

N° d'ordre:

RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE
MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE SCIENTIFIQUE
UNIVERSITÉ MOULOUD MAMMERRI DE TIZI OUZOU
FACULTÉ DES SCIENCES
DÉPARTEMENT DE MATHÉMATIQUES
LABORATOIRE LMPA



MÉMOIRE DE MASTER

Filière : Mathématiques
Spécialité : Probabilités & Statistique

Par

RAHMANI LOUIZ

APPROCHE BAYÉSIENNE DANS LES MODÈLES DE RUPTURE

Soutenue le Octobre 2022 devant le jury :

Dr. ATIL LYNDA	UMMTO	Présidente du jury
Dr. MERABET DALILA	UMMTO	Examinatrice
Dr. BELKACEM CHERIFA	UMMTO	Encadreur

Année Universitaire : 2021/2022

Dédicace

Je dédie ce travail :

À mon cher père MEZIANE , décédé trop tôt ,qui m'a toujours poussé et motivé dans mes études. Je ne saurais exprimer mon grand chagrin en ton absence .J'aurais aimé que tu sois a mes cotés ce jour. que ce travail sois une prière pour le repos de ton âme .

À ma chère mère TAOUES pour tous les sacrifices qu'elle a consentis pour me permettre de suivre mes études dans les meilleurs condition possible .Que Dieu lui donne une très longue vie et une sante de fer.

À mon cher oncle RAMDAN et sa femme OUIZA.

À mon grand frère SALIM et sa femme AREZIKA pour leur encouragement et leur soutiens .

À mes chers frères SIHMED , SOFIANE , AMINE , GAYA .

À mes chères sœurs CHAHIRA , MELISSA et son mari AZZDINE .

À mon cher oncle YAHIA , sa femme NORA et leurs enfants .

À mes chères tantes CHABHA , NADIA et leur famille .

À mon on oncle MOUH et sa femme FATIHA .

À mes chers grand parents paternels et maternels, que Dieu garde leurs âmes dans son vaste paradis .

À tout mes amis .

À tout les étudiants de la promotion PROBA-STATS année 2021 \2022 .

REMERCIEMENTS

JE remercie dieu le tout puissant de m'avoir donnée suffisamment la santé , la volonté et le courage afin de pouvoir accomplir ce modeste travail .

Tout d'abord , ce travail ne serait pas aussi riche et n'aurait pas pu avoir le jour sans l'aide et l'encadrement de M^{me} BELKACEM CHERIFA , on la remercie pour ces précieux conseils, sa rigueur et sa disponibilité durant notre préparation de ce mémoire.

Je remercie également les membres du jury, M^{me} ATIL LYNDA et M^{me} MERABET DALILA pour avoir accepté d'évaluer ce travail.

Je remercie sincèrement tous mes enseignants de Proba-Stat.

Enfin, je tien à remercier tous ceux qui m'ont apporter, de près ou de loin de l'aide et des conseils pour la réalisation de ce travail.

Tizi-Ouzou, le 29 octobre 2022.

TABLE DES MATIÈRES

TABLE DES MATIÈRES	iv
LISTE DES FIGURES	vi
LISTE DES TABLEAUX	vii
INTRODUCTION	1
1 LES MODÈLES DE RUPTURE	3
1.1 INTRODUCTION	3
1.2 DÉFINITION : POINT DE RUPTURE	3
1.3 DIFFÉRENTS TYPES DE RUPTURE	6
1.3.1 Rupture unique	6
1.3.2 Rupture multiple	6
1.3.3 Rupture épidémique	7
1.4 EXEMPLES DE RUPTURE DANS DIFFÉRENTS DOMAINES D'APPLICATION	8
1.4.1 Anomalie climatique annuelle	9
1.4.2 COVID-19 en France	11
2 INFÉRENCE DANS LES MODÈLES DE RUPTURE	19
2.1 INTRODUCTION	19
2.2 APPROCHE CLASSIQUE DE DÉTECTION DE RUPTURE	19
2.2.1 Tests de rupture non paramétriques	19
2.2.2 Tests de rupture paramétriques	24
2.2.3 Estimation d'un point de rupture	28
2.3 APPROCHE BAYÉSIENNE	30
2.3.1 Modèle gaussien	31
2.3.2 Modèle de régression :	33
2.4 AUTRES MÉTHODES DE DÉTECTION DE RUPTURE	37
2.4.1 La détection de rupture par procédure de contraste pénalisé .	37
2.4.2 La règle CUSUM	37
2.4.3 La règle de Shewhart	38
2.4.4 Détection hors ligne et en ligne de rupture	38
3 ESTIMATION BAYÉSIENNE D'UN POINT DE RUPTURE DANS UNE DISTRIBUTION EXPONENTIELLE	39
3.1 INTRODUCTION	39
3.2 PRÉSENTATION DU MODÈLE	39
3.3 ESTIMATION BAYÉSIENNE	40
3.3.1 La fonction de la vraisemblance	40
3.3.2 Les lois a priori	40
3.3.3 La densité a posteriori jointe :	41
3.3.4 Les densités a posteriori marginales	43

3.4	ESTIMATION BAYÉSIENNE DU POINT DE RUPTURE SOUS LES FONCTIONS COÛTS USUELLES.	44
3.4.1	La fonction coût quadratique	44
3.4.2	La fonction coût absolu	45
3.4.3	La fonction coût 0-1	45
3.5	ESTIMATION BAYÉSIENNE SOUS LA FONCTION COÛT LINEX	45
3.6	APPLICATION ET SIMULATION	46
3.6.1	Estimateur Bayésien sous les fonctions coûts quadratique	47
3.6.2	Estimateur Bayésien sous la fonction coût absolu	48
3.7	CONCLUSION	48
	CONCLUSION GÉNÉRALE	49
	BIBLIOGRAPHIE	50

LISTE DES FIGURES

1.1	Exemples de rupture	4
1.2	ARFIMA(τ) = (80)	5
1.3	ARFIMA(τ_1, τ_2) = (150, 275)	5
1.4	ARFIMA($\tau_1, \tau_2, \tau_3, \tau_4, \tau_5$) = (150, 400, 550, 700, 850)	6
1.5	La figure représente une rupture dans une loi normale.	6
1.6	La figure représente une séries ayant subi plusieurs ruptures par rapport au seuil adaptatif.	7
1.7	La figure indique une rupture épidémique.	7
1.8	Anomalies climatique mondiales de 1880 à 2015.	11
1.9	Taux de mortalité du COVID-19 en France du 15/02/2020 au 10/07/2020.	14
1.10	Nombre de décès quotidiens du COVID-19 en France du 27/02/2020 au 10/07/2020	16
1.11	Nouveaux cas quotidiens du COVID-19 en France du 25/02/2020 au 10/07/2020	18
2.1	Algorithme explicatif de test de pettitt.	21

LISTE DES TABLEAUX

2.1	Valeurs critiques de q_1 et $q_{1,T}/T^5$	22
2.2	Valeurs critique de q_1^* et $\frac{q_{1,T}^*}{T^4}$	23
2.3	Valeurs critiques de $q_{1,T}^0$ et $q_{1,T}^0/T_2$	24
2.4	Valeurs critiques de $J(T)$ et $J_1(T)$	25
2.5	valeurs critique avec $\alpha = 0,05$ des tests LR et XLW	28
3.1	Les observations générées a partir du modèle (3,1)	46
3.2	Les valeurs simulées de la distribution a posteriori marginale de m.	47
3.3	La valeur de l'estimateur m^*	47
3.4	La distribution a posteriori cumulée de m	48

INTRODUCTION GÉNÉRALE

Beaucoup de données sont issues des phénomènes susceptibles qui varient avec le temps. C'est par exemple le cas des données financières, hydrologiques, climatiques, économiques, agronomiques....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.

Les changements discrets dans un phénomène temporel peuvent se manifester par la présence de faibles ruptures dans les données recueillies. Afin de comprendre ce phénomène, le contrôler ou le prévoir, il peut s'avérer très important de détecter au préalable ces ruptures avant toute inférence statistique. La majeure partie de temps, la moyenne des données avant et après la rupture ne seront pas les mêmes.

Les premiers résultats sur la détection de rupture datent de Page (1954,1955), lorsqu'il utilisait les sommes partielles pour tester les changements dans la moyenne des observations indépendantes. Depuis lors, les recherches sur ce sujet ont considérablement avancé et ont été étendues aux séries chronologiques. Une liste non-exhaustive d'articles sur ce sujet comprend Härdle et Tsybakov (1997), Härdle et al. (1998), Bardet et al. (2009), Döring (2010, 2011), Ciuperca (2011), Bardet et Kengne (2014), Amano (2012), Yang et Song (2014), Kengne (2012), Chen et al. (2011), Dehling et al. (2013), Dehling et al. (2017), Wang et al. (2012), Francq et Zakoïan (2012), Bardet et al. (2012), Zhou et al. (2014), Fotopoulos et al. (2009), Huh (2010), Enikeeva et al. (2018), Gombay et Serban (2009). On trouve aussi des méthodes basées sur l'optimisation comme Lavielle (1998), Hawkins (2001) et Menne et Williams Jr (2005).

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 évidence un changement dans la moyenne. Broemeling (1985) et Broemeling et Tsurumi (1986) se sont intéressés à une analyse Bayésienne de quelques modèles linéaires.

Ce document comporte trois chapitres :

Le premier Chapitre comprend quelques généralités sur les modèles de rupture, nous faisons le point sur les différents types de rupture et nous abordons dans une partie deux exemples de rupture dans différents domaines d'application.

Dans le deuxième Chapitre nous nous intéressons ou différents méthodes de détection de rupture (tests d'hypothèse, estimation, approche bayésienne) nous rappelons aussi quelques résultats existantes dans ce domaine .

Dans le troisième Chapitre nous nous intéressons a l'estimation d'un point de rupture dans le paramètre d'une distribution exponentielle. Nous déterminons l'estimateur de ce point par l'approche bayésienne sur différentes fonctions coût .

En fin nous présentons des résultats numériques obtenus a partir de donnée simulés.

Nous terminons par une conclusion générale et quelques perspectives.

LES MODÈLES DE RUPTURE



1.1 INTRODUCTION

La fréquence des changements de situation dans plusieurs domaines scientifiques explique l'intérêt porté à l'analyse statistique des points de rupture et de l'estimation. En effet, dans la pratique, lorsque les structures de contrôle révèlent qu'il y'a des ruptures quelque part? dans l'évolution du phénomène étudié, il est naturel que l'on veuille localiser la position de ces ruptures qui est à l'origine du changement de régime, Du point de vue statistique, une rupture est un lieu ou un temps de sorte que les observations suivent une distribution jusqu'à ce point et suivent une autre distribution après celui-ci.

Une rupture d'un modèle paramétrique est définie par un changement de l'une ou plusieurs de ces caractéristiques. Il peut s'agir par exemple de la moyenne, de la variance, du moment d'ordre r , de la famille de la distribution... Plusieurs problèmes peuvent être définis similairement.

1.2 DÉFINITION : POINT DE RUPTURE

Définition 1.1 *Nous dirons qu'il y'a rupture dans une suite de variables aléatoires $X_1, X_2, X_3, \dots, X_n$ s'il existe un entier $m(1 \leq m \leq n - 1)$; tel que les variables aléatoires $X_1, X_2, X_3, \dots, X_m$ 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.

Définition 1.2 *un point de rupture est un instant donné dans une séquence d'observations où la distribution de probabilité des observations change .*

La notion de detection de rupture est la question liée à l'estimation de l'emplacement de ce point de rupture dans une série d'observations .

Nous considérons un phénomène aléatoire que nous observons à chaque instant à un moment donné, nous remarquonsqu'il change de comportement.

Mathématiquement , nous avons une série d'observations (séquence de données) :

$$\{x_t\}_{t=1, \dots, n} = \{ x_1, \dots, x_n \},$$

issus d'un échantillon de variable aléatoire X_1, \dots, X_n . n est le nombre d'observations et i est l'instant d'arrivée de l'observation x_t . On admet qu'au début, dans l'état "normal" (avant la rupture), l'échantillon X_1, \dots, X_n est identiquement distribué (pas nécessairement independant) selon la fonction de densité $f_0(\cdot)$. A un instant donné $1 < v \leq n$, supposé inconnu, une rupture se produit et la distribution des X_i change pour satisfaire la distribution de post-changement $f_1(\cdot) \neq f_0(\cdot)$.

Dans ce cas, l'échantillon X_1, X_2, \dots, X_{v-1} est identiquement distribué selon la densité de pré-changement $f_0(\cdot)$, et l'échantillon X_v, X_{v+1}, \dots, X_n est identiquement distribué selon une densité de post-changement $f_1(\cdot)$.

Les changements qui se produisent dans une séquence d'observations peuvent être divers et différents comme l'illustre la Figure 1.1.

Néanmoins, en général, ils sont principalement liés à deux caractéristiques statistiques, qui sont la moyenne et/ou la variance de la séquence. C'est-à-dire, une rupture existe si la moyenne et/ou la variance des deux segments $\{X_1, \dots, X_{v-1}\}$ et $\{X_v, \dots, X_n\}$ sont différentes.

Dans le cas d'existence de multiples ruptures, nous supposons qu'il existe un certain nombre m de points de rupture, aux instants $\{v_1, \dots, v_m\}$.

Chaque instant est un entier compris entre 2 et n . Les points de ruptures sont ordonnés de telle sorte que $v_i < v_j$ si $i < j$. Dans ce cas, la densité de distribution change à chaque instant de rupture $\{v_i\}_{i=1, \dots, m}$.

Il y a alors $(m + 1)$ "régimes" de distribution correspondant aux intervalles $[1, v_1[, [v_1, v_2[, \dots, [v_m, n[$ et aux paramètres μ_{j-1} ou σ_{j-1}^2 qui sont respectivement les moyennes et les variances des observations du $j^{\text{ème}}$ segment $]v_{j-1}, v_j[$, avec $j = 1, \dots, (m + 1)$.

La figure 1.1 montre dans (a) un exemple de rupture dans la moyenne à l'instant $v = 100$, dans (b) un exemple de rupture dans la variance et dans (c) il montre un exemple de rupture sur la moyenne et sur la variance. Les exemples (a), (b) et (c) représentent le cas où il existe deux états (régimes) :

état normal (régime pré-changement) et état anormal (régime post-changement). Et la rupture est l'instant de passage de l'état normal à l'état anormal. Le cas (d) est particulier car il montre un exemple de régression qui ne rentre pas dans le cadre précédent il montre un exemple de rupture de pente où le régime post-changement n'est pas stable et ne représente pas un seul état mais plusieurs, avec un passage progressif entre les états.

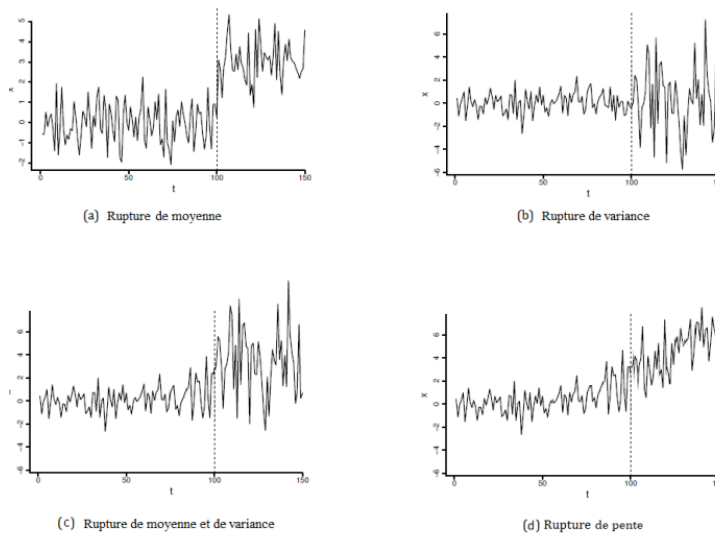


FIGURE 1.1 – Exemples de rupture

Définition 1.3 Dans un processus aléatoire $(X_t)_{t \in \mathbb{Z}}$, de paramètre d , $d \in \mathbb{R}$, une rupture $k \in \mathbb{N}$ se présente par :

$$(X_1; \dots; X_k)^{d_1} \text{ et } (X_{k+1}; \dots; X_n)^{d_2}$$

avec $1 < k < n$ et $d_1 \neq d_2$

Ceci veut dire que le processus $(X_1; \dots; X_k)$ dépend d'un paramètre d_1 par contre le deuxième processus $(X_{k+1}; \dots; X_n)$ dépend d'un autre paramètre d_2 , à l'instant k il se produit un changement dans la valeur de paramètre d , cet instant k représente une rupture. Ainsi, un processus aléatoire ayant plusieurs ruptures $\{k, \ell, \dots\}$ se présente par l'écriture,

$$(X_1; \dots; X_k)^{d_k}; (X_{k+1}; \dots; X_\ell)^{d_\ell}; \dots; (X_{\ell+1}; \dots; X_n)^{d_n};$$

avec $1 < k < \ell < n$ et $d_k \neq \dots \neq d_n$

les figures suivantes montrent un nombre différent de rupture $(\tau_i)_{i > 1}$ dans des processus aléatoires (les lignes rouges).

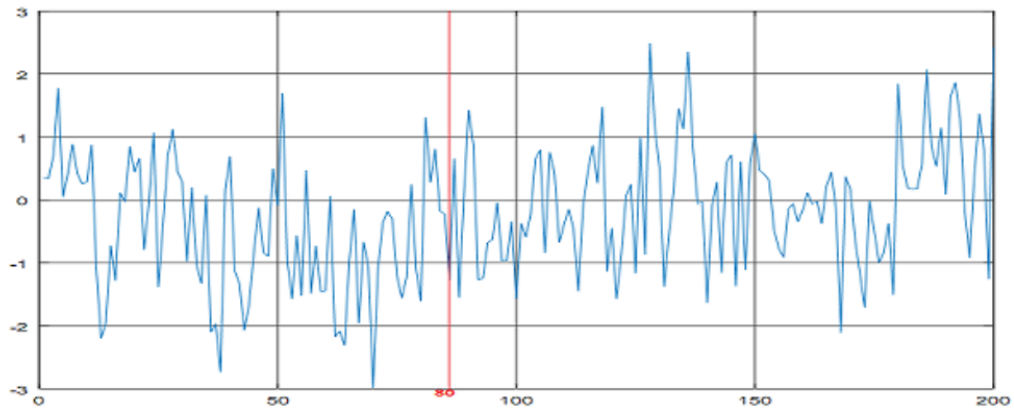


FIGURE 1.2 - $ARFIMA(\tau) = (80)$

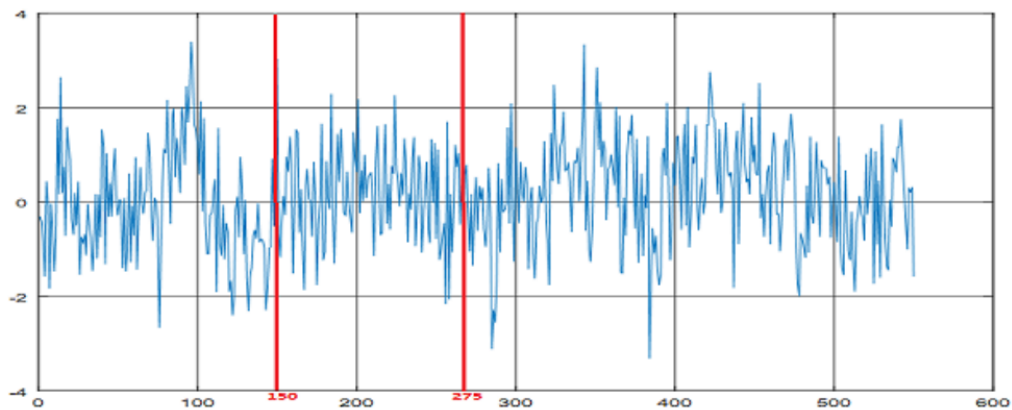


FIGURE 1.3 - $ARFIMA_{\tau_1, \tau_2} = (150, 275)$

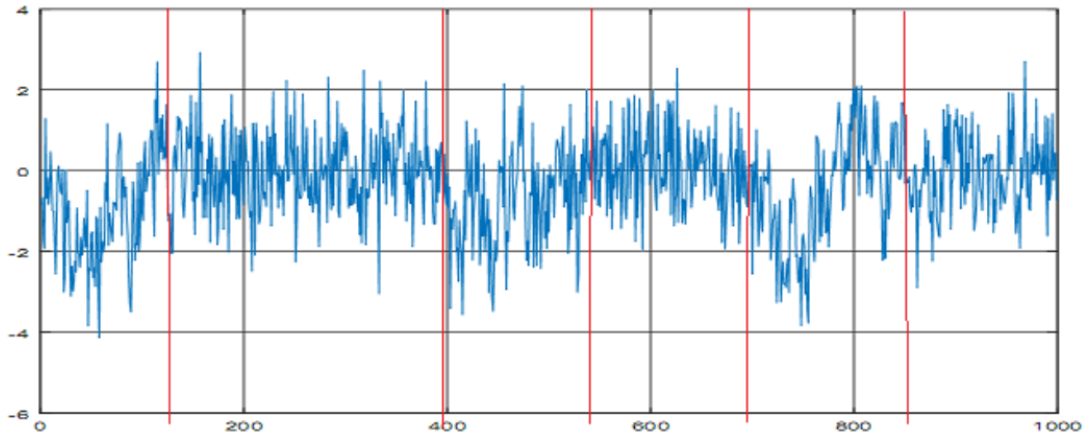


FIGURE 1.4 – $ARFIMA(\tau_1, \tau_2, \tau_3, \tau_4, \tau_5) = (150, 400, 550, 700, 850)$

ARFIMA :AutoRegressive Fractionally Integrated Moving Average (Moyenne mobile autorégressive à intégration fractionné)

1.3 DIFFÉRENTS TYPES DE RUPTURE

1.3.1 Rupture unique

Nous dirons qu'il y a rupture dans une suite de variables aléatoires X_1, \dots, X_n de fonction de répartition $F_i(x|\theta_i), i = \overline{1, n}$, s'il existe un entier $m(1 \leq m \leq n - 1)$ telque :

$$X_i = \begin{cases} F_1(x|\theta_1), & i = 1, \dots, m \\ F_2(x|\theta_2), & i = m + 1, \dots, n \end{cases}$$

Avec θ_1 et θ_2 sont des paramètres inconnus réels ou vectoriels ($\theta_1 \neq \theta_2$).

Exemple :

La figure suivante représente le cas d'une loi normale avec rupture unique dans la moyenne.

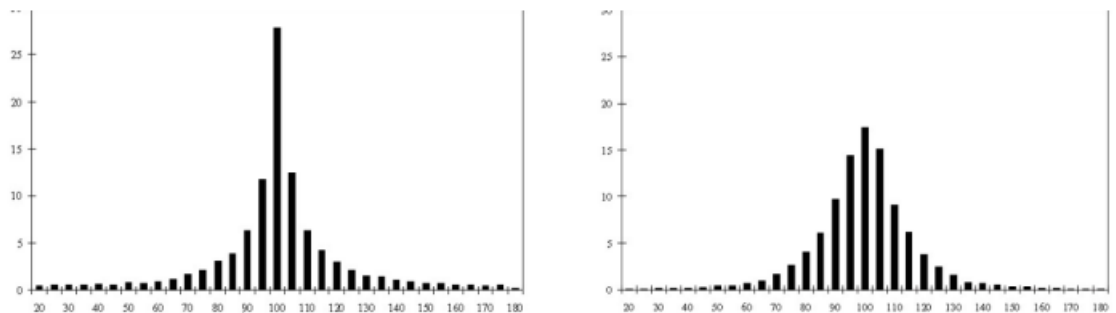


FIGURE 1.5 – La figure représente une rupture dans une loi normale.

1.3.2 Rupture multiple

Nous dirons qu'une suite de variables aléatoires admet des ruptures multiples s'il existe des entiers m, \dots, q , avec $(1 \leq m \leq \dots \leq q \leq n - 1)$ telque :

$$\theta_1 = \dots = \theta_m \neq \theta_{m+1} = \dots = \theta_q \neq \theta_{q+1} = \dots = \theta_n$$

Exemple :

La figure suivante représente une séries ayant subi plusieurs ruptures par rapport au seuil adaptatif.

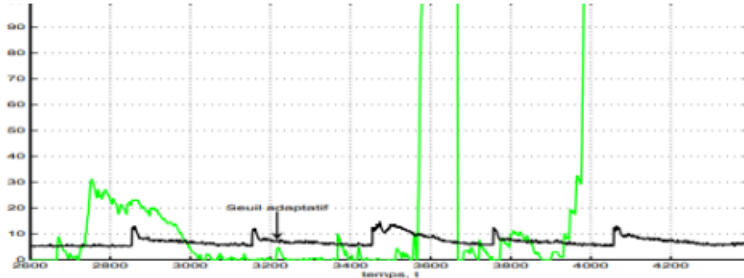


FIGURE 1.6 – La figure représente une séries ayant subi plusieurs ruptures par rapport au seuil adaptatif.

1.3.3 Rupture épidémique

Soit X_1, \dots, X_n des variables aléatoires réelles de répartition continues respective F_1, \dots, F_n .

Considérons le modèle épidémique suivant :

Il existe des entiers m et q ($1 \leq m \leq q \leq n$) tels que

$$F_1 = F_2 = \dots = F_m = F_{q+1} = \dots = F_n = F$$

$$F_{m+1} = \dots = F_q = G \neq F$$

Nous notons $l = q - m$ la longueur de l'épidémie.

Et m, q les bornes de localisation.

Parmi les chercheurs qui se sont intéressés au problèmes de rupture épidémique, citons Rackauskas et Suquet .

Exemple :

la figure suivante indique une rupture épidémique.

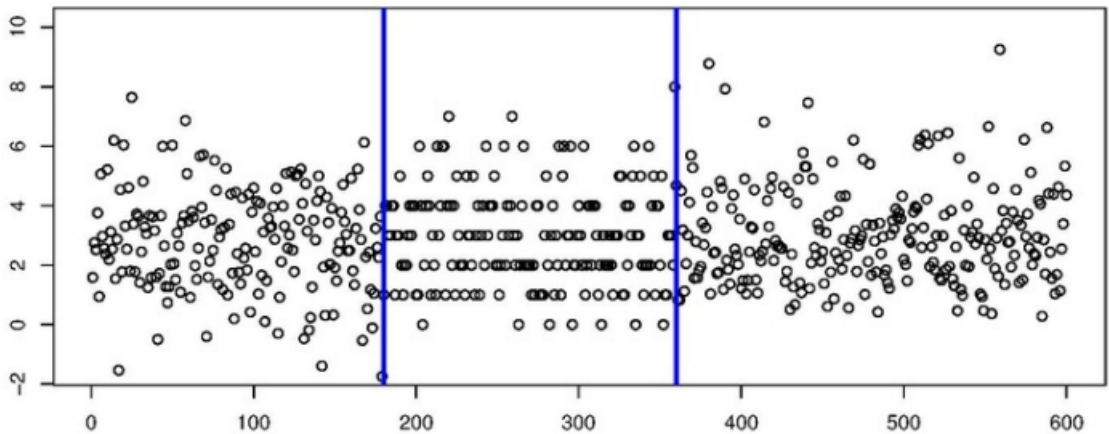


FIGURE 1.7 – La figure indique une rupture épidémique.

1.4 EXEMPLES DE RUPTURE DANS DIFFERENTS DOMAINES D'APPLICATION

Dans cette section, nous présentons deux exemples de rupture de données réelles.

Le premier jeu de données est relatif aux anomalies climatiques. ces données sont issues de la base de données de la NOAA. Les autres jeux de données sont relatifs à la première vague du COVID-19 en France. Ces données concernent le taux journalier de décès, le nombre journalier de décès et le nombre journalier de cas. Elles proviennent des bases de données de la Worldometers.

A fin d'appliquer les méthodes de détection de rupture utilisées par (Ngatchou-Wandjia et Marwa Ltaifa). Les deux séries sont modéliser comme suit :

Modélisation

Le chronogramme de chacune des séries brutes (Y_t) considérées dans cette étude semble présenter une tendance et aucune saisonnalité (voir Figures 1.8 (a), 1.9 (a), 1.10 (a) et 1.11 (a)). Les résultats établis dans Ngatchou-Wandji et Ltaifa (2021) ne peuvent s'appliquer directement à ces séries. Nous les décomposons alors en la somme de deux composantes comme suit :

$$Y_t = Z_t + X_t$$

où X_t est une série stationnaire par morceaux de moyenne μ et de variance σ sur chaque "morceau" et Z_t représente la composante tendancielle supposée continue. Il s'agit pour nous de considérer la série corrigée de la tendance, c'est-à-dire la série résiduelle qui exhibera les mêmes points de ruptures que la série brute. Il y a plusieurs méthodes d'estimation de la tendance d'une série chronologique. Nous utilisons ici la méthode des moyennes mobiles, en particulier les moyennes mobiles arithmétiques d'ordre 5. Car ce sont elles qui nous ont semblé produire les meilleures estimations. Nous estimons donc Z_t par :

$$\hat{Z}_t = \frac{1}{5} \sum_{j=-2}^2 Y_{t+j}$$

Cette estimation est obtenue en utilisant la commande "ma" du logiciel R. Ensuite, si on suppose qu'on a potentiels points de rupture qu'on note t_1, t_2, \dots, t_ℓ , tels que $t_1 < t_2 < \dots < t_\ell$, sur chaque intervalle $[t_{i-1}, t_i)$, où la série résiduelle semble stationnaire, nous ajustons à cette série un modèle de la forme $X_t = \mu + (\beta_i / \sqrt{n}) + \sigma_i \epsilon_t$, $i \in \{1, \dots, K + 1\}$, k le nombre des ruptures, n le nombre des données, $\mu + (\beta_i / \sqrt{n})$ et σ_i sont respectivement la moyenne et la variance de X_t sur l'intervalle $[t_{i-1}, t_i)$, $\beta_1 = 0$ et pour tout $i = 2, \dots, k + 1$, $\beta_i \in \mathbb{R}$, et ϵ_t est un bruit blanc gaussien centré réduit.

Cela correspond à notre problème de test pour $T(x) = \mu$, $V(x) = \sigma_i$ sur chaque intervalle $[t_{i-1}, t_i)$, $\gamma_0 = (\mu, \mu, \dots, \mu)^\top$, $\gamma_n = (\mu, \mu + (\beta_2 / \sqrt{n}), \mu + (\beta_3 / \sqrt{n}), \dots, \mu + (\beta_{k+1} / \sqrt{n}))$ et $\beta = (0, \beta_2, \beta_3, \dots, \beta_{k+1})$.

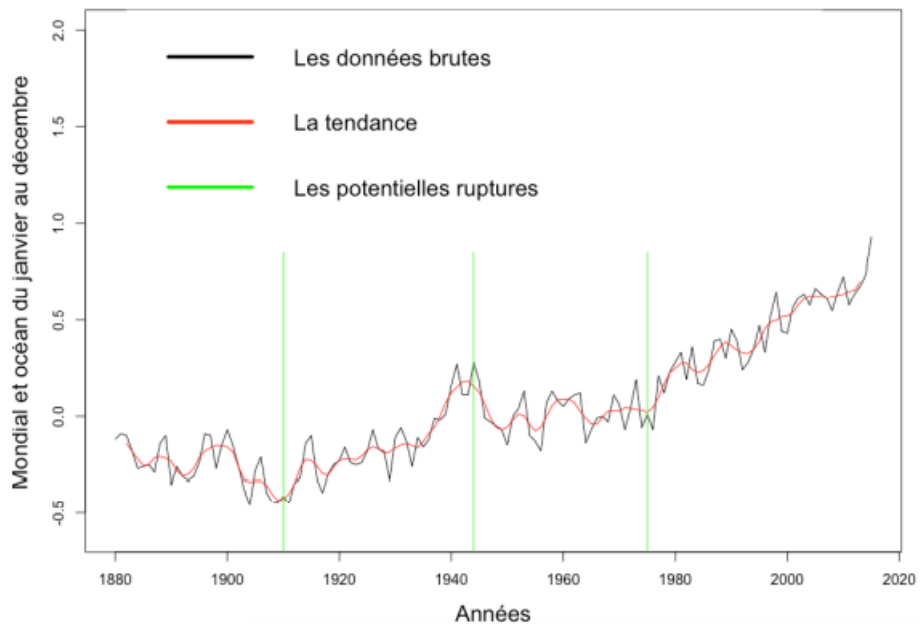
Nous commençons dans le paragraphe ci-dessous par l'étude des ruptures dans les données sur les anomalies climatiques annuelles.

1.4.1 Anomalie climatique annuelle

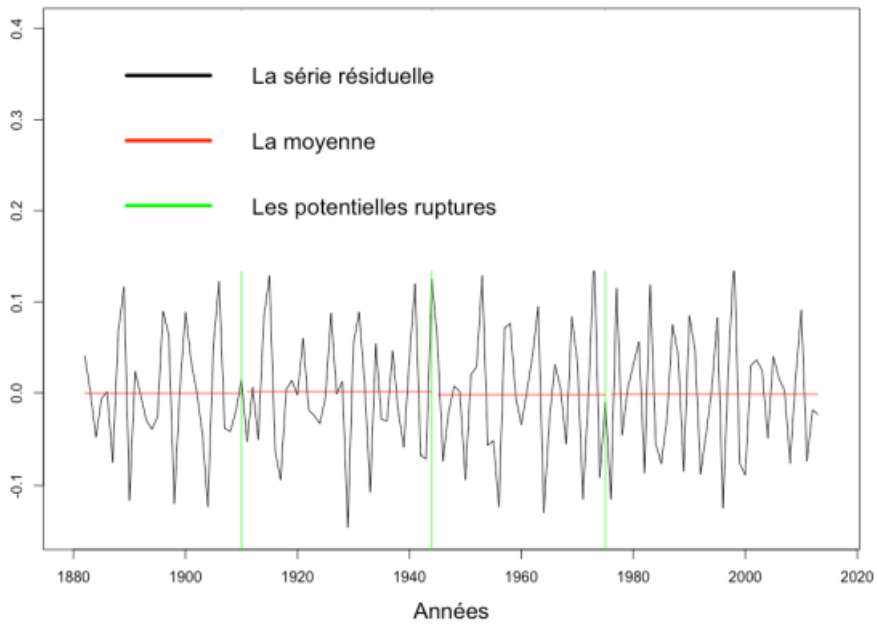
nous prenons les données sur les anomalies climatiques annuelles, de la période de 1880 à 2015. la figure 1.8 (a) présente le chronogramme des données brutes et semble exhiber 3 ruptures correspondants au années 1910, 1944 et 1975. Nous désignons par points de ruptures ici les années qui semblent indiquer un changement (une brusque augmentation ou une diminution) dans la moyenne des données par rapport aux années précédentes. Nous voulons savoir si les années potentielles ruptures présentent des point de ruptures ou si ces points sont dans leur voisinage. Nous commençons par estimer la tendance par la méthode de moyenne mobile. Cette estimation est représentée en rouge dans la figure 1.8 (a). Puis, nous prenons la série résiduelle représentée dans la figure 1.8 (b). En utilisant le test de normalité Shapiro-Wilk sur la série résiduelle et sur chaque morceau par l'utilisation de la commande Shapiro.test sous R, nous retenons l'hypothèse de normalité. Nous appliquant les tests de Box-Pierce et Box-Ljung en utilisant la commande de Box.test sous R sur ces données.

Nous trouvons p-valeur = 0,03858. Donc pour un seuil critique $\alpha = 0.05$ nous rejetons l'hypothèse iid des résidus. D'où l'hypothèse d'hétéroscédasticité. Ici nous prenons $k = 3$. $t_1 = 1910, t_2 = 1944, t_3 = 1975$, $\mu + (\beta_i / \sqrt{n})$ et σ_i sont respectivement la moyenne et la variance des données résiduelles sur l'intervalle $[t_{i-1}, t_i)$, pour $i = 1, 2, 3, 4$ où $t_0 = 1880$ et $t_4 = 2015$, $\beta_1 = 0$ et pour tout $i = 2, \dots, k + 1, \beta_i \in \mathbb{R}$ et $(\varepsilon_t)_t$ est un bruit blanc gaussien centré réduit.

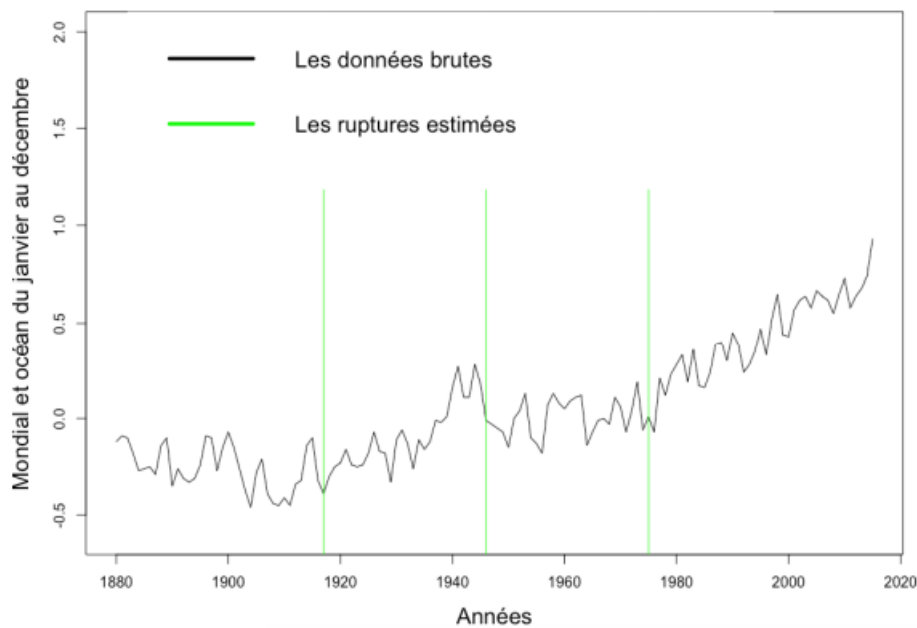
Ensuite, nous calculons la puissance de test autour de chaque points t_i Et nous prenons les années donnant la puissance maximale. Nous obtenons les points suivants : $\hat{t}_1 = 1915, \hat{t}_2 = 1944, \hat{t}_3 = 1975$ représentés dans la figure 1.8 (c). Ces points sont ceux considérés hormis le point \hat{t}_1 qui est proche de t_1 considéré. Ces points semblent coller à la réalité. En effet, dans le premier intervalle la tendance globale est à la baisse. Elle est à la hausse dans l'intervalle suivant, puis constante dans le suivant et à la hausse dans le dernier.



(a) Série brute



(b) Série résiduelle



(c) Dates de rupture estimées

FIGURE 1.8 – Anomalies climatiques mondiales de 1880 à 2015.

1.4.2 COVID-19 en France

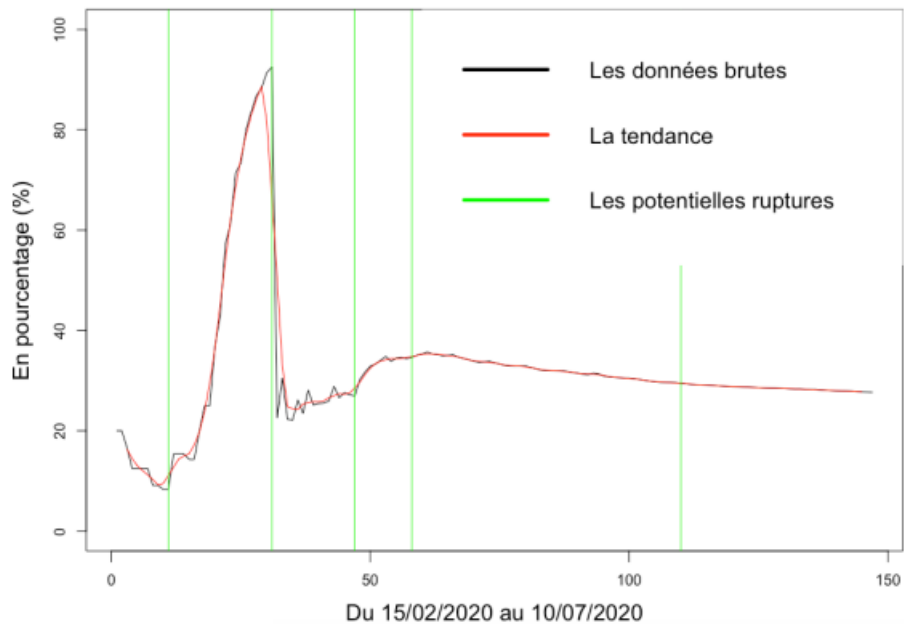
Dans ce paragraphe nous étudions trois séries de données journalières relatives à la première vague du Covid-19 en France en 2020. Nous désignons par les numéros $1, 2, \dots, n$ les jours des données dans l'ordre. C'est-à-dire que : 1 correspond au premier jour, 2 au deuxième jour, ..., n au dernier jour. Dans la première, la deuxième et la troisième applications, 1 correspond respectivement aux 15/02/2020, 27/02/2020 et 25/02/2020 et n correspond au 10/07/2020. Nous commençons d'abord par donner quelques éléments de la chronologie de cette première vague qui peuvent aider à l'interprétation des ruptures dans les séries étudiées.

- Le 14/02/2020 : Enregistrement du premier décès en France, celui d'un touriste chinois de 80 ans.
- Le 25/02/2020 : Enregistrement du premier décès français et cinq nouveaux cas.
- Le 15/03/2020 : Fermeture des lieux publics.
- Le 17/03/2020 : Mise en place du confinement.
- Le 11/05/2020 : Début de la phase 1 du dé-confinement.
- Le 02/06/2020 : Début de la phase 2 du dé-confinement.
- Le 22/06/2020 : Début de la phase 3 du dé-confinement.

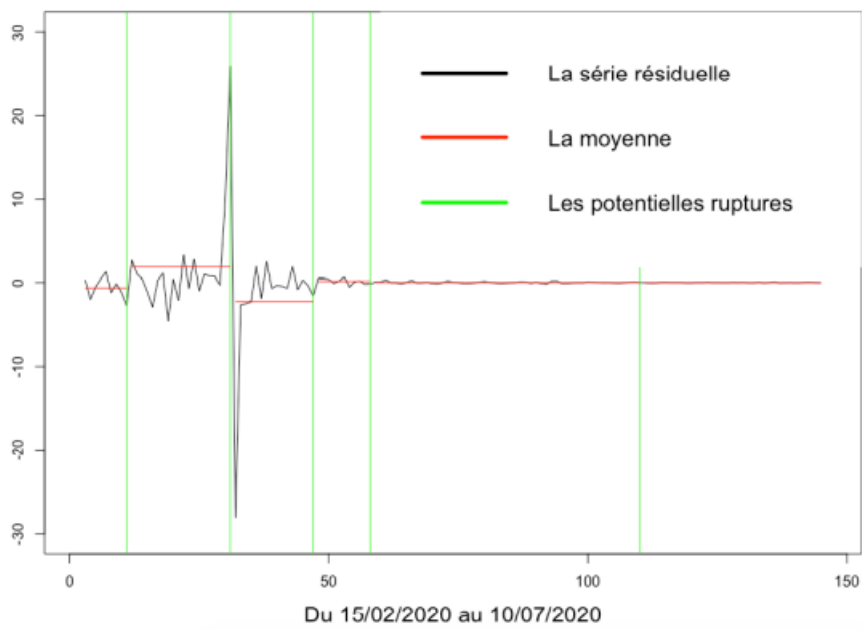
- **Taux de mortalité**

La première série étudiée est celle du taux de mortalité dans la période du 15/02/2020 au 10/07/2020. la figure 1.9 (a) correspond au chronogramme des données brutes. Ce graphique, semble présenter 5 potentiels points de ruptures correspondants aux jours $t_1 = 11$, $t_2 = 31$, $t_3 = 47$, $t_4 = 58$ et $t_5 = 110$, à compter à partir du 15/02. C'est-à-dire que les potentielles ruptures correspondent respectivement aux 25/02/2020, 16/03/2020, 01/04/2020, 12/04/2020 et 03/06/2020. Ici t_1 , t_2 , t_3 et t_4 nous semblent indiquer des instants de changement dans la moyenne de la série. En utilisant le test de normalité Shapiro-Wilk sur la série résiduelle et sur chaque morceau par l'utilisation de la commande `shapiro.test` sous R, nous retenons l'hypothèse de normalité sur chaque morceau. L'utilisation des tests de Box et Pierce nous conduit à rejeter l'hypothèse iid. Cela nous permet de justifier la présence des points de ruptures dans les données. En t_3 le changement de la moyenne est significatif mais ce point peut-être une fausse alarme. Nous pouvons remarquer très facilement que la série résiduelle représentée dans la figure 1.9(b) exhibe les mêmes points de rupture que la série des données brutes. Donc après l'estimation de la tendance, nous ajustons à la série corrigée de la tendance au modèle décrit dans (le paragraphe précédant) pour $k = 5$, $t_1 = 11$, $t_2 = 31$, $t_3 = 47$, $t_4 = 58$ et $t_5 = 110$, $\mu + (\beta_i / \sqrt{n})$ et σ_i sont respectivement la moyenne et la variance des données résiduelle sur l'intervalle $[t_{i-1}, t_i)$, pour $i = 1, \dots, n$ avec $t_0 = 1$ et $t_6 = n$, $\beta_1 = 0$ et pour tout $i = 2, \dots, k + 1$, $\beta_i \in \mathbb{R}$ est un bruit blanc gaussien centré réduit. Cet ajustement est dû au fait que $p - valeur = 0.002681$ lorsque nous avons appliqué les tests de Box sur la série résiduelle.

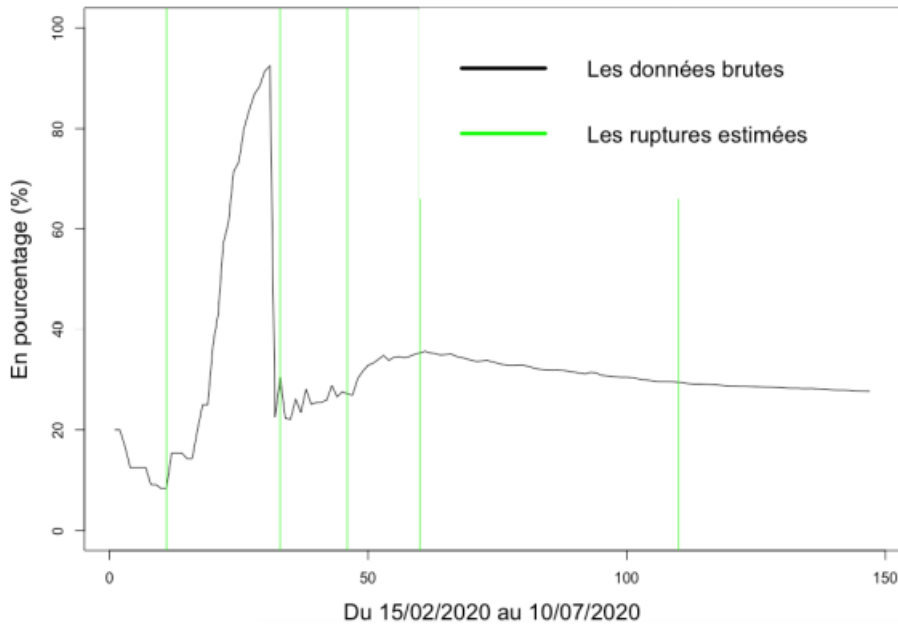
Nous calculons la puissance du test autour des points t_i ci-dessus et nous prenons les années donnant la puissance maximale. Nous obtenons les points $\hat{t}_1 = 11$, $\hat{t}_2 = 32$, $\hat{t}_3 = 46$, $\hat{t}_4 = 60$, $\hat{t}_5 = 110$ à compter à partir du 15/02/2020. Ils sont représentés dans la figure 1.9 (c). Nous remarquons que notre méthode fournie comme estimateurs des dates des ruptures, les ruptures potentielles. Pour interpréter ces estimations, nous dirons qu'autour du 25/02/2020, le taux de mortalité qui est à la baisse, va brusquement augmenter pour atteindre son pic le 15/03/2020 et chuter drastiquement le lendemain puis rester presque constant jusqu'au 30/03/2020, avant de remonter lentement jusqu'au 11/04/2020, puis redescendre lentement jusqu'au 03/06/2020 avant de se stabiliser jusqu'à la phase 3 du dé-confinement et même longtemps après.



(a) Série brute



(b) Série résiduelle

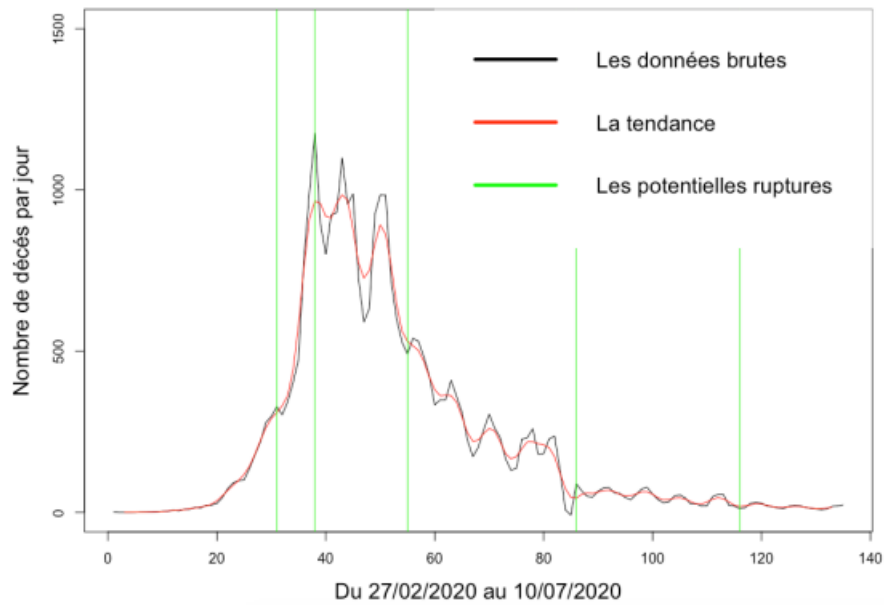


(c) Dates de rupture estimées

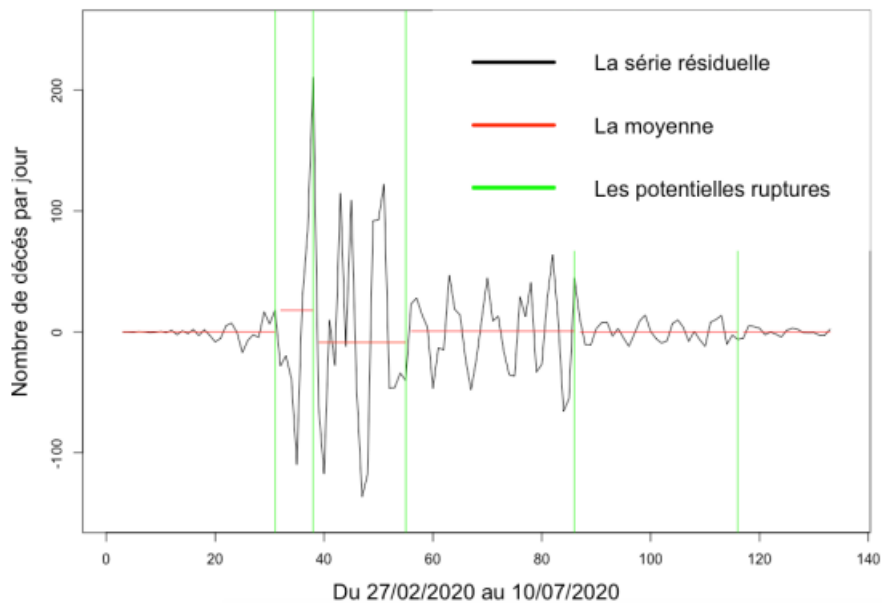
FIGURE 1.9 – Taux de mortalité du COVID-19 en France du 15/02/2020 au 10/07/2020.

- **Nombre de décès quotidiens**

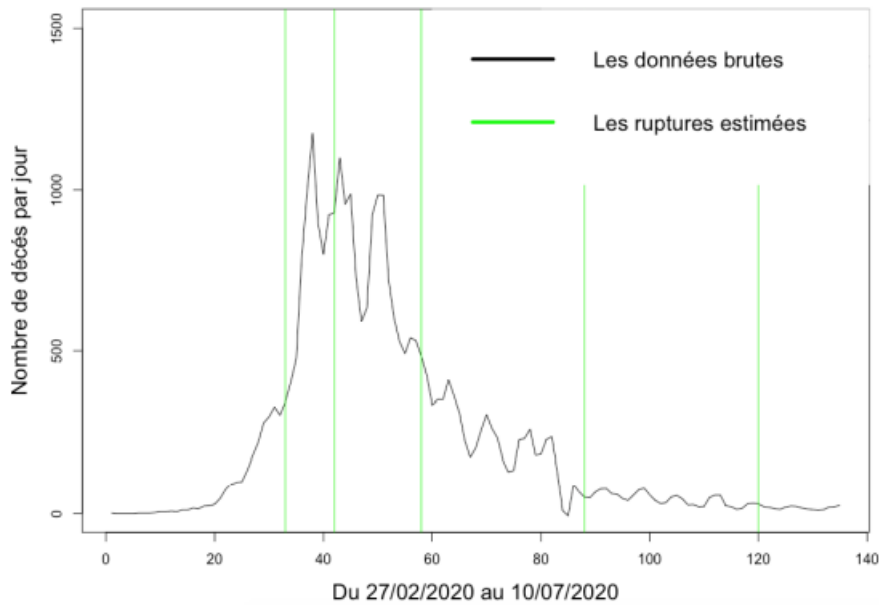
Nous cherchons dans cette partie les points de rupture dans la série des données correspondant aux nombres de décès quotidiens du COVID-19 en France dans la période du 27/02/2020 au 10/07/2020. la figure 1.10 (a) correspond au chronogramme des données brutes. Nous voulons savoir si les potentiels points de rupture dans cette série, tracés en vert dans la figure, représentent réellement des points de rupture. Ces points sont $t_1 = 31$, $t_2 = 38$, $t_3 = 55$, $t_4 = 86$ et $t_5 = 116$ correspondant respectivement aux dates suivantes : 28/03/2020, 04/04/2020, 21/04/2020, 22/05/2020 et 21/06/2020. Comme dans les cas précédents, nous estimons la tendance, puis nous ajustons à la série corrigée de la tendance le modèle décrit dans la section précédente, où 1 et n sont respectivement 27/02/2020 et 10/07/2020. Cette série est représentée dans la figure 1.10 (b). Nous calculons la puissance du test autour des t_i et nous prenons les dates donnant la plus grande puissance. Nous obtenons ainsi les dates : $\hat{t}_1 = 35$, $\hat{t}_2 = 40$, $\hat{t}_3 = 56$, $\hat{t}_4 = 88$ et $\hat{t}_5 = 117$ à compter à partir du 27/02/2020. Ces dates représentées dans la figure 1.10 (c) correspondent respectivement aux 02/03/2020, 06/04/2020, 22/04/2020, 21/05/2020 et 25/06/2020. Les dates estimées sont différentes des potentielles dates de rupture, mais elles leur sont très proches, et les résultats obtenus sont plus proches de la réalité. En effet, autour du $\hat{t}_1 = 02/03/2020$, le nombre de décès augmente drastiquement, atteint son pic et redescend autour du $\hat{t}_2 = 06/04/2020$, puis oscille significativement jusqu'aux environs du $\hat{t}_3 = 22/04/2020$, et un peu moins significativement entre les première et deuxième phases du déconfinement, jusqu'au $\hat{t}_4 = 21/05/2020$, date à partir de laquelle il se stabilise avant de se réduire considérablement à partir du $\hat{t}_5 = 25/06/2020$, peu après la troisième phase du déconfinement.



(a) Série brute



(b) Série résiduelle



(c) Dates de rupture estimées

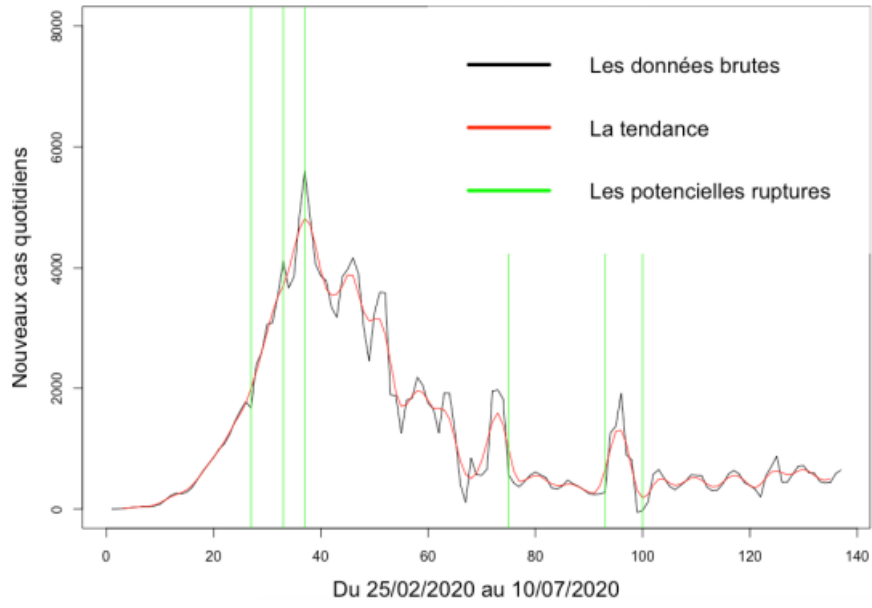
FIGURE 1.10 – Nombre de décès quotidiens du COVID-19 en France du 27/02/2020 au 10/07/2020

• Nouveaux cas quotidiens

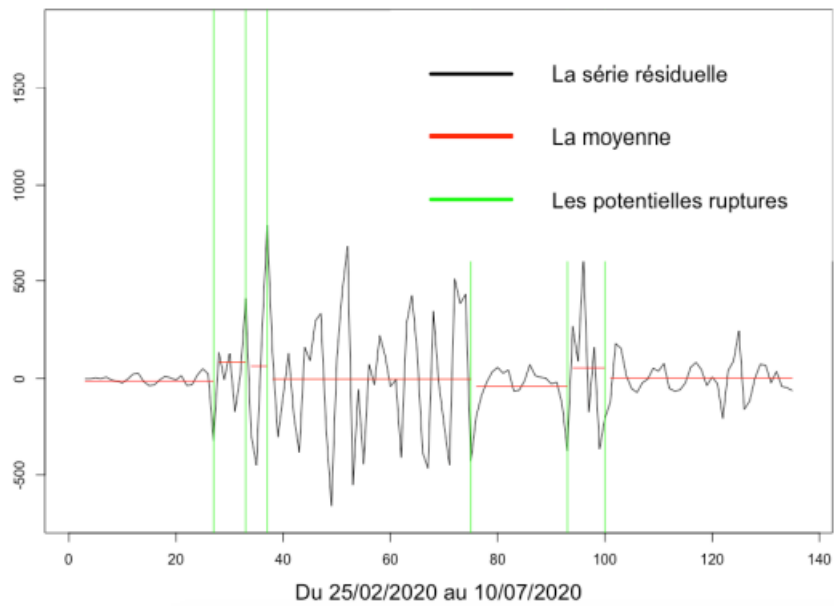
Nous prenons les données des nouveaux cas quotidiens du COVID-19 en France dans la période du 25/02/2020 au 10/07/2020 présentées dans la figure 1.11 (a). On dénombre 6 potentiels points de ruptures, représentés en vert. Ces points sont, $t_1 = 27$, $t_2 = 33$, $t_3 = 37$, $t_4 = 75$, $t_5 = 93$ et $t_6 = 100$ qui correspondent respectivement aux dates suivantes : 23/03/2020, 28/03/2020, 01/04/2020, 09/05/2020, 27/05/2020 et 03/06/2020.

Comme précédemment, nous estimons la tendance par la méthode de moyenne mobile d'ordre 5 et nous ajustons aux données résiduelles le modèle décrit dans (le paragraphe précédant) pour $k = 7$, $\mu + (\beta_i / \sqrt{n})$ et σ_i sont respectivement la moyenne et la variance des observations sur l'intervalle $[t_{i-1}, t_i)$, pour $i = 1, \dots, k + 1$, avec $t_0 = 1$ et $t_7 = n$, $\beta_1 = 0$ et pour tout $i = 2, \dots, k + 1$, $\beta_i \in \mathbb{R}$ et ε_t est un bruit blanc gaussien centré réduit.

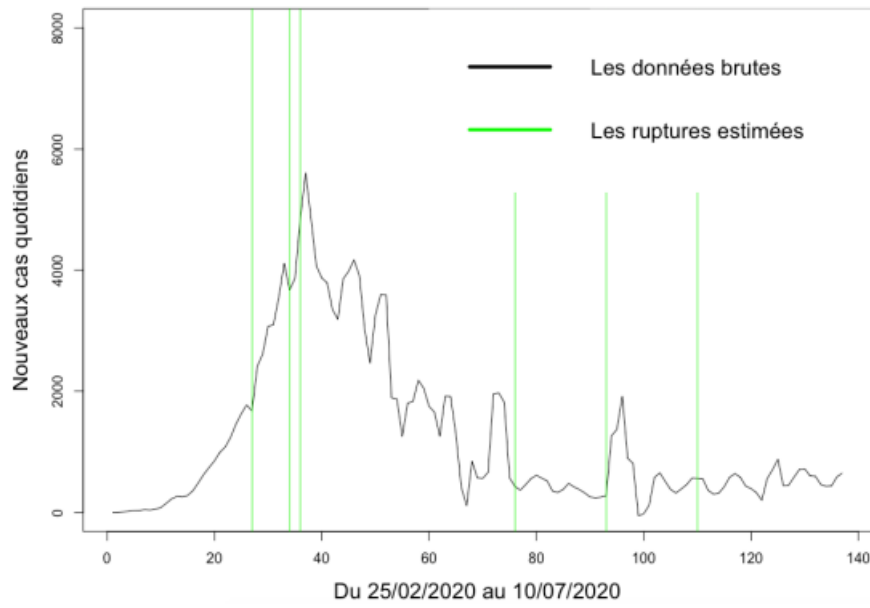
Ensuite, nous calculons la puissance comme dans les cas précédents. Nous obtenons les dates suivantes : $\hat{t}_1 = 26$, $\hat{t}_2 = 34$, $\hat{t}_3 = 40$, $\hat{t}_4 = 40$, $\hat{t}_5 = 93$ et $\hat{t}_6 = 110$ à compter à partir du 25/02/2020. Ces dates sont représentées dans la figure 1.11 (c). Elles correspondent respectivement aux 22/03/2020, 29/03/2020, 04/04/2020, 4/05/2020, 27/05/2020 et 13/06/2020. Les dates données par notre méthode sont différentes de celles que nous avons soupçonnées mais elles sont très proches



(a) Série brute



(b) Série résiduelle



(c) Dates de rupture estimées

FIGURE 1.11 – Nouveaux cas quotidiens du COVID-19 en France du 25/02/2020 au 10/07/2020

INFÉRENCE DANS LES MODÈLES DE RUPTURE

2

2.1 INTRODUCTION

L'étude de rupture de modèles est un passage incontournable 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 ya lieu le ou les points de rupture, identifier les paramètres sujets à un changement et en fin estimer leur amplitude.

Ainsi l'approche est double : soit se contenter de vérifier s'il ya rupture (souvent considéré comme un problème de test d'hypothèse) ou bien de localiser le point de rupture s'il ya lieu (vu par fois comme un problème d'estimation)

L'approche bayésienne à été largement sollicitée dans ce domaine.

2.2 APPROCHE CLASSIQUE DE DÉTECTION DE RUPTURE

La détection de rupture dans les signaux et système est un problème particulièrement important dans la mesure où, souvent c'est précisément dans le changement brusque que se concentre une part prépondérante de l'information.

Le plus souvent, les procédures considérées suppose que le point de rupture correspond à l'un des instants d'observations et tentent de répondre à la question suivante : y a t-il rupture dans le modèle considéré ou non ? Ces procédures sont en général basées sur les tests d'hypothèses.

2.2.1 Tests de rupture non paramétriques

Dans cette section, quelques tests de rupture déjà présents dans la littérature seront décrits. Ces tests servent, comme leur nom l'indique, à détecter si un jeu de données possède ou non une rupture.

Test de Pettitt (1979) :

Soit $(X_t)_{t \in \mathbb{Z}}$ un processus aléatoire et $\tau \in [1, N]$ l'instant de rupture. Nous allons présenter un test qui fait la détection de rupture au niveau de la fonction de répartition $F(x)$, le test de Pettitt (1979).

Ce test est considéré comme un test non paramétrique, il est basé sur le test de Mann-Whitney, qui est considéré comme un test de rang. La statistique de test est donnée par,

$$K_n = \max_{1 < t < n} |U_{t,N}|$$

$U_{t,N}$ est la statistique de test Mann-Whitney, le test a comme hypothèse nulle.

H_0 : Il n'existe pas de rupture.

C'est à dire que le processus admet une seule distribution $F(x)$ et $\tau = N$.

Contre l'hypothèse alternative,

H_1 : Il existe une seule rupture à l'instant τ , et le processus admet deux fonction de distribution tel que :

$$F_1(x)_{(X_1, \dots, X_\tau)} \neq F_2(x)_{(X_{\tau+1}, \dots, X_N)}$$

Pettitt(1979) a donné la probabilité de dépassement approximative d'une valeur k par :

$$p(K_n > k) \simeq 2 \exp\left(\frac{-6k^2}{N^3 + N^2}\right)$$

cette probabilité doit être comparée avec un risque α de première espèce, c'est à dire :

On rejette H_0 si la valeur de la statistique de test est inférieure à α , sinon on l'accepte.

La figure suivante explique la méthode pour obtenir les U_i .

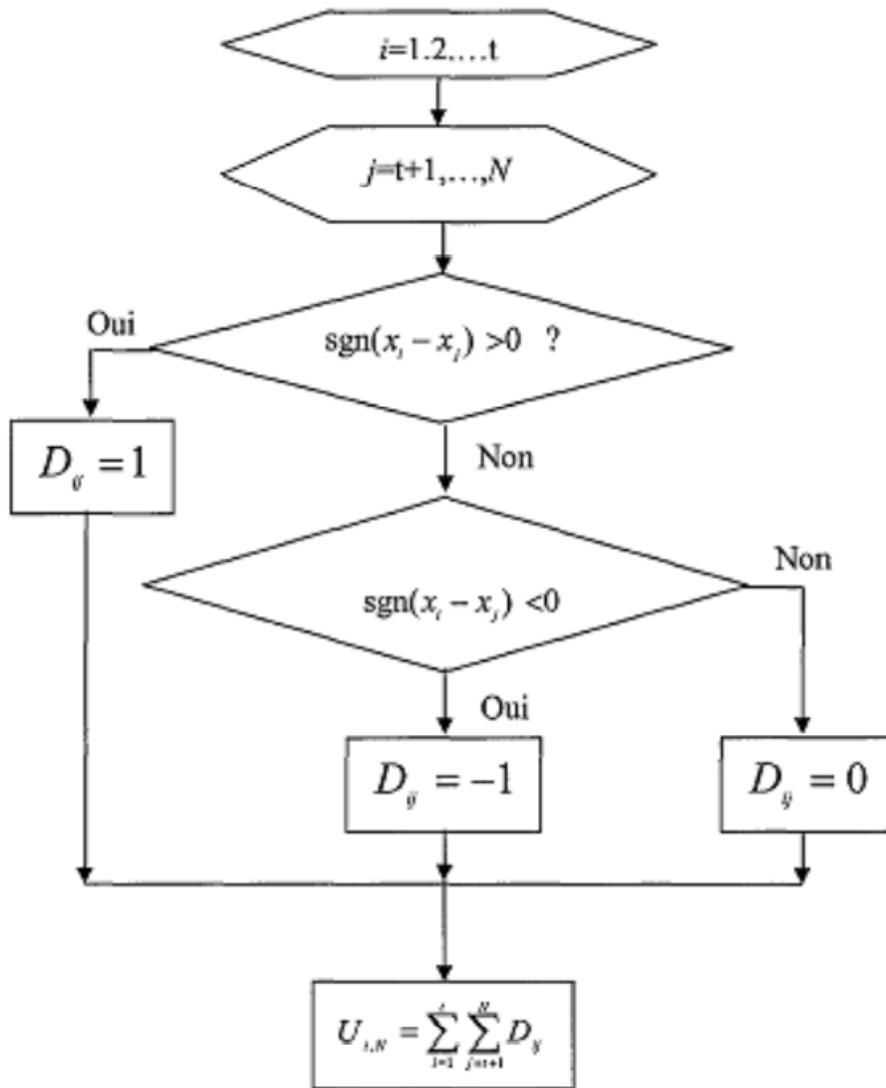


FIGURE 2.1 – Algorithme explicatif de test de pettitt.

Test de Lambard (1987)

Supposons que l'on a des variables aléatoires y_1, \dots, y_T indépendantes qui suivent chacune une distribution continue $F(y, \theta_1), \dots, F(y, \theta_T)$.

On essaie d'étudier s'il y a un changement tel que les paramètres θ_i ($i \in \overline{1, \tau}$) soient de la forme $\theta_1 = \dots = \theta_\tau \neq \theta_{\tau+1} = \dots = \theta_T$.

Mais on a besoin d'un modèle plus général si les paramètres θ changent graduellement dans un intervalle.

Alors, Lambard (1987) a introduit ce modèle :

$$\theta_i = \begin{cases} \rho & \text{si } i = 1, \dots, \tau_1; \\ \rho + \frac{i-\tau_1}{\tau_2-\tau_1} \delta & \text{si } i = \tau_1 + 1, \dots, \tau_2; \\ \rho + \delta & \text{si } i = \tau_2 + 1, \dots, T. \end{cases} \quad (2.1)$$

pour $\tau_1 < \tau_2$,
 ρ et δ inconnus .
 Lombard (1987) utilise l'hypothèse suivante :
 $H_0 : \delta = 0$
 versus
 $H_1 : \delta \neq 0$.

Pour commencer, voici quelques notations :

- i) r_i est le rang de la variable aléatoire y_i ,
- ii) ϕ représente une fonction de score avec :

$$0 < \int_0^1 \phi(w) dw < \infty,$$

$$\bar{\phi} = \frac{1}{T} \sum_{i=1}^T \phi\left(\frac{i}{T+1}\right),$$

$$A^2 = \frac{1}{T+1} \sum_{i=1}^T \left[\phi\left(\frac{i}{T+1}\right) - \bar{\phi} \right]^2$$

Alors, le score de classement de y_i , noté $s(r_i)$ est :

$$s(r_i) = \frac{1}{A} \left[\phi\left(\frac{r_i}{T+1}\right) - \bar{\phi} \right] \text{ pour } i = 1, \dots, T.$$

Donc Lombard (1987) a introduit le test statistique suivant :

$$v_{\tau_1, \tau_2} = \sum_{j=\tau_1+1}^{\tau_2} \sum_{i=1}^{j-1} S(r_i) = (\tau_2 - \tau_1) \sum_{i=1}^{\tau_1} s(r_i) + \sum_{i=\tau+1}^{\tau_2-1} (\tau_2 - i) s(r_i)$$

qui est le test de rang localement le plus puissant pour tester l'hypothèse nulle versus une hypothèse alternative unilatérale si τ_1 et τ_2 sont connus. La statistique dépend indirectement des y_i en utilisant le rang de $y_i, i = 1, \dots, T$. S'ils sont inconnus, Lombard propose le test suivant :

$$q_{1,T} = \sum_{t_1=1}^{T-1} \sum_{t_2=t_1+1}^T v_{t_1, t_2}^2.$$

Si on observe une grande valeur, on rejette l'hypothèse nulle. Lorsque T tend vers l'infini, il est à noter que :

$$\frac{q_{1,T}}{T^5} \rightarrow q_1 = \sum_{n=1}^{\infty} \frac{1}{(\eta\pi)^4} Z_n^2 \text{ sous } H_0.$$

où $Z = (Z_1, Z_2, \dots)^t$ est un vecteur de variables aléatoires indépendantes et identiquement distribuées de loi normale standard. Les points critiques pour différentes valeurs de α et pour toute valeur de T sont présentés dans le tableau 2.1.

α	0,1	0,075	0,05	0,025	0,01
q_1	0,0287	0,0334	0,0403	0,0525	0,0690
$q_{1,T} / T^5$	0,0289	0,0334	0,0402	0,0515	0,0662

TABLE 2.1 – Valeurs critiques de q_1 et $q_{1,T}/T^5$

Si on est intéressé à tester l'hypothèse nulle versus une hypothèse alternative unilatérale ($\delta > 0$ ou $\delta < 0$), Lombard propose trois différents scores :

i) Le score de Wilcoxon : $\phi_1(u) = 2u - 1$, pour tester des changements dans la position (une tendance centrale),

ii) Le score de Mood : $\phi_2(u) = (2u - 1)^2$, pour tester les changements d'échelle,

iii) Le score logarithmique : $\phi_3(u) = \log(1 - u)$ (les auteurs utilisent $-\log(1 - u)$), pour tester des changements d'échelle dans une distribution ayant pour domaine $[0, \infty)$.

Pour le score de Wilcoxon, la fonction de score $s(r_i)$ est équivalente à :

$$s(r_i) = \sqrt{\frac{12}{T(T+1)}} \left(r_i - \left[\frac{T+1}{2} \right] \right).$$

Lorsqu'on veut tester s'il y a l'apparition d'une tendance, c'est-à-dire que $\tau_2 = T$ et que $\tau_1 = \tau$, la statistique du test de Lombard est modifiée tel que $t_2 = T$:

$$q_{1,T}^* = \sum_{i=1}^{T-1} v_{i,T}^2,$$

et, lorsque T tend vers l'infini, sous l'hypothèse nulle :

$$\frac{q_{1,T}^*}{T^4} \longrightarrow q_1^* = \sum_{n=1}^{\infty} \gamma_n Z_n^2,$$

où $\gamma_1 > \gamma_2 > \dots > 0$ sont les solutions positives réelles de $\tan\left(\gamma^{\frac{-1}{4}}\right) + \tanh\left(\gamma^{\frac{-1}{4}}\right) = 0$.

Les points critiques pour différentes valeurs de α et pour toute valeur de T sont présentés dans le tableau 2.2.

α	0,1	0,075	0,05	0,025	0,01
q_1^*	0,0879	0,1027	0,1242	0,162	0,21 » 5
$\frac{q_{1,T}^*}{T^4}$	0,0882	0,1026	0,1242	0,158	0,2035

TABLE 2.2 – Valeurs critique de q_1^* et $\frac{q_{1,T}^*}{T^4}$

Lorsqu'on veut tester s'il y a un changement abrupte, c'est-à-dire que $\tau_1 = \tau$ et $\tau_2 = \tau + 1$, toujours en utilisant la fonction de score de Wilcoxon, la statistique de test de Lombard devient :

$$q_{1,T}^0 = \sum_{t=1}^{T-1} v_{t,t+1}^2 = \sum_{t=1}^{T-1} \left[\sum_{i=1}^t s(r_i) \right]^2$$

et, sous l'hypothèse nulle, la distribution limite de $\frac{q_{1,T}^0}{T^2}$ converge vers la distribution limite du test d'adéquation de Cramer von Mises (q_1^0). Une fois de plus, pour toute valeur de T , les points critiques sont présentés dans la tableau 2.3

α	0,1	0,075	0,05	0,025	0,01
$q_{1,T}^0$	0,3473	0,3939	0,4614	0,5608	0,7435
$q_{1,T}^0 / T^2$	0,3431	0,3870	0,4521	0,5596	0,7022

TABLE 2.3 – Valeurs critiques de $q_{1,T}^0$ et $q_{1,T}^0 / T^2$

2.2.2 Tests de rupture paramétriques

Modèle Gaussien :

a) Test de Jarusková (1997) :

Le test de Jarušková (1997) est un test paramétrique de type «maximum». Tout comme le test de Lombard (1987), supposons que l'on observe des variables aléatoires y_1, \dots, y_T . L'hypothèse nulle stipule que $H_0 : \mu_1 = \mu_2$. L'hypothèse alternative stipule qu'il existe un point dans le temps $\tau \in \{1, \dots, T - 1\}$ qui fait en sorte que le modèle est :

$$y_i \sim \begin{cases} N(\mu_1, \sigma^2) & \text{si } i = 1, \dots, \tau; \\ N(\mu_2, \sigma^2) & \text{si } i = \tau + 1, \dots, T; \end{cases} \quad (2.2)$$

où $\mu_1 \neq \mu_2$.

En supposant σ^2 inconnu, le test de Jarušková, noté $J(T)$ est le suivant :

$$J(T) = \max_{1 \leq \tau < T} |J_\tau| = \max_{1 \leq \tau < T} \frac{1}{S_\tau} \sqrt{\frac{(T - \tau)\tau}{T}} |\bar{y}_\tau - \bar{y}_\tau^*|,$$

où :

$$\bar{y}_\tau = \frac{\sum_{j=1}^{\tau} y_j}{\tau},$$

$$\bar{y}_\tau^* = \frac{\sum_{j=\tau+1}^T y_j}{T - \tau},$$

$$s_\tau = \sqrt{\frac{\sum_{j=1}^{\tau} (y_j - \bar{y}_\tau)^2 + \sum_{j=\tau+1}^T (y_j - \bar{y}_\tau^*)^2}{T - 2}}.$$

L'hypothèse nulle est rejetée lorsque la statistique $J(T)$ est plus grande qu'une certaine valeur critique (voir le tableau 2.4).

Si l'hypothèse nulle est vraie, alors \bar{y}_τ et \bar{y}_τ^* devraient être d'environ la même valeur. Jarušková (1997) obtient celles-ci par simulations. Une statistique tronquée peut aussi être utilisée :

$$J_1(T) = \max_{t_0 T \leq \tau \leq (1-t_0)T} |J_\tau|,$$

où :

$t_0 \in (0; 0,5)$ (Jarušková utilise $t_0 = 0,05$).

Jarušková (1997) obtient aussi les valeurs critiques par simulations et sont aussi présentées dans le tableau 2.4.

T	$\alpha = 0,05$		$\alpha = 0,01$	
	J(T)	J ₁ (T)	J(T)	J ₁ (T)
50	3,15	3,08	3,76	3,69
100	3,16	3,06	3,71	3,62
200	3,19	3,07	3,72	3,61
300	3,21	3,08	3,73	3,62
500	3,24	3,09	3,73	3,62

TABLE 2.4 – Valeurs critiques de J(T) et J₁(T)

b) Test 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 :

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

L'expression $\left[|KS_n/n - S| / \{K(1 - K/n)\}^{1/2} \right]$ 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 :

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

En s'intéressant au test unilatéral, on suppose que le signe de d 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 :

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

avec : $1 \leq n_0 < n_1 < n$

c) Test de pettit(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 :

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

d) Test de Brow 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 :

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

$$\text{où : } z_i = i / (i+1)^{1/2} (x_{i+1} - \bar{x}_i),$$

on remarque que sous H_0 , les z_n sont des variables aléatoires indépendantes de loi normale centrée et réduite.

e) Test de chernoff et Zacks,1964 :

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

dite "statistique quasi-bayésienne" qui est :

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

Modèle de regression :

Test de Reeves et al (2007) :

Les tests de Reeves et al. (2007) utilisent les sommes au carré des résidus et comparent leur statistique de test à une loi de Fisher dont les degrés de liberté changent dépendamment du test utilisé. Tout comme les autres tests, supposons que l'on a des variables aléatoires y_1, \dots, y_T et que celles-ci, sous l'hypothèse nulle, soient indépendantes et identiquement distribuées tandis que, sous l'hypothèse alternative, il existe un point τ tel que la moyenne des variable y_1, \dots, y_τ soit différente des variables $y_{\tau+1}, \dots, y_T$.

Reeves et al. (2007) proposent les tests (méthodes) LR (Modified Lund and Reeves TPR method) et XLW (Modified Wang's TPR method) où la méthode TPR (two-phase regression) avec un point de rupture au point τ est simplement :

$$y_i = \begin{cases} \mu_1 + \beta_1 x_i + \epsilon_i & \text{si } i = 1, \dots, \tau \\ \mu_2 + \beta_2 x_i + \epsilon_i & \text{si } i = \tau + 1, \dots, T \end{cases} \quad (2.3)$$

ainsi que $x_1 \leq \dots \leq x_T$, $\epsilon_i \sim N(0, \sigma^2)$ pour $i = 1, \dots, T$ et $\beta_1, \beta_2, \mu_1, \mu_2, \tau$ sont inconnus.

Il y a cependant une contrainte de continuité de la régression au point t se traduisant par $\mu_2 = \mu_1 + (\beta_1 - \beta_2)x_\tau$. Les deux tests suivants proposent quelques modifications au test TPR, tel que l'égalité de β_1 et β_2 .

• **Test LR :**

Lund et Reeves (2007) ont modifié le modèle TPR en cessant d'imposer la contrainte de continuité. Le modèle LR suppose que les y_i sont des entiers. Le modèle LR est donc :

$$y_i = \begin{cases} \mu_1 + \beta_1 t + \epsilon_i & \text{si } i = 1, \dots, \tau \\ \mu_2 + \beta_2 t + \epsilon_i & \text{si } i = \tau + 1, \dots, T \end{cases} \quad (2.4)$$

où on peut alors détecter une rupture dans la moyenne ($\mu_1 \neq \mu_2$) et dans la tendance ($\beta_1 \neq \beta_2$). Les erreurs suivent une loi normale de moyenne nulle et de variance inconnue. Ce qui revient donc à confronter les hypothèses suivantes :

$$H_0 = \mu_1 = \mu_2 \text{ et } \beta_1 = \beta_2,$$

$$H_1 = (\mu_1 \neq \mu_2) \text{ et } (\beta_1 \neq \beta_2).$$

Si le point de rupture τ est connu et fixé, le test devient simplement :

$$F_\tau = \frac{(SSE_0 - SSE_A / 2)}{SSE_A / (T - 4)} \sim F_{2, T-4},$$

où SSE_0 et SSE_A sont respectivement les sommes des erreurs au carré sous l'hypothèse nulle et alternative. Sous H_0 , la statistique de test devrait suivre une loi de Fisher à $(2, T - 4)$ degrés de liberté. On rejette l'hypothèse nulle pour de grandes valeurs de F_τ . Si le point de rupture t est inconnu, la statistique devient :

$$F_{max} = \max_{1 \leq \tau \leq T} F_\tau,$$

et on rejette H_0 pour une grande valeur de F_{max} . Comme cette statistique ne suit pas de loi connue, Lund et Reeves (2007) obtiennent les valeurs critiques par simulations et sont présentées dans le tableau 2.5.

• **Test XLW :**

Wang (2007) a apporté une petite modification au test LR car il trouvait que celui-ci expliquait mal les phénomènes climatiques. Son modèle devient donc :

$$y_i = \begin{cases} \mu_1 + \beta_1 t + \epsilon_i & \text{si } i = 1, \dots, \tau \\ \mu_2 + \beta_2 t + \epsilon_i & \text{si } i = \tau + 1, \dots, T \end{cases} \quad (2.5)$$

où les termes sont définis précédemment. Les hypothèses deviennent donc :

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

Si le point de rupture t est connu et fixé, le test devient simplement :

$$F_\tau = \left(\frac{SSE_0 - SSE_A}{SSE_A / (T - 3)} \right) \sim F_{1, T-3}$$

où les termes sont aussi définis précédemment. Si τ est inconnu la statistique F_{\max} est une fois de plus utilisée. Wang (2007) obtient aussi les valeurs critiques par simulations et sont présentées dans le tableau 2.5.

T	LT(F_{\max})	XLW(F_{\max})
25	11,67	7,37
50	11,07	6,92
75	11,06	6,88
100	11,09	6,91
200	11,21	7,01
500	11,54	7,24

TABLE 2.5 – valeurs critique avec $\alpha = 0,05$ des tests LR et XLW

Cependant, dans Dubé (2011), il est écrit sur le test T P R et ses dérivées que la convergence vers les valeurs critiques obtenues n'est pas fiable et que ces tests ne sont pas robustes par rapport au non-respect du postulat de normalité. Les splines étant plus flexibles qu'une régression linéaire, on s'attend à ce que l'introduction des splines au modèle LR soit une solution envisageable afin de réduire les erreurs. Ceci ne règlera peut-être pas le problème de non-respect du postulat de normalité, mais cela pourra peut-être améliorer les performances du test T P R.

2.2.3 Estimation d'un point de rupture

Dans cette section, nous nous intéressons à l'estimation de ce point de rupture.

Plusieurs approches sont utilisées dans ces problèmes . Soit l'approche classique de l'inférence fondée sur la fonction de vraisemblance , ou encore les méthodes non paramétrique basées sur les somme cumulatives.

Méthodes du maximum de vraisemblance dans le cas gaussien

• rupture dans la moyenne μ :

On considère un processus aléatoire (X_1, X_2, \dots, X_n) , qui suit une lois normale $N(\mu_0, \sigma_0^2)$, on suppose que les observations sont indépendantes les unes des autres .L'objectif est la détection d'un point de rupture $k < n$ au niveau de la moyenne μ après la $k^{\text{ème}}$ observation X_k . Pour cela nous mettrons l'hypothèse suivante :

pour un processus $X : X_t = \mu + \varepsilon_t$ et ,

$$\left\{ \begin{array}{l} H_0 : (X_1, X_2, \dots, X_n) \rightarrow N(\mu_0, \sigma_0^2) \\ \text{Contre l'alternative,} \\ H_1 : (X_1, X_2, \dots, X_k) \rightarrow N(\mu_1, \sigma_0^2) \text{ et } (X_{k+1}, \dots, X_n) \rightarrow N(\mu_2, \sigma_0^2). \end{array} \right. ,$$

A titre d'information, ce type de test utilisé dans le domaine de la fiabilité .
La densité de la loi normale est définie par,

$$f_{\mu, \sigma^2}(X) = \frac{1}{\left(\sqrt{2\pi\sigma_0^2}\right)^n} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right)$$

Nous définissons la fonction de vraisemblance sous l'hypothèse H_0 , et sous l'hypothèse H_1

$$L_{H_0}(X_1, X_2, \dots, X_n) = \frac{1}{\left(\sqrt{2\pi\sigma_1^2}\right)^k} \exp\left(-\frac{\sum_{i=1}^k (X_i - \mu_1)^2}{2\sigma_1^2}\right) \frac{1}{\left(\sqrt{2\pi\sigma_1^2}\right)^{n-k}} \exp\left(-\frac{\sum_{i=k+1}^n (X_i - \mu_2)^2}{2\sigma_1^2}\right)$$

Les estimateurs des deux paramètres sous l'hypothèse H_0 sont :

$$\begin{cases} \hat{\mu}_0 = \frac{1}{n} \sum_{i=1}^n X_i \\ \hat{\sigma}_0^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \hat{\mu}_0)^2. \end{cases}$$

Sous l'hypothèse H_1 où nous supposons qu'une rupture se produirait, les estimateurs de μ et σ sont définis par :

$$\begin{cases} \hat{\mu}_1(k) = \frac{1}{k} \sum_{i=1}^k X_i \\ \hat{\mu}_2(k) = \frac{1}{n-k} \sum_{i=k+1}^n X_i \\ \hat{\sigma}_1(k)^2 = \frac{1}{n} \left(\sum_{i=1}^k (X_i - \hat{\mu}_1(k))^2 + \sum_{i=k+1}^n (X_i - \hat{\mu}_2(k))^2 \right) \end{cases}$$

Notons aussi, que le rapport de la vraisemblance entre les deux modèle (avec et sans rupture) nous donne :

$$Rap_{vrs}(X_1, X_2, \dots, X_n, k) = \frac{L_{H_1}(X_1, X_2, \dots, X_n)}{L_{H_0}(X_1, X_2, \dots, X_n)} = \left(\frac{\hat{\sigma}_0}{\hat{\sigma}_1(k)} \right)^{\frac{n}{2}}$$

Et l'estimation de la rupture \hat{k} se définit par :

$$\hat{k} = \arg \max_{1 < k < n} (Rap_{vrs}(X_1, X_2, \dots, X_n, k)).$$

• **Rupture dans la variances σ^2 :**

Nous utilisons la même procédure, celle ci consiste à la (minimisation ou maximisation) du rapport entre les deux fonctions de vraisemblance sous l'hypothèse H_0 et sous l'alternative H_1 . Tout d'abord on rappelle que l'hypothèse du test est :

$$\begin{cases} H_0 : (X_1, X_2, \dots, X_n) \rightarrow N(\mu_0, \sigma_0^2) \\ \text{Contre l'alternative,} \\ H_1 : (X_1, X_2, \dots, X_k) \rightarrow N(\mu_1, \sigma_1^2) \text{ et } (X_{k+1}, \dots, X_n) \rightarrow N(\mu_2, \sigma_2^2). \end{cases}$$

La fonction de vraisemblance sous l'hypothèse \mathbf{H}_1 est :

$$L_{H_1}(X_1, \dots, X_k, \dots, X_n) = \left(\frac{\exp\left(\frac{-\sum_{i=1}^k (X_i - \mu_1)^2}{2\sigma_1^2}\right)}{(\sqrt{2\pi\sigma_1^2})^k} \right) \times \left(\frac{\exp\left(\frac{-\sum_{i=k+1}^n (X_i - \mu_1)^2}{2\sigma_2^2}\right)}{(\sqrt{2\pi\sigma_2^2})^{n-k}} \right)$$

Les estimateurs des différents paramètres μ et σ sont les solution de système suivantes :

$$\left\{ \begin{array}{l} \hat{\mu}_1(k) = \frac{1}{\hat{\sigma}_1^2(k)} \sum_{i=1}^k (X_i - \hat{\mu}_1(k))^2 + \frac{1}{\hat{\sigma}_2^2(k)} \sum_{i=k+1}^n (X_i - \hat{\mu}_1(k))^2 \\ \hat{\sigma}_1^2(k) = \frac{1}{k} \sum_{i=1}^k (X_i - \hat{\mu}_1(k))^2 \\ \hat{\sigma}_2^2(k) = \frac{1}{n-k} \sum_{i=k+1}^n (X_i - \hat{\mu}_1(k))^2 \end{array} \right.$$

Le rapport de vraisemblance nous donne :

$$Rap_{vors}(X_1, X_2, \dots, X_n, k) = \frac{L_{H_1}(X_1, X_2, \dots, X_n)}{L_{H_0}(X_1, X_2, \dots, X_n)} = \left(\frac{\hat{\sigma}_0^2}{\hat{\sigma}_1^2(k)} \right)^{\frac{k}{2}} \left(\frac{\hat{\sigma}_0^2}{\hat{\sigma}_1^2(k)} \right)^{\frac{(n-k)}{2}}$$

L'estimateur de la rupture \hat{k} est donc donné par :

$$\hat{k} = \arg \max_{1 < k < n} (Rap_{vors}(X_1, X_2, \dots, X_n, k))$$

2.3 APPROCHE BAYÉSIENNE

La détection de rupture a posteriori dans différents modèles stochastiques a été étudiée depuis les années 1950, une importante littérature existe sur le sujet, la monographie (Baseville, M., et Nikiforov) fait l'état sur plusieurs méthodes de détection de rupture .

Cependant, la prise en compte de la complexité des algorithmes et du temps de calcul informatique est récente. On peut la dater du début de 21^{ème} siècle simultanément à l'explosion de la taille des jeux de données due aux progrès technologiques et informatiques. Ce phénomène est popularisé sous le nom de Big Data, l'état de l'art avant Big Data était d'une complexité en temps de calcul en quantité de mémoire de l'ordre $O(n^2)$ (Pierre R. Bertrand). Avant de présenter cette méthode, nous allons donner quelques définitions de la distribution et de la densité a posteriori et a priori.

Soit (X_1, X_2, \dots, X_n) une suite de variables aléatoires indépendantes de paramètre θ .

La densité a posteriori de modèle (X_1, X_2, \dots, X_n) est définie par :

$$\pi(\theta \setminus X) = \frac{L(X_1, X_2, \dots, X_n / \theta) \cdot \pi(\theta)}{k(\theta)}$$

où :

$L(X_1, X_2, \dots, X_n)$ est la fonction de vraisemblance construite à partir des données,

$\pi(\theta)$ est la densité a priori du modèle. et

$$k(\theta) = \int_{\theta} L(X_1, X_2, \dots, X_n / \theta) \cdot \pi(\theta) d\theta.$$

La fonction $k(\theta)$ est appelée pour certains littératures une constante de normalisation, de tel sort que $\pi(\theta / X)$ soit une densité de probabilité. Une fois $\pi(\theta / X)$ est obtenue comme densité de probabilité, nous pourrions faire de l'estimation et de la détection de rupture sur le paramètre choisi, plusieurs méthodes connues peuvent être utilisées, citons par exemple l'algorithme MLR (mixture likelihood ratio), où on utilise la distribution a priori $\pi(\theta)$ pour la détection de la rupture, aussi l'algorithme GLR qui maximise la vraisemblance construite à partir de la distribution de paramètre. En utilisant différentes fonctions coûts, nous retrouvons les différents estimateurs Bayésien de ce point.

2.3.1 Modèle gaussien

cas où les variables sont indépendantes

Considérons le modèle suivant :

$$\begin{cases} X_i = \phi_0 + \epsilon_i; & i = 1, 2, \dots, m \\ X_i = \phi_1 + \epsilon_i; & i = m + 1, \dots, n \end{cases} \quad (2.6)$$

Où :

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

Les ϵ_i représentent les erreurs aléatoires, elles sont gaussiennes, indépendantes et identiquement distribuées de moyenne nulle et de variance constante et inconnue σ^2 ; ($\epsilon_i \sim N(0, \sigma^2)$).

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.

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 (L.D.Broemeling et D.Holbert, 1977) *Holbert et Broemeling [1977]*
 Étant donné le modèle (2.6)

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

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 :

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

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

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

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 :

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

où :

$$S_j^k = \sum_{i=j}^k (X_i - \bar{X}_i^k)^2$$

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

cas où les variables sont dépendantes

Considérons le modèle suivant :

$$\begin{cases} X_i = \phi_0 + \epsilon_i, & i = 1, 2, \dots, m; \\ X_i = \phi_1 + \epsilon_i, & i = m+1 \dots n \end{cases} \quad (2.7)$$

Où :

ϕ_0 et ϕ_1 des moyennes réelles et inconnues.

Les $\epsilon_i, i = 1, \dots, n$ suivent un processus auto-régressif d'ordre 1.

$\epsilon_i = \rho \epsilon_{i-1} + e_i; e_i \sim N(0, \sigma_i^2), i = 1, \dots, n$, avec ρ et σ sont inconnus.

Considérons le cas où la rupture est dans la moyenne .

Supposons que ϕ_0 et ϕ_1 du modèle sont différentes et inconnues et les erreurs $\epsilon_i \sim AR(1), i = 1, \dots, n$, tel que σ_i sont constantes et inconnues

En assignant des lois a priori impropres aux paramètres $\sigma_i^2, (\phi_0, \phi_1), \rho$ et m , le théorème suivant nous détermine la masse de probabilité marginale a posteriori du point de m

Théorème 2.2 (L.D.Broemeling et D.Holbert) *Holbert et Broemeling [1977]*
considérons le modèle (2.7) Si on suppose que la densité conjointe a priori des paramètres $\sigma_i^2, (\phi_0, \phi_1), \rho$ et m est :

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

alors

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

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

où :

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

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

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

$$\alpha_3 = \sum_{i=2}^n (X_i)^2,$$

$$\beta_1 = \rho, \beta_2 = (1 - \rho) \sum_{i=2}^n (X_i - \rho X_{i-1}) + (X_{m+1} - \rho X_m).$$

2.3.2 Modèle de régression :

1^{er} cas : les erreurs sont indépendantes

Le modèle considéré est le suivant :

$$\begin{cases} Y_i = X_i \beta_1 + \epsilon_i, & i = 1, 2, \dots, m \\ Y_i = X_i \beta_2 + \epsilon_i & i = m + 1, \dots, n \end{cases} \quad (2.8)$$

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 $(\beta_1 \neq \beta_2)$.

$\epsilon_i; i = 1, \dots, n$ 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 auto-corrélées, 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.

Le théorème suivant nous donne les densités a posteriori des différents paramètres du modèle (2.8).

Théorème 2.3 (Chin.Choy et Broemelling) *LD* [1980]

Étant donné le modèle (2.8).

m, β, σ_2 sont des paramètres inconnus avec $\beta = (\beta'_1, \dots, \beta'_2)$.

m est uniformément distribué sur l'ensemble $\{1, 2, \dots, n-1\}$

La loi conjointe a priori de β et de $R = 1/\sigma_i^2$ 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 normal-Gamma.

R suit une loi 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 donnée par :

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

Où : $a^* = a + n/2$.

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

ii) La densité de probabilité a posteriori du paramètre de régression

$\beta = \beta'_1, \beta'_2$ est :

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

où : $p(m) = \alpha^* / D(m) (x'(m) x(m) + \tau)$,

$t[\beta, p, a^*, \beta^*(m), p(m)]$ est une densité de student p -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 :

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

où β_i et β_i^* , $i = 1, 2$ sont $p \times 1$ et p_{ij} , $i, j = 1, 2$ sont des matrices de $p \times p$.

Alors :

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

$$\pi(\beta_1|y) = \sum_{m=1}^{n-1} \pi(m|y)t[\beta, p, a^*, \beta_1^*(m), p_1^*(m)],$$

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

iv) La densité de probabilité a postériori de β_2 est :

$$\pi(\beta_2|y) = \sum_{m=1}^{n-1} \pi(m|y)t[\beta, p, a^*, \beta_2^*(m), p_2^*(m)],$$

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

v) La densité de probabilité a postériori 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^{ème} cas : les erreurs sont auto-corrélées

Considérons le même modèle de régression et intéressons nous au cas où les erreurs sont régies par un processus autorégressif d'ordre un AR(1). Soit le théorème suivant :

Théorème 2.4 (Salazar et Broemeling ,1981) *Salazar et al. [1981]*

$$\begin{cases} Y_i = X_i\beta_1 + \mu_i, & i = 1, 2, \dots, m \\ Y_i = X_i\beta_2 + \mu_i & i = m + 1, \dots, n \end{cases} \quad (2.8)$$

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

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

$m, \beta_1, \beta_2, \rho,$ 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.

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

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 :

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

Où

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

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

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

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

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

$$H(\rho) = X'X + 1 = \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix}$$

$$= \begin{pmatrix} p_{11} + Z_1' Z_1 + \rho^2 X_m' & p_{12} - \rho^2 X_m' X_m \\ p_{21} - \rho^2 X_{m+1}' X_m & p_{22} + Z_2' Z_2 + \rho^2 X_{m+1}' X_m \end{pmatrix},$$

$$p = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix}, \tilde{B} = (\tilde{\beta}'_1, \tilde{\beta}'_2)'$$

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

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

$$P_\mu + X'_y = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} Z' & -\rho X_m' & 0 \\ 0 & X_{m+1}' & Z_2' \end{pmatrix}$$

$$= \begin{pmatrix} p_{11} \cdot \mu_1 + p_{12} \cdot \mu_2 + Z_1' V_1 - \rho X_m' (Y_{m+1} - \rho Y_m) \\ p_{21} \cdot \mu_1 + p_{22} \cdot \mu_2 + Z_2' V_2 + X_{m+1}' (Y_{m+1} - \rho Y_m) \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}$$

ii) La densité a posteriori marginale de β est :

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

Où $f(\beta)$ est une densité de Student à $2K$ dimensionnelle et à $(2a + n)$ degré 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 :

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

2.4 AUTRES MÉTHODES DE DÉTECTION DE RUPTURE

2.4.1 La détection de rupture par procédure de contraste pénalisé

C'est la méthode la plus courante pour la détection d'une où plusieurs ruptures, le principe de cette méthode est la minimisation ou maximisation d'un contraste construit à partir des données de processus $(X_t)_{t \in \mathbb{Z}}$, en rajoutant une fonction bien choisie qui dépend du nombre de paramètres de contraste.

Marc Lavielle et al (2000), dans leurs fameux travaux sur la détection de rupture au niveau de spectre d'un processus stationnaire se basent sur un contraste pénalisé pour détecter les ruptures, ce contraste construit a partir de la fonction Log-vraisemblance (contraste de Whittle), qui a fournit $O_p(n^{-1})$ comme vitesse de convergence. Bardet et al (2002) propose une autre vision sur le contraste utilisé, ils ont construit un contraste à partir de la régression de "Log-Log" de la variance des coefficients d'ondelettes, leur vitesse de convergence était acceptable mais pas bonne. La détection de rupture par programmation dynamique a été abordée aussi dans Bardet et al, le contraste pénalisé était construit à partir de la fonction de quasi-vraisemblance, qui a fournit une vitesse de convergence proche de $O_p(1)$. Dans notre travail nous adaptions le contraste pénalisé qui était utilisé dans Bardet et al avec plusieurs pénalités mais nous avons stabilisé sur une fonction de type $2 / \sqrt{n}$.

2.4.2 La règle CUSUM

Depuis son introduction par Page (1954), cette méthode a été développée par de nombreux littératures, grâce à son intérêt et sa simplicité. Plusieurs résultats d'optimalité étaient obtenus, citons Moustakides (1986) Lorden (1971), l'hypothèse fondamentale de la règle est l'indépendance entre les observations. D'autres travaux étaient proposés avec une dépendance entre les variables. Le principe de la règle de CUSUM est la maximisation du rapport entre la vraisemblance sous l'hypothèse H_0 et la vraisemblance sous l'hypothèse H_1 , qui sera comparé avec une valeur critique choisie par l'expérimentateur, le livre de Basseville, M. et Nikiforov est considéré comme une référence de CUSUM.

2.4.3 La règle de Shewhart

C'est le même principe de détection de rupture par la règle de CUSUM, le rapport de vraisemblance sur un échantillon de taille n est aussi utilisé ici, mais cette règle est considérée comme une méthode en ligne que nous abordons par la suite. Si le rapport de la vraisemblance calculé sous les deux hypothèses H_0 et H_1 est significatif par rapport à une valeur bien choisie d'un tableau statistique, alors une alarme sera déclenchée, dans ce cas on rejette l'hypothèse H_0 et on accepte l'hypothèse H_1 , c'est une sorte d'un temps d'arrêt.

Sinon on reconstruit un autre échantillon entre n et $t > n$ et on refait la même chose.

2.4.4 Détection hors ligne et en ligne de rupture

Il existe deux grandes axes qu'il faut signaler de la détection de rupture dans un processus aléatoire $(X_t)_{t \in \mathbb{Z}}$. La détection "hors ligne", qui veut dire qu'on doit prendre toutes les données du processus aléatoire $(X_t)_{t \in \mathbb{Z}}$, et on applique par la suite une des méthodes pour détecter le changement dans le paramètre sélectionné.

cette méthode est basée sur la maximisation ou la minimisation d'une fonction qui contient une partie ou bien toute l'information sur le paramètre sélectionné et l'instant de rupture correspondant à l'indice de maximum ou bien à l'indice de minimum. Cette méthode est très utilisée malgré les difficultés trouvées, soit au niveau de la description de la fonction qui contient l'information de paramètre (vraisemblance, quasi-vraisemblance, log-vraisemblances,...), ou bien au niveau de la sélection du point de rupture.

Une autre façon de repérer les ruptures est de prendre une partie d'un processus aléatoire est faire une sorte de surveillance Monitoring nous parlons ici de cadre "en ligne"; au lieu de prendre le processus aléatoire complet "hors ligne". Le traitement se fait à jour par apport à l'arrivée des données pour détecter un changement à l'intérieur du processus, les difficultés de cette procédure se représentent dans le fait de tomber sur un mauvais point de rupture ou bien un faux point de rupture, car nous traitons les données on fonction de temps de leurs arrivées, d'un autre coté cette procédure se base sur un test de rupture qui fait la différence entre les observations déjà testées et les observations qui arrivent, en parallèle nous devons fixer un temps d'arrêt pour ce 'test'.

Cette procédure se base sur un critère qui contient une fonction spécifique et bien choisie, nous signalons que certains littérateurs remplaçant cette fonction par une constante pour simplifier la procédure, un passage vers les mouvements browniens est nécessaire pour traiter la partie théorique.

Cette méthode contient une grande partie expérimentale par rapport à la partie théorique car son but est de tester à chaque fois les observations arrivées par rapport celles du passé (estimer, comparer les estimateurs obtenus à chaque fois, détecter la rupture, vérifier la fausse rupture), aussi elle prend du temps par rapport à la détection "hors ligne" mais elle est aussi en quelque sort pratique et rentable, de nombreux travaux sur la détection en ligne de rupture on été réalisés.

ESTIMATION BAYÉSIENNE D'UN POINT DE RUPTURE DANS UNE DISTRIBUTION EXPONENTIELLE

3.1 INTRODUCTION

La distribution exponentielle joue un rôle important dans les applications pratiques, comme le temps de défaillance d'un article spécial. Dans certains systèmes de fabrication, les produits sont soumis à des changements brusques dans la fonction de taux de défaillance, qui sont observés en raison des actions de maintenance.

Dans tels cas, il est important de déterminer le moment du changement de taux de défaillance des articles et le point observé est connu sous le nom de point de rupture.

Dans ce chapitre basé sur l'article de MS.Fallahnezhad, B.Rasti, MH.Aboodie (2013), nous nous intéressons à l'estimation Bayésienne d'un point de rupture du paramètre d'une distribution exponentielle. Pour cela nous cherchons la densité a posteriori des paramètres du modèle étudié, puis nous déduisons l'estimateur de Bayes de ce point en utilisant différentes fonctions coûts.

3.2 PRÉSENTATION DU MODÈLE

soit (X_1, \dots, X_n) une suite de variables aléatoires de loi exponentielle de paramètre λ_i , $\lambda_i > 0$, $i = \overline{1, 2}$,

avec :

$$X_1, \dots, X_m \sim \exp(\lambda_1)$$

$$X_{m+1}, \dots, X_n \sim \exp(\lambda_2)$$

avec : $\lambda_1 \neq \lambda_2$

Le paramètre m est un point de rupture, c'est un paramètre aléatoire prenant les valeurs $1, 2, \dots, n - 1$, n étant la taille de l'échantillon.

3.3 ESTIMATION BAYÉSIENNE

La méthode du maximum de vraisemblance ainsi que d'autres approches classiques sont basées uniquement sur les informations empiriques fournies par les données. Cependant, quand il y'a une connaissance technique des paramètres de la distribution ; la méthode Bayésienne apparaît comme une approche attrayante devant les difficultés théoriques des approches classiques.

La procédure de Bayes est basée sur la densité a posteriori jointe des différents paramètres $\pi(\lambda_1, \lambda_2, m|X)$, qui est proportionnelle au produit de la fonction de vraisemblance et la densité jointe a priori :

$$\pi(\lambda_1, \lambda_2, m|X) = \frac{L(\lambda_1, \lambda_2, m|X) g(\lambda_1, \lambda_2, m)}{\sum_{m=1}^{n-1} \int_0^{+\infty} \int_0^{+\infty} L(\lambda_1, \lambda_2; |X) g(\lambda_1, \lambda_2, m) d\lambda_1 d\lambda_2}$$

Avec :

$L(\lambda_1, \lambda_2, m|X)$ la fonction de vraisemblance des différents paramètres.

et

$g(\lambda_1, \lambda_2, m)$ la densité jointe a priori de m , λ_1 et λ_2

3.3.1 La fonction de la vraisemblance

comme :

$$\begin{cases} X_i \sim \exp(\lambda_1), & i = 1, \dots, m \\ X_i \sim \exp(\lambda_2), & i = m + 1, \dots, n \end{cases}$$

alors

$$f(X_i) = \begin{cases} \lambda_1 e^{-\lambda_1 x_i} & , \quad i = 1, \dots, m \\ \lambda_2 e^{-\lambda_2 x_i} & , \quad i = m + 1, \dots, n \end{cases}$$

Donc la fonction de vraisemblance est donnée par :

$$L(\lambda_1, \lambda_2, m|X) = \lambda_1 e^{-\lambda_1 x_1} \dots \lambda_1 e^{-\lambda_1 x_m} \cdot \lambda_2 e^{-\lambda_2 x_{m+1}} \dots \lambda_2 e^{-\lambda_2 x_n}$$

Ce qui nous donne :

$$L(\lambda_1, \lambda_2, m|X) = \lambda_1^m e^{-\lambda_1(x_1 + \dots + x_m)} \lambda_2^{n-m} e^{-\lambda(x_{m+1} + \dots + x_n)}$$

Alors :

$$L(\lambda_1, \lambda_2, m|X) = \lambda_1^m e^{-\lambda_1 \sum_{i=1}^m x_i} \lambda_2^{n-m} e^{-\lambda \sum_{i=m+1}^n x_i}$$

3.3.2 Les lois a priori

Les distributions a priori des paramètres inconnus sont attribuées comme suit :

Une loi gamma pour les λ_i avec $i = 1, 2$. de paramètres a_i et b_i , avec a_i et b_i sont deux réels strictement positifs.

Une loi uniforme discrète sur l'intervalle $[1, n - 1]$ pour m .

De plus nous supposons que les paramètres λ_1, λ_2 , et m sont indépendants.

Ce qui nous donne :

$$\begin{cases} g_1(\lambda_1) = \frac{b_1^{a_1}}{\Gamma(a_1)} \lambda_1^{a_1-1} e^{-b_1 \lambda_1} ; & a_1, b_1 > 0 \\ g_2(\lambda_2) = \frac{b_2^{a_2}}{\Gamma(a_2)} \lambda_2^{a_2-1} e^{-b_2 \lambda_2} ; & a_2, b_2 > 0 \\ g_3(m) = \frac{1}{n-1} \end{cases}$$

La densité a priori jointe :

Comme les paramètres λ_1, λ_2, m sont indépendants alors la densité a posteriori jointe de λ_1, λ_2 et m est donnée par :

$$\begin{aligned} g(\lambda_1, \lambda_2, m) &= g_1(\lambda_1) * g_2(\lambda_2) * g_3(m) \\ &= \frac{b_1^{a_1}}{\Gamma(a_1)} \lambda_1^{a_1-1} e^{-b_1 \lambda_1} \cdot \frac{b_2^{a_2}}{\Gamma(a_2)} \lambda_2^{a_2-1} e^{-b_2 \lambda_2} \cdot \frac{1}{n-1} \\ &= \frac{1}{n-1} \cdot \frac{b_1^{a_1} b_2^{a_2}}{\Gamma(a_1) \Gamma(a_2)} \cdot \lambda_1^{a_1-1} e^{-b_1 \lambda_1} \cdot \lambda_2^{a_2-1} e^{-b_2 \lambda_2} \end{aligned}$$

donc :

$$g(\lambda_1, \lambda_2, m) = \frac{1}{n-1} \cdot \frac{b_1^{a_1} b_2^{a_2}}{\Gamma(a_1) \Gamma(a_2)} \cdot \lambda_1^{a_1-1} e^{-b_1 \lambda_1} \cdot \lambda_2^{a_2-1} e^{-b_2 \lambda_2}.$$

3.3.3 La densité a posteriori jointe :

En vertu du théorème de Bayes nous avons :

$$\pi(\lambda_1, \lambda_2, m|X) = \frac{L(\lambda_1, \lambda_2, m|X) g(\lambda_1, \lambda_2, m)}{\sum_{m=1}^{n-1} \int_0^{+\infty} \int_0^{+\infty} L(\lambda_1, \lambda_2; |X) g(\lambda_1, \lambda_2, m) d\lambda_1 d\lambda_2}$$

Posons :

$$h(X) = L(\lambda_1, \lambda_2, m|X) \cdot g(\lambda_1, \lambda_2, m)$$

et

$$F(X) = \sum_{m=1}^{n-1} \int_0^{+\infty} \int_0^{+\infty} L(\lambda_1, \lambda_2; |X) g(\lambda_1, \lambda_2, m) d\lambda_1 d\lambda_2$$

calcul de h(X)

$$\begin{aligned} h(X) &= \lambda_1^m e^{-\lambda_1 \sum_{i=1}^m x_i} \cdot \lambda_2^{n-m} e^{-\lambda_2 \sum_{i=m+1}^n x_i} \cdot \frac{1}{n-1} \cdot \frac{b_1^{a_1} b_2^{a_2}}{\Gamma(a_1) \Gamma(a_2)} \cdot \lambda_1^{a_1-1} e^{-b_1 \lambda_1} \cdot \lambda_2^{a_2-1} e^{-b_2 \lambda_2} \\ &= \frac{1}{n-1} \cdot \frac{b_1^{a_1} b_2^{a_2}}{\Gamma(a_1) \Gamma(a_2)} \lambda_1^{m+a_1-1} \cdot e^{-\lambda_1 (\sum_{i=1}^m x_i + b_1)} \cdot \lambda_2^{n-m+a_2-1} \cdot e^{-\lambda_2 (\sum_{i=m+1}^n x_i + b_2)}. \end{aligned}$$

calcul de F(X)

$$F(X) = \sum_{m=1}^{n-1} \int_0^{+\infty} \int_0^{+\infty} \frac{1}{n-1} \cdot \frac{b_1^{a_1} b_2^{a_2}}{\Gamma(a_1) \Gamma(a_2)} \lambda_1^{m+a_1-1} \cdot e^{-\lambda_1 (\sum_{i=1}^m x_i + b_1)} \cdot \lambda_2^{n-m+a_2-1} \cdot e^{-\lambda_2 (\sum_{i=m+1}^n x_i + b_2)} \cdot d\lambda_1 \cdot d\lambda_2$$

$$F(X) = \sum_{m=1}^{n-1} \frac{1}{n-1} \cdot \frac{b_1^{a_1} b_2^{a_2}}{\Gamma(a_1) \Gamma(a_2)} \int_0^{+\infty} \lambda_1^{m+a_1-1} \cdot e^{-\lambda_1 (\sum_{i=1}^m x_i + b_1)} d\lambda_1 \cdot \int_0^{+\infty} \lambda_2^{n-m+a_2-1} \cdot e^{-\lambda_2 (\sum_{i=m+1}^n x_i + b_2)} d\lambda_2$$

posons :

$$I_1 = \int_0^{+\infty} \lambda_1^{m+a_1-1} \cdot e^{-\lambda_1 (\sum_{i=1}^m x_i + b_1)} d\lambda_1$$

et

$$I_2 = \int_0^{+\infty} \lambda_2^{n-m+a_2-1} \cdot e^{-\lambda_2 (\sum_{i=m+1}^n x_i + b_2)} d\lambda_2$$

Calcul de l'intégrale I_1

$$I_1 = \frac{\Gamma(m+a_1)}{(\sum_{i=1}^m x_i + b_1)^{m+a_1}} \cdot \int_0^{+\infty} \frac{(\sum_{i=1}^m x_i + b_1)^{m+a_1}}{\Gamma(m+a_1)} \cdot \lambda_1^{m+a_1-1} \cdot e^{-\lambda_1 (\sum_{i=1}^m x_i + b_1)} d\lambda_1$$

d'où

$$I_1 = \frac{\Gamma(m+a_1)}{(\sum_{i=1}^m x_i + b_1)^{m+a_1}}$$

Calcul de l'intégrale I_2

$$I_2 = \frac{\Gamma(n-m+a_2)}{(\sum_{i=m+1}^n x_i + b_2)^{n-m+a_2}} \int_0^{+\infty} \frac{(\sum_{i=m+1}^n x_i + b_2)^{n-m+a_2}}{\Gamma(n-m+a_2)} \cdot \lambda_2^{n-m+a_2-1} \cdot e^{-\lambda_2 (\sum_{i=m+1}^n x_i + b_2)} \cdot d\lambda_2$$

d'où

$$I_2 = \frac{\Gamma(n-m+a_2)}{(\sum_{i=m+1}^n x_i + b_2)^{n-m+a_2}}$$

ce qui nous donne:

$$F(X) = \sum_{m=1}^{n-1} \frac{1}{n-1} \cdot \frac{b_1^{a_1} b_2^{a_2}}{\Gamma(a_1) \Gamma(a_2)} \cdot \frac{\Gamma(m+a_1)}{(\sum_{i=1}^m x_i + b_1)^{m+a_1}} \cdot \frac{\Gamma(n-m+a_2)}{(\sum_{i=m+1}^n x_i + b_2)^{n-m+a_2}}$$

donc la densité jointe a posteriori de λ_1 et λ_2 et m est donnée par:

$$\pi(\lambda_1, \lambda_2, m | X) = \frac{\frac{1}{n-1} \cdot \frac{b_1^{a_1} b_2^{a_2}}{\Gamma(a_1) \Gamma(a_2)} \lambda_1^{m+a_1-1} \cdot e^{-\lambda_1 (\sum_{i=1}^m x_i + b_1)} \cdot \lambda_2^{n-m+a_2-1} \cdot e^{-\lambda_2 (\sum_{i=m+1}^n x_i + b_2)}}{\sum_{m=1}^{n-1} \frac{1}{n-1} \cdot \frac{b_1^{a_1} b_2^{a_2}}{\Gamma(a_1) \Gamma(a_2)} \cdot \frac{\Gamma(m+a_1)}{(\sum_{i=1}^m x_i + b_1)^{m+a_1}} \cdot \frac{\Gamma(n-m+a_2)}{(\sum_{i=m+1}^n x_i + b_2)^{n-m+a_2}}}$$

$$\pi(\lambda_1, \lambda_2, m|X) = \frac{\frac{1}{n-1} \cdot \frac{b_1^{a_1} b_2^{a_2}}{\Gamma(a_1)\Gamma(a_2)} \lambda_1^{m+a_1-1} \cdot e^{-\lambda_1(\sum_{i=1}^m x_i + b_1)} \cdot \lambda_2^{n-m+a_2-1} \cdot e^{-\lambda_2(\sum_{i=m+1}^n x_i + b_2)}}{\frac{1}{n-1} \cdot \frac{b_1^{a_1} b_2^{a_2}}{\Gamma(a_1)\Gamma(a_2)} \sum_{m=1}^{n-1} \frac{\Gamma(m+a_1)}{(\sum_{i=1}^m x_i + b_1)^{m+a_1}} \cdot \frac{\Gamma(n-m+a_2)}{(\sum_{i=m+1}^n x_i + b_2)^{n-m+a_2}}}$$

$$\pi(\lambda_1, \lambda_2, m|X) = \frac{\lambda_1^{m+a_1-1} \cdot e^{-\lambda_1(\sum_{i=1}^m x_i + b_1)} \cdot \lambda_2^{n-m+a_2-1} \cdot e^{-\lambda_2(\sum_{i=m+1}^n x_i + b_2)}}{\sum_{m=1}^{n-1} \frac{\Gamma(m+a_1)}{(\sum_{i=1}^m x_i + b_1)^{m+a_1}} \cdot \frac{\Gamma(n-m+a_2)}{(\sum_{i=m+1}^n x_i + b_2)^{n-m+a_2}}}$$

d'où:

$$\pi(\lambda_1, \lambda_2, m|X) = \frac{\lambda_1^{m+a_1-1} \cdot e^{-\lambda_1(\sum_{i=1}^m x_i + b_1)} \cdot \lambda_2^{n-m+a_2-1} \cdot e^{-\lambda_2(\sum_{i=m+1}^n x_i + b_2)}}{\Psi(a_1, a_2, b_1, b_2, m, n)}$$

avec:

$$\Psi(a_1, a_2, b_1, b_2, m, n) = \sum_{m=1}^{n-1} \frac{\Gamma(m+a_1)}{(\sum_{i=1}^m x_i + b_1)^{m+a_1}} \cdot \frac{\Gamma(n-m+a_2)}{(\sum_{i=m+1}^n x_i + b_2)^{n-m+a_2}}$$

3.3.4 Les densités a posteriori marginales

La distribution a posteriori marginale du point λ_1

$$\pi(\lambda_1|X) = \sum_{m=1}^n \int_0^{+\infty} \pi(\lambda_1, \lambda_2, m|X) d\lambda_2$$

d'où

$$\pi(\lambda_1 | X) = \frac{\sum_{m=1}^n \left(\lambda_1^{m+a_1-1} \cdot e^{-\lambda_1(b_1 + \sum_{i=1}^m x_i)} \frac{\Gamma(n-m+a_2)}{(b_2 + \sum_{i=m+1}^n x_i)^{n-m+a_2}} \right)}{\Psi(a_1, a_2, b_1, b_2, m, n)}$$

La distribution a posteriori marginale du point λ_2

$$\pi(\lambda_2|X) = \sum_{m=1}^n \int_0^{+\infty} \pi(\lambda_1, \lambda_2, m|X) d\lambda_1$$

d'où

$$\pi(\lambda_2 | X) = \frac{\sum_{m=1}^n \left(\lambda_2^{n-m+a_2-1} \cdot e^{-\lambda_2(\sum_{i=m+1}^n x_i + b_2)} \frac{\Gamma(m+a_1)}{(b_1 + \sum_{i=1}^m x_i)^{m+a_1}} \right)}{\Psi(a_1, a_2, b_1, b_2, m, n)}$$

La distribution a posteriori marginale du point m

$$\pi(m | X) = \int_0^{+\infty} \int_0^{+\infty} \pi(\lambda_1, \lambda_2, m | X) d\lambda_1 d\lambda_2$$

$$\pi(m | X) = \int_0^{+\infty} \int_0^{+\infty} \frac{\lambda_1^{m+a_1-1} \cdot e^{-\lambda_1(\sum_{i=1}^m x_i + b_1)} \cdot \lambda_2^{n-m+a_2-1} \cdot e^{-\lambda_2(\sum_{i=m+1}^n x_i + b_2)}}{\Psi(a_1, a_2, b_1, b_2, m, n)} d\lambda_1 d\lambda_2$$

$$\pi(m | X) = \frac{1}{\Psi(a_1, a_2, b_1, b_2, m, n)} \int_0^{+\infty} \int_0^{+\infty} \lambda_1^{m+a_1-1} \cdot e^{-\lambda_1(\sum_{i=1}^m x_i + b_1)} \cdot \lambda_2^{n-m+a_2-1} \cdot e^{-\lambda_2(\sum_{i=m+1}^n x_i + b_2)}$$

donc:

$$\pi(m | X) = \frac{\frac{\Gamma(m+a_1)}{(\sum_{i=1}^m x_i + b_1)^{m+a_1}} \cdot \frac{\Gamma(n-m+a_2)}{(\sum_{i=m+1}^n x_i + b_2)^{n-m+a_2}}}{\Psi(a_1, a_2, b_1, b_2, m, n)}$$

posons :

$$T(m) = \frac{\Gamma(m+a_1)}{(\sum_{i=1}^m x_i + b_1)^{m+a_1}} \cdot \frac{\Gamma(n-m+a_2)}{(\sum_{i=m+1}^n x_i + b_2)^{n-m+a_2}}$$

ce qui nous donne :

$$\pi(m | X) = \frac{T(m)}{\sum_{m=1}^{n-1} T(m)}$$

3.4 ESTIMATION BAYÉSIIENNE DU POINT DE RUPTURE SOUS LES FONCTIONS COÛTS USUELLES.

3.4.1 La fonction coût quadratique

La fonction coût quadratique est définie par :

$$L_1(\alpha, d) = (\alpha - d)^2$$

L'estimateur de Bayes m^* associé à la distribution marginale $\pi(m | X)$ sous la fonction coût quadratique est la moyenne a posteriori de m .

L'estimateur de Bayes est :

$$m^* = E(m) = \sum_{m=1}^{n-1} m \cdot \pi(m | X)$$

$$= \sum_{m=1}^{n-1} m \frac{\frac{\Gamma(m+a_1)}{(\sum_{i=1}^m x_i + b_1)^{m+a_1}} \cdot \frac{\Gamma(n-m+a_2)}{(\sum_{i=m+1}^n x_i + b_2)^{n-m+a_2}}}{\Psi(a_1, a_2, b_1, b_2, m, n)}$$

avec :

$$\Psi(a_1, a_2, b_1, b_2, m, n) = \sum_{m=1}^{n-1} \frac{\Gamma(m+a_1)}{(\sum_{i=1}^m x_i + b_1)^{m+a_1}} \cdot \frac{\Gamma(n-m+a_2)}{(\sum_{i=m+1}^n x_i + b_2)^{n-m+a_2}}$$

donc :

$$m^* = \sum_{m=1}^{n-1} m \frac{T(m)}{\sum_{m=1}^{n-1} T(m)}$$

avec :

$$T(m) = \frac{\Gamma(m+a_1)}{(\sum_{i=1}^m x_i + b_1)^{m+a_1}} \cdot \frac{\Gamma(n-m+a_2)}{(\sum_{i=m+1}^n x_i + b_2)^{n-m+a_2}}$$

3.4.2 La fonction coût absolu

La fonction coût absolu est la fonction définie par :

$$L_2(\alpha, d) = |\alpha - d|$$

L'estimateur de Bayes de point m sous la fonction coût absolu est la médiane de la distribution a posteriori $\pi(m|X)$.

3.4.3 La fonction coût 0-1

La fonction coût 0-1 est la fonction définie par :

$$\begin{cases} 0 & |\alpha - d| < \varepsilon, \varepsilon > 0 \\ 1 & \text{sinon} \end{cases}$$

L'estimateur de Bayes de m sous cette fonction coût 0-1 est le mode de la distribution a posteriori $\pi(m|X)$.

3.5 ESTIMATION BAYÉSIENNE SOUS LA FONCTION COÛT LINEX

Dans cette section, nous nous intéressons à l'estimateur de Bayes de m sous la fonction coût Linex. Cette fonction est définie par :

$$L_c(\alpha, d) = \exp\{c(\alpha, d)\} - c(\alpha, d) - 1$$

avec : $c \neq 0$

L'estimateur Bayésien sous cette fonction coût est donné par :

$$\hat{d} = -\frac{1}{c} \ln \left\{ E_{\pi(\alpha|X)}(e^{-c\alpha}) \right\}$$

avec :

$E_{\pi(\alpha|X)}$ est l'espérance a posteriori. Dans notre cas on cherche à estimer le point de rupture m :

soit \hat{m} l'estimateur Bayésien de m sous la fonction coût Linex.

$$\hat{m} = -\frac{1}{c} \ln \left\{ E_{\pi(m|X)}(e^{-cm}) \right\}$$

on a :

$$E_{\pi(m|X)}(e^{-cm}) = \sum_{m=1}^{n-1} (e^{-cm}) \cdot \pi(m|X)$$

$$E_{\pi(m|X)}(e^{-cm}) = \sum_{m=1}^{n-1} (e^{-cm}) \cdot \left(\frac{T(m)}{\sum_{m=1}^{n-1} T(m)} \right)$$

avec :

$$T(m) = \frac{\Gamma(m+a_1)}{(\sum_{i=1}^m x_i + b_1)^{m+a_1}} \cdot \frac{\Gamma(n-m+a_2)}{(\sum_{i=m+1}^n x_i + b_2)^{n-m+a_2}}$$

donc :

$$\hat{m} = -\frac{1}{c} \ln \left\{ \sum_{m=1}^{n-1} (e^{-cm}) \cdot \left(\frac{T(m)}{\sum_{m=1}^{n-1} T(m)} \right) \right\}$$

3.6 APPLICATION ET SIMULATION

Considérons le modèle exponentiel suivant :

$$\begin{cases} X_i \sim \exp(\lambda_1), & i = 1, \dots, 10 \\ X_i \sim \exp(\lambda_2), & i = 10, \dots, 20 \end{cases} \quad (3,1)$$

Nous allons généré 20 observations issues de ce modèle avec $\lambda_1 = 2$ et $\lambda_2 = 3$.

Les valeurs des X_i retrouvées pour $i = \overline{1,20}$ sont données dans le tableau suivant :

(Le programme est donné en annexe)

I	X_i	I	X_i
1	0,9205063	11	0,20096753
2	0,8343803	12	0,24967753
3	0,8950511	13	0,76086410
4	0,2099929	14	0,37254216
5	0,6773078	15	0,51465687
6	0,1428695	16	0,66184728
7	0,3192391	17	0,22724731
8	0,3546899	18	0,07151217
9	0,8412991	19	0,06996478
10	0,7855346	20	0,11964631

TABLE 3.1 – Les observations générées a partir du modèle (3,1)

supposons maintenant que les $\lambda_i, i = \overline{1,2}$ sont distribués suivant une loi gamma de paramètres respectivement $a_1 = 1,5$ et $a_2 = 1,8, b_1 = 1,5$ et $b_2 = 2,0$

$$\begin{cases} \lambda_1 \sim \Gamma(a_1, b_1) \\ \lambda_2 \sim \Gamma(a_2, b_2) \end{cases}$$

La densité a posteriori de m est donnée par :

$$\pi(m|X) = \frac{T(m)}{\sum_{m=1}^{n-1} T(m)}$$

avec :

$$T(m) = \frac{\Gamma(m+a_1)}{(\sum_{i=1}^m x_i + b_1)^{m+a_1}} \cdot \frac{\Gamma(n-m+a_2)}{(\sum_{i=m+1}^n x_i + b_2)^{n-m+a_2}}$$

Les valeurs simulées de la distribution a posteriori marginale de m sont données par le tableau suivant :

m	$\pi(m X)$	m	$\pi(m X)$
1	0,07554930	11	0,02931353
2	0,07358928	12	0,02326958
3	0,09444903	13	0,03600483
4	0,04890792	14	0,03458956
5	0,05309451	15	0,04499595
6	0,03080065	16	0,11927080
7	0,02408565	17	0,09917026
8	0,02042946	18	0,04824509
9	0,02814173	19	0,03157229
10	0,04162351	20	0,04289706

TABLE 3.2 – Les valeurs simulées de la distribution a posteriori marginale de m.

3.6.1 Estimateur Bayésien sous les fonctions coûts quadratique

$$m^* = \sum_{m=1}^{n-1} m\pi(m /X) = 10,05736$$

Le tableau suivant nous donne la valeur de l'estimation m^* pour différentes valeurs de a_1, b_1 et a_2, b_2

(a_1, b_1)	(a_2, b_2)	m^*
(1,5 ; 1,5)	(1,8 ; 2,0)	10,05736
(1,75 ; 2)	(2 ; 2,25)	10,31827
(2 ; 2,5)	(2,20 ; 2,5)	9,206811
(0,001 ; 0,001)	(0,001 ; 0,001)	13,66451
(1 ; 0,59)	(1 ; 0,54)	11,56921
(1 ; 0,5)	(1 ; 0,33)	10,58278
(1 ; \bar{X}_1)	(1 ; \bar{X}_2)	10,03177

TABLE 3.3 – La valeur de l'estimateur m^*

3.6.2 Estimateur Bayésien sous la fonction coût absolu

A fin de calculer la médiane a posteriori, nous allons calculer les distributions a posteriori cumulées de m . Elles sont donnée dans le tableau suivant :

m	$F(m X)$	m	$F(m X)$
1	0,755493	11	0,5199846
2	0,1491386	12	0,5432542
3	0,2435876	13	0,5792590
4	0,2924955	14	0,6138485
5	0,3455900	15	0,6588544
6	0,3763907	16	0,7781153
7	0,4004763	17	0,8772856
8	0,4209058	18	0,9255306
9	0,4490475	19	0,9571029
10	0,4906710	20	1,0000000

TABLE 3.4 – La distribution a posteriori cumulée de m

D'après le tableau des distributions a posteriori cumulées de m :

La médiane a posteriori est égale à 10.

3.7 CONCLUSION

D'après les résultats de simulation retrouvés, nous remarquons que la méthode Bayésienne estime bien la valeur du point de rupture, nous remarquons également que peu importe le choix de la fonction coût (quadratique ou absolu), nous retrouvons des très bons estimateurs.

Nous avons fait varier les valeurs des paramètres a_i et b_i pour $i = \overline{1,2}$ nous remarquons que les résultats sont meilleurs avec :

$$a_1=a_2=1 \text{ et } b_1 = \overline{X_1} \text{ et } b_2 = \overline{X_2}$$

$$\text{avec } : \overline{X_1} = \frac{1}{1-m} \sum_{i=m+1}^n x_i \text{ et } \overline{X_2} = \frac{1}{1-m} \sum_{i=m+1}^n x_i$$

CONCLUSION GÉNÉRALE

Pour faire un bref retour sur les chapitres de ce mémoire, nous avons vu quelques généralités sur les modèles de rupture. Et nous avons donné deux exemples dans différents domaines d'application.

En suite nous avons aborder quelques méthodes de détection de rupture parmi ces méthodes, nous retrouvons les tests statistiques et quelques méthodes d'estimation.

La méthodologie bayésienne apporte une grande souplesse dans les méthodologies statistique, elle a été largement sollicitée dans des problèmes de détection de rupture.

Dans la partie application nous avons traité le cas d'une rupture dans une séquence d'observations d'une loi exponentielle.

On adoptons l'approche Bayésienne nous avons calculer l'estimateur de point de rupture sous différentes fonctions coût.

Comme perspectives il serais intéressant d'étendre ce travail on utilisant d'autres fonctions coût et pour quoi pas d'autres lois a priori plus générales a fin d'améliorer l'estimation.

BIBLIOGRAPHIE

- Emad-Eldin AA Aly et Sana S BuHamra. Rank tests for two change points. *Computational statistics & data analysis*, 22(4) :363–372, 1996.
- Osamu Amano, Kenichi Mizobe, Yasuhiko Bando, et Koji Sakiyama. Anatomy and histology of rodent and human major salivary glands—overview of the japan salivary gland society-sponsored workshop—. *Acta histochemica et cytochemica*, 45(5) :241–250, 2012.
- Fabrice Bardet et Roser Cussó. Les essais randomisés contrôlés, révolution des politiques de développement? une évaluation par la banque mondiale de l’empowerment au bangladesh. *Revue française de socio-économie*, (2) :175–198, 2012.
- Jean-Marc Bardet et William Kengne. Monitoring procedure for parameter change in causal time series. *Journal of Multivariate Analysis*, 125 :204–221, 2014.
- Eduard Belitser et Farida Enikeeva. Empirical bayesian test of the smoothness. *Mathematical Methods of Statistics*, 17(1) :1–18, 2008.
- Gustavo Cavalcanti, Marlene Doring, Marilene Rodrigues Portella, Emanuely Casal Bortoluzzi, Andreia Mascarelo, et Marcos Paulo Dellani. Multimorbidity associated with polypharmacy and negative self-perception of health. *Revista Brasileira de Geriatria e Gerontologia*, 20 :634–642, 2017.
- Calvin Yu-Chian Chen. Tcm database@ taiwan : the world’s largest traditional chinese medicine database for drug screening in silico. *PloS one*, 6(1) :e15939, 2011.
- Belkacem Cherifa. *Stabilité des modèles de rupture sous contamination*. PhD thesis, Université de Tizi Ouzou-Mouloud Mammeri, 2013.
- Herman Chernoff et Shelemyahu Zacks. Estimating the current mean of a normal distribution which is subjected to changes in time. *The Annals of Mathematical Statistics*, 35(3) :999–1018, 1964.
- Alin Ciuperca, George A Elliott, et Luis Santiago. On inductive limits of type- i c^* -algebras with one-dimensional spectrum. *International Mathematics Research Notices*, 2011(11) :2577–2615, 2011.
- Bo Dalsgaard, Matthias Schleuning, Pietro K Maruyama, D Matthias Dehling, Jesper Sonne, Jeferson Vizentin-Bugoni, Thais B Zanata, Jon Fjeldså, Katrin Böhning-Gaese, et Carsten Rahbek. Opposed latitudinal patterns of network-derived and dietary specialization in avian plant–frugivore interaction systems. *Ecography*, 40(12) :1395–1401, 2017.
- Herold Dehling, Aeneas Roach, et Murad S Taqqu. Non-parametric change-point tests for long-range dependent data. *Scandinavian Journal of Statistics*, 40(1) :153–173, 2013.

- Timothy F Elsmore, Dennis L Reeves, et Andrea N Reeves. The ares® test system for palm os handheld computers. *Archives of Clinical Neuropsychology*, 22(Suppl_1) :S135–S144, 2007.
- M Fallahnezhad, B Rasti, et M Abooie. Improving the performance of bayesian estimation methods in estimations of shift point and comparison with maximum likelihood estimation approach. *a a*, 1(2) :1–1, 2013.
- Christos B Fotopoulos et Evangelos L Psomas. The impact of “soft” and “hard” tqm elements on quality management results. *International Journal of Quality & Reliability Management*, 2009.
- Christian Francq et Jean-Michel Zakoïan. Strict stationarity testing and estimation of explosive and stationary generalized autoregressive conditional heteroscedasticity models. *Econometrica*, 80(2) :821–861, 2012.
- Edit Gombay et Daniel Serban. Monitoring parameter change in ar (p) time series models. *Journal of Multivariate Analysis*, 100(4) :715–725, 2009.
- Abdellatif Guenaizi. *Estimation semi-paramétrique des ruptures offline et online d’un processus longue mémoire*. PhD thesis, 2021.
- Wolfgang Härdle, Enno Mammen, et Marlene Müller. Testing parametric versus semiparametric modeling in generalized linear models. *Journal of the American Statistical Association*, 93(444) :1461–1474, 1998.
- Wolfgang Härdle et Alexandre Tsybakov. Local polynomial estimators of the volatility function in nonparametric autoregression. *Journal of econometrics*, 81(1) :223–242, 1997.
- Roger Hawkins. *Second language syntax : A generative introduction*. Wiley-Blackwell, 2001.
- Don Holbert et Lyle D Broemeling. Bayesian inferences related to shifting sequences and two-phase regression . 1977.
- Cédric Hufnagel, Patrick Chambres, Pierre R Bertrand, et Frédéric Dutheil. The need for objective measures of stress in autism. *Frontiers in psychology*, 8 :64, 2017.
- Dongeun Huh, Benjamin D Matthews, Akiko Mammoto, Martín Montoya-Zavala, Hong Yuan Hsin, et Donald E Ingber. Reconstituting organ-level lung functions on a chip. *Science*, 328(5986) :1662–1668, 2010.
- Daniela Jarušková. Some problems with application of change-point detection methods to environmental data. *Environmetrics : The official journal of the International Environmetrics Society*, 8(5) :469–483, 1997.
- Rebecca Killick et Idris Eckley. changepoint : An r package for changepoint analysis. *Journal of statistical software*, 58(3) :1–19, 2014.
- Marc Lavielle. Optimal segmentation of random processes. *IEEE Transactions on Signal Processing*, 46(5) :1365–1373, 1998.
- Marc Lavielle et Eric Moulines. Least-squares estimation of an unknown number of shifts in a time series. *Journal of time series analysis*, 21(1) :33–59, 2000.
- Chin Choy LD, and Broemeling. Some bayesian inferences for a changing linear model. *Technometrics*, 22(1) :71–78, 1980.

- Yves Le Bissonnais, Jacques Thorette, Cécile Bardet, Joël Daroussin, et al. L'érosion hydrique des sols en France. *Rapport INRA, IFEN*, 106, 2002.
- Alexandre Leroux. Méthode bayésienne de détection de rupture et/ou de tendance pour des données temporelles. 2016.
- Gary Lorden. Procedures for reacting to a change in distribution. *The Annals of Mathematical Statistics*, pages 1897–1908, 1971.
- George V Moustakides. Optimal stopping times for detecting changes in distributions. *The Annals of Statistics*, 14(4) :1379–1387, 1986.
- Joseph Ngatchou-Wandji et Marwa Ltaifa. Détection de rupture dans des données climatiques et dans des données de santé. *Biostatistiques et sciences de la santé*, 2(1).
- Joseph Ngatchou-Wandji et Marwa Ltaifa. On detecting weak changes in the mean of Charn models. *arXiv preprint arXiv :2101.08597*, 2021.
- ES Page. A test for a change in a parameter occurring at an unknown point. *Biometrika*, 42(3/4) :523–527, 1955.
- Alfredas Račkauskas et Charles Suquet. Hölder norm test statistics for epidemic change. *Journal of Statistical Planning and Inference*, 126(2) :495–520, 2004.
- F Mark Rafaty, Richard L Goode, et Willard E Fee. The brow-lift operation. *Archives of Otolaryngology*, 101(8) :467–468, 1975.
- Jaxk Reeves, Jien Chen, Xiaolan L Wang, Robert Lund, et Qi Qi Lu. A review and comparison of changepoint detection techniques for climate data. *Journal of Applied Meteorology and Climatology*, 46(6) :900–915, 2007.
- Diego Salazar, Lyle Broemeling, et Albert Chi. Parameter changes in a regression model with autocorrelated errors. *Communications in Statistics-Theory and Methods*, 10(17) :1751–1758, 1981.
- Yang Song. What should economists know about the current Chinese hukou system? *China Economic Review*, 29 :200–212, 2014.
- Teruhiko Terasawa, Joseph Lau, Stéphane Bardet, Olivier Couturier, Tomomitsu Hotta, Martin Hutchings, Takashi Nishishi, et Hirokazu Nagai. Fluorine-18-fluorodeoxyglucose positron emission tomography for interim response assessment of advanced-stage Hodgkin's lymphoma and diffuse large B-cell lymphoma : a systematic review. *J Clin Oncol*, 27(11) :1906–14, 2009.
- Marie-Françoise Thorel, Denise Auclair, Huguette Petit, et Jacqueline Vandevelde. Tuberculose de la chèvre : diagnostic biologique. Dans *Annales de Recherches Veterinaires*, volume 11, pages 251–257, 1980.
- Susan Van Dieren, JWJ Beulens, AP Kengne, LM Peelen, GEHM Rutten, Mark Woodward, YT Van der Schouw, et KGM Moons. Prediction models for the risk of cardiovascular disease in patients with type 2 diabetes : a systematic review. *Heart*, 98(5) :360–369, 2012.
- Xiaolan L Wang, Qiuzi H Wen, et Yuehua Wu. Penalized maximal t test for detecting undocumented mean change in climate data series. *Journal of Applied Meteorology and Climatology*, 46(6) :916–931, 2007.

Zheng Wang, John M Tchernev, et Tyler Solloway. A dynamic longitudinal examination of social media use, needs, and gratifications among college students. *Computers in human behavior*, 28(5) :1829–1839, 2012.

Jan R Williams, Susan Frances Haka, Mark S Bettner, et Joseph V Carcello. *Financial and managerial accounting*. China Machine Press, 2005.

Jian Zhou. *Multicatalyst system in asymmetric catalysis*. John Wiley & Sons, 2014.

ملخص

يتم تعريف الفاصل على أنه تغيير في البيانات من أ نقطة محددة. متوسط البيانات في معظم الأحيان قبل وبعد الانفصال لن يكون هو نفسه. في هذا العمل نهتم باكتشاف التمزق بواسطة نهج بايز الذي يستخدم على نطاق واسع في هذا المجال.

Résumé

La rupture est définie comme un changement des données a partir d'un certain point . la majeure partie du temps ,la moyenne des données avant et après la rupture ne seront pas les mêmes .

Dans ce travail, nous nous intéressons à la détection de rupture par l'approche Bayésienne qui est largement sollicité dans ce domaine.

Abstract

The break is defined as a change in the data from a certain point. point. Most of the time, the average of the data before and after before and after the break will not be the same.

In this work, we are interested in the detection of rupture by the Bayesian approach Bayesian approach which is widely used in this field.