

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
UNIVERSITE MOULOU D MAMMERI, TIZI-OUZOU
FACULTE DES SCIENCES
Département de Mathématiques



Mémoire de Master

Domaine : Mathématiques

Option : Probabilités et statistique

Rupture épidémique dans les processus autorégressifs d'ordre 1 à innovations faiblement dépendantes

Présenté par
Melle Hideur Lydia

Devant le jury d'examen composé de :

M. Mehiri Mohammed Maître assistant A, U.M.M.T.O; Président
M. Graiche Farid Maître de conférence B, U.M.M.T.O; Encadreur
M. Yousfi Smail Maître de Conférence B, U.M.M.T.O, Examinateur

Remerciements

Je remercie Dieu le Tout-Puissant de m'avoir accordé la patience, la force et la volonté de mener à bien ce mémoire.

Je remercie tout d'abord les membres du jury M.Mehiri Mohammed et M.Yousfi Smail, pour le temps que vous avez consacré à écouter et évaluer ce travail.

Un immense merci à mon encadreur M.Graïche Farid, pour sa disponibilité, sa patience et ses précieux conseils tout au long de ce travail. Votre accompagnement bienveillant et vos orientations m'ont été d'une aide inestimable pour structurer mes idées et avancer avec clarté.

Je remercie sincèrement l'ensemble des enseignants de la faculté des Sciences, pour la qualité de l'enseignement qu'ils nous ont dispensé tout au long de notre formation.

Dédicaces

Je dédie ce modeste travail :

À mes chers parents,

Pour votre amour infini, vos sacrifices et votre soutien inconditionnel. Vous êtes ma première source d'inspiration et ma plus grande fierté.

Merci de m'avoir toujours soutenue avec amour, patience et dévouement.

Merci d'avoir cru en moi bien avant que je n'y croie moi-même.

À ma sœur chérie,

Complice de mes rires et témoin de mes larmes, ton soutien discret mais constant a été comme une mélodie douce accompagnant chaque page de ce mémoire.

À toute ma famille, mes oncles, tantes et mes cousins, ainsi que ma grand-mère maternelle. Leur soutien tout au long de cette année était très précieux pour moi.

À ma meilleure amie d'enfance Sarah,

Toi qui as gardé vivants les souvenirs de nos jeux insouciantes, merci d'avoir su, même adulte, garder cette précieuse innocence qui m'a rappelé l'essentiel quand le sérieux prenait trop de place.

À mon fiancé et sa merveilleuse famille, merci de m'avoir accueillie comme l'une des vôtres. Votre bienveillance et vos encouragements ont été un cadeau précieux.

À vous, mes chers camarades de la promotion Master 2 Probabilités et Statistique (Amel.K, Amel.L, Ryma, Soumia, Lila, Rachida, Hanane, Lisa, Thanina, Karima), merci pour cette aventure partagée. Nos échanges, nos doutes surmontés ensemble et nos moments de complicité ont rendu ce parcours plus riche et moins solitaire. Que nos chemins se croisent encore longtemps !.

À la mémoire de ma chère grand-mère, dont l'amour, la tendresse et les prières m'accompagnent encore aujourd'hui. Ton souvenir reste vivant dans mon cœur, et ce travail te rend hommage. Puisse ton âme reposer en paix.

Table des matières

Introduction Générale	5
1 Généralités sur les Processus ARMA	7
1.1 Rappel sur les Séries Chronologiques	7
1.1.1 Définition	7
1.1.2 Bruit Blanc	7
1.1.3 Stationnarité	8
1.1.4 Causalité et inversibilité	9
1.1.5 Opérateur retard	10
1.1.6 Fonction d'autocovariance et d'autocorrélation	10
1.2 Processus autorégressif d'ordre p (AR(p))	10
1.2.1 Définition	11
1.2.2 Conditions de Stationnarité	11
1.2.3 Moments d'un Processus AR(p)	12
1.2.4 Causalité et inversibilité	13
1.2.5 Estimation des paramètres	13
1.3 Processus moyenne mobile	18
1.3.1 Définition	18
1.3.2 Moments d'un processus MA(q)	19
1.3.3 Propriétés	19
1.3.4 Estimation des paramètres	20
1.4 Processus autorégressifs moyenne mobile	21
1.4.1 Définition	21
1.4.2 Moments des processus ARMA	21
1.4.3 Propriétés	22
1.4.4 Estimation des paramètres	22
2 Tests de rupture épidémique dans la moyenne de variables aléatoires faiblement dépendantes	24
2.1 Introduction	24
2.2 Statistiques de tests classiques	25
2.2.1 Statistique de Chen et Zhou	25
2.2.2 Statistique de Levin et Kline	26
2.3 Statistiques de tests Hölderiennes	26
2.3.1 Espaces de Hölder	26
2.3.2 Statistiques de Račkauskas et Suquet	27
2.4 Convergence et consistence	28

2.4.1 Cas de variables aléatoires indépendantes	28
2.4.2 Cas de variables aléatoires faiblement dépendantes	28
2.5 Étude de simulation	29
2.5.1 Cas de variables aléatoires indépendantes	29
2.5.2 Cas de variables aléatoires faiblement dépendantes	31
2.6 Application aux données réelles	32
3 Tests de rupture épidémique dans un processus AR(1) à innovations faiblement dépendantes	34
3.1 Introduction	34
3.2 Convergence et consistance	35
3.3 Étude de simulation	36
3.4 Application aux données réelles	40
Conclusion Générale	42
Annexe A	43
A.1 Mouvement brownien	43
A.1.1 Définition	43
A.1.2 Les accroissements du mouvement brownien	43
A.1.3 Propriétés du mouvement brownien	44
A.2 Pont brownien	45
Annexe B	47
B.1 Programme de simulation des statistiques	47
B.2 Programme de simulation des puissances	50

Introduction Générale

Dans de nombreux phénomènes évoluant dans le temps, il arrive que le comportement du système change **brusquement** à un moment donné. En statistique, ces changements, appelés **ruptures**, correspondent à des modifications significatives des propriétés statistiques du processus, telles que la moyenne, la variance ou encore la structure de dépendance temporelle. Le moment précis où survient une telle modification est désigné par le terme de **point de rupture** (change point).

Parmi les différentes formes de ruptures étudiées en statistique, la rupture épidémique occupe une place particulière. Elle se caractérise par un changement temporaire dans le comportement du processus, suivi d'un retour à son état initial. Plusieurs travaux dans la littérature ont été consacrés à la détection de ruptures épidémiques, notamment ceux de Levine et Kline [11], Gombay [8], Yao [16], Avery et Henderson [1], ainsi que Račkauskas et Suquet [14].

Le travail de ce mémoire porte sur les tests statistiques pour la détection de rupture épidémique dans la moyenne d'un processus autorégressif d'ordre 1 (AR(1)), dans le cas où les innovations sont faiblement dépendantes. Il est structuré en trois chapitres et deux annexes :

Le **premier** chapitre est consacré à un rappel sur les séries chronologiques. Nous présentons d'abord les notions fondamentales nécessaires à l'analyse de ce type de modèles, notamment la stationnarité, ainsi que les outils descriptifs tels que les fonctions d'autocovariance, d'autocorrélation et d'autocorrélation partielle. Nous introduisons ensuite les principaux modèles linéaires classiques, en particulier les modèles autorégressifs AR(p), les modèles à moyenne mobile MA(q), ainsi que les modèles mixtes ARMA(p,q). Pour chaque modèle, nous donnons la définition formelle, les principales propriétés, les méthodes d'estimation, ainsi qu'un exemple d'application sur des données réelles qui peuvent être modélisées par chacun de ces modèles.

Le **deuxième** chapitre est consacré à l'étude des tests statistiques de détection de rupture épidémique dans la moyenne d'une suite de variables aléatoires faiblement dépendantes. Nous présentons d'abord les statistiques classiques, notamment celles proposées par Levine et Kline [11], ainsi que par Chen et Zhou [4]. Nous abordons ensuite les statistiques hölderiennes, proposées par Račkauskas et Suquet [14], en introduisant brièvement les espaces de Hölder sur lesquels ces statistiques sont définies. L'intérêt des statistiques hölderiennes est qu'elles détectent des épidémies plus courtes que les statistiques classiques. Une étude de simulation est ensuite réalisée afin de comparer les performances et les puissances de ces différentes statistiques dans le cas de variables indépendantes et faiblement dépendantes. Le chapitre se conclut par une application sur des données réelles.

Le **troisième** chapitre est consacré à la détection de rupture épidémique dans la moyenne d'un processus autorégressif d'ordre 1, dans le cas où les innovations sont faiblement dépendantes. Nous appliquons

les statistiques hölderiennes présentées dans le chapitre 2, en les adaptant au contexte spécifique des processus AR(1). Afin d'évaluer la performance et la puissance de ces tests, nous réalisons une étude de simulation en considérant deux types d'innovations : gaussiennes et uniformes sur $[-1,1]$, ainsi que pour deux valeurs différentes du coefficient autorégressif ϕ . Nous concluons ce chapitre par une application sur des données réelles.

L'annexe A présente quelques propriétés du mouvement brownien et du pont brownien, utiles pour la compréhension théorique des tests étudiés dans ce mémoire.

L'annexe B regroupe les programmes de simulation utilisés pour la mise en œuvre des tests et le calcul de leurs puissances. Ces programmes, développés en langage R, permettent de reproduire les résultats numériques présentés dans les chapitres 2 et 3.

Le mémoire se termine par une conclusion et quelques perspectives de recherche.

Chapitre 1

Généralités sur les Processus ARMA

1.1 Rappels sur les Séries Chronologiques

1.1.1 Définition

On appelle une série chronologique (ou série temporelle) une suite d'observations d'une famille de variables aléatoires évoluant dans le temps, qu'on note $\{X_t, t \in I\}$, où I est un intervalle de temps.

Les séries chronologiques univariées apparaissent dans de nombreux domaines scientifiques et techniques. Voici quelques exemples :

- **Finance et économétrie** : évolution des indices boursiers, des prix, des données économiques d'entreprises, des ventes.
- **Médecine et biologie** : Analyse de signaux physiologiques (rythme cardiaque..), nombre de cas d'une maladie (comme Covid-19).
- **Sciences de la terre et de l'espace** : variations météorologiques, taches solaires, phénomènes naturels.
- **Traitement des données** : mesures successives de position ou de direction d'un objet mobile (trajectographie).

1.1.2 Bruit Blanc

Le bruit blanc joue un rôle clé dans la modélisation, car il représente l'erreur aléatoire non expliquée par un modèle déterministe. On distingue deux types :

Bruit blanc faible

On dit que la suite de variables aléatoires $(\varepsilon_t)_{t \in \mathbb{Z}}$ constitue un bruit blanc **faible** si elle vérifie les conditions suivantes :

- $E(\varepsilon_t) = 0, \forall t \in \mathbb{Z}$.
- $Var(\varepsilon_t) = \sigma^2 < \infty$.

- $Cov(\varepsilon_t, \varepsilon_s) = 0, \forall t \neq s$.

Bruit blanc fort

C'est une suite de variables aléatoires $(\varepsilon_t)_{t \in \mathbb{Z}}$ qui sont indépendantes et identiquement distribuées (i.i.d), avec une moyenne nulle et une variance finie. On écrit : $\varepsilon_t \sim \text{i.i.d.}(0, \sigma^2)$.

Remarque Un bruit blanc est dit **gaussien** s'il suit une loi normale de moyenne nulle et de variance σ^2 , on note $\varepsilon_t \sim N(0, \sigma^2)$.

1.1.3 Stationnarité

C'est une propriété importante et nécessaire pour appliquer certains théorèmes des séries chronologiques, il existe deux types de stationnarité : stationnarité faible et stationnarité forte.

Stationnarité Faible (ou stationnarité du second ordre)

Un processus $\{X_t, t \in \mathbb{Z}\}$ est dit stationnaire du second ordre (faiblement stationnaire) si les propriétés suivantes sont vérifiées :

- $E(X_t) = \mu < \infty, \forall t \in \mathbb{Z}$.
- $E(X_t^2) = \sigma^2 < \infty$.
- $\forall h > 0, Cov(X_t, X_{t+h}) = \gamma(h)$ indépendant de t .

Stationnarité Forte (ou stricte)

On dit qu'un processus $\{X_t, t \in \mathbb{Z}\}$ est fortement stationnaire si et seulement si la loi du vecteur $(X_{t_1}, X_{t_2}, \dots, X_{t_k})$ est identique à celle de $(X_{t_1+h}, X_{t_2+h}, \dots, X_{t_k+h})$, $\forall (t_1, t_2, \dots, t_k) \in \mathbb{Z}^k, \forall h \in \mathbb{Z}$.

Remarques

- La stationnarité forte implique la stationnarité faible, mais la réciproque n'est pas toujours vraie.
- Un processus stationnaire est homoscedastique (c'est à dire la variance est constante dans le temps).

Exemple d'un processus stationnaire

Un bruit blanc est un processus stationnaire.

Exemple d'un processus non stationnaire : Une marche aléatoire

Le processus $\{X_t, t \in \mathbb{Z}\}$ est une marche aléatoire lorsqu'il est défini par la relation suivante :

$$X_t = X_{t-1} + \varepsilon_t, \quad t \in \mathbb{Z}^*,$$

avec $(\varepsilon_t)_{t \in \mathbb{Z}^*}$ une suite de variables aléatoires de moyenne nulle et de variance σ^2 et X_0 une valeur initiale.

On peut écrire :

$$X_t = X_0 + \sum_{i=1}^t \varepsilon_i.$$

En vérifiant les conditions de stationnarité, on aura :

- $E[X_t] = E[X_0]$. L'espérance est donc constante dans le temps.
- $Var(X_t) = \sum_{i=1}^t Var(\varepsilon_i) = t \cdot \sigma^2$.
- $Cov(X_t, X_{t+h}) = Cov\left(\sum_{i=1}^t \varepsilon_i, \sum_{j=1}^{t+h} \varepsilon_j\right) = t \cdot \sigma^2$

Alors la marche aléatoire n'est pas stationnaire car la variance et la covariance dépendent de t .

Dans la suite, nous utiliserons le terme stationnaire pour un processus faiblement stationnaire.

1.1.4 Causalité et inversibilité

1. Un processus $\{X_t, t \in \mathbb{Z}\}$ est dit **causal** s'il existe une suite réelle $\{a_k\}$ et un autre processus $\{\varepsilon_t, t \in \mathbb{Z}\}$ telle que :

$$X_t = \sum_{k=0}^{\infty} a_k \varepsilon_{t-k}, \quad \text{avec} \quad \sum_{k=0}^{\infty} |a_k| < \infty,$$

c'est à dire $\{X_t\}$ peut s'écrire en fonction des valeurs passées d'un autre processus.

2. Un processus $\{X_t, t \in \mathbb{Z}\}$ est dit **inversible** s'il existe une suite réelle $\{b_k\}$ et un autre processus $\{\varepsilon_t, t \in \mathbb{Z}\}$ telle que :

$$\varepsilon_t = \sum_{k=0}^{\infty} b_k X_{t-k}, \quad \text{avec} \quad \sum_{k=0}^{\infty} |b_k| < \infty.$$

1.1.5 Opérateur retard

L'opérateur retard est un opérateur qui transforme une variable X_t en sa valeur passée. Il est désigné par la lettre B (ou encore L) et est défini par :

$$\begin{cases} BX_t = X_{t-1}, & \text{(processus décalé d'une unité de temps)} \\ B^2X_t = B(BX_t) = BX_{t-1} = X_{t-2}, \\ \vdots \\ B^dX_t = X_{t-d}, & \text{(processus décalé de } d \text{ unités de temps)} \end{cases}$$

1.1.6 Fonction d'autocovariance et d'autocorrélation

Définition 1.1. Soit $\{X_t, t \in \mathbb{Z}\}$ un processus de moyenne $E(X_t) = \mu$, on appelle fonction d'autocovariance la fonction γ définie par :

$$\gamma(h) = \text{Cov}(X_t, X_{t+h}) = E[(X_{t+h} - \mu)(X_t - \mu)], \quad \forall h \in \mathbb{Z}.$$

Remarque La fonction d'autocovariance d'un processus stationnaire vérifie

1. $\forall h \in \mathbb{Z}, \gamma(-h) = \gamma(h)$. La fonction est symétrique.
2. $\gamma(0) = \text{Var}(X_t)$.
3. $\forall h \in \mathbb{Z}, |\gamma(h)| \leq \gamma(0)$.

Définition 1.2. La fonction d'autocorrélation ACF (Autocorrelation Function) du processus $\{X_t, t \in \mathbb{Z}\}$ notée ρ est donnée par :

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \frac{\text{Cov}(X_t, X_{t+h})}{\text{Var}(X_t)}, \quad \forall h \in \mathbb{Z}.$$

Remarque La fonction d'autocorrélation d'un processus stationnaire vérifie

1. $\forall h \in \mathbb{Z}, \rho(-h) = \rho(h)$.
2. $\rho(0) = 1$.
3. $\forall h \in \mathbb{Z}, |\rho(h)| \leq 1$.

1.2 Processus autorégressif d'ordre p (AR(p))

Les modèles autorégressifs forment une classe importante de modèles utilisés pour analyser les séries temporelles. Ils reposent sur l'hypothèse que la valeur présente d'une série dépend linéairement d'un certain nombre de valeurs précédentes, avec un terme d'erreur aléatoire.

1.2.1 Définition

On dit que $\{X_t, t \in \mathbb{Z}\}$ est un processus autorégressif d'ordre p , noté AR (p), s'il est défini par l'équation suivante :

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + c + \varepsilon_t, \quad t \in \mathbb{Z}, \quad (1.1)$$

en d'autres termes :

$$X_t = \sum_{i=1}^p \phi_i X_{t-i} + \varepsilon_t + c, \quad t \in \mathbb{Z},$$

où

- X_t est la valeur de la série temporelle à l'instant t .
- $\phi_1, \phi_2, \dots, \phi_p$ sont des coefficients réels fixés.
- ε_t est un bruit blanc indépendant de X_{t-k} .

1.2.2 Conditions de Stationnarité

En appliquant l'opérateur retard B , on peut réécrire l'équation (1.1) sous la forme :

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) X_t = \varepsilon_t,$$

soit $\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$.

$\{X_t, t \in \mathbb{Z}\}$ est faiblement stationnaire si les racines du polynôme caractéristique $\Phi(B) = 0$ sont de modules strictement supérieurs à 1.

Remarques

Pour $p = 1$: Soit $\{X_t, t \in \mathbb{Z}\}$ un processus AR(1) défini par :

$$X_t = \phi_1 X_{t-1} + \varepsilon_t, \quad t \in \mathbb{Z}.$$

Ce processus est dit stationnaire si $|\phi_1| < 1$.

Pour $p = 2$: Soit $\{X_t\}$ un processus AR(2) défini par :

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t, \quad t \in \mathbb{Z}.$$

Ce processus est stationnaire s'il vérifie les conditions suivantes :

$$\begin{cases} |\phi_2| < 1 \\ \phi_1 + \phi_2 < 1 \\ \phi_1 - \phi_2 < 1 \end{cases}$$

1.2.3 Moments d'un Processus AR(p)

On suppose que $\{X_t, t \in \mathbb{Z}\}$ est un processus stationnaire et que les bruits blancs sont forts ($\varepsilon_t \sim \text{i.i.d.}(0, \sigma^2)$).

Espérance

En prenant l'espérance de l'équation (1.1) on obtient :

$$E[X_t] = \phi_1 E[X_{t-1}] + \phi_2 E[X_{t-2}] + \cdots + \phi_p E[X_{t-p}] + c + E[\varepsilon_t]$$

Puisque le processus $\{X_t\}$ est stationnaire et que $E[\varepsilon_t] = 0$, l'équation devient :

$$\mu = (\phi_1 + \phi_2 + \cdots + \phi_p)\mu + c, \quad \text{avec } \mu = E[X_t].$$

On obtient

$$\mu = \frac{c}{1 - (\phi_1 + \phi_2 + \cdots + \phi_p)}.$$

Variance et autocovariance

La fonction d'autocovariance d'un processus AR(p) satisfait une relation de récurrence de la forme :

$$\gamma(h) = \begin{cases} \phi_1 \gamma(1) + \phi_2 \gamma(2) + \cdots + \phi_p \gamma(p) + \sigma_\varepsilon^2 & \text{si } h = 0, \\ \phi_1 \gamma(h-1) + \phi_2 \gamma(h-2) + \cdots + \phi_p \gamma(h-p) & \text{si } h > 0, \end{cases}$$

Pour $h = 1, 2, \dots, p$, nous obtenons un système d'équations :

$$\begin{cases} \gamma(1) = \phi_1 \gamma(0) + \phi_2 \gamma(1) + \cdots + \phi_p \gamma(p-1), \\ \gamma(2) = \phi_1 \gamma(1) + \phi_2 \gamma(0) + \cdots + \phi_p \gamma(p-2), \\ \vdots \\ \gamma(p) = \phi_1 \gamma(p-1) + \phi_2 \gamma(p-2) + \cdots + \phi_p \gamma(0). \end{cases}$$

Pour $h = 0$, on obtient la variance d'un AR(p).

Fonction d'autocorrélation

La fonction d'autocorrélation d'un processus AR(p) est donnée par :

$$\rho(h) = \begin{cases} 1 & \text{si } h = 0, \\ \phi_1 \rho(h-1) + \phi_2 \rho(h-2) + \cdots + \phi_p \rho(h-p) & \text{si } h > 0, \end{cases}$$

En divisant les équations du système obtenu précédemment par $\gamma(0)$, on obtient les équations de Yule-Walker suivantes :

$$\begin{cases} \rho(1) = \phi_1 + \phi_2\rho(1) + \dots + \phi_p\rho(p-1), \\ \rho(2) = \phi_1\rho(1) + \phi_2 + \dots + \phi_p\rho(p-2), \\ \vdots \\ \rho(p) = \phi_1\rho(p-1) + \phi_2\rho(p-2) + \dots + \phi_p. \end{cases}$$

Fonction d'autocorrélation partielle

La fonction d'autocorrélation partielle PACF (Partial Autocorrelation Function) notée Φ_{kk} peut être définie comme suit

$$\phi_{kk} = \frac{\det \begin{pmatrix} 1 & \rho(1) & \dots & \rho(1) \\ \rho(1) & 1 & \dots & \rho(2) \\ \vdots & \vdots & \ddots & \vdots \\ \rho(k-1) & \rho(k-2) & \dots & \rho(k) \end{pmatrix}}{\det \begin{pmatrix} 1 & \rho(1) & \dots & \rho(k-1) \\ \rho(1) & 1 & \dots & \rho(k-2) \\ \vdots & \vdots & \ddots & \vdots \\ \rho(k-1) & \rho(k-2) & \dots & 1 \end{pmatrix}}$$

1.2.4 Causalité et inversibilité

1. La condition nécessaire et suffisante pour qu'un processus $AR(p)$ soit causal est que toutes les racines du polynôme caractéristique soient de modules strictement supérieurs à 1.
2. Un processus $AR(p)$ est toujours inversible.

1.2.5 Estimation des paramètres

Méthode des Moindres Carrés

Un modèle autorégressif d'ordre p peut s'écrire sous la forme suivante :

$$X_t = Z'\Phi + \varepsilon_t,$$

où

- ε_t est un bruit blanc.
- $Z' = (X_{t-1}, X_{t-2}, \dots, X_{t-p})$.
- $\Phi' = (\phi_1, \phi_2, \dots, \phi_p)$ est le vecteur des paramètres.

L'estimation des paramètres du modèle précédent par la méthode des moindres carrés est donnée par :

$$\hat{\Phi} = (ZZ')^{-1}ZX.$$

Cas particulier ($p = 1$) :

On considère le modèle AR(1) suivant :

$$X_t = \phi X_{t-1} + \varepsilon_t, \quad t \in \mathbb{Z}.$$

On utilise la méthode des moindres carrés pour estimer le paramètre ϕ . L'objectif est de minimiser la somme des carrés des erreurs SSE (Sum of Squared Errors), c'est-à-dire :

$$SSE(\phi) = \sum_{t=2}^n \varepsilon_t^2 = \sum_{t=2}^n (X_t - \phi X_{t-1})^2.$$

Maintenant, on dérive $SSE(\phi)$ par rapport à ϕ :

$$SSE'(\phi) = 2 \sum_{t=2}^n (X_t - \phi X_{t-1})(-X_{t-1}) = 2 \left(\phi \sum_{t=2}^n (X_{t-1})^2 - \sum_{t=2}^n (X_t X_{t-1}) \right).$$

Pour minimiser $SSE(\phi)$, en égalant la dérivée à zéro :

$$SSE'(\phi) = 0 \implies 2 \left(\phi \sum_{t=2}^n (X_{t-1})^2 - \sum_{t=2}^n (X_t X_{t-1}) \right) = 0.$$

Finalement, on obtient l'estimateur suivant :

$$\hat{\phi} = \frac{\sum_{t=2}^n (X_t X_{t-1})}{\sum_{t=2}^n (X_{t-1})^2}.$$

Propriétés de l'estimateur des moindres carrés

– **Consistance** : quand $T \rightarrow \infty$, on aura

$$\hat{\phi} \xrightarrow{p} \phi.$$

– **Normalité** :

$$\sqrt{T}(\hat{\phi} - \phi) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1 - \phi^2).$$

Méthode du Maximum de Vraisemblance

D'une manière générale la fonction de vraisemblance de Φ au vu des observations (x_1, \dots, x_T) est donnée par :

$$L(x_1, \dots, x_T; \phi) = \prod_{t=1}^T f(x_t, \phi).$$

Avec la fonction f définie par :

- Dans le cas discret, on pose $f(x, \phi) = P_{\Phi}(X = x)$.
- Dans le cas continu, on pose $f(x, \phi) = f_{\Phi}(X)$, la densité de X .

Lorsque $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$, la fonction de vraisemblance est donnée par :

$$L(\Phi, \sigma^2) = \prod_{t=p+1}^T \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\varepsilon_t^2}{2\sigma^2}\right).$$

En maximisant le logarithme de cette vraisemblance, on obtient les mêmes estimateurs que ceux de la méthode des moindres carrés.

Exemple d'un AR(1)

L'étude porte sur l'évolution du taux de chômage mensuel aux États-Unis entre 2010 et 2023. Les données sont disponibles sur

<https://data.bls.gov/pdq/SurveyOutputServlet>.

Le tableau ci-dessous résume les principales caractéristiques des données.

Moyenne	Écart-type	Maximum	Minimum	Mediane
6,718	1,898	9,90	4,00	6,650

Tableau 1.1 : Résumé des données

Ces données se modélisent selon un modèle autorégressif d'ordre 1, dont l'équation est la suivante :

$$X_t = 0.918X_{t-1} + 6.886 + \varepsilon_t, \quad t = 2, \dots, 156.$$

La figure ci-dessous montre le graphe de la fonction d'autocorrélation partielle, utilisée pour identifier le nombre de termes autorégressifs significatifs dans la modélisation de ces données.

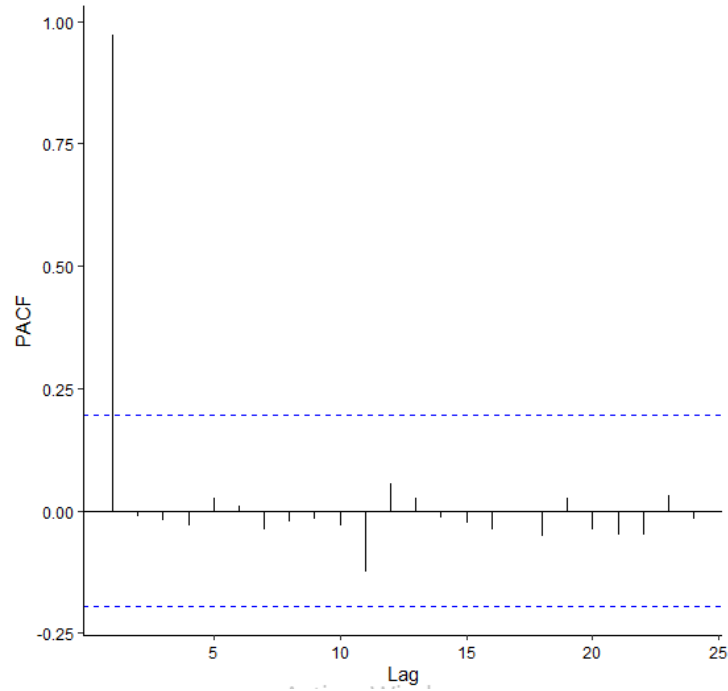


FIG. 1.1 : Fonction d'autocorrélation partielle

On observe un seul pic significatif au premier lag sur le graphique d'autocorrélation partielle (PACF), ce qui signifie que ces données suivent un processus AR(1).

Le graphe suivant illustre la comparaison entre les taux de chômage observés et les valeurs ajustées à l'aide du modèle AR(1).

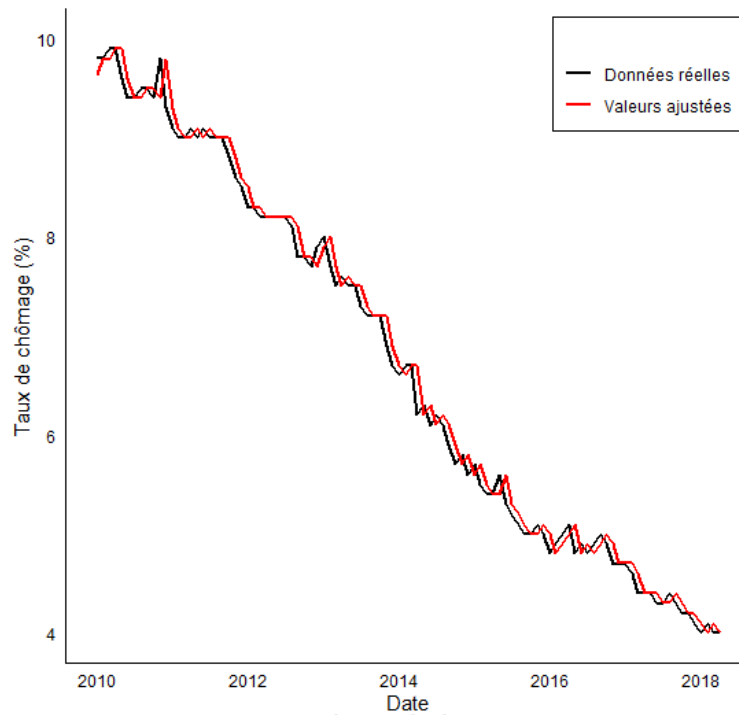


FIG. 1.2 : Comparaison des Courbes des Données Réelles et Ajustées

Le graphe montre que les deux courbes sont proches, ce qui veut dire que le modèle ajuste bien ces données.

Exemple d'un AR(2)

L'étude porte sur l'évolution du taux de chômage mensuel en France entre janvier 1983 et janvier 2024. Les données sont extraites de la base FRED (<https://fred.stlouis-fed.org/series/LRHUTTTT>).

Le tableau ci-dessous résume les principales caractéristiques des données.

Moyenne	Écart-type	Maximum	Minimum	Mediane
9,634	1,421	12,400	7,317	9,488

Tableau 1.2 : Résumé des données

Ces données se modélisent selon un modèle autorégressif d'ordre 2, dont l'équation est la suivante :

$$X_t = 1.369X_{t-1} - 0.512X_{t-2} + 9.377 + \varepsilon_t, \quad t = 3, \dots, 492.$$

La figure ci-dessous montre le graphe de la fonction d'autocorrélation partielle, utilisée pour identifier le nombre de termes autorégressifs significatifs dans la modélisation de ces données.

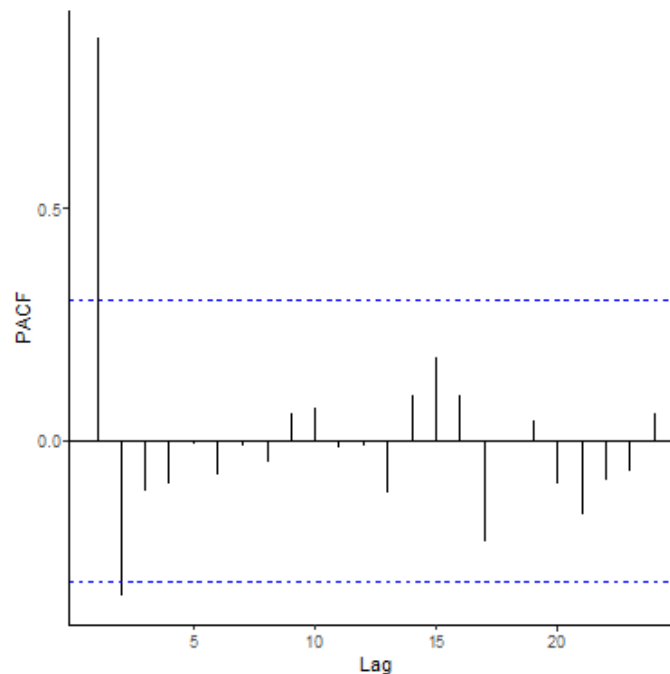


FIG. 1.3 : Fonction d'autocorrélation partielle

On observe deux pics significatifs au premier et au deuxième lag sur le graphique d'autocorrélation partielle (PACF), ce qui signifie que ces données suivent un processus AR(2).

Le graphe suivant illustre la comparaison entre les taux de chômage observés et les valeurs ajustées à l'aide du modèle AR(2).

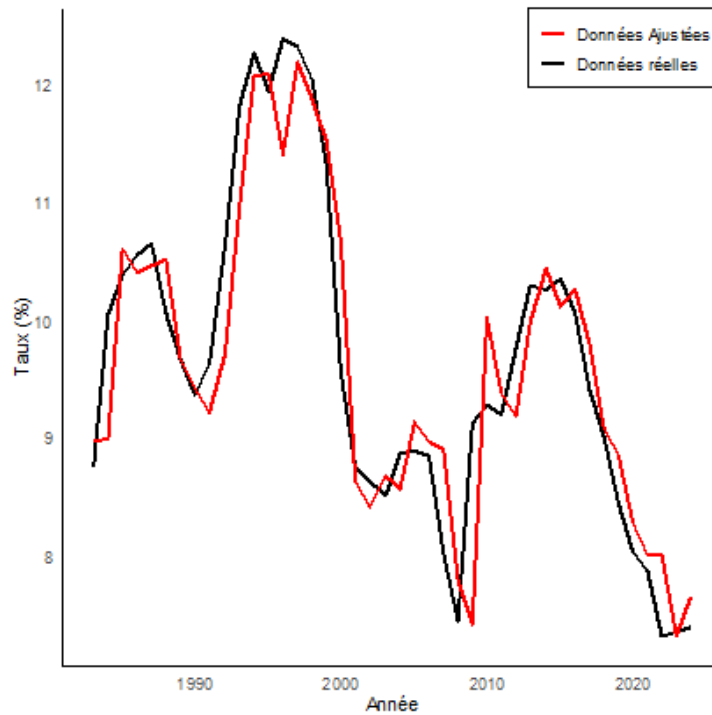


FIG. 1.4 : Comparaison des Courbes des Données Réelles et Ajustées

Le graphe montre que les deux courbes sont proches, ce qui veut dire que le modèle ajuste bien ces données.

1.3 Processus moyenne mobile

1.3.1 Définition

On dit que $\{X_t, t \in \mathbb{Z}\}$ est un processus moyenne mobile d'ordre q , usuellement noté MA(q) (moving average), s'il s'écrit sous la forme suivante :

$$X_t = \sum_{k=1}^q \theta_k \varepsilon_{t-k} + \varepsilon_t, \quad t \in \mathbb{Z}, \tag{1.2}$$

où $\theta_1, \theta_2, \dots, \theta_q$ sont les coefficients du modèle et ε_t est un bruit blanc de variance σ^2 .

L'équation (1.2) peut s'écrire aussi

$$X_t = \Theta(B)\varepsilon_t \quad \text{avec} \quad \Theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q.$$

1.3.2 Moments d'un processus MA(q)

Espérance

En prenant l'espérance de l'équation (1.2), on aura :

$$E[X_t] = \theta_1 E[\varepsilon_{t-1}] + \theta_2 E[\varepsilon_{t-2}] + \cdots + \theta_q E[\varepsilon_{t-q}] + E[\varepsilon_t].$$

Puisque ε_t est un bruit blanc, alors $E[\varepsilon_t] = 0$ et par conséquent :

$$E[X_t] = 0.$$

Variance

La variance d'un processus MA(q) est donnée par :

$$\text{Var}[X_t] = E[X_t^2] - (E[X_t])^2 = E[X_t^2].$$

En appliquant cette relation sur l'équation (1.2), on aura :

$$\begin{aligned} \text{Var}[X_t] &= E[(\theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t)^2] \\ &= E[\theta_1^2 \varepsilon_{t-1}^2] + E[\theta_2^2 \varepsilon_{t-2}^2] + \cdots + E[\theta_q^2 \varepsilon_{t-q}^2] + E[\varepsilon_t^2] \\ &= \sigma^2(1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2). \end{aligned}$$

Autocovariance

La fonction d'autocovariance d'un MA(q) est donnée par :

$$\gamma(h) = \begin{cases} \sigma^2 (1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2), & h = 0, \\ \sigma^2 (\theta_h + \theta_1 \theta_{h+1} + \theta_2 \theta_{h+2} + \cdots + \theta_{h+q} \theta_q), & 1 \leq h \leq q, \\ 0 & h > q. \end{cases}$$

Fonction d'autocorrélation

La fonction d'autocorrélation est donnée par :

$$\rho(h) = \begin{cases} 1 & \text{si } h = 0, \\ \frac{\theta_h + \theta_{h+1} \theta_1 + \theta_{h+2} \theta_2 + \cdots + \theta_q \theta_{q+h}}{1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2} & \text{si } 1 \leq h \leq q, \\ 0 & \text{si } h > q. \end{cases}$$

1.3.3 Propriétés

1. Tout processus MA(q) est centré et stationnaire.
2. Tout processus MA(q) est causal.

3. Un processus MA(q) est inversible si les racines du polynôme caractéristique $\Theta(B)$ soient de modules strictement supérieurs à 1.

1.3.4 Estimation des paramètres

Méthode des Moindres Carrés

On cherche à minimiser la somme des carrés des erreurs :

$$SSE(\theta_1, \dots, \theta_q) = \sum_{t=1}^n \varepsilon_t^2 = \sum_{t=1}^n (X_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q})^2.$$

Pour minimiser cette somme, on la dérive par rapport à chaque paramètre, puis en égalant la dérivée à zéro.

Finalement, on obtient l'estimateur suivant :

$$\hat{\theta} = \frac{\sum_{t=1}^n \varepsilon_{t-q} X_t}{\sum_{t=1}^n \varepsilon_{t-q}^2},$$

avec $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_q)$.

Exemple d'un MA(1)

Le tableau suivant présente les données mensuelles enregistrées entre janvier 2005 et décembre 2008 sur le nombre de nouveaux cas de la maladie de Newcastle (ND) dans une province chinoise. Ces données, extraites de l'article (Fangge Li, Peixian Luan [6]).

Date	2005	2006	2007	2008
Janvier	0	15	7	7
Février	7	10	12	7
Mars	6	0	12	8
Avril	9	10	6	7
Mai	12	11	8	7
Juin	15	6	8	7
Juillet	11	10	4	7
Août	9	11	5	4
Septembre	9	9	8	5
Octobre	11	11	10	5
Novembre	10	9	8	6
Décembre	15	8	6	4

Tableau 1.3 : Nombre de nouveaux cas de la maladie de Newcastle en CHINE

Ces données se modélisent selon un MA(1), dont l'équation est la suivante :

$$X_t = -0.682\varepsilon_{t-1} - 0.015, \quad t = 2, \dots, 48.$$

1.4 Processus autorégressifs moyennes mobiles

Les processus autorégressifs moyennes mobiles, notés ARMA(p,q) généralisent les modèles AR(p) et MA(q).

1.4.1 Définition

Un processus ARMA(p,q) est défini par l'équation :

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} + c = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}, \quad \forall t \in \mathbb{Z}, \quad (1.3)$$

où ϕ_1, \dots, ϕ_p sont les coefficients autorégressifs et $\theta_1, \dots, \theta_q$ les coefficients moyennes mobiles.

Avec les notations précédentes, (1.3) peut s'écrire :

$$\Phi(B)X_t = \Theta(B)\varepsilon_t.$$

Remarque

Si $q = 0$, le modèle devient un processus AR(p) et si $p = 0$, le modèle devient un processus MA(q).

1.4.2 Moments des processus ARMA

Espérance

En prenant l'espérance des deux cotés de l'équation (1.3), avec $\{X_t, t \in \mathbb{Z}\}$ stationnaire et $E[\varepsilon_t] = 0$, on obtient :

$$E[X_t](1 - \phi_1 - \phi_2 - \dots - \phi_p) = c.$$

En posant $E[X_t] = \mu$, on aura :

$$\mu = \frac{c}{1 - \phi_1 - \phi_2 - \dots - \phi_p}.$$

Variance

$$Var[X_t] = \gamma(0) = \sigma^2(1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2) + \phi_1 \gamma(1) + \phi_2 \gamma(2) + \dots + \phi_p \gamma(p).$$

$\sigma^2(1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2)$ représente la variance d'un MA(q) et $\phi_1 \gamma(1) + \phi_2 \gamma(2) + \dots + \phi_p \gamma(p)$ représente la variance de la partie autorégressive.

Autocovariance

En multipliant l'équation (1.3) par X_{t-h} , puis en prenant l'espérance, on obtient :

$$\begin{cases} \gamma(h) - \phi_1\gamma(h-1) - \dots - \phi_p\gamma(h-p) = 0, & \forall h \geq q+1, \\ \gamma(h) - \phi_1\gamma(h-1) - \dots - \phi_p\gamma(h-p) = \sigma^2(\theta_h + \theta_1\theta_{h-1} + \dots + \theta_{h-q}\theta_q), & 0 \leq h \leq q. \end{cases}$$

Remarque

L'autocovariance d'un ARMA(p,q) satisfait une relation de récurrence d'ordre p (équation de Yule-Walker) à partir du rang q+1.

1.4.3 Propriétés

1. Le processus ARMA(p, q) est stationnaire si les racines du polynôme $\Phi(Z) = 0$ sont de modules strictement supérieurs à 1.
2. Il est inversible si les racines du polynôme $\Theta(Z) = 0$ sont de modules strictement supérieurs à 1.

1.4.4 Estimation des paramètres

Un modèle ARMA(p, q) peut être défini comme suit :

$$X_t = Z'\Phi + W'\Theta + \varepsilon_t,$$

où

- $Z' = (x_{t-1}, x_{t-2}, \dots, x_{t-p})$,
- $W' = (\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-q})$,
- $\Phi' = (\phi_1, \phi_2, \dots, \phi_p)$ est le vecteur des paramètres autorégressifs,
- $\Theta' = (\theta_1, \theta_2, \dots, \theta_q)$ est le vecteur des paramètres de moyenne mobile.

L'estimateur des paramètres par la méthode des moindres carrés est donné par :

$$\hat{\beta} = (MM')^{-1}MX$$

où $M = \begin{bmatrix} Z & W \end{bmatrix}$, représente la matrice qui contient les retards de X_t et de ε_t .

Exemple d'un ARMA(1,2)

La série étudiée présente des données sur le taux effectif des fonds fédéraux entre 1970 et 2008. Ces données ont été extraites dans le cours de Pelgrin F [13].

Le tableau ci-dessous résume les principales caractéristiques des données.

Moyenne	écart-type	Maximum	Minimum	mediane
6,28	3,5548	19,10	0,11	5,55

Tableau 1.4 : Résumé des données

Ces données se modélisent selon un ARMA(1,2) , dont l'équation est la suivante :

$$X_t = 0.9721X_{t-1} + \varepsilon_t + 0.4903\varepsilon_{t-1} + 0.0822\varepsilon_{t-2} + 6.0076, \quad t = 3, \dots, 456.$$

Le graphe suivant présente la comparaison entre ces données et les données ajustées selon ARMA(1,2).

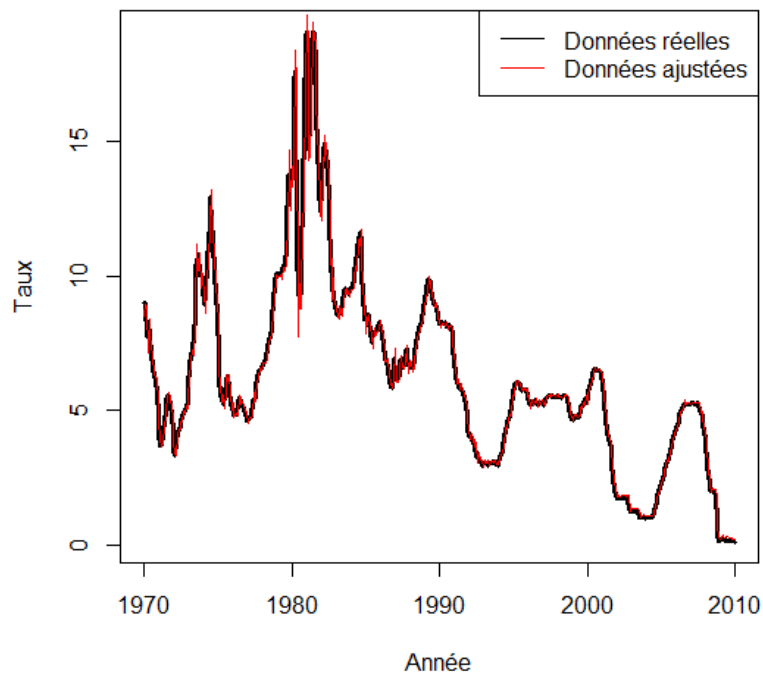


FIG. 1.5 : Comparaison des courbes des données réelles et ajustées.

On remarque que les deux courbes sont proches, ce qui veut dire que le modèle ajuste bien ces données.

Chapitre 2

Tests de rupture épidémique dans la moyenne de variables aléatoires faiblement dépendantes

2.1 Introduction

Dans l'analyse des séries temporelles et des suites de variables aléatoires, il est fréquent que les données soient soumises à des perturbations imprévues qui modifient temporairement leur comportement habituel. La détection de tels changements, appelés **ruptures**, permet non seulement de mieux comprendre l'évolution des phénomènes observés, mais aussi d'ajuster les modèles statistiques afin d'en améliorer la précision. Parmi les différentes formes de ruptures, on distingue notamment les ruptures **épidémiques**, qui se caractérisent par une variation temporaire d'un paramètre de la série sur un intervalle de temps limité, puis la série retrouve son comportement initial.

Soit un échantillon (X_1, X_2, \dots, X_n) de variables aléatoires indépendantes de moyennes μ_i et de variance σ_i^2 , $i = 1, \dots, n$.

On dit qu'il y a une **rupture épidémique** dans la moyenne s'il existe deux indices $1 < p < q < n$ et un réel $\delta \neq 0$ tels que :

$$\mu_i = \begin{cases} \mu_0 & \text{si } i = 1, 2, \dots, p, \\ \mu_0 + \delta & \text{si } i = p + 1, \dots, q, \\ \mu_0 & \text{si } i = q + 1, \dots, n, \end{cases}$$

où p et q sont les **bornes** de la rupture et $l^* = q - p$ sa **longueur**.

Le problème central est de tester l'hypothèse nulle :

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_n = \mu_0$$

contre l'hypothèse alternative :

$$H_A : \exists 1 < p < q < n \text{ tels que } \begin{cases} \mu_i = \mu_0 & \text{pour } i \leq p \text{ ou } i > q, \\ \mu_i = \mu_0 + \delta & \text{pour } p < i \leq q. \end{cases}$$

Ici, les paramètres μ_0 , p , q , et δ sont supposés inconnus.

Dans la suite de ce chapitre, nous allons présenter quelques statistiques pour tester H_0 contre H_A . On définit d'abord les sommes cumulées des observations :

$$S(t) = \sum_{k=1}^t X_k, \quad \text{avec } S(0) = 0.$$

2.2 Statistiques de tests classiques

2.2.1 Statistique de Chen et Zhou

Soit (X_1, X_2, \dots, X_n) un échantillon de variables aléatoires indépendantes et i et j deux entiers tels que $1 \leq i < j \leq n$, on définit les sommes suivantes

$$U_{i,j} = \sum_{k \in I_{i,j}} X_k, \quad \text{et} \quad V_{i,j} = \sum_{k \in I_{i,j}^c} X_k.$$

avec $I_{i,j} = \{i+1, \dots, j\}$ et $I_{i,j}^c = \{1, \dots, i, j+1, \dots, n\}$.

On définit

$$R_{i,j} = \frac{(n - (j - i))U_{i,j} - (j - i)V_{i,j}}{n}.$$

Soient i_0 et j_0 vérifiant

$$R_{i_0, j_0} = \max_{1 \leq i < j \leq n} R_{i,j}.$$

i_0 et j_0 peuvent être considérés comme des estimations pour les bornes de la rupture k^* et l^* .

Soit $X_1^{(i_0, j_0)} < X_2^{(i_0, j_0)} < \dots < X_{n-j_0+i_0}^{(i_0, j_0)}$, les éléments ordonnés de

l'ensemble $\{X_k, k \in I_{(i_0, j_0)}^c\}$ et

$$m = \left\lceil \frac{n - (j_0 - i_0)}{2} \right\rceil$$

La statistique de test proposée par Chen et Zhou [4] est alors :

$$Z_n = \frac{mR_{i_0, j_0}}{\sum_{k=n-j_0+i_0-m+1}^{n-j_0+i_0} X_k^{(i_0, j_0)} - \sum_{k=1}^m X_k^{(i_0, j_0)}}.$$

Cette statistique est ensuite comparée à une valeur critique C_α obtenue par simulation de Monte Carlo sous l'hypothèse nulle (voir Chen et Zhou [4]). On rejette l'hypothèse nulle d'absence de rupture épidémique au niveau de signification α , si $Z_n > C_\alpha$.

2.2.2 Statistique de Levin et Kline

La statistique de Levin et Kline est définie comme :

$$Q_n(X_1, X_2, \dots, X_n) = \max_{1 \leq i < j \leq n} \left| S(j) - S(i) - \frac{j-i}{n} S(n) \right|.$$

Sous H_0 , on a $n^{-1/2}Q_n$ converge dans $C[0, 1]$ vers $\sup_{0 < t < 1} |B(t)|$, où $B(t) = W(t) - tW(1)$ est le pont brownien correspondant au mouvement brownien W .

Sous l'hypothèse alternative, la statistique $n^{-1/2}Q_n$ tend vers l'infini quand $n \rightarrow \infty$, donc le test basé sur Q_n est consistant.

2.3 Statistiques de tests hölderiennes

2.3.1 Espaces de Hölder

L'espace de Hölder d'ordre α avec $0 < \alpha \leq 1$, noté $H_\alpha[0, 1]$, est l'espace des fonctions f définies sur $[0, 1]$, nulles en 0, telles que :

$$\|f\|_\alpha = \sup_{0 < |t-s| \leq 1} \frac{|f(t) - f(s)|}{|t-s|^\alpha} < +\infty.$$

On définit également le module de continuité hölderienne de f , noté $\omega_\alpha(f, \delta)$, par

$$\omega_\alpha(f, \delta) = \sup_{0 < |t-s| \leq \delta} \frac{|f(t) - f(s)|}{|t-s|^\alpha}.$$

On définit le sous-espace $H_\alpha^0[0, 1]$ de $H_\alpha[0, 1]$ par :

$$f \in H_\alpha^0[0, 1] \iff f \in H_\alpha[0, 1] \text{ et } \lim_{\delta \rightarrow 0} \omega_\alpha(f, \delta) = 0.$$

L'espace $(H_\alpha, \|\cdot\|_\alpha)$ est un espace de Banach non séparable. En revanche, $(H_\alpha, \|\cdot\|_\beta)$ est séparable pour tout $0 < \beta < \alpha$ et H_α s'injecte continûment dans H_β .

$(H_\alpha^0, \|\cdot\|_\alpha)$ est un sous-espace fermé séparable de $(H_\alpha, \|\cdot\|_\alpha)$.

2.3.2 Statistiques de Račkauskas et Suquet

Afin de détecter des courtes épidémies, Račkauskas et Suquet (2004) ont proposé la statistique $U_{\alpha,n}$, avec $0 < \alpha < \frac{1}{2}$, définie par :

$$U_{\alpha,n}(X_1, X_2, \dots, X_n) = \max_{1 \leq i < j \leq n} \frac{\left| S(j) - S(i) - S(n) \frac{j-i}{n} \right|}{\left[\left(\frac{j-i}{n} \right) \left(1 - \frac{j-i}{n} \right) \right]^\alpha}.$$

La loi limite de $U_{\alpha,n}$ est définie par :

$$U_{\alpha,\infty} = \sup_{0 < t-s < 1} \frac{|B(t) - B(s)|}{[(t-s)(1-(t-s))]^\alpha}.$$

Soit D_j l'ensemble des nombres dyadiques sur $[0, 1]$ de niveau j :

$$D_0 = \{0, 1\}, \quad D_j = \{(2\ell - 1)2^{-j} : 1 \leq \ell \leq 2^{j-1}\}, \quad j \geq 1.$$

On note :

$$D = \bigcup_{j \geq 0} D_j, \quad D^* = D \setminus \{0\}.$$

Pour $r \in D_j$, $j \geq 0$, on définit :

$$r^- = r - 2^{-j}, \quad r^+ = r + 2^{-j}.$$

Račkauskas et Suquet (2004) ont également défini la statistique $D_{\alpha,n}$ par :

$$D_{\alpha,n}(X_1, X_2, \dots, X_n) = \max_{1 \leq j \leq \log_2(n)} \frac{1}{2^{-j\alpha}} \max_{r \in D_j} \left| S(nr) - \frac{1}{2}S(nr^+) - \frac{1}{2}S(nr^-) \right|.$$

La loi limite de $D_{\alpha,n}$ est donnée par :

$$D_{\alpha,\infty} = \sup_{j \geq 1} \frac{1}{2^{-j\alpha}} \max_{r \in D_j} \left| W(r) - \frac{1}{2}W(r^+) - \frac{1}{2}W(r^-) \right|.$$

La valeur seuil lue dans la table de U_α pour $\alpha = 0.25$ et au risque 0.05 est 2.27 (voir Imeçaoudene et al.[7]).

La valeur seuil lue dans la table de D_α pour $\alpha = 0.25$ et au risque 0.05 est 1.32 (voir Račkauskas et Suquet [14]).

Les statistiques hölderiennes sont particulièrement utiles pour détecter des **ruptures épidémiques** de courte durée. La statistique Q_n , par exemple, détecte des ruptures de longueur d'au moins $n^{1/2}$, tandis que les statistiques $U_{\alpha,n}$ et $D_{\alpha,n}$ détectent des ruptures dont la longueur est de l'ordre n^δ , avec $\delta = \frac{1-2\alpha}{2-2\alpha}$.

Les statistiques $D_{\alpha,n}$ présentent un comportement asymptotique similaire à celui des statistiques $U_{\alpha,n}$, mais elles sont plus utiles car leurs lois limites sont connues.

2.4 Convergence et consistance

2.4.1 Cas de variables aléatoires indépendantes

Soit $(X_i)_{i \geq 1}$ une suite de variables aléatoires indépendantes. Si les (X_i) sont identiquement distribuées de variance σ^2 , alors la convergence sous H_0 des statistiques $U_{\alpha,n}$ et $D_{\alpha,n}$ dans l'espace de H_0^α est donnée par le résultat suivant.

Théorème 2.4. (Račkauskas et Suquet [14]) On suppose que

$$\mathbb{P}(|X_1| > t) = o(t^{-p}), \quad \text{où } p = \frac{1}{\frac{1}{2} - \alpha}.$$

Alors pour tout $0 < \alpha < \frac{1}{2}$

$$\sigma^{-1}n^{-1/2}U_{\alpha,n} \xrightarrow{\mathcal{L}} U_{\alpha,\infty}$$

et

$$\sigma^{-1}n^{-1/2}D_{\alpha,n} \xrightarrow{\mathcal{L}} D_{\alpha,\infty}.$$

La consistance de rejeter H_0 contre H_A est donnée par le résultat suivant.

Théorème 2.5. (Račkauskas et Suquet [14]) Si $\sup_k \mathbb{V}(X_k) < +\infty$ et

$$\lim_{n \rightarrow \infty} h_n n^{1-2\alpha} = +\infty, \quad \text{avec } h_n = \frac{l^*}{n} \left(1 - \frac{l^*}{n}\right)$$

Alors

$$n^{-1/2}U_{\alpha,n} \xrightarrow{P} +\infty$$

et

$$n^{-1/2}D_{\alpha,n} \xrightarrow{P} +\infty.$$

2.4.2 Cas de variables aléatoires faiblement dépendantes

Soit (X_k) une suite de variables aléatoires vérifiant $X_k = \varepsilon_k + \sum_{i=1}^q \theta_i \varepsilon_{k-i} + a_k$, où les ε_k sont i.i.d. centrées et de variance finie σ^2 et a_k est une suite de nombres réels. On va

tester l'hypothèse nulle

$$H'_0 : a_1 = a_2 = \dots = a_n = 0$$

contre

$$H'_A : \exists 1 < p < q < n \text{ tels que } \begin{cases} a_1 = \dots = a_p = a_{q+1} = \dots = a_n = 0, \\ a_{p+1} = \dots = a_q = \delta \end{cases} \quad \delta \neq 0.$$

La convergence et la consistance des statistiques $U_{\alpha,n}$ et $D_{\alpha,n}$ dans l'espace de H'_0 sont données par les résultats suivants.

Théorème 2.6.(Graiche [10]) Supposons que

$$\lim_{t \rightarrow 0} t^{\frac{1}{2}-\alpha} \mathbb{P}(|\varepsilon_1| > t) = 0, \quad \text{pour un } 0 < \alpha < \frac{1}{2}. \quad (2.1)$$

Alors sous (H'_0)

$$\sigma^{-1} n^{-\frac{1}{2}} D_{\alpha,n} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} D_\alpha$$

et

$$\sigma^{-1} n^{-\frac{1}{2}} U_{\alpha,n} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} U_\alpha.$$

Théorème 2.7.(Graiche [10]) Si la condition (2.1) est satisfaite et $\lim_{n \rightarrow \infty} n^{\alpha-\frac{1}{2}} (l^*)^{1-\alpha} = +\infty$.

Alors sous H'_A

$$n^{-\frac{1}{2}} D_{\alpha,n} \xrightarrow[n \rightarrow \infty]{P} +\infty$$

et

$$n^{-\frac{1}{2}} U_{\alpha,n} \xrightarrow[n \rightarrow \infty]{P} +\infty.$$

2.5 Étude de simulation

2.5.1 Cas de variables aléatoires indépendantes

Pour comparer les performances des quatre statistiques $D_n = \sigma^{-1} n^{-1/2} D_{a,n}$, $U_n = \sigma^{-1} n^{-1/2} U_{a,n}$, $R_n = \sigma^{-1} n^{-1/2} Q_n$ et Z_n , avec $a = 0.25$, nous simulons 1000 échantillons de taille $n = 60$ de variables aléatoires indépendantes suivant une loi normale $N(0,1)$ puis une loi uniforme $U[-1,1]$, avec un seuil de significativité de 0.05.

Sous H'_A , le tableau suivant donne les valeurs des quatre statistiques pour différentes valeurs de l^* et δ .

l^*	δ	$X \sim N(0, 1)$				$X \sim U[-1, 1]$			
		D_n	U_n	R_n	Z_n	D_n	U_n	R_n	Z_n
3	2	1.0766	1.7285	1.3446	3.8621	1.4161	2.9093	1.7178	6.069
	4	1.1482	2.3668	1.3864	7.7632	/	/	/	15.0924
	5	1.3409	/	1.5445	9.7004	/	/	/	/
5	2	1.3065	2.9099	1.4068	6.2321	1.8941	2.0696	2.3284	9.7684
	4	1.3933	/	1.7341	12.5665	/	5.5957	/	/
7	2	1.3095	2.9861	1.8486	8.3726	2.4078	2.5613	2.9627	13.2251
	3	1.5108	/	/	12.5704	/	/	/	/
10	1	1.2904	1.7534	1.5502	5.4198	1.8218	2.9526	2.1289	8.5114
	2	1.6485	2.6646	/	10.6906	/	/	/	/
15	1	1.9615	2.0424	1.7922	7.3534	1.7585	2.9954	2.6871	12.1042
	2	/	2.9925	/	15.4189	/	/	/	/
Valeur critique c		1.32	2.27	1.50	8.38	1.32	2.27	1.50	8.38

Tableau 1.1 : Comparaison entre D_n , U_n , R_n et Z_n .

On note que quand l^* augmente, les statistiques détectent plus rapidement l'épidémie pour un δ plus petit. On remarque aussi qu'il est plus intéressant d'utiliser la statistique U_n , pour de courtes épidémies ($l^* < 7$).

Le tableau suivant présente les valeurs des puissances des tests basés sur D_n , U_n , R_n et Z_n .

l^*	δ	$X \sim N(0, 1)$				$X \sim U[-1, 1]$			
		D_n	U_n	R_n	Z_n	D_n	U_n	R_n	Z_n
3	2	0.9165	0.9317	0.9279	0.90	0.9581	0.992	0.9736	0.9028
	4	0.9774	1	0.9983	0.9304	1	1	1	1
	5	0.998	1	1	1	1	1	1	1
5	2	0.9435	0.973	0.9639	0.9087	0.9959	1	0.9998	0.9826
	4	0.999	1	1	0.9987	1	1	1	1
7	2	0.9693	0.992	0.9882	0.949	0.9998	1	1	0.9998
	3	0.9987	1	0.9999	0.9988	1	1	1	1
10	1	0.9424	0.9506	0.9534	0.9072	0.9885	0.996	0.9949	0.9515
	2	0.9956	0.9989	0.9988	0.9866	1	1	1	1
15	1	0.9368	0.9738	0.9767	0.941	0.9876	0.9994	0.9995	0.9953
	2	0.9963	1	1	0.9999	1	1	1	1

Tableau 1.2 : Puissance des tests basés sur D_n , U_n , R_n et Z_n .

On remarque que pour les courtes épidémies ($l^* < 7$), le test basé sur la statistique U_n est plus puissant.

2.5.2 Cas de variables aléatoires faiblement dépendantes

Pour comparer les performances des statistiques D_n et U_n avec $a = 0.25$, nous simulons 1000 échantillons de taille $n = 60$ de variables aléatoires faiblement dépendantes, générées par un processus moyen mobile d'ordre 2, défini par la relation suivante :

$$X_k = \varepsilon_k + 0.5 \varepsilon_{k-1} + 0.3 \varepsilon_{k-2} + a_k, \quad k = 3, \dots, 60,$$

où les ε_k suivent une loi normale $N(0, 1)$ puis une loi uniforme $U[-1, 1]$.

Sous H'_A , le tableau suivant nous donne les valeurs des statistiques pour différentes valeurs de l^* et δ .

l^*	δ	$\varepsilon_i \sim N(0, 1)$		$\varepsilon_i \sim U[-1, 1]$	
		D_n	U_n	D_n	U_n
3	2	1.2771	2.0832	1.2991	2.2208
	4	1.2984	2.3327	1.3123	2.5317
	6	1.3615	/	1.3930	/
5	2	1.3730	2.2461	1.4324	2.6130
	4	/	2.7912	/	/
7	1	1.3332	2.1148	1.4710	2.3028
	3	/	2.7920	/	/
10	1	1.4301	2.1861	1.6378	2.5059
	2	/	2.6417	/	/
Valeur critique c		1.32	2.27	1.32	2.27

Tableau 1.3 : Comparaison entre D_n et U_n .

On remarque que pour très courtes épidémies (quand $l^* < 5$), la statistique U_n détecte plus rapidement l'épidémie.

Nous nous intéressons maintenant à la puissance des tests. Le tableau suivant présente les valeurs des puissances des tests basés sur D_n et U_n .

l^*	δ	$\varepsilon_i \sim N(0, 1)$		$\varepsilon_i \sim U[-1, 1]$	
		D_n	U_n	D_n	U_n
3	2	0.9372	0.9576	0.932	0.9706
	4	0.9398	0.9763	0.951	0.9998
	6	0.9435	1	0.9667	1
5	2	0.9502	0.9722	0.9641	0.995
	4	0.967	1	0.9875	1
7	1	0.9418	0.9601	0.9565	0.9765
	3	0.976	0.9976	0.9959	1
10	1	0.9556	0.9689	0.9781	0.9887
	2	0.9861	0.9927	0.9996	0.9999

Tableau 1.4 : Puissance des tests basés sur D_n et U_n .

On remarque que le test basé sur la statistique U_n est le plus puissant dans les deux cas.

2.6 Application aux données réelles

Nous prenons les données présentées dans le chapitre 1 (Exemple 1.5), qui se modélisent selon le modèle MA(1) suivant

$$X_k = -0.682\varepsilon_{k-1} - 0.015, \quad k = 2, \dots, 48.$$

Nous avons d'abord testé la normalité des innovations ε_k à l'aide du test de **Shapiro-Wilk**. La p-value obtenue est $0.0964 > 0.05$, ce qui veut dire qu'on ne peut pas rejeter l'hypothèse de normalité (les innovations sont gaussiennes).

Afin de tester s'il y a une rupture épidémique dans la moyenne de ces données, on calcule les valeurs des statistiques U_n et D_n et on obtient le tableau suivant.

Les statistique	Valeurs obtenues	Valeurs critiques
D_n	1.6085	$C_1 = 1.32$
U_n	2.4097	$C_2 = 2.27$

Tableau 1.5 : Statistiques D_n et U_n pour les données de la maladie de Newcastle

On remarque que les deux statistiques D_n et U_n dépassent leurs valeurs seuil. On rejette donc l'hypothèse nulle H_0 , ce qui implique qu'il y a une rupture épidémique dans la moyenne de ces données.

Pour estimer les bornes p et q de l'épidémie, nous avons appliqué la méthode proposée par Chen et Zhou basée sur la statistique $R_{i,j}$. Cette approche permet d'identifier le segment $[i_0, j_0]$ à l'aide du maximum du $R_{i,j}$.

Pour ces données, le maximum de $R_{i,j}$ est atteint pour $i_0 = 3$ (correspondant à mars 2005) et $j_0 = 27$ (correspondant à mars 2007).

Ceci implique que la rupture épidémique a eu lieu entre avril 2005 et mars 2007.

La Figure ci-dessous présente les trois box-plots correspondants aux trois phases (avant, pendant et après l'épidémie).

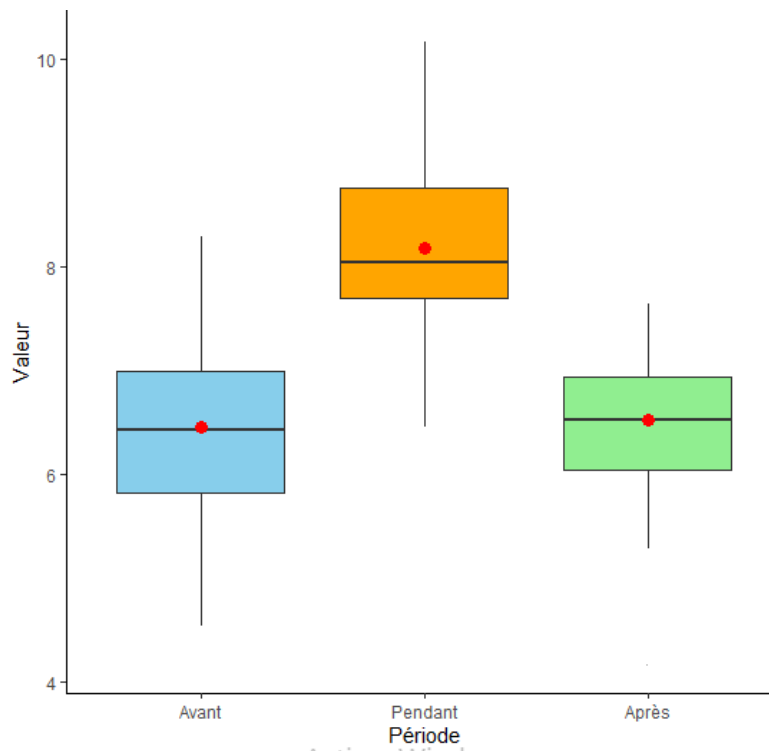


FIG. 2.1 : Boxplots

On observe un changement marqué de la moyenne du processus (le point rouge) vers des valeurs plus élevées pendant la période correspondant à l'épidémie. Cette hausse est suivie d'un retour à la moyenne initiale, ce qui indique la fin de la phase épidémique et un retour à la situation initiale.

Chapitre 3

Tests de rupture épidémique dans un processus AR(1) à innovations faiblement dépendantes

3.1 Introduction

Une rupture dans la moyenne des innovations se produit lorsque l'espérance des innovations change de manière significative à un ou plusieurs instants.

Dans notre cas, on s'intéresse à la rupture épidémique dans la moyenne d'un AR(1) à innovations faiblement dépendantes donnée par :

$$X_t = \phi X_{t-1} + \varepsilon_t + a_k,$$

Le modèle précédent est étudié par Markevičiūtė [12] dans le cas d'innovations i.i.d. Dans ce travail, on s'intéresse au cas où les innovations sont définies par

$$\varepsilon_t = e_t + \sum_{i=1}^q \theta_i e_{t-i}, \quad X_0 = 0, \quad |\phi| < 1$$

et (e_t) sont i.i.d. centrées de variance finie σ^2 .

On veut tester l'hypothèse nulle

$$H'_0 : a_1 = a_2 = \dots = a_n = 0.$$

contre l'hypothèse alternative

$$H'_A : \exists 1 < p < q < n \quad \begin{cases} a_1 = a_2 = \dots = a_n = 0. \\ a_{p+1} = \dots = a_q = \delta, \quad \delta \neq 0. \end{cases}$$

On considère les statistiques hölderiennes définies dans le chapitre 2

$$\tilde{U}_{\alpha,n} = U_{\alpha,n}(X_1, \dots, X_n).$$

$$\tilde{D}_{\alpha,n} = D_{\alpha,n}(X_1, \dots, X_n).$$

3.2 Convergence et consistance

La convergence des statistiques $\tilde{U}_{\alpha,n}$ et $\tilde{D}_{\alpha,n}$ est donnée par le résultat suivant.

Théorème 3.1 On suppose que $\forall 0 < \alpha < \frac{1}{2}$,

$$\lim_{t \rightarrow 0} t^{\frac{1}{1/2-\alpha}} P(|e_1| > t) = 0. \quad (3.1)$$

Alors sous H'_0 et pour $|\phi| < 1$

$$\sigma^{-1}n^{-\frac{1}{2}}(1-\phi)\tilde{U}_{\alpha,n} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} U_\alpha. \quad (a)$$

et

$$\sigma^{-1}n^{-\frac{1}{2}}(1-\phi)\tilde{D}_{\alpha,n} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} D_\alpha. \quad (b)$$

Idées de preuve

On définit d'abord le processus

$$\tilde{Z}_n(t) = \sum_{i=1}^{nt} X_{i-1} + (nt - \lfloor nt \rfloor)X_{(nt)+1}, \quad t \in [0, 1].$$

On montre ensuite que la condition (3.1) implique

$$\sigma^{-1}n^{-1/2}(1-\phi)\tilde{Z}_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} w \quad \text{dans } H_\alpha^0. \quad (3.2)$$

Pour montrer (a) et (b), on va vérifier que

$$\sigma^{-1}n^{-1/2}(1-\phi)\tilde{U}_{\alpha,n} = g(\sigma^{-1}n^{-1/2}(1-\phi)\tilde{Z}_n) + o_p(1).$$

et

$$\sigma^{-1}n^{-1/2}(1-\phi)\tilde{D}_{\alpha,n} = h(\sigma^{-1}n^{-1/2}(1-\phi)\tilde{Z}_n) + o_p(1),$$

avec g et h deux fonctions continues vérifiant $g(w) = U_{\alpha,\infty}$ et $h(w) = D_{\alpha,\infty}$.

De (3.2) et comme g et h sont continues, on déduit

$$\sigma^{-1}n^{-1/2}(1-\phi)\tilde{U}_{\alpha,n} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} U_\alpha.$$

et

$$\sigma^{-1}n^{-1/2}(1-\phi)\tilde{D}_{\alpha,n} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} D_\alpha.$$

La consistance de rejeter H'_0 contre H'_A est donnée par le résultat suivant.

Théorème 3.2. Supposons que la condition (3.1) est satisfaite et

$$\lim_{n \rightarrow +\infty} n^{-\frac{1}{2}+\alpha}(l^*)^{1-\alpha} = +\infty.$$

Alors sous (H'_A)

$$n^{-\frac{1}{2}}(1-\phi)\tilde{U}_{\alpha,n} \xrightarrow[n \rightarrow +\infty]{P} +\infty.$$

et

$$n^{-\frac{1}{2}}(1-\phi)\tilde{D}_{\alpha,n} \xrightarrow[n \rightarrow +\infty]{P} +\infty.$$

Idées de preuve

Soit $I_n = \{k^* + 1, \dots, n^+\}$ et $I_n^c = \{1, 2, \dots, n\} \setminus I_n$.

Sous H'_A , on a

$$X_k = \begin{cases} X'_k + \frac{\delta}{1-\phi} & \text{si } k \in I_n \\ X'_k & \text{si } k \in I_n^c, \end{cases}$$

avec X'_k satisfaisant H'_0 .

Ensuite, on montre que

$$n^{-\frac{1}{2}}(1-\phi)\tilde{U}_{\alpha,n} \geq \frac{1}{2}|\delta|n^{\alpha-\frac{1}{2}}(l^*)^{1-\alpha} - n^{-\frac{1}{2}}(1-\phi)\tilde{U}'_{\alpha,n}$$

et

$$n^{-\frac{1}{2}}(1-\phi)\tilde{D}_{\alpha,n} \geq \frac{1}{2^{2\alpha+2}}|\delta|n^{\alpha-\frac{1}{2}}(l^*)^{1-\alpha} - n^{-\frac{1}{2}}(1-\phi)\tilde{D}'_{\alpha,n},$$

avec $\tilde{U}'_{\alpha,n}$ et $\tilde{D}'_{\alpha,n}$ sont les statistiques basées sur les X'_k .

Du théorème 3.1 et en utilisant la condition (3.1), on déduit

$$n^{-1/2}(1-\phi)\tilde{U}_{\alpha,n} \xrightarrow[n \rightarrow \infty]{P} +\infty.$$

et

$$n^{-1/2}(1-\phi)\tilde{D}_{\alpha,n} \xrightarrow[n \rightarrow \infty]{P} +\infty.$$

3.3 Étude de simulation

Nous examinons les performances des statistiques $\tilde{D}_n = n^{-\frac{1}{2}}(1-\phi)\tilde{D}_{\alpha,n}$ et $\tilde{U}_n = n^{-\frac{1}{2}}(1-\phi)\tilde{U}_{\alpha,n}$, avec $\alpha = 0.25$ pour deux types de variables aléatoires e_i : gaussiennes et uniformes sur l'intervalle $[-1,1]$ et pour deux valeurs du coefficient ϕ au seuil de significativité de 0,05.

1^{er} cas : Les e_i sont gaussiennes

Le tableau suivant nous donne les valeurs des statistiques pour différentes valeurs de l^* et δ et ϕ .

l^*	δ	$\phi = 0.5$		$\phi = 0.8$	
		\tilde{D}_n	\tilde{U}_n	\tilde{D}_n	\tilde{U}_n
2	2	0.9177	1.4264	0.9825	1.3568
	4	1.2456	2.2716	1.0416	1.5608
	6	1.3980	/	1.3447	1.5952
	7.7	/	/	/	2.2813
3	2	1.0522	1.6412	1.0153	1.3792
	4	1.2625	2.5311	1.2404	1.7471
	4.6	1.6453	/	1.3935	2.0319
	5	/	/	/	2.2966
5	1	1.0236	1.5141	1.0517	1.4059
	2	1.3047	2.0556	1.1024	1.4611
	4	1.7307	2.8026	1.7736	2.6244
7	1	1.1693	1.6150	1.2253	1.6636
	3	2.0421	3.0177	1.8030	2.6969
10	1	1.3165	1.7799	1.3690	2.1040
	2	2.2738	3.5603	/	2.9449
15	1	1.7971	3.3140	1.1139	1.9624
	2	/	/	1.5555	3.0685
Valeur critique c		1.32	2.27	1.32	2.27

Tableau 1.1 : Comparaison des statistiques \tilde{D}_n et \tilde{U}_n .

On note que quand l^* augmente, les deux statistiques détectent plus rapidement l'épidémie.

Quand $\phi = 0.5$, on remarque que pour les courtes épidémies ($l^* < 5$), la statistique \tilde{U}_n détecte plus rapidement cette épidémie.

Quand $\phi = 0.8$, il est plus intéressant d'utiliser \tilde{D}_n .

Le tableau suivant présente les valeurs des puissances des tests basés sur \tilde{D}_n et \tilde{U}_n .

l^*	δ	$\phi = 0.5$		$\phi = 0.8$	
		\tilde{D}_n	\tilde{U}_n	\tilde{D}_n	\tilde{U}_n
2	2	0.9152	0.9161	0.9056	0.9048
	4	0.9218	0.9319	0.9091	0.9105
	6	0.9389	0.9894	0.9399	0.9246
	7.7	0.9948	1	0.9649	0.9713
3	2	0.9217	0.9178	0.9100	0.9024
	4	0.9382	0.9874	0.9335	0.9196
	4.6	0.9751	0.9988	0.9497	0.9431
	5	0.9798	0.9996	0.9571	0.9524
5	1	0.9191	0.9152	0.9080	0.9022
	2	0.9424	0.9506	0.9251	0.9146
	4	0.9998	1	0.9937	0.9931
7	1	0.9296	0.9161	0.9108	0.9031
	3	1	1	0.9945	1
10	1	0.9411	0.9255	0.9108	0.9488
	2	0.9997	0.9999	0.9661	0.9825
15	1	0.9334	0.9545	0.9094	0.9143
	2	0.9963	1	0.9858	0.9999

Tableau 1.2 : Les puissances de test basés sur les statistiques D_n et U_n .

On remarque que quand $\phi = 0.5$, le test basé sur la statistique \tilde{U}_n est plus puissant. Quand $\phi = 0.8$, le test basé sur la statistique \tilde{D}_n est plus puissant.

2^{ème} cas : Les e_i sont uniformes sur $[-1,1]$

l^*	δ	$\phi = 0.5$		$\phi = 0.8$	
		\tilde{D}_n	\tilde{U}_n	\tilde{D}_n	\tilde{U}_n
2	2	1.1713	1.7763	0.9104	1.2594
	4.6	1.2929	2.3110	1.4805	2.2313
	5	1.7420	/	/	2.4620
3	2	1.2854	2.2241	1.1267	1.5477
	3	1.3087	3.1529	1.4746	2.1787
	5	1.6138	/	/	3.5499
5	1	1.2583	1.8424	1.0116	1.3391
	2	1.3146	3.1942	1.5901	2.2792
	4	1.9430	/	/	/
7	1	2.0249	2.1704	1.2015	1.6119
	3	/	3.6474	1.9449	3.8939
10	1	2.4744	2.1732	1.3193	2.0091
	2	/	3.8437	2.2971	3.9364
Valeur critique c		1.32	2.27	1.32	2.27

Tableau 1.3 : Comparaison des statistiques \tilde{D}_n et \tilde{U}_n .

On remarque que quand $\phi = 0.5$, pour les courtes épidémies ($l^* < 7$) la statistique \tilde{U}_n détecte plus rapidement cette épidémie.

Quand $\phi = 0.8$, la statistique \tilde{D}_n détecte plus rapidement l'épidémie.

Le tableau suivant présente les valeurs des puissances des tests basés sur \tilde{D}_n et \tilde{U}_n .

l^*	δ	$\phi = 0.5$		$\phi = 0.8$	
		\tilde{D}_n	\tilde{U}_n	\tilde{D}_n	\tilde{U}_n
2	2	0.9277	0.9231	0.9142	0.9056
	4.6	0.9964	1	0.9632	0.9423
	5	1	1	0.9731	0.9715
3	2	0.9505	0.9571	0.9274	0.9118
	3	0.9999	1	0.9937	0.9638
	5	0.9852	1	1	1
5	1	0.9367	0.9275	0.9159	0.9051
	2	0.9963	0.9889	0.9771	0.9764
	4	0.9985	1	0.9992	0.9994
7	1	0.9661	0.9582	0.9300	0.9141
	3	0.9998	0.9999	1	1
10	1	0.9967	0.9991	0.9390	0.9495
	2	0.9996	1	0.9974	0.9999

Tableau 1.4 : Les puissances de test basés sur les statistiques \tilde{D}_n et \tilde{U}_n .

On remarque que pour les courtes épidémies, quand $\phi = 0.5$, le test basé sur la statistique \tilde{U}_n est plus puissant. Par contre, quand $\phi = 0.8$, c'est le test basé sur la statistique \tilde{D}_n qui est le plus puissant.

3.4 Application aux données réelles

Nous prenons les données présentées dans le chapitre 1 (Exemple 1.6), qui se modélisent selon le modèle ARMA(1,2) suivant :

$$X_t = 0.9721X_{t-1} + \varepsilon_t, \quad t = 3, \dots, 481,$$

avec $\varepsilon_t = e_t + 0.4903e_{t-1} + 0.0822e_{t-2} + 6.0076$.

Afin de tester s'il y a une rupture épidémique dans la moyenne des innovations ε_t , on calcule les valeurs des statistiques \tilde{D}_n et \tilde{U}_n basées sur les données et on obtient le tableau suivant.

Les statistique	Valeurs obtenues	Valeurs critiques
\tilde{D}_n	1.4358	$C_1 = 1.32$
\tilde{U}_n	1.5115	$C_2 = 2.27$

Tableau 1.5 : Les valeurs des statistiques \tilde{D}_n et \tilde{U}_n

On remarque que $\tilde{D}_n > C_1$, ce qui indique qu'il y a une rupture épidémique dans la moyenne des innovations, qu'on peut voir sur le graphe suivant.

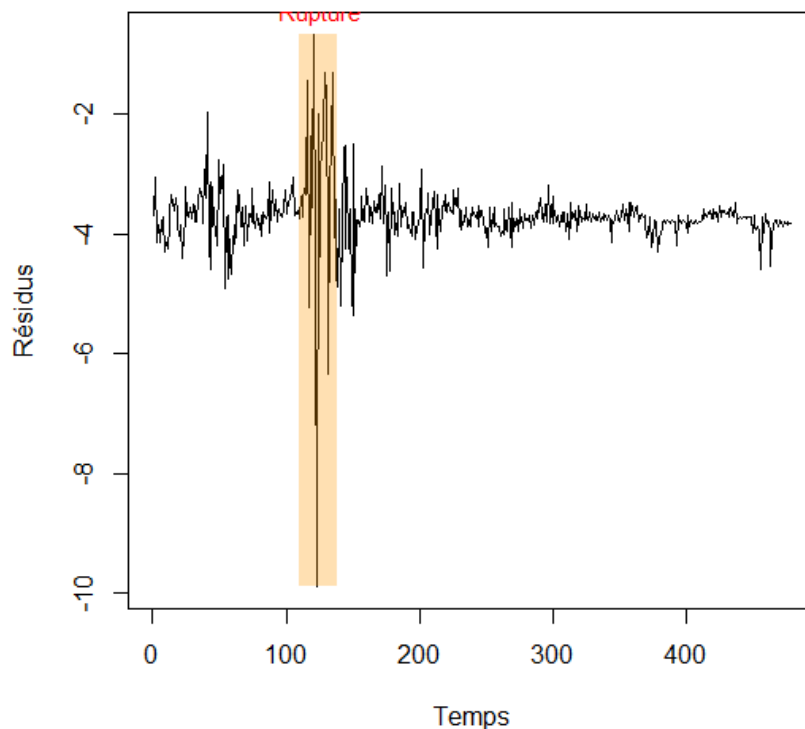


FIG. 3.1 : Graphe des innovations

On note que $\tilde{U}_n < C_2$, ce qui implique que la statistique \tilde{U}_n n'a pas détecté cette épidémie.

Pour estimer le début p et la fin q de cette épidémie, nous avons calculé le maximum du $R_{i,j}$, qui est atteint pour $i_0 = 110$ et $j_0 = 139$.

Ceci implique que la rupture épidémique a lieu entre avril 1979 et juillet 1981.

Figure ci-dessous présente les trois box-plots correspondants à ces trois phases :

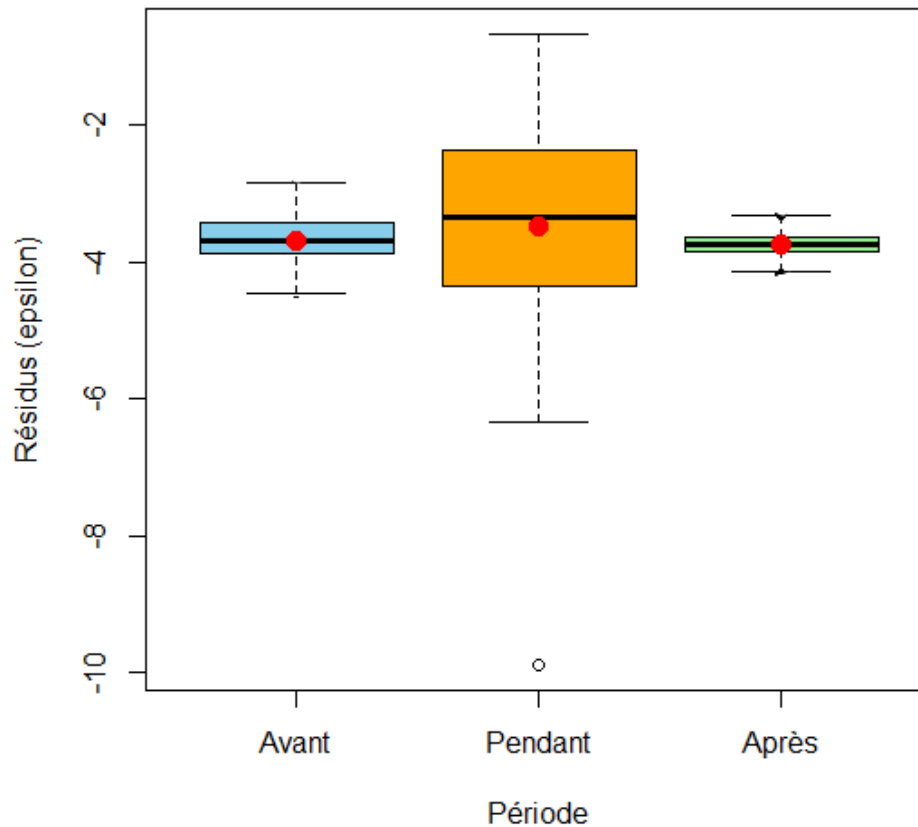


FIG. 3.2 : Boxplots

On remarque qu'il n'y a pas de différence significative dans la moyenne des innovations entre les périodes avant et après l'épidémie, par contre, il y a une différence significative pendant l'épidémie, ce qui montre qu'il y a un changement dans la moyenne puis un retour à l'état initial.

Conclusion Générale

Ce travail a porté sur la détection de ruptures épidémiques dans la moyenne d'un processus autorégressif d'ordre 1 avec des innovations faiblement dépendantes.

Après un rappel théorique sur les séries chronologiques et les principaux modèles linéaires (AR, MA, ARMA), nous avons étudié quelques statistiques de test afin de détecter une rupture épidémique dans la moyenne des variables aléatoires, notamment les statistiques classiques et les statistiques hölderiennes.

Nous avons ensuite adapté les statistiques hölderiennes au cadre spécifique des processus AR(1), à travers des simulations utilisant différents types d'innovations (gaussiennes et uniformes) et différentes valeurs du paramètre autorégressif ϕ . Enfin, une application sur des données réelles a été présentée.

Ce travail ouvre plusieurs perspectives de recherche, notamment l'extension à des modèles plus complexes ou présentant une dépendance forte, ainsi que le développement de méthodes alternatives, telles que les approches non paramétriques ou bayésiennes.

Annexe A

Mouvement brownien et pont brownien

A.1 Mouvement brownien

A.1.1 Définition

On dit que le processus $(\xi_t)_{t \geq 0}$ est un **mouvement brownien** s'il s'agit d'un processus gaussien continu, tel que :

$$\mathbb{E}[\xi_t] = \mu t,$$

et

$$\text{Cov}(\xi_t, \xi_s) = \sigma^2(s \wedge t), \quad \forall s, t \geq 0,$$

où $s \wedge t = \min(s, t)$.

Il est possible de ramener ce mouvement brownien à un mouvement brownien canonique (ou standard) en effectuant le changement de variable suivant :

$$W_t = \frac{\xi_t - \mu t}{\sigma}$$

Le processus $(\zeta_t)_{t \geq 0}$ obtenu est alors centré et de covariance standard :

$$\mathbb{E}[W_t] = 0, \quad \text{et} \quad \text{Cov}(W_t, W_s) = s \wedge t$$

A.1.2 Les accroissements du mouvement brownien

Proposition A.1 Soit $(\xi_t)_{t \geq 0}$ un mouvement brownien. Pour tous $0 \leq s < t$, l'accroissement $\xi_t - \xi_s$ suit une loi normale $\xi_t - \xi_s \sim \mathcal{N}(0, t - s)$.

Cette proposition implique que le mouvement brownien a des accroissements stationnaires.

Proposition A.2 Pour tout $0 = t_0 \leq t_1 \leq t_2 \leq \dots \leq t_k$, les accroissements successifs :

$$\xi_{t_1} - \xi_{t_0}, \xi_{t_2} - \xi_{t_1}, \dots, \xi_{t_k} - \xi_{t_{k-1}}$$

sont indépendantes. Cela signifie que le mouvement brownien a des accroissements indépendants.

A.1.3 Propriétés du mouvement brownien

Propriété de Markov

Le mouvement brownien est un processus de Markov, c'est-à-dire

Pour tout $t \leq s$ et $A \in \mathcal{B}(\mathbb{R})$, avec $\mathcal{F}_t = \sigma(W_\tau, \tau \leq t)$,

$$P(W_s \in A \mid \mathcal{F}_t) = P(W_s \in A \mid W_t).$$

Démonstration : Il suffit de vérifier que pour tout $t_1 \leq t_2 \leq \dots \leq t_k \leq s$, $\forall k$ on a :

$$P(W_s \in A \mid W_{t_1}, W_{t_2}, \dots, W_{t_k}) = P(W_s \in A \mid W_{t_k}).$$

En effet, les accroissements $(W_s - W_{t_k}), (W_{t_k} - W_{t_{k-1}}), \dots, (W_{t_2} - W_{t_1}), (W_{t_1} - W_0)$ sont indépendants.

Ainsi, la loi de W_s conditionnellement aux : $W_{t_k} - W_{t_{k-1}}, \dots, W_{t_2} - W_{t_1}, W_{t_1} - W_0, W_{t_1}$ est une loi normale, de variance $s - t_k$, ce qui implique :

$$P(W_s \in A \mid \mathcal{F}_t) = P(W_s \in A \mid W_t),$$

Propriété des martingales

Soit $(W_t)_{t \geq 0}$ un mouvement brownien, alors les processus suivants sont des martingales par rapport à (\mathcal{F}_t) :

1. $(W_t)_{t \geq 0}$.
2. $(W_t^2 - t)_{t \geq 0}$.
3. $\left(\exp \left(\lambda W_t - \frac{\lambda^2 t}{2} \right) \right)_{t \geq 0}$.

Invariance du mouvement brownien

Soit W_t un mouvement brownien, alors les processus suivants sont également des mouvements browniens :

1. $-W_t$ (c'est la symétrie du mouvement brownien).
2. $\frac{1}{\sqrt{c}} W_{ct}$, $c > 0$ (propriété d'autosimilarité du mouvement brownien).

3. $W_{t+s} - W_s$, avec $s \geq 0$ (propriété d'invariance par translation du mouvement brownien).
4. W'_t défini par : $W'_0 = 0$, $W'_t = tW_{\frac{1}{t}}$ pour $t > 0$ (propriété d'invariance par inversion du temps du mouvement brownien).

Régularité trajectorielle

On définit un mouvement brownien standard qu'on notera W_t comme étant un processus gaussien centré, à accroissements indépendants et stationnaires et à trajectoires continues et γ -höldérienne, pour tout $\gamma \in]0, \frac{1}{2}[$.

Notons que $W_0 = 0$ p.s (car la loi de W_0 est $\mathcal{N}(0, 0)$) et $W_t \sim \mathcal{N}(0, t)$ car $\mathbb{E}(W_t) = 0$ et $\text{Var}(W_t) = \Gamma(t, t) = t$.

Irrégularité des trajectoires

On a déjà vu que les trajectoires sont γ -höldériennes en tout point et pour tout $0 \leq \gamma < \frac{1}{2}$.

Nous allons maintenant voir que presque toutes les trajectoires sont nulle part γ -höldériennes pour $\gamma \geq \frac{1}{2}$.

Pour presque tout ω , $(W_t(\omega))_{t \geq 0}$ n'est en aucun point γ -höldérienne pour $\gamma \geq \frac{1}{2}$.

Comme conséquence de ce résultat, on peut déduire que :

1. Presque toute trajectoire est nulle part dérivable.
2. Presque toute trajectoire est à variation non bornée.

A.2 Pont brownien

Le pont brownien est un processus gaussien centré, noté $\{B_t, t > 0\}$, tel que

- i) $\mathbb{E}[B_t] = 0$,
- ii) $\mathbb{E}[B_s B_t] = s(1 - t)$, $\forall s < t$.

On le définit de la façon suivante.

$$B_t = W_t - tW_1.$$

Comme W_t et W_1 sont des éléments aléatoires gaussiens, B_t est aussi un élément aléatoire gaussien. De plus, on a :

$$B_0 = B_1 = 0, \quad \mathbb{E}[B_t] = 0,$$

$$\mathbb{E}[(B_t - B_s)^2] = (t - s)(1 - (t - s)) \quad \text{si } s < t,$$

$$\mathbb{E}[(B_{s_2} - B_{s_1})(B_{t_2} - B_{t_1})] = -(s_2 - s_1)t_2 \quad \text{si } s_1 \leq s_2 \leq t_1 \leq t_2.$$

Le pont brownien admet les mêmes propriétés que le mouvement brownien.

Annexe B

Programmes de simulation

Dans cette annexe, nous présentons le programme de simulation des deux statistiques de test présentées dans le chapitre 3 ainsi que le programme de calcul de leurs puissances codés en langage R.

B.1 Programme de simulation des statistiques

```
n <- 60      #taille d'échantillon
a <- 0.25
p <- 20
l <- 3
q <- p + 1
delta <- 5
phi <- 0.5   #coefficient

k <- floor(log2(n))
n_simulations <- 1000
x <- numeric(n)
eps <- numeric(n)
Dn <- numeric(n_simulations)
Un <- numeric(n_simulations)

for (s in 1:n_simulations) {
  e <- rnorm(n)      #innovations

  eps[1] <- e[1]
  eps[2] <- e[2] + 0.5 * e[1]
  for (i in 3:p) {
    eps[i] <- e[i] + 0.5 * e[i - 1] + 0.3 * e[i - 2]
  }

  x[1] <- eps[1]
  x[2] <- eps[2] + phi * x[1]
  for (i in 3:p) {
    x[i] <- eps[i] + phi * x[i - 1]
  }
  for (i in (p + 1):q) {
    x[i] <- eps[i] + phi * x[i - 1] + delta
  }
  for (i in (q + 1):n) {
    x[i] <- eps[i] + phi * x[i - 1]
  }
}
```

```

x <- x[1:n]
S <- cumsum(x)      #somme cumulée

U <- matrix(0, nrow = n - 2, ncol = n - 1)
V <- matrix(0, nrow = n - 2, ncol = n - 1)
R <- matrix(0, nrow = n - 2, ncol = n - 1)

for (j in 2:(n - 1)) {
  for (i in 1:(j - 1)) {
    U[i, j] <- S[j] - S[i]
    V[i, j] <- S[n] - U[i, j]
    R[i, j] <- ((n - (j - i)) * U[i, j] - (j - i) * V[i, j]) / n
  }
}

M <- R[p, q]      #pour trouver les bornes d'épidémie
#la statistique D_n
D1 <- matrix(0, nrow = k, ncol = 2^(k - 1))
for (j in 2:k) {
  for (l in 2:(2^(j - 1))) {
    S1 <- S[floor(n * ((2 * l - 1) / (2^j)))]
    S2 <- S[floor((2 * n * (l - 1)) / (2^j))]
    S3 <- S[floor((2 * n * l) / (2^j))]
    D1[j, l] <- (2^(j * a)) * abs(S1 - (1 / 2) * S2 - (1 / 2) * S3)
  }
}

D2 <- max(D1, na.rm = TRUE)
D3 <- max(D2, na.rm = TRUE)
S11 <- numeric(k)
S12 <- numeric(k)
D4 <- numeric(k)

for (j in 1:k) {
  S11[j] <- S[floor(n / (2^j))]
  S12[j] <- S[floor((2 * n) / (2^j))]
  D4[j] <- (2^(j * a)) * abs(S11[j] - (1 / 2) * S12[j])
}

D5 <- max(D4)
D_a <- max(D3, D5)
#la statistique U_n définie dans le cours

U1 <- matrix(0, nrow = n, ncol = n - 1)
for (i in 2:n) {
  for (j in 1:(i - 1)) {
    U1[i, j] <- abs(S[j] - S[i] - S[n] * ((j - i) / n)) /
      (abs(((j - i) / n) * (1 - ((j - i) / n)))^a)
  }
}
U_a <- max(U1)
#la convergence
Dn[s] <- (1 / sd(e)) * n^(-0.5) * (1 - phi) * D_a
Un[s] <- (1 / sd(e)) * n^(-0.5) * (1 - phi) * U_a

cat("Simulation", s, ": Dn =", Dn[s], ", Un =", Un[s], "\n")
}
#les moyennes
Dn_mean <- mean(Dn)

```

```
Un_mean <- mean(Un)
```

```
Dn_mean
```

```
Un_mean
```

B.2 Programme de simulation des puissances

```

n <- 60
a <- 0.25
p <- 20
l <- 10
q <- p + 1
delta <- 2
  b1 <- 1.32
  b2 <- 2.27
i_0 <- p
j_0 <- q
n_simulations <- 1000
x <- numeric(n)
eps <- numeric(n)
#les statistiques
Dn <- numeric(n_simulations)
Un <- numeric(n_simulations)

for (s in 1:n_simulations) {

e <- rnorm(n)    #les innovations du modele ARMA
eps[1] <- e[1]
  eps[2] <- e[2] + 0.5 * e[1]
  for (i in 3:p) {
    eps[i] <- e[i] + 0.5 * e[i - 1] + 0.3 * e[i - 2]
  }
x[1] <- eps[1]
x[2] <- eps[2] + phi * x[1]
  for (i in 3:p) {
    x[i] <- eps[i] + phi * x[i - 1]
  }
  for (i in (p + 1):q) {
    x[i] <- eps[i] + phi * x[i - 1] + delta
  }
  for (i in (q + 1):n) {
    x[i] <- eps[i] + phi * x[i - 1]
  }

x<-x[1:n]
  S <- cumsum(x)
  k <- floor(log2(n))
  U <- matrix(0, nrow = n - 2, ncol = n - 1)
  V <- matrix(0, nrow = n - 2, ncol = n - 1)
  R <- matrix(0, nrow = n - 2, ncol = n - 1)
for (j in 2:(n-1)) {
  for (i in 1:(j-1)) {
    U[i, j] <- S[j] - S[i]
    V[i, j] <- S[n] - U[i, j]
    R[i, j] <- ((n - (j - i)) * U[i, j] - (j - i) * V[i, j]) / n
  }
}
max(R)    #pour trouver les bornes d'épidémie

#la statistiques D_n

D1 <- matrix(0, nrow = k, ncol = 2^(k - 1))
for (j in 2:k) {
  for (l in 2:(2^(j - 1))) {

```

```
S1 <- S[floor(n * ((2 * 1 - 1) / (2^j)))]
S2 <- S[floor((2 * n * (1 - 1)) / (2^j))]
S3 <- S[floor((2 * n * 1) / (2^j))]
D1[j, 1] <- (2^(j * a)) * abs(S1 - (1/2) * S2 - (1/2) * S3)
}
}
D2 <- max(D1)
D3 <- max(D2)
S11 <- numeric(k)
S12 <- numeric(k)
D4 <- numeric(k)
for (j in 1:k) {
  S11[j] <- S[floor(n / (2^j))]
  S12[j] <- S[floor((2 * n) / (2^j))]
  D4[j] <- (2^(j * a)) * abs(S11[j] - (1/2) * S12[j])
}
D5 <- max(D4)
DY <- max(D3, D5)

#la statistique U_n
U1 <- matrix(0, nrow = n, ncol = n - 1)
for (i in 2:n) {
  for (j in 1:(i - 1)) {
    U1[i,j] <- abs(S[i] - S[j] - S[n] * ((i - j) / n)) / abs(((i - j) / n) * (1 - ((i - j) / n)))^a
  }
}
U2 <- max(U1)
U2<-max(U2)

Dn[s] <- (1/sd(e))* n^(-1 / 2) * DY
Un[s] <- (1/sd(e))* n^((-1 / 2)) * U2

}
Power_Dn <- sum(Dn <= b1)
Power_Un <- sum(Un <= b2)
B1=1-(Power_Dn/10000)
B2=1-(Power_Un/10000)
cat("Puissance du test Dn,Un: ", Power_Dn, " ", Power_Un, "\n")
B1
B2
```

Bibliographie

- [1] Avery P.J. et Henderson D.A. (1999). Detecting a changed segment in DNA sequences. *J. Roy. Statist. Soc. Ser. C*, 48(4), 489–503.
- [2] Brockwell P.J. et Davis R.A. (1987). *Time Series : Theory and Methods*, Springer-Verlag, New York.
- [3] Chen Z., Li Z. et Zhou M. (2016). *Detecting change-points in epidemic models*, Journal of Advanced Statistics, vol. 1, no. 4, pp. 181–190.
- [4] Chen Z., Li Z. et Zhou M. (2020), *Detecting Change-Points in Epidemic Models*.
- [5] Djebbar R. (2016). *Polycopié du cours : Statistique des processus, Université 8 mai 1945 Guelma*.
- [6] Fangge L. et Peixian L. (2011). *ARMA Model for Predicting the Number of New Outbreaks of Newcastle Disease during the Month*, p 660–662.
- [7] Imeçaoudene K. et Hamadouche D. (2017). *Gaussian approximation for residuals of stationary autoregressive process in Hölder norm*, *Theory of Stochastic Processes*, no. 2, 19-33.
- [8] Gombay E. (1994). *Testing for change-points with rank and sign statistics*. *Statistics and Probability Letters*, 20(1), 49–55.
- [9] Graiche F., Merabet D. et Hamadouche D. (2016). *Testing change in the variance with epidemic alternatives*. *Communications in Statistics – Theory and Methods*, vol. 45, no. 13, pp. 3822–3837.
- [10] Graiche F. (2025). *Testing epidemic change in the mean of weakly dependent random variables*, hal-05056614.
- [11] Levin B. et Kline J. (1985). *The CUSUM test of homogeneity with an application in spontaneous abortion epidemiology*, *Statistics in Medicine*, vol. 4, no 4, pp. 469–488.
- [12] Markevičiūtė J. (2016). *Epidemic change tests for mean of innovations of an AR(1) process*, *statistics and probability Letters* 112,79-91.
- [13] Pelgrin F. (2011-2012). *ARMA(p,q) models (part 3)*, University of Lausanne, Ecole des HEC, Department of Mathematics (IMEA-Nice).

- [14] Rackauskas A. et Suquet Ch. (2004). *Hölder norm test statistics for epidemic change*, *J. Statist. Plann. Infer.* 126, 495-520.
- [15] Rackauskas A. et Suquet Ch. (2011). *On the distribution of sequential Hölder norm of the Brownian motion*, *Pub. IRMA*, Lille, 71, no. 9, 1-29.
- [16] Yao Q. (1993). *Tests for change-points with epidemic alternatives*, *Biometrika*, vol. 80, no 1, p. 179–191.

Résumé

Dans ce travail, on s'intéresse à la détection de rupture épidémique dans la moyenne du processus autorégressif d'ordre 1 à innovations faiblement dépendantes. Après une présentation des concepts fondamentaux des séries chronologiques et des modèles ARMA, nous étudions quelques statistiques de test classiques et hölderiennes, pour détecter une rupture épidémique dans la moyenne des variables aléatoires. Nous adaptons ensuite les statistiques hölderiennes au cas de données suivant un autorégressif d'ordre 1 et évaluons leurs performances à travers des simulations et une application sur des données réelles.

Abstract

In this work, we're focusing on detecting epidemic change-points in the mean of an autoregressive process of order 1 with weakly dependent innovations. After presenting the fundamental concepts of time series and ARMA models, we examine several statistical tests - both classical and Hölderian - to detect these epidemic changes in the mean of random variables. We then adapt the Hölderian statistics to the autoregressive of order 1 case and evaluate their performance through simulations and an application to real data.

الملخص

يتناول هذا العمل مسألة الكشف عن التغيرات الوبائية في متوسط عملية انحدار ذاتي من الدرجة الأولى مع ابتكارات ضعيفة الارتباط. بعد تقديم نظري لمفاهيم السلاسل الزمنية والنماذج الخطية الكلاسيكية، قمنا بدراسة عدة اختبارات إحصائية، منها التقليدية ومنها الهولدرية. ثم قمنا بتكييف هذه الاختبارات مع نموذج انحدار من الدرجة الأولى وتم تقييم فعاليتها من خلال المحاكاة وتطبيقات على بيانات حقيقية.