

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE  
SCIENTIFIQUE  
UNIVERSITE MOULOUD MAMMERRI, TIZI-OUZOU  
FACULTE DES SCIENCES  
DEPARTEMENT DE MATHEMATIQUES



MEMOIRE DE MASTER

SPECIALITÉ : MATHÉMATIQUES

OPTION : PROBABILITÉS ET STATISTIQUE

Sujet :

# Rupture épidémique dans les innovations d'un processus autorégressif d'ordre 1.

Présenté par

M<sup>elle</sup> KRIM FARAH

Devant le jury d'examen composé de :

M. Mehiri Mohammed	Maître Assistant A ;	U.M.M.T.O ;	Président
M. Graïche Farid	Maître de Conférences B ;	U.M.M.T.O ;	Rapporteur
Mme. Merabet Dalila	Maître de Conférences B ;	U.M.M.T.O ;	Examinatrice

Soutenu : Le 30/06/2024

# Résumé

Dans ce travail, on s'intéresse à la rupture épidémique dans les processus autorégressifs d'ordre 1. On présente d'abord quelques statistiques classiques et hölderiennes pour tester la rupture épidémique dans la moyenne des variables aléatoires, puis on utilise les statistiques hölderiennes pour détecter la rupture épidémique dans la moyenne des innovations du processus AR(1).

# Abstract

In this work, our aim is to study the epidemic change in autoregressive processes of order 1. We first present some classical and Hölderian statistics for testing the epidemic change in the mean of random variables, and then use Hölderian statistics to detect an epidemic change in the mean of innovations of the AR(1) process.

## ملخص

هدفنا في هذا العمل هو دراسة التغير الوبائي في عمليات الانحدار التلقائي من الرتبة 1. نقدم أولاً بعض الإحصاءات الكلاسيكية وإحصاءات هولدر لاختبار التغير الوبائي في متوسط المتغيرات العشوائية، ثم نستخدم إحصاءات هولدر للكشف عن التغير الوبائي في متوسط ابتكارات عملية الانحدار التلقائي من الرتبة الأولى.

# Table des matières

<b>Introduction</b>	<b>4</b>
<b>1 Processus stationnaires et processus autorégressifs</b>	<b>6</b>
1.1 Processus Stochastiques . . . . .	6
1.2 Processus stationnaire . . . . .	7
1.2.1 Propriétés . . . . .	8
1.2.2 Exemple d'un processus stationnaire . . . . .	8
1.2.3 Exemple d'un processus non stationnaire: Marche aléatoire . . . . .	9
1.3 Processus autorégressif d'ordre un . . . . .	9
1.3.1 Définition . . . . .	9
1.3.2 Propriétés des processus AR(1) . . . . .	10
1.3.3 Moments d'un processus AR(1) . . . . .	10
1.3.4 Estimation du paramètre autorégressif . . . . .	12
1.3.5 Inversibilité et causalité d'un processus AR(1) . . . . .	13
1.3.6 Exemple d'un AR(1) . . . . .	13
1.4 Processus autorégressifs d'ordre p . . . . .	14
1.4.1 Définition . . . . .	14
1.4.2 Conditions de stationnarité d'un AR(p) . . . . .	15
1.4.3 Condition de stationnarité d'un processus AR(2) . . . . .	15
1.4.4 Moments d'un processus AR(p) . . . . .	15
1.4.5 Estimation des paramètres autorégressifs . . . . .	18
1.4.6 Inversibilité et causalité d'un processus AR(p) . . . . .	18
1.4.7 Exemple d'un AR(2) . . . . .	18
<b>2 Tests de rupture épidémique dans la moyenne des variables aléatoires.</b>	<b>20</b>
2.1 Introduction . . . . .	20
2.2 Statistiques de tests classiques . . . . .	21
2.2.1 Statistique de Levin et Kline . . . . .	21
2.2.2 Statistique du rapport de vraisemblance . . . . .	21
2.3 Statistiques de tests hölderiennes . . . . .	22
2.3.1 Espaces de Hölder . . . . .	22

2.3.2	Statistiques de Rakauskas et Suquet . . . . .	23
2.3.3	Convergence et consistance . . . . .	24
2.4	Étude de simulation . . . . .	25
2.5	Application aux données réelles . . . . .	27
<b>3</b>	<b>Tests de rupture épidémique dans la moyenne des innovations d'un AR(1)</b>	<b>28</b>
3.1	Introduction . . . . .	28
3.2	Statistiques de tests basés sur les résidus . . . . .	29
3.2.1	Étude de simulation . . . . .	30
3.3	Statistiques de tests basés sur les observations . . . . .	34
3.3.1	Étude de simulation . . . . .	35
3.4	Application aux données réelles . . . . .	39
	<b>Conclusion</b>	<b>40</b>
	<b>Annexes</b> . . . . .	<b>40</b>
<b>A</b>	<b>Mouvement Brownien et pont Brownien</b>	<b>41</b>
A.1	Mouvement brownien . . . . .	41
A.1.1	Définition du mouvement brownien . . . . .	41
A.1.2	Propriétés du mouvement brownien . . . . .	41
A.2	Pont brownien . . . . .	43
<b>B</b>	<b>Programmes de simulation</b>	<b>45</b>
	<b>Bibliographie</b>	<b>50</b>

# Introduction générale

Un point de rupture est un moment où les paramètres d'une distribution de données, tels que la moyenne, la variance, ou d'autres changent.

Parmi les premiers chercheurs à avoir étudié les problèmes de rupture, on trouve Chernoff et Rubin [5], Chernoff et Zacks [6], Sen et Srivastava [23], Pettitt [20], Gombay et Horváth [8], ainsi que plusieurs autres.

Levin et Kline [17], Bromeling et Tsurumi [3], Yao [25], Avery et Henderson [1] ainsi que Rakauskas et Suquet [22], ont étudié les problèmes de rupture épidémique, un type caractérisé par une variation significative dans un paramètre de la distribution des variables aléatoires observées, suivie d'un retour à la situation initiale.

En général, l'inférence statistique sur les points de rupture comporte deux aspects : tester la présence d'une rupture dans la séquence des variables aléatoires observées, et estimer le nombre de points de rupture ainsi que leurs positions.

Dans ce travail, nous nous intéressons aux tests de rupture épidémique dans la moyenne des innovations d'un processus autorégressif d'ordre 1. Ce type de processus est couramment utilisé pour modéliser les séries chronologiques dans divers domaines tels que la médecine, la finance, la météorologie... ( voir par exemple Brockwell et Davis [2]).

Ce mémoire est composé de trois chapitres et trois annexes.

Dans le premier chapitre, nous introduisons d'abord les processus autorégressifs d'ordre 1 (AR(1)) et nous présentons quelques propriétés. Ensuite, nous donnons un exemple de données réelles récentes qui se modélise selon un AR(1). Nous terminons par une généralisation aux processus autorégressifs d'ordre  $p$  ( $AR(p), p > 1$ ).

Le deuxième chapitre est consacré aux tests statistiques pour la détection de la rupture épidémique dans la moyenne des variables aléatoires. Nous débutons par la présentation de quelques statistiques classiques, telles que la statistique de Levin et Kline

[17], la statistique du rapport de vraisemblance proposé par Siegmund [24], et enfin les statistiques hölderiennes proposées par Rakauskas et Suquet [22].

Pour comparer les performances et la puissance de ces statistiques, nous réalisons une étude de simulation. Nous concluons ce chapitre par une application sur des données réelles.

Dans le troisième chapitre, nous testons la rupture épidémique dans la moyenne des innovations d'un AR(1) en utilisant les statistiques hölderiennes définies au chapitre 2, basées d'abord sur les résidus et ensuite sur les observations du processus AR(1). Nous effectuons une étude de simulation pour deux types d'innovations : gaussiennes et uniformes sur  $[-1, 1]$  et pour deux valeurs du paramètre autorégressif . Nous terminons par un exemple d'application sur les données réelles.

Les propriétés du mouvement brownien et du pont brownien sont présentées dans l'annexe A. Dans l'annexe B, on donne quelques programmes de simulation utilisés dans ce mémoire codés en Language R.

# Chapitre 1

## Processus stationnaires et processus autorégressifs

### 1.1 Processus Stochastiques

#### Définition 1.1

Un processus stochastique  $(Y_t)_{t \in T}$  est une famille de variables aléatoires  $Y_t$  indexée par un ensemble  $T$ . En général  $T = \mathbb{R}_+$  ou  $\mathbb{R}$ . Si  $T = \mathbb{N}$  alors le processus est une suite de variables aléatoires. Plus généralement quand  $T \subseteq \mathbb{Z}$ , le processus est dit discret ou une série temporelle. Pour  $T \subseteq \mathbb{R}^d$ , on parle de champ aléatoire (drap quand  $d = 2$ ). Si  $T$  n'est pas dénombrable, on a un processus continu.

Un processus dépend de deux paramètres :  $Y_t(\omega)$  dépend de  $t$  (en général le temps) et de l'aléatoire  $\omega \in \Omega$  :

- Pour  $t \in T$  fixé,  $\omega \mapsto Y_t(\omega)$  est une variable aléatoire sur l'espace de probabilité  $(\Omega, \mathcal{F}, \mathbb{P})$ .
- Pour  $\omega$  fixé,  $t \mapsto Y_t(\omega)$  est une fonction à valeurs réelles, appelée trajectoire du processus.

#### Définition 1.2

Un processus  $(Y_t)$  est dit à accroissements stationnaires si la loi des accroissements  $Y_{t+h} - Y_t$  ne dépend pas de  $t > 0$ , c'est-à-dire  $Y_{t+h} - Y_t \stackrel{\mathcal{L}}{=} Y_h$ .

#### Définition 1.3

Un processus  $(Y_t)$  est dit à accroissements indépendants si pour tout  $p \geq 1$  et  $0 < t_1 < t_2 < \dots < t_p$ , les variables aléatoires  $Y_{t_1}, Y_{t_2} - Y_{t_1}, \dots, Y_{t_p} - Y_{t_{p-1}}$  sont indépendantes.

**Définition 1.4**

Un processus  $(Y_t, t \in \mathbb{Z})$  admet une représentation causale s'il peut s'écrire comme combinaison linéaire des valeurs passées d'un autre processus, c'est-à-dire, qu'il existe une suite  $(\psi_i, i \in \mathbb{N})$  et un processus  $(X_t, t \in \mathbb{Z})$  tels que

$$Y_t = \sum_{i \in \mathbb{N}} \psi_i X_{t-i}, \quad \forall t \in \mathbb{Z}.$$

**Définition 1.5**

Un processus  $(Y_t, t \in \mathbb{Z})$  admet une représentation inversible s'il peut s'écrire comme combinaison linéaire des valeurs d'un autre processus, c'est-à-dire, qu'il existe une suite  $(\psi_i, i \in \mathbb{Z})$  et un processus  $(X_t, t \in \mathbb{Z})$  tels que

$$X_t = \sum_{i \in \mathbb{Z}} \psi_i Y_{t-i}, \quad \forall t \in \mathbb{Z}.$$

**Définition 1.6**

Un processus gaussien à temps discret  $(Y_t)_{t \in \mathbb{Z}}$  est un processus discret telle que la loi de nimporte quel vecteur extrait est gaussien, c'est-à-dire

$$\forall n \in \mathbb{N}^*, \forall (t_1, \dots, t_n) \in \mathbb{Z}^n : (Y_{t_1}, \dots, Y_{t_n}) \text{ est un vecteur gaussien.}$$

## 1.2 Processus stationnaire

**Définition 1.7**

Un processus est dit (strictement) stationnaire si pour tout  $h \in \mathbb{Z}$ ,  $(Y_{t+h})_{t \in \mathbb{Z}} \stackrel{\mathcal{L}}{=} (Y_t)_{t \in \mathbb{Z}}$  ne dépend pas de  $h$ , c'est-à-dire pour tout  $h \in \mathbb{Z}$  et tout  $t_1, \dots, t_p \in \mathbb{Z}$ , on a

$$(Y_{t_1+h}, \dots, Y_{t_p+h}) \stackrel{\mathcal{L}}{=} (Y_{t_1}, \dots, Y_{t_p}).$$

**Définition 1.8**

Le processus  $\{Y_t; t \in \mathbb{Z}\}$  est dit faiblement stationnaire (ou stationnaire du second ordre), si les trois conditions suivantes sont satisfaites :

(i)  $\forall t \in \mathbb{Z}, \quad E(Y_t) = \mu$ , indépendante de  $t$ .

(ii)  $\forall t \in \mathbb{Z}, \quad V(Y_t) = \sigma^2 < \infty$ , indépendante de  $t$ .

(iii)  $\forall (t, h) \in \mathbb{Z}^2, \quad \text{Cov}(Y_t, Y_{t+h}) = E[(Y_{t+h} - \mu)(Y_t - \mu)] = \gamma_h$ , ne dépend que de  $h$ .

$\gamma_h$  est dite l'auto-covariance d'ordre  $h$  de  $y_t$ .

**1.2.1 Propriétés**

1. De (iii) si  $h = 0$  on aura,  $\gamma(0) = \text{Var}(Y_t) = \sigma^2$ .
2. La stationnarité stricte implique la stationnarité faible (la réciproque n'est vraie que pour les processus gaussiens).

**1.2.2 Exemple d'un processus stationnaire**

Un exemple classique de processus stationnaire est le bruit blanc.

**Définition 1.9** Un bruit blanc fort est une suite de variables aléatoires indépendantes et identiquement distribuées centrées et de variance finie.

$$\mathbb{E}(\varepsilon_t) = 0. \quad \text{et} \quad V(\varepsilon_t) = \sigma^2.$$

On note  $\varepsilon_t \sim \text{i.i.d}(0, \sigma^2)$ .

**Définition 1.10** Un bruit blanc faible est une suite de variables aléatoires non-corrélées centrées et de variance finie.

$$\mathbb{E}(\varepsilon_t) = 0.$$

$$\gamma(h) = \mathbb{E}(\varepsilon_t \cdot \varepsilon_{t+h}) = \begin{cases} \sigma^2 & \text{si } h = 0. \\ 0 & \text{si } h \neq 0. \end{cases}$$

On note  $\varepsilon_t \sim \text{BB}(0, \sigma^2)$ .

**Définition 1.11** Un bruit blanc gaussien est un bruit blanc fort pour lequel la loi commune des variables aléatoires est une distribution normale de moyenne nulle et de variance  $\sigma^2$ , on note  $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$ .

### 1.2.3 Exemple d'un processus non stationnaire : Marche aléatoire

La suite  $(Y_t, t \in \mathbb{Z})$  est une marche aléatoire (*random walk*) lorsqu'elle s'écrit sous la forme suivante :

$$Y_t = Y_{t-1} + \varepsilon_t.$$

où  $\varepsilon_t$  est un bruit blanc de moyenne nulle et de variance  $\sigma^2$  et  $Y_0$  dénote une valeur initiale quelconque. On note que son espérance  $\mathbb{E}(Y_t)$  est constante et égale à  $\mathbb{E}(Y_0)$  pour tout  $t$ .

De plus

$$\begin{aligned} \text{Cov}(Y_t, Y_{t+h}) &= \text{Cov} \left( \sum_{i=1}^{t+h} \varepsilon_i, \sum_{j=1}^t \varepsilon_j \right) \\ &= \sum_{i=1}^{t+h} \sum_{j=1}^t \varepsilon_i \varepsilon_j \\ &= \sum_{i=1}^t \sigma_\varepsilon^2 = t \cdot \sigma_\varepsilon^2. \end{aligned}$$

Donc une marche aléatoire n'est pas stationnaire car la suite des covariances dépend de  $t$ .

Dans la section suivante, nous nous intéressons à la stationnarité faible.

## 1.3 Processus autorégressif d'ordre un

### 1.3.1 Définition

Un processus autorégressif d'ordre 1 est un modèle de série temporelle où la valeur actuelle dépend linéairement de sa valeur précédente. L'équation d'un processus AR(1) est donnée par

$$y_t = c + \phi y_{t-1} + \varepsilon_t. \tag{1.1}$$

Où :

$y_t$  est la valeur de la variable de série temporelle à l'instant  $t$ .

$c$  est une constante.

$\phi$  est le paramètre autorégressif qui représente la dépendance de la valeur courante par rapport à la valeur précédente.

$\varepsilon_t$  dit processus des innovations, ( $\varepsilon_t \sim BB(0, \sigma^2)$ ) indépendant de  $y_{t-1}, y_{t-2} \dots$

### 1.3.2 Propriétés des processus AR(1)

1. Si  $|\phi| = 1$ , le processus  $(y_t, t \in \mathbb{Z})$  est une marche aléatoire.
2. Si  $|\phi| < 1$ , le processus  $(y_t, t \in \mathbb{Z})$  est stationnaire.
3. Si  $|\phi| > 1$ , le processus  $(y_t, t \in \mathbb{Z})$  est explosif.

### 1.3.3 Moments d'un processus AR(1)

Pour calculer les différents moments d'un processus autorégressif d'ordre 1, On va supposer que les bruits blancs sont des bruits blancs forts ( $\varepsilon_t \sim \text{i.i.d.}(0, \sigma^2)$ ).

#### Espérance

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

$$\mathbb{E}[y_t] = c + \phi\mathbb{E}[y_{t-1}] + \mathbb{E}[\varepsilon_t]. \quad (1.2)$$

En supposant que  $|\phi| < 1$ , on a la stationnarité de  $(y_t)$ , donc

$$\mathbb{E}[y_t] = \mathbb{E}[y_{t-1}] = \mu$$

Comme  $\mathbb{E}[\varepsilon_t] = 0$ , l'équation (1.2) devient  $(1 - \phi)\mu = c$ ,

ce qui implique

$$\mathbb{E}[y_t] = \mu = \frac{c}{1 - \phi}. \quad (1.3)$$

#### Variance

En remplaçant  $c$  par  $\mu(1 - \phi)$  dans l'équation (1.1), on aura :

$$y_t = \mu(1 - \phi) + \phi y_{t-1} + \varepsilon_t.$$

D'où :

$$(y_t - \mu) = \phi(y_{t-1} - \mu) + \varepsilon_t. \quad (1.4)$$

En prenant l'espérance du carré de l'équation (1.4), on aura :

$$\mathbb{E}[(y_t - \mu)^2] = \phi^2 \mathbb{E}[(y_{t-1} - \mu)^2] + 2\phi \mathbb{E}[(y_{t-1} - \mu)\varepsilon_t] + \mathbb{E}[(\varepsilon_t)^2].$$

On a  $\varepsilon_t$  et  $y_{t-1}$  qui sont indépendants et  $\mathbb{E}[\varepsilon_t] = 0$ , ce qui implique

$$\mathbb{E}[(y_{t-1} - \mu)\varepsilon_t] = 0.$$

Comme  $(y_t)$  est stationnaire, on a bien :

$$\mathbb{E}[(y_t - \mu)^2] = \mathbb{E}[(y_{t-1} - \mu)^2] = \text{Var}[y_t].$$

Ce qui implique :

$$\text{Var}[y_t] = \phi^2 \text{Var}[y_t] + \sigma^2.$$

Par conséquent,

$$\text{Var}[y_t] = \frac{\sigma^2}{1 - \phi^2}.$$

### Autocovariance

En multipliant l'équation (1.4) par  $(y_{t-1} - \mu)$ , on obtient :

$$(y_t - \mu)(y_{t-1} - \mu) = \phi(y_{t-1} - \mu)^2 + \varepsilon_t(y_{t-1} - \mu) \tag{1.5}$$

En appliquant l'espérance à l'équation (1.5), on obtient :

$$\begin{aligned} \text{Cov}(y_t, y_{t-1}) &= \mathbb{E}[(y_t - \mu)(y_{t-1} - \mu)] \\ &= \phi \mathbb{E}[(y_{t-1} - \mu)^2] + \mathbb{E}[\varepsilon_t(y_{t-1} - \mu)] \\ &= \phi \text{Var}(y_t) \\ &= \frac{\phi \sigma^2}{1 - \phi^2}. \end{aligned}$$

### Autocorrélation

$$\begin{aligned} \text{Corr}(y_t, y_{t-1}) &= \frac{\text{Cov}(y_t, y_{t-1})}{\text{Var}(y_t)} \\ &= \frac{\phi \text{Var}(y_t)}{\text{Var}(y_t)} \\ &= \phi. \end{aligned}$$

### Autocorrélation partielle

La fonction d'autocorrélation partielle (PACF) à l'ordre  $h$ , notée  $\phi_{hh}$ , est définie par

$$\phi_{hh} = \text{corr}(y_h - P(y_h|y_{h-1}, \dots, y_1), y_0 - P(y_0|y_1, \dots, y_{h-1})), h \geq 2.$$

ou  $P(y_h|y_{h-1}, \dots, y_1)$  est la régression de  $y_h$  sur l'ensemble  $y_{h-1}, \dots, y_1$ .

De même,  $P(y_0|y_1, \dots, y_{h-1})$  est la régression de  $y_0$  sur l'ensemble  $y_1, \dots, y_{h-1}$ .

Pour l'ordre 1 :

$$\phi_{11} = \text{Corr}(y_1, y_0) = \rho(1).$$

Pour un processus AR(1) :

- La fonction d'autocorrélation partielle à l'ordre 1 est égale au paramètre autorégressif :  $\phi_{11} = \phi$
- La fonction d'autocorrélation partielle est nulle au-delà du l'ordre 1 :  $\phi_{hh} = 0$  pour  $h \geq 2$ .

### 1.3.4 Estimation du paramètre autorégressif

Il existe plusieurs méthodes pour estimer le paramètre autorégressif  $\phi$  telles que la méthode du maximum de vraisemblance, méthode des moindres carrés, etc.

On présente ici la méthode des moindres carrés qu'on va utiliser dans le chapitre 3.

#### Méthode des moindres carrés

On considère le modèle

$$y_t = \phi y_{t-1} + \varepsilon_t, \quad t = 1, 2, \dots \tag{1.6}$$

La méthode des moindres carrés pour le modèle (1.6) consiste à minimiser la fonction

$$S(\phi) = \sum_{t=2}^n \varepsilon_t^2 = \sum_{t=2}^n (y_t - \phi y_{t-1})^2$$

On dérive la fonction  $S(\phi)$  on obtient :

$$\begin{aligned} S'(\phi) &= 2 \sum_{t=2}^n (y_t - \phi y_{t-1})(-y_{t-1}) \\ &= 2 \left( \phi \sum_{t=2}^n (y_{t-1})^2 - \sum_{t=2}^n (y_t y_{t-1}) \right) \\ S'(\phi) = 0 &\text{ implique } \phi = \frac{\sum_{t=2}^n (y_t y_{t-1})}{\sum_{t=2}^n (y_{t-1})^2}. \end{aligned}$$

On peut facilement vérifier que

$$S''(\phi) = 2 \sum_{t=2}^n (y_{t-1})^2 > 0.$$

On obtient l'estimateur suivant :

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

### 1.3.5 Inversibilité et causalité d'un processus AR(1)

Un processus AR(1) est toujours inversible.

Un processus AR(1) est causal si  $|\phi| < 1$ .

### 1.3.6 Exemple d'un AR(1)

Nous avons analysé les températures journalières de la ville d'Oran sur deux mois du 01 avril au 30 mai 2023 disponible sur (<https://power.larc.nasa.gov/data-access-viewer>).

Le tableau suivant donne un résumé des données.

Moyenne	Écart-type	Minimum	Maximum
20.02	2.65	15.79	26.41

TABLE 1.1 – Résumé des données.

Nous avons constaté que ces données se modélisent selon un processus AR(1), qui s'écrit sous la forme suivante :

$$y_t = 0.78y_{t-1} + 3.81378 + \varepsilon_t. \tag{1.7}$$

Le graphe de la fonction d'autocorrélation partielle est donné dans la figure suivante.

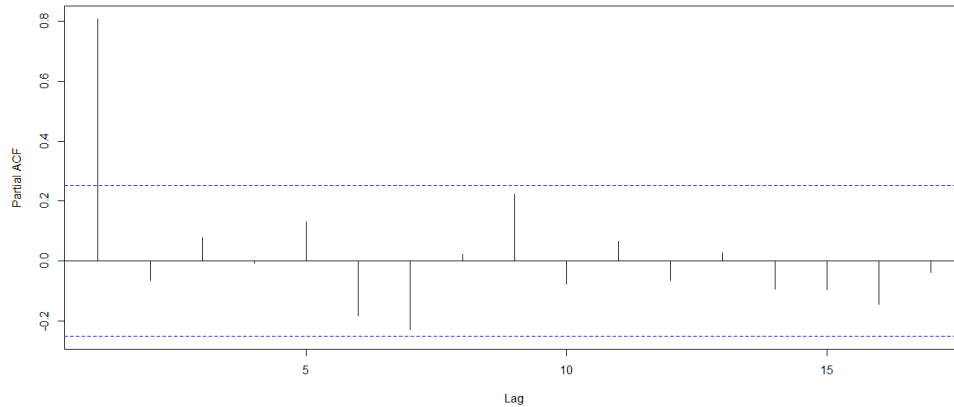


FIGURE 1.1 – Fonction d'autocorrélation partielle.

On voit un seul pic significatif au premier lag sur le graphique d'autocorrélation partielle. Cela suggère que ces données suivent un processus AR(1).

Le graphe suivant donne les courbes des données réelles et les données ajustées selon le modèle (1.7).

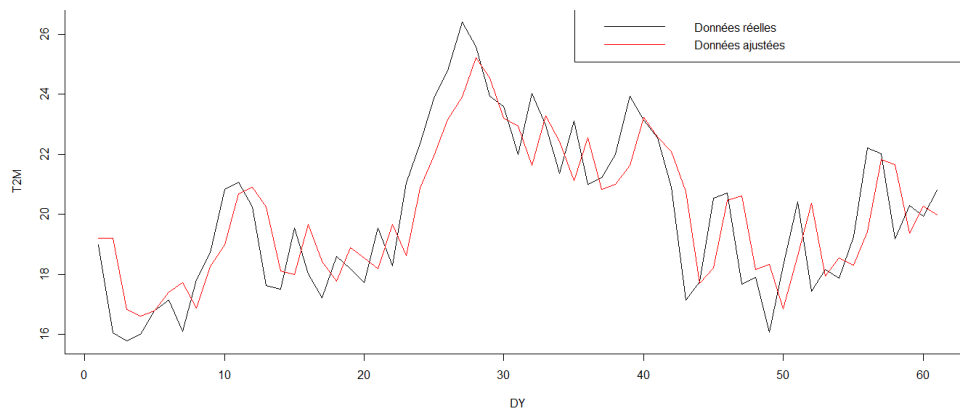


FIGURE 1.2 – Comparaison des Températures Réelles et Ajustées.

On remarque que les deux courbes sont proches ce qui veut dire que le modèle (1.7) ajuste bien les données.

## 1.4 Processus autorégressifs d'ordre $p$

### 1.4.1 Définition

On dit que  $(y_t, t \in \mathbb{Z})$  est un processus autorégressif d'ordre  $p$ , noté AR( $p$ ), s'il s'écrit sous la forme

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t. \quad (1.8)$$

ou encore

$$y_t = \sum_{i=1}^p \phi_i y_{t-i} + \varepsilon_t. \quad (1.9)$$

pour tout  $t \in \mathbb{Z}$ , où  $\phi_1, \phi_2, \dots, \phi_p \neq 0$  sont des réels,  $(\varepsilon_t, t \in \mathbb{Z})$  est un bruit blanc de variance  $\sigma^2$ .

### 1.4.2 Conditions de stationnarité d'un AR(p)

l'opérateur qui transforme une variable  $y_t$  en sa valeur passée est appelé opérateur retard. Il est désigné par la lettre B (ou L) et tel que :

$$\left\{ \begin{array}{l} By_t = y_{t-1} \quad (\text{processus décalé d'une unité de temps}) \\ B^2 y_t = B(By_t) = Bx_{t-1} = y_{t-2} \quad \text{ou } x_{t-1} = By_t \\ \vdots \\ B^d y_t = y_{t-d} \quad (\text{processus décalé de } d \text{ unités de temps}) \end{array} \right.$$

Le modèle (1.9) peut s'écrire

$$\Phi(B)y_t = \varepsilon_t \quad \text{où} \quad \Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p^p B$$

Le processus  $(y_t)$  est stationnaire si toutes les racines du polynôme  $\Phi(Z) = 0$  sont de module strictement supérieur à 1.

### 1.4.3 Condition de stationnarité d'un processus AR(2)

Soit  $(y_t, t \in \mathbb{Z})$  un processus AR(2) défini par

$$Y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t$$

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

$$\left\{ \begin{array}{l} |\phi_2| < 1 \\ \phi_1 + \phi_2 < 1 \\ \phi_2 - \phi_1 < 1 \end{array} \right.$$

### 1.4.4 Moments d'un processus AR(p)

Pour calculer les différents moments d'un processus AR(p), On va supposer que les bruits blancs sont des bruits blancs forts ( $\varepsilon_t \sim \text{i.i.d.}(0, \sigma^2)$ ) et que le processus AR(p) est stationnaire.

### Espérance

En prenant l'espérance de l'équation(1.8), on obtient

$$\mathbb{E}[y_t] = \phi_1\mathbb{E}[y_{t-1}] + \phi_2\mathbb{E}[y_{t-2}] + \cdots + \phi_p\mathbb{E}[y_{t-p}] + c + \mathbb{E}[\varepsilon_t] \quad (1.10)$$

Puisque on a la stationnarité de  $(y_t)$  et comme  $\mathbb{E}[\varepsilon_t] = 0$ , l'équation (1.10) devient

$$\mu = (\phi_1 + \phi_2 + \cdots + \phi_p)\mu + c.$$

Ce qui implique

$$\mathbb{E}[y_t] = \mu = \frac{c}{1 - (\phi_1 + \phi_2 + \cdots + \phi_p)}.$$

### Variance et autocovariance

Sans perte de généralité, supposant que  $c = 0$ , l'équation (1.8) devient

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t \quad (1.11)$$

En multipliant l'équation (1.11) par  $y_{t-k}$ , puis en prenant l'espérance on aura

$$\gamma(k) = \mathbb{E}[y_t y_{t-k}] = \phi_1 \gamma(k-1) + \cdots + \phi_p \gamma(k-p) + \mathbb{E}[\varepsilon_t y_{t-k}] \quad (1.12)$$

Pour  $k = 0, 1, \dots, p$ , nous avons un système de  $p + 1$  équations avec  $\gamma(0), \gamma(1), \dots, \gamma(p)$  inconnues

$$\gamma(0) = \phi_1 \gamma(1) + \phi_2 \gamma(2) + \cdots + \phi_p \gamma(p) + \sigma^2 \quad (1.13)$$

$$\gamma(1) = \phi_1 \gamma(0) + \phi_2 \gamma(1) + \cdots + \phi_p \gamma(p-1)$$

⋮

$$\gamma(p) = \phi_1 \gamma(p-1) + \phi_2 \gamma(p-2) + \cdots + \phi_p \gamma(0)$$

⋮

$$\gamma(k) = \phi_1 \gamma(k-1) + \phi_2 \gamma(k-2) + \cdots + \phi_p \gamma(k-p) \quad (1.14)$$

### Autocorrélation

En divisant  $\gamma(k)$  sur  $\gamma(0)$ , on obtient les équations de Yule-Walker suivantes

$$\begin{aligned}\rho(1) &= \phi_1 + \phi_2\rho(1) + \dots + \phi_p\rho(p-1) \\ \rho(2) &= \phi_1\rho(1) + \phi_2 + \dots + \phi_p\rho(p-2) \\ &\vdots \\ \rho(p) &= \phi_1\rho(p-1) + \phi_2\rho(p-2) + \dots + \phi_p \\ &\vdots \\ \rho(k) &= \phi_1\rho(k-1) + \phi_2\rho(k-2) + \dots + \phi_p\rho(k-p)\end{aligned}$$

### Autocorrélation Partielle

Soit

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t.$$

La PACF d'ordre  $k$  est donnée par

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

La PACF est nulle pour  $k > p$ .

### 1.4.5 Estimation des paramètres autorégressifs

Un modèle AR(p) s'écrit

$$y_t = Z_t' \phi + \varepsilon_t$$

où  $\varepsilon_t$  est un bruit blanc,  $Z_t' = (Y_{t-1}, Y_{t-2}, \dots, Y_{t-p})$  et  $\phi' = (\phi_1, \phi_2, \dots, \phi_p)$ .

L'estimation des paramètres du modèle  $Y_t = Z_t' \phi + \varepsilon_t$  par la méthode des moindres carrés est donnée par

$$\hat{\phi} = (Z_t' Z_t)^{-1} Z_t' Y_t$$

#### Proposition

Si les racines du polynôme caractéristique (racines de  $\phi(z) = 0$ ) sont à l'extérieur du disque unité, alors

$$\hat{\phi} \xrightarrow{P} \phi \quad \text{et} \quad \sqrt{n}(\hat{\phi} - \phi) \xrightarrow{\mathcal{L}} N(0, \sigma_\varepsilon^2 V),$$

où  $V = p \lim_{n \rightarrow \infty} \frac{1}{n} Z_t' Z_t$ .

### 1.4.6 Inversibilité et causalité d'un processus AR(p)

Un AR(p) est toujours inversible.

On dit qu'un processus AR(p) est causal si les racines de son polynôme caractéristique sont de module strictement supérieur à 1.

### 1.4.7 Exemple d'un AR(2)

Nous avons analysé également les températures de la ville d'Oran du premier trimestre de l'année 2021 disponible sur (<https://power.larc.nasa.gov/data-access-viewer>).

Le tableau ci-dessous nous donne un récapitulatif des données.

Moyenne	Écart-type	Minimum	Maximum
13.64	2.55	8.10	19.51

TABLE 1.2 – Résumé des données.

Nous avons vérifié que ces données se modélisent selon un processus AR(2), ayant la forme suivante :

$$y_t = 1.1143y_{t-1} - 0.3191y_{t-2} + 5.70813 + \varepsilon_t. \quad (1.15)$$

Le graphe de la fonction d'autocorrélation partielle est donné dans la figure suivante.

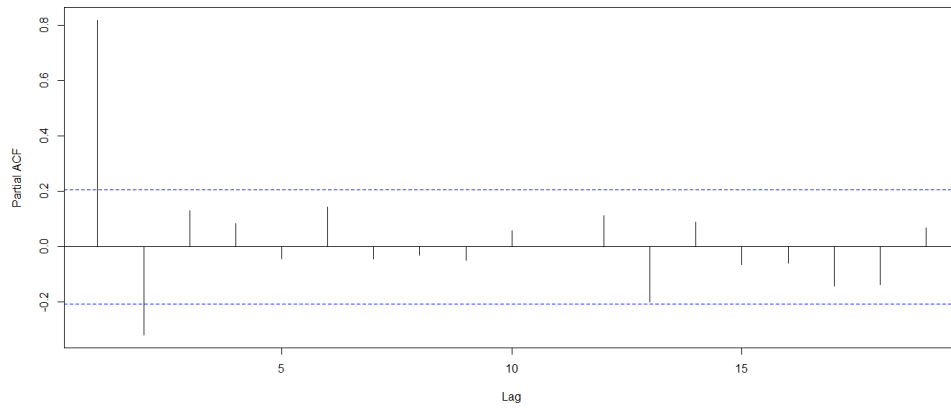


FIGURE 1.3 – Fonction d'autocorrélation partielle.

On remarque deux pics significatifs au premier et au deuxième lag sur le graphique d'autocorrélation partielle, ce qui montre que ces données suivent un processus AR(2). Le graphe suivant donne les courbes des données réelles et les données ajustées selon le modèle (1.15).

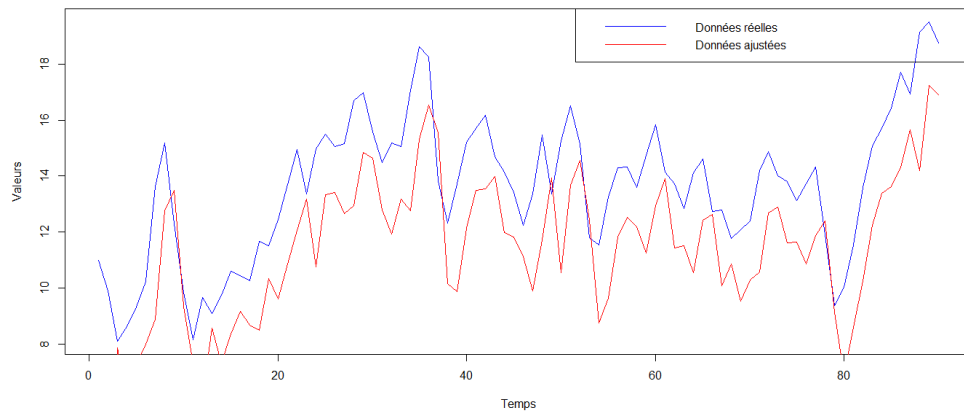


FIGURE 1.4 – Comparaison des Courbes de Températures Réelles et Ajustées.

On voit que les courbes sont proches, ce qui montre que le modèle (1.15) ajuste bien les données.

# Chapitre 2

## Tests de rupture épidémique dans la moyenne des variables aléatoires.

### 2.1 Introduction

Soit  $(X_1, X_2, \dots, X_n)$  un échantillon de variables aléatoires indépendantes de moyennes  $\mu_1, \mu_2, \dots, \mu_n$  respectivement.

On dit qu'une rupture est présente dans la moyenne d'un échantillon s'il existe  $1 < p < n$  et  $\delta \neq 0$  tel que

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

$p$  est dit point de rupture .

Une rupture épidémique est observée dans la moyenne d'un échantillon s'il existe  $1 < p < q < n$  et  $\delta \neq 0$  tel que

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

Dans la suite, on va s'intéresser seulement à la rupture épidémique dans la moyenne des variables aléatoires.

On veut tester au risque  $\alpha$  l'hypothèse nulle

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

contre l'hypothèse alternative

$$H_A : \exists 1 < p < q < n \text{ tels que } \mu_1 = \mu_2 = \dots = \mu_p = \mu_{q+1} = \dots = \mu_n = \mu_0, \\ \mu_{p+1} = \dots = \mu_q = \mu_0 + \delta, \quad \delta \neq 0.$$

On suppose que  $\mu_0, p, q$  et  $\delta$  sont inconnus.

On note  $l^* = q - p$  la longueur de l'épidémie et on suppose que  $l^*$  et  $n - l^*$  tendent vers l'infini quand  $n$  tend vers l'infini.

Soit  $S(t) = \sum_{k \leq t} X_k$ , ( $0 < t \leq n$ ) une suite des sommes partielles des  $X_i$  et  $S(0) = 0$ .

Dans la section suivante, nous allons voir quelques statistiques de tests pour tester  $H_0$  contre  $H_A$ .

## 2.2 Statistiques de tests classiques

### 2.2.1 Statistique de Levin et Kline

Levin and Kline [16] ont proposé la statistique suivante

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

Sous  $H_0$ , on a  $n^{-\frac{1}{2}}Q_n$  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  $n^{-\frac{1}{2}}Q_n$  tend vers l'infini quand  $n \rightarrow \infty$ , donc le test basé sur  $Q_n$  est consistant.

### 2.2.2 Statistique du rapport de vraisemblance

Pour  $1 \leq n_0 < n_1 \leq n$ , Siegmund [23] a proposé la statistique

$$Z_n(X_1, X_2, \dots, X_n) = \max_{n_0 \leq j-i \leq n_1} \frac{\left( S(j) - S(i) - \left( \frac{j-i}{n} \right) S(n) \right)}{\sqrt{(j-i) \left( 1 - \frac{j-i}{n} \right)}}.$$

Une approximation des grandes déviations du niveau de signification du test statistique basé sur  $Z_n$  a ensuite été développée par Siegmund [23] et Yao [25].

En effet, en supposant que  $b = \sqrt{n}$  et  $c$  est une constante positive, avec  $n_0 = nt_0$  et

$n_1 = nt_1$  où  $0 < t_0 \leq t \leq t_1 \leq 1$ , alors quand  $n \rightarrow \infty$ , nous avons :

$$P_{H_0}(Z_n \leq b) \sim 1 - \frac{1}{4}b^3\Phi(b) \int_{t_0}^{t_1} \frac{1}{(1-t)t^2} \left( v \left[ \frac{c}{t(1-t)^{\frac{1}{2}}} \right] \right)^2 dt,$$

où  $\Phi$  est la fonction de densité d'une loi normale standard et  $v$  est donnée par

$$v(x) = 2x^{-1} \exp \left\{ -2 \sum_{n=1}^{\infty} n^{-1} \Phi\left(-\frac{1}{2}xn^{\frac{1}{2}}\right) \right\}, \quad x > 0.$$

## 2.3 Statistiques de tests hölderiennes

### 2.3.1 Espaces de Hölder

L'espace de Hölder d'ordre  $\alpha$  ( $0 < \alpha \leq 1$ ), 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 note  $\omega_\alpha(f, \delta)$  le module de continuité hölderien de  $f$

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

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

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

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

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

## Convergence hölderienne

**Proposition A.1. (Hamadouche [11])** La convergence en loi dans  $H_\alpha$  d'une suite de processus  $(\xi_n, n \geq 1)$  équivaut à léquitenion sur  $H_\alpha^0$  de la suite des lois  $P_n = P\xi_n^{-1}$  des éléments aléatoires  $\xi_n$  et à la convergence des lois fini-dimensionnelles de  $\xi_n$ .

## Equitension hölderienne

**Théorème A.1. (Kerkyacharian, Roynette [15])** Soit  $(\xi_n)_{n \geq 1}$  une suite de processus nuls en zéro et vérifiant pour des constantes  $\gamma > 0$ ,  $\delta > 0$  et  $c > 0$  :

$$\forall \lambda > 0, P(|\xi_n(t) - \xi_n(s)| > \lambda) \leq \frac{c}{\lambda^\gamma} |t - s|^{1+\delta}.$$

Alors la suite des lois  $P_n$  des processus  $\xi_n$  est équitendue dans  $H_0^\alpha$  pour  $0 < \alpha < \frac{\delta}{\gamma}$ .

La version des moments de ce théorème est obtenue par linéarité de Markov. Elle s'énonce par le corollaire suivant.

**Corollaire A.1. (Lamperti [16])** Soit  $(\xi_n)_{n \geq 1}$  une suite de processus nuls en zéro et vérifiant pour des constantes  $\gamma > 0$ ,  $\delta > 0$  et  $c > 0$

$$E|\xi_n(t) - \xi_n(s)|^\gamma \leq c|t - s|^{1+\delta}.$$

Alors la suite des lois  $P_n$  des processus  $\xi_n$  est équitendue dans  $H_\alpha^0$  pour  $0 < \alpha < \frac{\delta}{\gamma}$ .

**Théorème A.2. (Kolmogorov)** Soit  $(\xi_t, t \in [0, 1])$  un processus défini sur un espace de probabilité  $(\Omega, \mathcal{B}, P)$  et supposons qu'il existe  $\delta > 0$ ,  $\gamma > 0$  et  $c > 0$  tels que

$$\forall \lambda > 0, P(|\xi_t - \xi_s| > \lambda) \leq \frac{c}{\lambda^\gamma} |t - s|^{1+\delta}.$$

Alors il existe une version de  $\xi$  à trajectoires dans  $H_\alpha^0$  pour tout  $0 < \alpha < \frac{\delta}{\gamma}$ .

**Théorème A.3. (Hamadouche [11])** Soit  $(\xi_n)_{n \geq 1}$  une suite de processus à trajectoire dans  $H_\alpha^0$ , vérifiant les conditions suivantes :

1. Il existe des constantes  $a > 0$ ,  $b > 0$ ,  $c > 0$  et une suite de nombres positifs  $(a_n) \searrow 0$  telles que

$$E|\xi_n(t) - \xi_n(s)|^a \leq c|t - s|^b, \quad \text{du théorème A.2.}$$

pour tout  $n$  et tout  $s, t$  tels que  $|t - s| \geq a_n$ .

2.  $\forall \epsilon > 0, \lim_{n \rightarrow \infty} P\{w_\alpha(\xi, a_n) > \epsilon\} = 0$ .

Alors pour tout  $\alpha < a^{-1}(\min(a, b) - 1)$ , la suite  $\xi_n$  est équitendue dans  $H_\alpha^0$ .

### 2.3.2 Statistiques de Rakauskas et Suquet

Rakauskas et Suquet [22] ont proposé la statistique  $U_{\alpha, n}$ ,  $0 < \alpha < \frac{1}{2}$  définie par

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

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  $j$  :

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

On note  $D = \bigcup_{j>0} D_j$  et  $D^* = D \setminus \{0\}$ . Pour  $r \in D_j$ ,  $j \geq 0$ ,  $r^- = r - 2^{-j}$  et  $r^+ = r + 2^{-j}$ .

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

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

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

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

### 2.3.3 Convergence et consistance

Soit  $(X_i)_{i \geq 1}$  une suite de variables aléatoires indépendantes identiquement distribuées de variance  $\sigma^2$ . La convergence des statistiques  $U_{\alpha,n}$  et  $D_{\alpha,n}$  dans l'espace de  $H_\alpha^0$  sous  $H_0$  est donné par le résultat suivant.

**Théorème 2.1. (Rakauskas et Suquet [22])** On suppose que

$$P(|X_1| > t) = o(t^{-p}), \quad p = \frac{1}{\frac{1}{2} - \alpha}. \quad (2.1)$$

Alors

$$\sigma^{-1}n^{-\frac{1}{2}}U_{\alpha,n} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} U_{\alpha,n} \quad \forall 0 < \alpha < \frac{1}{2}.$$

et

$$\sigma^{-1}n^{-\frac{1}{2}}D_{\alpha,n} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} D_{\alpha,n} \quad \forall 0 < \alpha < \frac{1}{2}.$$

La consistance de rejeter  $H_0$  contre  $H_A$  dans  $H_\alpha^0$  est donnée par le résultat suivant.

**Théorème 2.2. (Rakauskas et Suquet [22])** Si  $\sup_{k \geq 1} V(X_k) < +\infty$  et

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

alors

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

et

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

La convergence et la consistance de  $D_{\alpha,n}$  dans  $H_\alpha^0$  est étendue par Graiche, Merabet et Hamadouche (2012 et 2016) aux variables aléatoires indépendantes non stationnaires et aux variables dépendantes ( $\alpha$ -mélangeantes) sous certaines conditions sur le coefficient de mélange.

Les statistiques Hölderiennes sont particulièrement utiles pour détecter des ruptures épidémiques de courtes durées. La statistique  $Q_n$ , par exemple, détecte des ruptures de longueur d'au moins  $n^{\frac{1}{2}}$ , tandis que la statistique  $U_{\alpha,n}$  détecte des ruptures dont la longueur est de l'ordre  $n^\delta$  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 Étude de simulation

Pour comparer les performances des quatre statistiques  $D_n = \sigma^{-1}n^{-1/2}D_{a,n}$ ,  $U_n = \sigma^{-1}n^{-1/2}U_{a,n}$  avec  $a = 0.25$ ,  $R_n = \sigma^{-1}n^{-1/2}Q_n$  et  $Z_n$ , nous simulons 1000 échantillons de taille  $n = 60$  à partir d'une loi normale de variance 1 au un seuil de significativité de 0,05.

Sous  $H_A$  et pour  $\mu_0 = 0$ , le tableau suivant nous donne les valeurs des statistiques pour différentes valeurs de  $l^*$  et  $\delta$ .

$l^*$	$\delta$	$D_n$	$R_n$	$U_n$	$Z_n$
3	5	1.2211	1.4438	2.6875	3.1167
	7	1.2880	1.5161	3.0150	3.8257
	10	1.3312	1.5759	3.2676	5.1613
5	3	1.2822	1.5922	2.6923	2.9726
	5	1.4703	1.8193	3.3004	4.1752
7	3	1.5236	1.8667	3.1007	3.5724
	7	1.7482	2.1533	3.7220	5.2838
10	1	1.181	1.4208	2.1384	2.3976
	3	1.9156	2.2215	3.5473	4.3237
15	1	1.1474	1.6474	2.4310	3.5724
	2	1.4627	2.2842	3.4196	3.9053
		c=1.32	c=1.50	c=2.27	c=3.60

TABLE 2.1 – Comparaison entre  $D_n$ ,  $R_n$ ,  $U_n$  et  $Z_n$ .

$c$  représente la valeur critique de chaque statistique au seuil 0.05.

Quand l'épidémie est courte, on remarque que la statistique  $U_n$  détecte plus rapidement la rupture.

Nous nous intéressons maintenant à la puissance des tests, pour cela nous simulons 10000 échantillons de taille  $n = 60$  provenant d'une distribution normale de variance 1. Le tableau suivant présente les valeurs des puissances des tests basés sur  $D_n$ ,  $R_n$ ,  $U_n$  et  $Z_n$ .

$l^*$	$\delta$	$D_n$	$R_n$	$U_n$	$Z_n$
3	5	0.2458	0.3335	0.9688	0.2095
	7	0.3761	0.5129	1	0.589
	10	0.5437	0.9463	1	0.9956
5	3	0.4054	0.6737	0.8634	0.1694
	5	0.7548	0.9987	1	0.7702
7	3	0.761	0.9843	0.994	0.4179
	5	0.9853	1	1	0.9974
10	1	0.3018	0.4	0.1398	0.037
	3	0.9984	1	1	0.8426
15	1	0.225	0.6613	0.2951	0.0786
	2	0.7109	0.9994	0.9902	0.6615

TABLE 2.2 – Puissance des tests basés sur  $D_n$ ,  $R_n$ ,  $U_n$  et  $Z_n$ .

On remarque que pour les courtes épidémies, le test basé sur la statistique  $U_n$  est le plus puissant.

## 2.5 Application aux données réelles

Le tableau suivant présente les données hebdomadaires sur le nombre de cas de grippe au Mexique, tirées de l'article de Chen , Li et Zhou [4].

Date	Nombre des cas	Date	Nombre des cas
2005-06-05	740	2005-12-04	1407
2005-06-12	710	2005-12-11	1393
2005-06-19	713	2005-12-18	1472
2005-06-26	654	2005-12-25	1540
2005-07-03	727	2006-01-01	1397
2005-07-10	690	2006-01-08	1383
2005-07-17	748	2006-01-15	1475
2005-07-24	664	2006-01-22	1310
2005-07-31	690	2006-01-29	1222
2005-08-07	698	2006-02-05	1169
2005-08-14	641	2006-02-12	1064
2005-08-21	641	2006-02-19	1137
2005-08-28	630	2006-02-26	1131
2005-09-04	811	2006-03-05	1069
2005-09-11	786	2006-03-12	1088
2005-09-18	855	2006-03-19	1038
2005-09-25	845	2006-03-26	950
2005-10-02	822	2006-04-02	980
2005-10-09	877	2006-04-09	880
2005-10-16	948	2006-04-16	933
2005-10-23	1027	2006-04-23	844
2005-10-30	1180	2006-04-30	855
2005-11-06	1294	2006-05-07	869
2005-11-13	1279	2006-05-14	909
2005-11-20	1311	2006-05-21	894
2005-11-27	1304	2006-05-28	823

TABLE 2.3 – Nombre de cas de grippe au Mexique

Afin de tester s'il y a une rupture épidémiques dans la moyenne de ces données, on calcule les valeurs des 4 statistiques et on obtient le tableau suivant.

$D_n$	$R_n$	$U_n$	$Z_n$
$c = 1.32$	$c = 1.62$	$c = 2.27$	$c = 3.60$
2.1655	3.0901	4.3964	1405.7

TABLE 2.4 – valeurs des statistiques pour les données de grippe.

On remarque que toutes les statistiques dépassent leurs valeurs seuils ce qui illustre clairement qu'il y a une rupture épidémique dans la moyenne des données.

# Chapitre 3

## Tests de rupture épidémique dans la moyenne des innovations d'un AR(1)

### 3.1 Introduction

Dans cette partie, nous nous concentrons spécifiquement sur la détection de rupture épidémique dans la moyenne des innovations d'un modèle autorégressif d'ordre 1, qui est de la forme suivante

$$y_t = \phi y_{t-1} + \varepsilon_t, \quad (3.1)$$

où  $\varepsilon_t$  sont des variables aléatoires indépendantes,  $y_0 = 0$  et  $|\phi| < 1$ .

On veut tester au risque  $\alpha$  l'hypothèse nulle

$$H'_0 : \mathbb{E}(\varepsilon_1) = \mathbb{E}(\varepsilon_2) = \dots = \mathbb{E}(\varepsilon_n) = 0.$$

contre l'hypothèse alternative

$$H'_A : \exists 1 < p < q < n \quad \mathbb{E}(\varepsilon_1) = \dots = \mathbb{E}(\varepsilon_p) = \mathbb{E}(\varepsilon_{q+1}) = \dots = \mathbb{E}(\varepsilon_n) = 0 \\ \mathbb{E}(\varepsilon_{p+1}) = \dots = \mathbb{E}(\varepsilon_q) = \delta \quad \delta \neq 0.$$

Les résidus associés au modèle (3.1) sont donnés par

$$\hat{\varepsilon}_t = (\phi - \hat{\phi})y_{t-1} + \varepsilon_t, \quad (3.2)$$

avec  $\hat{\phi}$  l'estimateur de  $\phi$  donné par la méthode des moindres carrés.

Sous  $H'_0$ ,  $\hat{\phi}$  est défini par

$$\hat{\phi} = \frac{\sum_{t=2}^n y_t y_{t-1}}{\sum_{t=2}^n y_{t-1}^2}.$$

Sous  $H'_A$ ,  $\hat{\phi}$  est défini par

$$\hat{\phi} = \frac{\sum_{t=2}^n y_t y_{t-1} - \delta \sum_{t=2}^n y_{t-1}}{\sum_{t=2}^n y_{t-1}^2}.$$

Soit  $S(t) = \sum_{i=1}^t \hat{\varepsilon}_i$ ;  $0 < t \leq n$  la suite des sommes partielles des  $\hat{\varepsilon}_i$  définie dans (3.2).

## 3.2 Statistiques de tests basés sur les résidus

On considère les statistiques hölderiennes définies dans le chapitre 2 basées sur les résidus du processus AR(1) suivantes.

$$\hat{U}_{\alpha,n} = U_{\alpha,n}(\hat{\varepsilon}_1, \hat{\varepsilon}_2, \dots, \hat{\varepsilon}_n).$$

$$\hat{D}_{\alpha,n} = D_{\alpha,n}(\hat{\varepsilon}_1, \hat{\varepsilon}_2, \dots, \hat{\varepsilon}_n).$$

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

**Théorème 3.1.** Supposons pour tout  $0 < \alpha < \frac{1}{2}$ ,

$$\lim_{t \rightarrow \infty} t^{\frac{1}{2}-\alpha} P(|\varepsilon_1| > t) = 0 \quad (3.3)$$

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

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

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

### Idées de la preuve

(a) On définit d'abord le processus de sommes partielles lissé polygonalement basé sur les résidus  $\hat{\varepsilon}$  par

$$\hat{V}_n(t) = \sum_{i=1}^{[nt]} \hat{\varepsilon}_i + (nt - [nt]) \hat{\varepsilon}_{[nt]+1}$$

D'après Rakauskas et Rasten [21], si la condition (3.3) est vérifiée, alors sous  $H'_0$  et pour tout  $|\phi| < 1$

$$\sigma^{-1} n^{-\frac{1}{2}} \hat{V}_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} W \quad \text{dans } H_{\alpha}^0, \quad \text{pour tout } 0 < \alpha < \frac{1}{2}. \quad (3.4)$$

Suivant les mêmes étapes du théorème 2.1. de Račkauskas et Suquet [21], avec  $X_k = \hat{\varepsilon}_k$ , on aura

$$\sigma^{-1} n^{-\frac{1}{2}} \hat{D}_{\alpha,n} = g(\sigma^{-1} n^{-\frac{1}{2}} \hat{V}_n) + o_P(1),$$

avec

$$g(x) = \sup_{j \geq 1} \max_{r \in D_j} \frac{|x(r) - \frac{1}{2}x(r^-) - \frac{1}{2}x(r^+)|}{(r - r^-)^\alpha}.$$

De (3.4) et la continuité de  $g$ , on déduit

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

(b) On utilise les mêmes étapes que la partie (a) avec

$$g(x) = \sup_{0 < s < t < 1} \frac{|x(t) - x(s) - (t - s)x(1)|}{(|t - s|)^\alpha}.$$

### 3.2.1 Étude de simulation

Dans la suite, nous évaluons les performances des statistiques  $\hat{T}_n = \sigma^{-1}n^{-1/2}\hat{D}_{a,n}$  et  $\hat{U}_n = \sigma^{-1}n^{-1/2}\hat{U}_{a,n}$  avec  $a = 0.25$  pour deux types d'innovations gaussiennes et uniformes sur  $[-1, 1]$ , pour deux valeurs de  $\phi$  au seuil de significativité de 0.05.

#### 1<sup>er</sup> cas : Innovations gaussiennes

Sous  $H'_A$  et pour  $\mu_0 = 0$ , nous simulons 1000 échantillons de taille  $n = 60$  provenant d'une loi normale de variance 1 en faisant varier  $l^*$  et  $\delta$ .

Le tableau suivant, nous donne les valeurs des statistiques basées sur les résidus.

		$\phi = 0.5$		$\phi = 0.8$	
$l^*$	$\delta$	$\hat{D}_n$	$\hat{U}_n$	$\hat{D}_n$	$\hat{U}_n$
3	0.8	0.9464	1.7756	1.0520	1.9751
	1.2	1.0288	1.9373	1.2458	2.3291
	1.6	1.1011	2.0842	1.4213	2.6662
	2	1.2050	2.3007	1.5982	3.0005
	2.2	1.2630	2.4691	1.7545	3.2407
	2.5	1.3472	2.6288	1.8895	3.5180
5	0.8	0.9796	1.8326	1.1238	2.0963
	1.2	1.0723	2.0566	1.3254	2.4880
	1.4	1.1669	2.2438	1.4472	2.7312
	1.6	1.206	2.3402	1.5439	2.9026
	2	1.3611	2.6658	1.8099	3.3878
7	0.8	1.0126	1.9104	1.1721	2.1932
	1	1.0847	2.0559	1.3046	2.4231
	1.2	1.1849	2.2506	1.4300	2.6446
	1.6	1.3656	2.6203	1.7101	3.1709
10	0.8	1.1350	2.0695	1.2543	2.3126
	1	1.2682	2.2813	1.4271	2.5947
	1.2	1.3737	2.4633	1.6104	2.8870
		c=1.32	c=2.27	c=1.32	c=2.27

TABLE 3.1 – valeurs de  $\hat{D}_n$  et  $\hat{U}_n$ .

On remarque que pour les deux valeurs de  $\phi$ , la statistique  $\hat{U}_n$  détecte la rupture plus rapidement que la statistique  $\hat{D}_n$ .

Quand la longueur de l'épidémie augmente, les deux statistiques détectent plus rapidement cette rupture ( $\delta$  plus petit).

Le tableau ci-dessous présente la puissance des tests. Ces résultats sont issus d'une simulation de 10 000 échantillons d'une loi normale de taille  $n = 60$ .

		$\phi = 0.5$		$\phi = 0.8$	
$l^*$	$\delta$	$\hat{D}_n$	$\hat{U}_n$	$\hat{D}_n$	$\hat{U}_n$
3	0.8	0.0636	0.1279	0.2016	0.3315
	1.2	0.1719	0.2701	0.3894	0.5162
	1.6	0.2583	0.394	0.5304	0.6313
	2	0.3616	0.5263	0.6352	0.715
	2.8	0.5457	0.712	0.7817	0.8405
5	0.8	0.1029	0.1651	0.2479	0.4021
	1.2	0.2177	0.3585	0.48	0.6269
	1.6	0.3742	0.5699	0.665	0.7715
	2	0.5097	0.7188	0.7761	0.8595
	2.5	0.6551	0.8359	0.8801	0.9389
7	0.8	0.1365	0.2091	0.2949	0.479
	1	0.2175	0.3477	0.4542	0.628
	1.2	0.309	0.4868	0.5915	0.7304
	1.6	0.5052	0.7156	0.7868	0.8727
10	0.8	0.2287	0.3091	0.4124	0.5966
	1	0.3793	0.5056	0.6203	0.7678
	1.2	0.5544	0.6908	0.7829	0.8656

TABLE 3.2 – Puissance de tests basés sur  $\hat{D}_n$  et  $\hat{U}_n$

On voit que le test basé sur  $\hat{U}_n$  est plus puissant que celui basé sur  $\hat{D}_n$  pour les deux valeurs de  $\phi$ .

### 2<sup>ème</sup> cas : Innovations uniformes

Sous  $H'_A$  et pour  $\mu_0 = 0$ , nous simulons 1000 échantillons de taille  $n = 60$  à partir d'une distribution uniforme sur  $[-1, 1]$  en faisant varier  $l^*$  et  $\delta$ .

		$\phi = 0.5$		$\phi = 0.8$	
$l^*$	$\delta$	$\hat{D}_n$	$\hat{U}_n$	$\hat{D}_n$	$\hat{U}_n$
3	0.8	1.0384	1.9826	1.2959	2.4384
	1	1.1321	2.1690	1.4471	2.7295
	1.4	1.3081	2.5906	1.7949	3.3703
	1.6	1.4216	2.8005	2.0705	3.7781
5	0.6	1.0443	2.0230	1.2410	2.3098
	0.8	1.1226	2.1727	1.4224	2.6853
	1	1.2588	2.4239	1.6284	3.0793
	1.2	1.4141	2.7684	1.8457	3.4596
7	0.6	1.0974	2.1057	1.3220	2.4907
	0.8	1.2321	2.3798	1.5302	2.8787
	1	1.4112	2.7278	1.8416	3.4111
10	0.6	1.2630	2.3051	1.4555	2.6324
	0.8	1.5019	2.7032	1.7398	3.1619
		c=1.32	c=2.27	c=1.32	c=2.27

TABLE 3.3 – valeurs de  $\hat{D}_n$  et  $\hat{U}_n$ .

On remarque que la statistique  $\hat{U}_n$  détecte la rupture plus rapidement que  $\hat{D}_n$ . Le tableau ci-dessous présente la puissance des tests pour 10 000 échantillons à partir d'une loi uniforme sur  $[-1, 1]$  de taille  $n = 60$ .

		$\phi = 0.5$		$\phi = 0.8$	
$l^*$	$\delta$	$\hat{D}_n$	$\hat{U}_n$	$\hat{D}_n$	$\hat{U}_n$
3	0.8	0.1953	0.3306	0.4531	0.5787
	1	0.2914	0.4516	0.5663	0.612
	1.4	0.4613	0.6383	0.7141	0.786
	2	0.659	0.8182	0.8539	0.9124
5	1	0.4128	0.6206	0.7108	0.809
	2	0.828	0.9524	0.9635	0.9901
7	0.8	0.3937	0.6081	0.6931	0.8162
	1	0.5623	0.7772	0.8308	0.9098
10	0.6	0.4072	0.5547	0.6559	0.803
	0.8	0.67	0.8028	0.8821	0.9325

TABLE 3.4 – Puissance de tests basés sur  $\hat{D}_n$  et  $\hat{U}_n$ .

Le test basé sur  $\hat{U}_n$  présente une puissance supérieure à celui basé sur  $\hat{D}_n$ .

### 3.3 Statistiques de tests basés sur les observations

Comme les innovations ne sont pas observées, Markeviit [17] a proposé d'utiliser les statistiques  $U_{\alpha,n}$  et  $D_{\alpha,n}$  basées sur les observations  $y_1, \dots, y_n$  d'un processus AR(1).

$$\tilde{U}_{\alpha,n} = U_{\alpha,n}(y_1, \dots, y_n).$$

$$\tilde{D}_{\alpha,n} = D_{\alpha,n}(y_1, \dots, y_n).$$

La convergence des statistiques  $\tilde{U}_{\alpha,n}$  et  $\tilde{D}_{\alpha,n}$  est donnée par les résultats suivants

**Théorème 3.2. (Markeviit [18])** Supposons que les  $(\varepsilon_i)$  sont i.i.d. centrées et vérifiant la condition (3.3) alors sous  $H'_0$ , on a

a) Pour  $|\phi| < 1$

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

b) Pour  $\phi = 1$

$$n^{-3/2+\alpha}\sigma^{-1}\tilde{U}_{\alpha,n} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} U_{\alpha,\infty}(X),$$

où  $X$  est un processus de Wiener intégré défini par  $X(t) := \int_0^t W(s) ds$ .

**Théorème 3.3. (Markeviit [18])** Supposons que les  $(\varepsilon_i)$  sont i.i.d. centrées et satisfaisant pour un certain  $p \geq 2$  la condition (3.3) alors sous  $H'_0$ , on a

a) Pour  $|\phi| < 1$

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

b) Pour  $\phi = 1$

$$n^{-3/2}\sigma^{-1}\tilde{D}_{\alpha,n} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} D_{\alpha,\infty}(X),$$

où  $X$  est un processus de Wiener intégré défini par  $X(t) := \int_0^t W(s) ds$ .

La consistance de rejeter  $H'_0$  contre  $H'_A$  est donnée par les résultats suivants.

**Théorème 3.4. (Markeviit [18])** Supposons que les  $(\varepsilon_i)$  sont i.i.d. centrées et satisfaisant pour un certain  $p \geq 2$  la condition (3.3) alors, sous  $H'_A$  on a

1.  $|\phi| < 1$  :

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

si  $n^{-1/2+\alpha}\ell^{*1-\alpha}|\delta| \rightarrow \infty$  et  $\limsup_{n \rightarrow \infty} \frac{\ell^*}{n} < 1$ .

2.  $\phi = 1$  :

$$n^{-3/2+\alpha}\tilde{U}_{\alpha,n} \xrightarrow[n \rightarrow \infty]{P} \infty,$$

si  $n^{-3/2+\alpha}\ell^{*(2-\alpha)}|\delta| \rightarrow \infty$  et  $\liminf_{n \rightarrow \infty} \left| \frac{1}{2} - \frac{m^*}{n} \right| > 0$ .

**Théorème 3.5. (Markeviit [18])** Supposons que les  $(\varepsilon_i)$  sont i.i.d. centrées et satisfaisant pour un certain  $p \geq 2$  la condition (3.3) et pour  $\ell^*/n \leq 1/2$  pour tout  $n$

1. Pour  $|\phi| < 1$  :

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

si  $n^{-1/2+\alpha}\ell^{*1-\alpha}|\delta| \rightarrow \infty$ .

2. Pour  $\phi = 1$  :

$$n^{-3/2}\sigma^{-1}\tilde{D}_{\alpha,n} \xrightarrow[n \rightarrow \infty]{P} \infty$$

si  $n^{-3/2+\alpha}\ell^{*(2-\alpha)}|\delta| \rightarrow \infty$ .

### 3.3.1 Étude de simulation

Nous examinons les performances des statistiques  $\tilde{D}_n = \sigma^{-1}n^{-1/2}(1-\phi)\tilde{D}_{a,n}$  et  $\tilde{U}_n = \sigma^{-1}n^{-1/2}(1-\phi)\tilde{U}_{a,n}$  avec  $a = 0.25$  pour deux types d'innovations gaussiennes et uniformes sur l'intervalle  $[-1, 1]$  et pour deux valeurs du coefficient autorégressif  $\phi$  au seuil de significativité de 0.05.

#### 1<sup>er</sup> cas : Innovations gaussiennes

Ce tableau présente les valeurs des statistiques issues d'une simulation de 1000 échantillons de taille  $n = 60$  d'une distribution normale.

		$\phi = 0.5$		$\phi = 0.8$	
$l^*$	$\delta$	$\tilde{D}_n$	$\tilde{U}_n$	$\tilde{D}_n$	$\tilde{U}_n$
3	1	0.8079	1.4803	0.6282	1.0817
	2	0.9539	1.7709	0.7558	1.3122
	3	1.1477	2.1910	0.9348	1.6253
	3.5	1.2583	2.4120	1.0207	1.7887
	4	1.3795	2.6967	1.1241	1.9835
	5	1.6222	3.2559	1.3332	2.3794
5	1	0.9171	1.6436	0.7028	1.1957
	1.5	1.0485	1.8809	0.8146	1.4079
	2	1.2235	2.2424	0.9635	1.6605
	2.4	1.3717	2.5372	1.0979	1.9071
	2.6	1.4642	2.7124	1.1759	2.0461
	2.8	1.5376	2.8570	1.2155	2.1452
	3	1.6094	3.0371	1.2952	2.2752
	3.2	1.6882	3.2127	1.3642	2.4122
7	1	0.9941	1.7660	0.7865	1.3731
	1.4	1.2079	2.1422	0.9271	1.6204
	1.6	1.3062	2.3051	0.9968	1.7382
	1.8	1.4044	2.5107	1.1076	1.9432
	2	1.5277	2.7243	1.1925	2.1061
	2.2	1.6492	2.9652	1.2611	2.2473
	2.4	1.7446	3.1610	1.3602	2.4278
10	1	1.1591	2.0557	0.8235	1.5384
	1.2	1.2889	2.2776	0.9273	1.7664
	1.4	1.4317	2.5356	1.0383	1.9917
	1.6	1.5638	2.8063	1.1292	2.1899
	1.8	1.7243	3.1171	1.2454	2.4306
	2	1.8554	3.3590	1.3575	2.6738
		c=1.32	c=2.27	c=1.32	c=2.27

TABLE 3.5 – Valeurs de  $\tilde{D}_n$  et  $\tilde{U}_n$ .

On remarque que  $\tilde{U}_n$  détecte plus rapidement la rupture que  $\tilde{D}_n$ .

Quand la longueur de l'épidémie augmente, les deux statistiques détectent plus rapidement cette rupture.

Le tableau suivant nous donne la puissance des tests pour 10000 échantillons de taille  $n = 60$ .

		$\phi = 0.5$		$\phi = 0.8$	
$l^*$	$\delta$	$\tilde{D}_n$	$\tilde{U}_n$	$\tilde{D}_n$	$\tilde{U}_n$
3	2	0.1099	0.1135	0.0357	0.0102
	3	0.2666	0.4214	0.1023	0.0639
	4	0.5082	0.8034	0.2534	0.2478
5	2	0.349	0.4497	0.1114	0.0959
	3	0.7754	0.9228	0.444	0.5016
	4	0.9768	0.9988	0.8488	0.9141
7	2	0.5082	0.8034	0.2887	0.3401
	3	0.9764	0.996	0.8545	0.9
	4	0.9997	1	0.9966	0.998
10	1	0.2804	0.3028	0.0404	0.0658
	2	0.9326	0.9647	0.5228	0.765

TABLE 3.6 – Puissance de tests basés sur  $\tilde{D}_n$  et  $\tilde{U}_n$

On voit que le test basé sur  $\tilde{U}_n$  est plus puissant que celui basé sur  $\tilde{D}_n$ .

### 2<sup>ème</sup> cas : Innovations uniformes

Les valeurs présentées dans le tableau suivant proviennent d'une simulation réalisée sur 1000 échantillons de taille  $n = 60$ , générés à partir d'une distribution uniforme sur l'intervalle  $[-1, 1]$ .

		$\phi = 0.5$		$\phi = 0.8$	
$l^*$	$\delta$	$\tilde{D}_n$	$\tilde{U}_n$	$\tilde{D}_n$	$\tilde{U}_n$
3	1	0.8972	1.6527	0.7159	1.2328
	2	1.2299	2.3894	1.0177	1.7787
	2.2	1.3232	2.5829	1.0714	1.8829
	2.4	1.4000	2.7658	1.1190	1.9903
	2.6	1.4541	2.9406	1.2147	2.1524
	2.8	1.5724	3.1539	1.2981	2.3117
	3	1.6547	3.3634	1.3449	2.4232
5	1	1.1362	2.0619	0.8709	1.4936
	1.2	1.2552	2.2743	0.9909	1.6976
	1.4	1.3749	2.5443	1.1159	1.9247
	1.6	1.5221	2.8162	1.2301	2.1387
	1.8	1.6571	3.1123	1.3213	2.3260
7	0.8	1.1903	2.1045	0.9209	1.5941
	1	1.3576	2.4045	1.0453	1.8377
	1.2	1.5219	2.7516	1.2146	2.1259
	1.4	1.7581	3.1876	1.3753	2.4629
10	0.6	1.1525	2.0491	0.8559	1.6139
	0.8	1.3987	2.4972	1.0162	1.9558
	1	1.6565	2.9561	1.1940	2.3063
	1.2	1.9135	3.4793	1.3965	2.7806
		c=1.32	c=2.27	c=1.32	c=2.27

TABLE 3.7 – Valeurs de  $\tilde{D}_n$  et  $\tilde{U}_n$

On remarque que  $\tilde{U}_n$  détecte plus rapidement la rupture que  $\tilde{D}_n$ .

Quand la longueur de l'épidémie augmente, les deux statistiques détectent plus rapidement cette rupture.

Le tableau suivant présente la puissance des tests obtenue à partir d'une simulation de 10 000 échantillons, de taille  $n = 60$ , extraits d'une distribution uniforme sur l'intervalle  $[-1, 1]$ .

		$\phi = 0.5$		$\phi = 0.8$	
$l^*$	$\delta$	$\tilde{D}_n$	$\tilde{U}_n$	$\tilde{D}_n$	$\tilde{U}_n$
3	2	0.3684	0.5959	0.1461	0.1115
	3	0.7933	0.9904	0.5178	0.6291
5	1	0.2476	0.2988	0.0699	0.0453
	2	0.9102	0.9888	0.6482	0.7355
7	1	0.5256	0.6107	0.159	0.1771
	2	0.998	1	0.9729	0.9857
10	1	0.822	0.8905	0.3008	0.543
	2	1	1	0.9995	1

TABLE 3.8 – Puissance de tests basés sur  $\tilde{D}_n$  et  $\tilde{U}_n$

Le test basé sur  $\tilde{U}_n$  est plus puissant par rapport a celui basé sur  $\tilde{D}_n$ .

### 3.4 Application aux données réelles

On considère les températures journalières sur 2 mois (du 1 avril au 30 mai 2023) présentées dans le chapitre 1, qui s'ajustent selon le modèle suivant

$$y_t = 0.78y_{t-1} + 3.81378 + \varepsilon_t$$

Les valeurs des innovations  $\varepsilon_i$  calculées à partir de ces données sont résumées dans le tableau suivant

Moyenne	Écart-type	Minimum	Maximum
-2.67114	1.4558	-6.1004	-0.0295

TABLE 3.9 – Résumé des innovations.

Pour examiner la normalité des innovations, nous avons utilisé le test de Shapiro-Wilk. La  $p$ -value obtenue est 0.237, qui est supérieure au seuil de significativité de 0.05. Par conséquent, on ne peut pas rejeter l'hypothèse nulle selon laquelle les innovations suivent une distribution normale. Ce qui implique que la condition (3.3) est vérifiée.

Pour vérifier s'il y a une rupture dans la moyenne des innovations, on calcule d'abord les statistiques basées sur les résidus et on obtient  $\hat{D}_n = 1.2836$  et  $\hat{U}_n = 2.1096$ . Ces valeurs indiquent qu'il y a pas de rupture dans la moyenne des innovations du processus AR(1). En calculant les statistiques basées sur observations, on obtient les valeurs suivantes  $\tilde{D}_n = 0.8886$  et  $\tilde{U}_n = 1.4336$  ce qui indique l'absence de rupture dans la moyenne des innovations du processus AR(1).

# Conclusion

Le travail de ce mémoire porte sur les tests de rupture épidémique dans la moyenne des innovations d'un processus autorégressif d'ordre 1.

Après avoir défini les processus autorégressifs et rappelé leurs propriétés, nous avons étudié quelques statistiques de tests classiques et hölderiennes pour détecter une rupture épidémique dans la moyenne des variables aléatoires.

Ensuite, nous avons utilisé les statistiques hölderiennes basées d'abord sur les résidus, puis celles basées sur les observations pour tester la rupture épidémique dans la moyenne des innovations d'un processus AR(1). Une application numérique sur des données réelles et simulées est aussi présentée dans cette partie.

Pour les perspectives de ce travail, il serait intéressant de proposer d'autres statistiques pour détecter des épidémies plus courtes. De plus, on peut s'intéresser aux innovations dépendantes comme les innovations  $\alpha$ -mélangeantes ou associées.

Enfin, une extension des résultats précédents aux processus AR(p) peut être envisagée.

# Annexe A

## Mouvement Brownien et pont Brownien

### A.1 Mouvement brownien

#### A.1.1 Définition du mouvement brownien

Pour des  $\mu$  et  $\sigma^2$  fixés, il existe un processus gaussien  $(X_t)_{t \geq 0}$  tel que :

$$\begin{aligned} E(X_t) &= t\mu, \\ \text{Cov}(X_s, X_t) &= \Gamma(s, t) = \sigma^2(s \wedge t), \quad \forall s, t \geq 0. \end{aligned}$$

Tout processus de cette famille est dit mouvement brownien. En faisant le changement de variable  $Y_t = \frac{X_t - \mu t}{\sigma}$ , on aura

$$\begin{aligned} E(Y_t) &= 0 \\ \Gamma(s, t) &= s \wedge t, \quad \forall s, t \geq 0. \end{aligned}$$

Un tel processus  $(Y_t)$  est alors appelé mouvement brownien canonique ou standard.

**Remarque :** Le mouvement brownien est également appelé processus de Wiener.

#### A.1.2 Propriétés du mouvement brownien

##### a) Indépendance et stationnarité des accroissements

Le mouvement brownien  $(X_t)$  est a accroissements indépendants, ce qui signifie que pour toute suite  $t_1 \leq t_2 \leq \dots \leq D_n$  ( $n \geq 2$ ), les variables aléatoires

$X_{t_2} - X_{t_1}, X_{t_3} - X_{t_2}, \dots, X_{D_n} - X_{t_{n-1}}$  sont indépendantes.

Le mouvement brownien est a accroissements stationnaires signifie que la loi de la variable aléatoire  $X_{t+s} - X_t$  ne dépend pas de  $t$ .

**b) Régularité trajectorielle**

Dans la suite, on définit un mouvement brownien standard qu'on notera  $W_t$  comme étant un processus gaussien centré, à accroissement indépendants et stationnaires et à trajectoires continues et  $\gamma$  – h lderienne, pour tout  $\gamma \in ]0, \frac{1}{2}[$ .

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

**c) Mesurabilit  du mouvement Brownien**

Dans ce qui a pr ced , on a suppos  seulement que  $\omega \rightarrow W_t(\omega)$  est mesurable    $t$  fix .

Gr ce   la continuit  des trajectoires, on peut  noncer le r sultat suivant

L'application  $W : \Omega \times \mathbb{R}^+ \rightarrow \mathbb{R}$

$$(\omega, t) \rightarrow W(\omega, t) = W_t(\omega) \quad \text{est mesurable pour } (A \otimes \mathcal{B}(\mathbb{R}_+), \mathcal{B}(\mathbb{R})).$$

**d) Invariance du mouvement brownien**

Soit  $(W_t)_{t \geq 0}$  un mouvement brownien. Les processus suivants sont  galement des mouvements browniens.

$X(t) = |W_t|$  (mouvement brownien r fl chi   l'origine).

Pour tout  $t \geq 0$   $(W_{t+s} - W_t)$  est un mouvement brownien standard ind pendant de  $\{W_u\}_{u < t}$ .

$(-W_t)_{t \geq 0}$  est un mouvement brownien.

Pour tout  $c \geq 0$ ,  $(\frac{1}{\sqrt{c}}W_{ct})_{t \geq 0}$  est un mouvement brownien standard.

$$X(t) = \begin{cases} 0 & \text{si } t = 0 \\ tW_{\frac{1}{t}} & \text{si } t > 0 \end{cases}$$

$X(t)$  est un mouvement brownien standard.

$$Y(t) = \begin{cases} W_t & \text{si } t \leq T \\ 0 & \text{si } t \geq T \end{cases} \quad T \text{ est le premier instant o  } W_t \text{ atteint la valeur } 0.$$

$Y(t)$  est appel  mouvement brownien absorb    l'origine.

$G_t = e^{W_t}, t \geq 0$  est un mouvement brownien g om trique.

**e) Propri t  de Markov**

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

$\forall t \leq s$  et  $A \in \mathcal{B}(R)$ ,

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

avec  $\mathcal{F}_t = \sigma(W_\tau, \tau \leq t)$

**f) Propriété de martingale**

Soit  $(\Omega, \mathcal{F}, \mathbb{P})$  un espace probabilisé,  $(\mathcal{F}_t)_{t \in T}$  une famille de sous-tribus de  $\mathcal{F}$ .

$(X_t)_{t \in T} = X$  est un processus réel adapté à la famille  $(\mathcal{F}_t)_{t \in T}$ .

On dit que  $X$  est une martingale par rapport à la famille  $(\mathcal{F}_t)_{t \in T}$  si :

1. Chaque variable aléatoire  $X_t$  est intégrable.
2.  $\forall s, t \in T$  tel que  $s \leq t$  :

$$\mathbb{E}(X_t \mid \mathcal{F}_s) = X_s, \quad \text{p.s.}$$

Soit  $(W_t)$  un mouvement brownien, alors les processus suivants sont des martingales par rapport à la filtration canonique  $(\mathcal{F}_t)_{t \geq 0}$  engendrée par les variables  $W_s$ ,  $(0 \leq s \leq t)$  :

$$(W_t), \quad (W_t^2 - t) \quad \text{et} \quad \exp(aW_t - \frac{1}{2}a^2t), \quad \forall a > 0.$$

**g) Irrégularité des trajectoires**

On a déjà vu que les trajectoires sont  $\gamma$ -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  $\gamma$ -höldériennes pour  $\gamma \geq \frac{1}{2}$ .

Pour presque tout  $\omega$ ,  $(W_t(\omega))_{t \geq 0}$  n'est en aucun point  $\gamma$ -höldériennes 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, défini sur  $T = [0, 1]$  tel que

$$\text{Cov}(B_s, B_t) = s(1 - t), \quad \forall s \leq t.$$

On le note  $B_t$ , et on le définit de la façon suivante.

$$B_t = W_t - tW_1,$$

où  $W_t$  est le mouvement brownien standard.

**Remarques :**

1.  $B_1 = 0$  p.s et cela signifie que toutes les courbes viennent p.s de 0 au temps  $t = 0$ , vers 0 au temps  $t = 1$ . C'est la raison pour laquelle on l'appelle pont brownien.

On peut considérer le pont brownien  $B_t^y$  entre 0 et  $y$  sous la forme suivante.

$$B_t^y = B_t - t(B_1 - y) = B_t + ty, 0 \leq t \leq 1,$$

où  $B_t$  est le pont brownien standard venant de 0 vers 0.

$B_t^y$  est appelé le pont brownien fractionnaire.

2.  $B = (B_t)$  possède pratiquement les mêmes propriétés que le mouvement brownien.

# Annexe B

## Programmes de simulation

Dans cette annexe, nous présentons le programme de simulation des quatres statistiques de test présentées dans le chapitre 2 ainsi que le programme de calcul des puissances des tests basés sur ces statistiques codés en langage R.

### Programme de simulation des statistiques de test

```
n <- 60
n_1 <- 59
n_0 <- 1
a <- 0.25
p <- 20
q <- p + 1
l_star <- 15
delta <- 2
k <- floor(log2(n))
n_simulations <- 1000
Dn <- numeric(n_simulations)
Rn <- numeric(n_simulations)
Zn <- numeric(n_simulations)
Un <- numeric(n_simulations)
for (s in 1:n_simulations) {
  x <- numeric(n)
  for (i in 1:p) {
    x[i] <- rnorm(1) }
  for (i in (p + 1):q) {
    x[i] <- rnorm(1) + delta }
  for (i in (q + 1):n) {
    x[i] <- rnorm(1) }
  S <- cumsum(x)
```

```
#Calcul de D_a,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)
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)
D_a,n <- max(D3, D5)
#calcul de Qn
Q1 <- matrix(0, nrow = n, ncol = n - 1)
for (i in 2:n) {
  for (j in 1:(i - 1)) {
    Q1[i,j] <- abs(S[i] - S[j] - ((i - j) / n) * S[n]) }}
Qn <- max(Q1)
#Calcul de Zn
Z1 <- matrix(0, nrow = n_1, ncol = n_1 - 1)
for (i in 2:n_1) {for (j in n_0:(i - 1)) {
  Z1[i,j] <- abs((S[i] - S[j] - ((i - j) / n) * S[n])
  / sqrt((i - j) * (1 - ((i - j) / n)))) } }
Zn <- max(Z1)
#Calcul de U_a,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[j] - S[i] - S[n] * ((j - i) / n))
  / abs(((j - i) / n) * (1 - ((j - i) / n)) )^a }}
U_a,n <- max(U1)
Dn[s] <- (1/sd(x)) * n^(-1/2) * D_a,n
Rn[s] <- (1/sd(x)) * n^(-1/2) * Qn
```

```
Un[s] <- (1/sd(x)) * n^(-1/2) * U_a,n
Zn[s] <- Zn
Dn_mean <- mean(Dn)
Rn_mean <- mean(Rn)
Un_mean <- mean(Un)
Zn_mean <- mean(Zn)
cat("Moy de Dn =", Dn_mean, ", Rn =", Rn_mean, ", Un=", Un_mean,",Zn=",Zn_mean,"\n")
```

**Programme de simulation de la puissance des tests statistiques.**

```
n <- 60
n_0 <- 1
n_1 <- 59
a <- 0.25
b1 <- 1.32
b2 <- 1.50
b3 <- 2.27
b4 <- 3.60
p <- 20
l_star <- 15
q <- p + 1
delta <- 2
n_simulations<- 10000
Dn <- numeric(n_simulations)
Rn <- numeric(n_simulations)
Un <- numeric(n_simulations)
Zn <- numeric(n_simulations)
for (s in 1:n_simulations) {
  x <- numeric(n)
  for (i in 1:p) {
    x[i] <- rnorm(1)}
  for (i in (p +1):q) {
    x[i] <- rnorm(1) + delta }
  for (i in (q + 1):n) {
    x[i] <- rnorm(1) }
  S <- cumsum(x)
  k <- floor(log2(n))
  #Calcul de D_a,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)
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)
D_a,n <- max(D3, D5)
#Calcul de Qn
Q1 <- matrix(0, nrow = n, ncol = n - 1)
for (i in 2:n) {
for (j in 1:(i - 1)) {
Q1[i,j] <- abs(S[i] - S[j] - ((i - j) / n) * S[n]) }}
Qn <- max(Q1)
#Calcul de Zn
Z1 =0
for (j in n_0:n_1){
Z1[j] <- abs(((j/ n) * S[n]) - S[j]) / sqrt(j*(1(j/n)))}
Zn <- max(Z1)
#Calcul de U_a,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[j] - S[i] - S[n] * ((j - i) / n)) /
abs(((j - i) / n) * (1 - ((j - i) / n)) ^a) }}
U_a,n <- max(U1)
Dn[s] <- (1/sd(x)) * n^(-1/2) * D_a,n
Rn[s] <- (1/sd(x)) * n^(-1/2) * Qn
Un[s] <- (1/sd(x)) * n^(-1/2) * U_a,n
Zn[s] <- Zn}
Power_Dn <- sum(Dn <= b1)
```

```
Power_Rn <- sum(Rn <= b2)
Power_Un <- sum(Un <= b3)
Power_Zn <- sum(Zn <= b4)
B1=1-(Power_Dn/10000)
B2=1-(Power_Rn/10000)
B3=1-(Power_Un/10000)
B4=1-(Power_Zn/10000)
cat("Puissance du test Dn, Rn, Un, Zn: ", Power_Dn, " ", Power_Rn, " "
, Power_Un, " ", Power_Zn, "\n")
B1
B2
B3
B4
```

# Bibliographie

- [1] Avery P.J., Henderson D.A. (1999). Detecting a changed segment in DNA sequences. *J. Roy. Statist. Soc. Ser. C*, 48(4) :489-503.
- [2] Brockwell P., Davis R. (2009). *Time Series Theory and Methods* (Springer Series in Statistics), Springer science and business media.
- [3] Broemling L.D., Tsurumi H. (1987). *Econometrics and Structural Change*, Serie Statistics, 74, Marcel Dekker, New York.
- [4] Chen Z., Li Z. et Zhou M. (2016). Detecting change-points in epidemic models, *Journal of advanced statistics*, vol.1,no.4, p 181-190.
- [5] Chernoff H. , Rubin H. (1956). The estimation of the location of a discontinuity in density. In : *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1 : Contributions to the Theory of Statistics*. University of California Press, p. 19-38.
- [6] Chernoff H., Zacks S. (1964). Estimating the current mean of a normal distribution which is subjected to changes in time. *The Annals of Mathematical Statistics*, vol. 35, no 3, p. 999-1018.
- [7] Gombay E. (1994). Testing for change-points with rank and sign statistics. *Statistics and Probability Letters*, vol. 20, no 1, p. 49-55.
- [8] Gombay E., Horvath L. (1994). An application of the maximum likelihood test to the change-point problem. *Stochastic Processes and their Applications*, vol. 50, no 1, p. 161-171.
- [9] Graiche F. (2012). *Convergence hölderienne des processus empiriques quantiles et applications*, thèse de Doctorat. Université Mouloud Mammeri Tizi-Ouzou.
- [10] 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, p. 3822-3837.
- [11] Hamadouche D. (2000). Invariance principles in Hölder spaces, *Portugaliae Mathematica*, vol. 57, no 2, p. 127-152.
- [12] Hamilton J. (1994). *Time Series Analysis*-Princeton University Press.

- [13] Imeçaoudene K. (2012). Contribution aux tests statistiques pour la détection de rupture épidémique, mémoire de Master. Université Mouloud Mammeri Tizi-Ouzou.
- [14] Imeçaoudene K., Hamadouche D. (2017). Gaussian approximation for residuals of stationary autoregressive process in Holder norm. *Theory of Stochastic Processes*, vol. 22, no 2, p. 19-33.
- [15] Kerkyacharian G., Roynette B. (1991). Une démonstration simples des théorèmes de Kolmogorov Donsker et Ito-Nisio. *Comptes rendus de l'Académie des sciences. Série 1, Mathématique*, vol. 312, no 11, p. 877-882.
- [16] Lamperti J. (1962). On convergence of stochastic processes, *Transactions of the American Mathematical Society*, vol. 104, no 3, p. 430-435.
- [17] Levin B., Kline J. (1985). The CUSUM test of homogeneity with an application in spontaneous abortion epidemiology, *Statistics in Medicine*, vol. 4, no 4, p. 469-488.
- [18] Markeviit J. (2016). Epidemic change tests for the mean of innovations of an AR(1) process. *Statistics and Probability Letters*, vol. 112, p. 79-91.
- [19] Merabet D. (2012). Le processus empirique sous dépendance, thèse de Doctorat. Université Mouloud Mammeri Tizi-Ouzou.
- [20] Pettitt A. N. (1980). A simple cumulative sum type statistic for the change-point problem with zero-one observations. *Biometrika*, vol. 67, no 1, p. 79-84.
- [21] Rakauskas A., Rasten I. (2008). Hölder convergence of autoregression residuals partial sum processes. *Lithuanian Mathematical Journal*, 48(4), 438-450.
- [22] Rakauskas A., Suquet Ch. (2004). Hölder norm test statistics for epidemic change, *Journal of statistical planning and inference*, vol. 126, no 2, p. 495-520.
- [23] Sen A., Srivastava M. S. (1975). On tests for detecting change in mean. *The Annals of statistics*, p. 98-108.
- [24] Siegmund D. (1988). Confidence sets in change-point problems, *International Statistical Review*, Vol. 56, No. 1 pp. 31-48.
- [25] Yao Q. (1993). Tests for change-points with epidemic alternatives, *Biometrika*, vol. 80, no 1, p. 179-191.