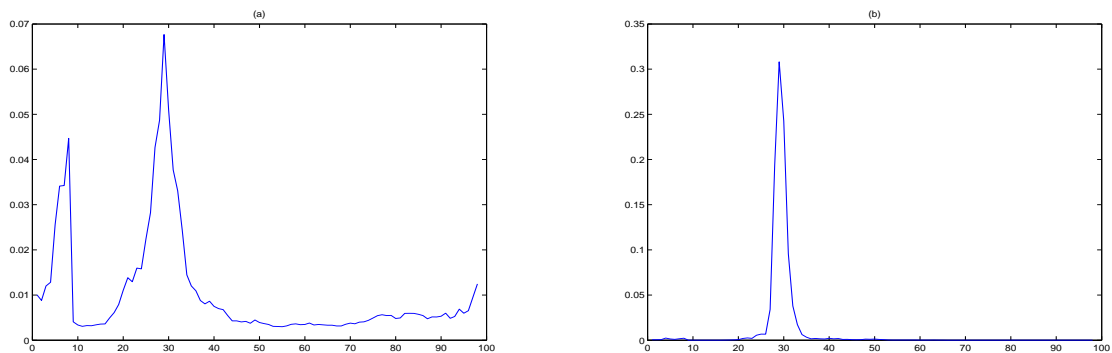


FIGURE 3.6 – Graphe d'une série simulée à partir du modèle (3,1), pour $n=100$, $k=10$ et $m=30$ (a) : $\Delta = 15$, (b) $\Delta = -15$

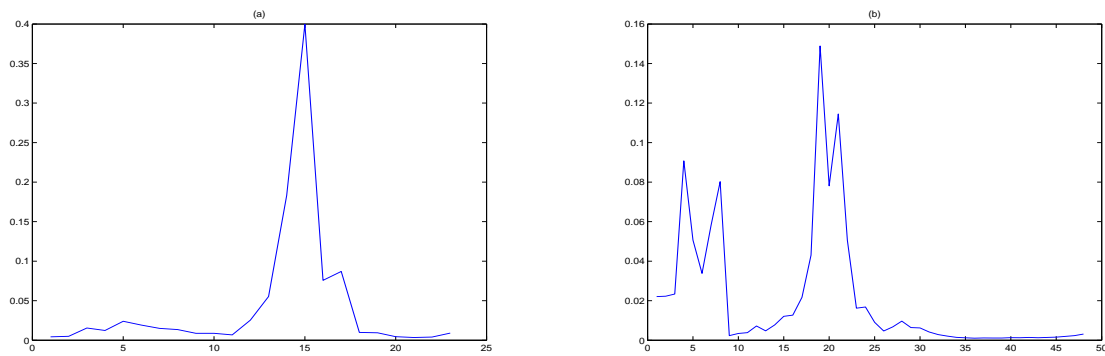


FIGURE 3.7 – Graphe d'une série simulée à partir du modèle (3,1), pour $n=25$, $k=5$ et $m=15$ (a) : $\Delta = 0$, (b) $\Delta = 4$

la taille de l'échantillon est grande, cette méthode d'estimation est insensible aux petites valeurs de Δ , (voir les figures (*FIG.3.5*)et(*FIG.3.6*)). Mais pour un échantillon de taille petite, la méthode perd sa performance pour des petites valeurs du contaminant, ce qu'on peut voir dans les figures (*FIG.3.7*). Donc plus la taille de l'échantillon est grande plus l'effet de la contamination est réduit.

3.6 conclusion

Ces résultats montrent que la performance de cette méthode n'est pas influencée par l'emplacement du point de rupture, ni par la position du contaminant. Elle est aussi insensible aux petites valeurs de l'amplitude Δ .

Nous remarquons également que la taille de l'échantillon a une influence sur la variation de la méthode en fonction de l'amplitude du contaminant, en effet plus la taille de l'échantillon est grande plus l'effet du contaminant est réduit.

Conclusion générale

Le domaine que nous avons élaboré répond à des situations tant économiques que sociales (embargo pétrolier, crash économique, mouvement de protection de consommateurs contre un produit pharmaceutique particulier, etc...). En effet, l'analyse de rupture dans des séquences de variables aléatoires a de multiples applications, notamment en contrôle de qualité, en traitement de signal, en économie et en finances.

Contrôle de qualité : Les techniques de surveillance et de maîtrise de la qualité font en général appel aux détecteurs séquentiel de changements. Ces outils sont utilisés pour détecter un dérèglement dans un contexte de production. L'estimation de l'instant et de l'amplitude d'un changement est donc d'un grand intérêt pour le responsable de la qualité.

Traitement de signal : Le problème de détection de changement dans les caractéristiques d'un signal est courant en traitement d'image, en médecine, en géophysique ou encore en sismologie.

Santé : Dans le domaine de la santé, on rencontre de nombreux problèmes de rupture, principalement dans des situations de surveillance au cours du temps (études longitudinales).

Economie et finances : La littérature sur les changement structureux en économétrie est abondante. Evidemment, la détection d'une rupture en économie et en finances trouve de nombreuses applications : changement de tendance du marché, crash boursier, etc.

La méthodologie Bayésienne apporte une grande souplesse dans les méthodologies statistiques, elle a été largement sollicitée dans les problèmes de détection de rupture. Afin d'étudier l'influence d'un contaminant sur la performance de cette méthode d'estimation de rupture, nous avons considéré un modèle Gaussien avec un changement dans la moyenne à un instant inconnu. Nous avons déterminé la densité a posteriori conjointe de ce point en présence d'un outlier. Une étude de simulation a été conduite, elle reflète une bonne estimation du point de rupture, par le mode a posteriori en présence d'une valeur aberrante.

Il serait intéressant d'étendre ce travail dans le cas où les variables sont dépendantes, à des modèles de régression et au cas où plusieurs ruptures à des instants inconnus apparaissent. Aussi, peut être est il possible d'améliorer l'estimation en choisissant une loi a priori plus générales.

Bibliographie

- [1] Abdelli Z., *Problème de rupture dans les séries chronologiques*, Thèse de magister, usthb, (1993).
- [2] Abramowitz M. et Stegun I., *Handbook of Mathematical Functions*, Dover, New York, (1964).
- [3] Andrews, Donald W. K. ; Lee, Inpyo ; Ploberger, Werner. *Optimal changepoint tests for normal linear regression*, J. Econometrics, 70 , no. 1, 9-38.(1996).
- [4] Belkacem C. et Fellag H., *Bayesian change-point estimation in the presence of a single outlier*, Journal Afrika Statistika, Vol.7, 381-390, (2012).
- [5] Berkes I., Gombay E., Horvath L. et Kokoszka P.S., *Sequential change-point detection in GARCH (p,q) models*, Econometric Theory, 20, 1140-1167, (2004).
- [6] Bernardo J., *Reference posterior distributions for Bayesian inference (with discussion)* , J.Royal statist. Soc. Séries B,41, 113-147, (1979).
- [7] Blackwell D. and Girshick M., *Theory of Games and Statistical Decisions* , John Wiley, New york, (1954).
- [8] Broemeling L.D. , *The Bayesian analysis of linear model*, Marcel Dekker Inc Eds, New York,(1985).
- [9] Broemeling L.D. et Holbert D., *Bayesian inference related to shifting sequences and two-phases regression*, Commun.Stat.Theo.Meth. A6(3), 265-275, (1977).
- [10] Broemeling, L.D. and Tsurumi, H. *Econometrics and structural changes*, Marcel Dekker Inc Eds, New York (1986).
- [11] Brown R.L., *Notes of statistical decision theory of conditional confidence procedures* , Ann. statist.,6, 59-71, (1976).
- [12] Brown R.L., Durbin J. et Evans J.M., *Techniques for testing the constancy of regression relationships over time (with discussion)* , J.R. statist., Soc., B 37, 149-92, (1975).

- [13] Chernoff, H. et Zacks, S., *Estimating the current mean of a normal distribution which is subjected to changes in time*, *Annal of Mathematical Statistics*, 35, pp 999-1018.(1964)
- [14] Chen M., Shao Q., et Ibrahim J., *Monte Carlo methods in Bayesian computation*, Springer-Verlag, New York (2000).
- [15] Chin Choy J.H. et Broemeling L.D., *Some Bayesian inferences for a changing linear model*, *Thechnometrics*, 22,1, 71-80. (1980).
- [16] Chu C.S.J., *Detecting parameter shift in GARCH models*, *Econometric Reviews*, 14, 241-266, (1995).
- [17] Cobb G.W., *The problem of the Nile : conditional solution to a change-point problem*, *Biometrika.*,62, 243-51, (1978).
- [18] Csörgo M. et Horvath L., *Nonparametric methods for the change-point problem* , In *Handbook of statistics.*,7, P.R. Krish naiah and C.R. Rao (Eds), John Wiley, New York, 403-425, (1988).
- [19] Deshayes J. et Picard D., *Off-line statistical analysis of change-point models using non parametric and likelihood methods*, *Lecture notes in control and information sciences*, 77 M, Basseville et al. (Eds), Spinger-Verlag, New York, 103-168, (1986).
- [20] Ferreira P.E.. *Bayesian analysis of a switching regression model : known number of regimes*,*Amer. Journ. Stat.Association*, 70, pp 370-374.(1975) .
- [21] Fearnhead, P., *Exact and efficient Bayesian inference for multiple change point problems. Statistics and computing*,16, 203–213. (2006).
- [22] Fisher R., *Statistical Methods and Scientific Inference* , Oliver and Boyd, Edinburgh, (1956).
- [23] Fotopoulos, S.B, Jandhyala,V.K., Khapalova, E. *Exact asymptotic distribution of change-point model for change in the mean of Gaussian sequences.. Annals of Applied Statistics*, 4, No. 2, 1081-1104. (2010).
- [24] Gauss C., *Méthodes des moindres carrés* , Mémoire sur la combinaison des observations , Mallet-Bachelier, Paris. transl. J. Bertrand, (1810).
- [25] Geman S. et Geman D., *Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images*, *IEEE trans. Pattern Anal. Mach, Intell.*, 6,721-741, (1984).
- [26] Hasting W., *Monte Carlo simpling methods using Markov chains and their application* , *Biometrika*, 57, 97-109, (1970).

- [27] Hills S. et Smith A., *Parametrization issues in Bayesian inference*. In Bernardo J., Berger J., Dawid A., et Smith A., éditeurs, *Bayesian Statistics 4*, pages 641–649. Oxford University Press, Oxford (1992).
- [28] Hsu D.A., *Test for variance shift at an unknown time point*, Journal of american statistical association, Ser.C, 26,PP, 279-284, (1977).
- [29] Hsu D.A., *Detecting shifts of parameter in gamma sequences with applications to stock price and air traffic flow analysis*, Journal of american statistical association, 74,PP, 31-40, (1979(a)).
- [30] Hsu D.A., *Test for variance shift at an unknown time point*, Journal of american statistical association, 77,PP, 29-39, (1982).
- [31] Holbert D., *A Bayesian analysis of switching linear model*, Journal of econometrics, 19,PP, 77-87, (1982).
- [32] Hwang J.et Pemantle R., *Confidence sets recentered at janes- stein estimators- a surprise concerning the unknown variance case* , Econometrics, 60(1-2), 145-156, (1994).
- [33] Ilmakunnas P. et Tsurami H., *Testing for parameter shifts in a regression model with two regimes of autocorrelated errors*, Economic Studies quarterley, vol.37,PP, 46-56, (1984).
- [34] kander, Z. and zachs, S. *Test procedures for possible changes in parameters of statistical distribution occuring at unknown time points*,. Ann. Math. Stat., 37, pp 1196-1210(1966).
- [35] Kass R. et Steffey D., *Approximate Bayesian inference in conditionally independent hierarchical models (parametric empirical Bayes models)*, J. American Statist. Assoc., 87,717-726, (1989).
- [36] Kokoszka P.S. et Leipus R., *Change-point estimation in ARCH models*, Bernoulli,6, 513-539, (2000).
- [37] Kokoszka P.S. et Leipus R., *Testing for parameter changes in ARCH models*, Lithuanian Mathematical Journal, 39, 231-247, (1999).
- [38] Koopman B., *On distributions admitting a sufficient statistic*, Trans. Amer. Math. Soc., 39,399-409, (1936).
- [39] Le cam L., *Asymptotic Methods in Statistical Decision Theory* , Springer- Verlag, New york, (1986).

- [40] Legendre A., *Nouvelles méthodes pour la détermination des orbites des comètes* , Courcier,Paris, (1805).
- [41] Lindley D., *Marking Decision* , John Wiley, New york, (1985).
- [42] Liséo B., *Elimination of nuisance parameters with reference priors* , *Biometrika*, 80(2), 295-304, (1993).
- [43] Menzefricke,U D., *A Bayesian Analysis of a change in the precision of a sequence of independent normal random variable at an unknown time point*, *Applied Statistics*, 30, pp 141-146.(1981).
- [44] Metropolis N. et Ulam S., *The Monte Carlo method*. *J. American Statist. Assoc.*, 44, (335–341), (1949).
- [45] Metropolis N., Rosenbluth A., Rosembluth M.,Teller A. et Taller E., *Equations of state calculations by fast computing machines* . *J. Chem. Phys.*, 21, (1087–1092), (1953).
- [46] Moustakides G., *Optimal procedures for detecting changes in distributions*. *Ann. Statist.*, 14, (1379–1387), (1986).
- [47] Naylor J. et Smith A., *Application of a method for the efficient computation of posterior distributions*, *Applied Statistics*, 31, 214–225, (1982).
- [48] Page, E.S. *Continuous inspection shemes*,*Biometrika*, 41, pp 100-115.(1954).
- [49] Page, E.S. *A test for a change in parameter occuring at unknown point*,*Biometrika*, 42, pp 523-526.(1955).
- [50] Page, E.S. *On problems in which a change in a parameter occurs at an unknown time point*,*Biometrika*, 44, pp 523-527.(1957).
- [51] Pandya M., Bhatt K. et Pandya H. *Bayesian estimation of change-point autoregressive process*, *IJRRAS.*, Vol 13(1), 41-52, (2012).
- [52] Parent E., Bernier J., *Le raisonnement Bayésien - Modélisation et inférence*, Springer, (2007).
- [53] Pettitt A.N., *A simple cumulative sum type statistic for the change-point problem with zero-one observations*, *Biometrika.*, 67, 79-84, (1980).
- [54] Pitman E., *Sufficient statistics and intrinsic accuracy* , *Proc. Cambridge Philos. Soc.*, 32,567-579, (1936).
- [55] Quandt R.E., *Tests of the hypothesis that a linear regression system obeys two separate regimes*, *J.A.S. Association*, vol.55, PP.523-527, (1960).

-
- [56] Raiffa H., Schlaifer R., *Applied statistical decision theory* , Technical report, Division of Research, Graduate School of Business Administration, Havard Univ, (1961).
- [57] Rigaiil , G, Lebarbier, E. and Robin, S., *Exact posterior distributions and model selection criteria for multiple change-point detection problems*. Statistics and computing. (2012).
- [58] Robert C.P., *L'Analyse Statistique Bayésienne*, Economica, (1992).
- [59] Robert C.P. et Casella G., *Monte Carlo Statistical Methods*, Springer-Verlag, New York, seconde édition, (2004).
- [60] Robert C.P., *Le Choix Bayésien - Principes et pratique*, Springer, (2006).
- [61] Salazar D., Broemeling L.D. et Chi A., *Parameter changes in regression model with autocorrelated errors*, Commun.Stat.Theo.Meth, A10 (17), 1751-1758, (1981).
- [62] Sen A. et Srivastava M.S., *On tests for detecting changes in means*, Ann. statist.,3, 98-103, (1975).
- [63] Severeni T., *On the relationship between Bayesian and non-Bayesian interval estimates* , J.Royal statist. Soc. Séries B,53, 611-618, (1991).
- [64] Strasser H., *Mathematical Theory of Statistics*,W. de Grayter, Berlin, (1985).
- [65] Tierney L., Kass R. et Kadane J., *Fully exponential Laplace approximations to expectations and variances of non-positive functions*, J. American statist. Assoc., 84,710-716, (1989).
- [66] Tsay R., *Outlier, level shifts and variance changes in time series*, J. Forecasting, 7, 1-20, (1988).
- [67] Tsurumi H., *A Bayesian test of a parameter shift and an application*, Econometrics, 6, 371-380, (1977).
- [68] Von Neumann J., *Various techniques used in connection with random digits*, J. Resources of the National Bureau of Standards-Applied Mathematics Series, 12,36-38, (1951).
- [69] Wald A. et Wolfowitz J., *Optimum character of the sequential probability ratio test*, Ann.Math.Statist., 19, (326–339), (1948).