

RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE
MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA
RECHERCHE SCIENTIFIQUE
UNIVERSITÉ ABOU BEKR BELKAID TLEMCCEN



Faculté des sciences
Département de Mathématiques
MÉMOIRE DE MASTER

présenté par

Zerriahen El Hassan

Soutenu le : 6 juillet 2022

Tests statistiques de détection de ruptures

Soutenu devant le jury composé de :

Mr. BOUKHARI FAKHR EDDINE	Professeur, Université de Tlemcen	Président
Mme. BENYAHIA OUAHIBA	M C B, Université de Tlemcen	Examineur
Mme. BOUKHIAR SOUAD	M C B, Université de Tlemcen	Examineur
Mr. ALLAM ABDELAZIZ	M C A, Université de Tlemcen	Encadrant

Année universitaire : 2021 - 2022

Table des matières

Introduction	4
1 Préliminaires	6
1.1 Stationnarité	6
1.2 Processus autorégressif	7
1.2.1 Estimateur de l'espérance, des auto-covariances et des autocorrélations	9
1.3 La loi du logarithme itéré	10
1.4 Mouvement Brownien	11
1.5 Test binomial	14
2 Test du changement de la moyenne à un point inconnu	17
2.1 Problème	17
2.2 Test proposé (Test d'hypothèses séquentiel)	17
2.3 Propriété du test	20
2.3.1 L'erreur de type 1	24
2.3.2 Puissance du test	25
2.3.3 Simulation :	26
2.3.4 Exemple de simulation du test	28
3 Méthodes statistique de détection de ruptures	29
3.1 Introduction :	29
3.2 La règle de Shewart et CUSUM	30
3.2.1 Le rapport de vraisemblance	30
3.2.2 La règle de Shewhart (1931) :	31
3.2.3 La règle du CUSUM	32

4	Détection des changements des paramètres dans un modèle	
	AR(p)	38
4.1	Introduction	38
4.2	Construction du test	39
4.3	Résultats d'approximation :	40
4.4	Simulation :	55
4.4.1	Détection du changement dans la moyenne :	55
4.4.2	Détection du changement dans la variance :	57
4.4.3	Détection du changement des paramètres d'un modèle autorégressif :	58
	Conclusion	60

Remerciement :

La réalisation de ce mémoire a été possible grâce au concours de plusieurs personnes à qui je voudrais témoigner toute ma gratitude.

Tout d'abord je tiens à remercier mes parents pour leurs soutiens et leurs sacrifices .

Je remercie également Mr.Allam pour la qualité de son encadrement exceptionnel, pour sa patience, sa disponibilité durant la préparation de ce mémoire.

Et enfin je remercie toute l'équipe pédagogique du département de mathématique de l'Université de Tlemcen et les intervenants professionnels responsables de ma formation, pour avoir assuré la partie théorique de celle-ci.

Introduction :

Lorsque des observations x_1, x_2, \dots ayant une distribution de paramètre θ , de valeur initiale égale à θ_0 , sont prises dans l'ordre, il peut se produire un changement dans le paramètre à un certain moment, c'est ce qu'on appelle une rupture. Par exemple, dans le contrôle de la qualité, les mesures sont prises séquentiellement sur le produit, et la valeur initiale θ_0 décrit le processus de fabrication lorsqu'il est performant. Un écart par rapport à celui-ci indique un problème de qualité et que des mesures correctives doivent être prises. On rencontre ce problème dans beaucoup de domaines, tels que l'économie, l'industrie.

L'étude statistique des détections de changement est devenue d'un intérêt majeur pour les statisticiens et a pris de plus en plus d'importance depuis les années 60.

La première contribution forte dans l'étude des détections de rupture est dû à Page (1954) qui a introduit une méthode basée sur la statistique des sommes cumulées CUSUM. Depuis, cette méthode est devenue un outil statistique standard pour tester et surveiller les ruptures dans les modèles de séries chronologiques.

Dans la théorie de la détection de rupture, il y a deux approches : "offline" et "online". Dans le cadre offline, l'ensemble des données est fixe, c'est-à-dire qu'il est observé et traité en une fois. Dans ce cas, on s'intéresse à la détection de toutes les ruptures, le plus précisément possible. Dans le cadre online, appelé aussi séquentiel, les données arrivent en temps réel, soit par point ou par lot. L'ajout et le traitement de données sont effectués instantanément avant l'arrivée de nouvelles données. Dans ce cas, on s'intéresse à la détection la plus rapide de la plus récente rupture. Notons que dans un test séquentiel, la détermination du seuil optimal de la région critique adapté à l'erreur de première espèce est assez compliquée, c'est pour cela qu'on utilise souvent une version tronquée du test.

Ce mémoire se base essentiellement sur les articles suivants : [8] et [14]. Il est réparti en quatre chapitres. Dans le premier, nous rappelons quelques notions et résultats que nous utilisons par la suite et qui sont relatifs, en particulier, à la classe des processus autorégressifs, au principe d'invariance forte, la loi du logarithme itéré et le test binomial.

Dans le deuxième chapitre, nous présentons la méthode proposée par Page [14] qui constitue le point de départ dans l'étude statistique de détection de rupture. Nous introduisons la statistique des sommes cumulées ainsi que ses propriétés. Des simulations illustrant les résultats théoriques sont données.

Le troisième chapitre est consacré à quelques tests de détection de rupture. Nous présentons dans une première étape, le test de Shewart qui a étudié le changement d'un paramètre θ d'une distribution en se basant sur le rapport de vraisemblance. Ensuite, nous développons le test CUSUM des rapports de vraisemblance dans le cas des données indépendantes. Une généralisation de ce test aux données dépendantes est donnée en utilisant les distributions conditionnelles et données dépendantes. Nous terminons ce chapitre par des simulations numériques qui mettent en évidence l'intérêt du test CUSUM.

Dans le quatrième chapitre, nous développons les résultats obtenus dans [8]. Les auteurs s'y intéressent au test CUSUM pour étudier le changement des paramètres d'un modèle autorégressif d'ordre p . Ils ont construit un test avec une statistique à l'aide des composantes du vecteur score relatives au paramètre d'intérêt et des estimateurs des paramètres de nuisance. Ils ont aussi établi des théorèmes limites avec les vitesses de convergence pour ces estimateurs et ont obtenu une approximation de la statistique test par un mouvement Brownien. Nous illustrons les résultats théoriques de ce chapitre par des simulations numériques sur le changement des trois paramètres (la moyenne, la variance et le coefficient d'autorégression) d'un processus autorégressif d'ordre 1.

Chapitre 1

Préliminaires

1.1 Stationnarité

Une des grandes questions dans l'étude de séries temporelles est de savoir si celles-ci suivent un processus stationnaire. La notion de stationnarité caractérise la capacité d'un processus à se décorrélérer totalement de l'indice temporel.

On distingue généralement la stationnarité au sens strict et la stationnarité au sens faible.

Définition 1.1.1. *Un processus aléatoire $(X_t)_{t \in \mathbb{Z}}$ est dit fortement stationnaire ou strictement stationnaire si pour quel que soit le n – uplet du temps $t_1, t_2, \dots, t_n \in \mathbb{Z}$ avec $t_1 < t_2 < \dots < t_n$ et pour tout $h \in \mathbb{Z}$, le vecteur $(X_{t_1+h}, X_{t_2+h}, \dots, X_{t_n+h})$ possède la même loi que le vecteur $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$.*

Cependant, cette notion est difficile à vérifier en pratique, alors que la notion suivante peut sembler moins exigeante car elle n'impose des contraintes qu'aux deux premiers moments des variables $X_t, t \in \mathbb{Z}$.

Définition 1.1.2. *Un processus aléatoire $(X_t)_{t \in \mathbb{Z}}$ est dit faiblement stationnaire si pour $t, h \in \mathbb{Z}$,*

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

Définition 1.1.3. *Soit $(X_t)_{t \in \mathbb{Z}}$ un processus stationnaire.*

1. On appelle fonction d'autocovariance associée à $(X_t)_{t \in \mathbb{Z}}$, la fonction γ définie pour tout $k \in \mathbb{Z}$ par $\gamma(h) = \text{con}(X_0, X_h)$.
2. On appelle fonction d'autocorrélation associée à $(X_t)_{t \in \mathbb{Z}}$, la fonction ρ donnée par $\rho(h) = \frac{\gamma(h)}{\gamma(0)}$.

1.2 Processus autorégressif

La classe des processus autorégressifs et plus généralement les modèles linéaires occupent une place centrale dans la théorie et le traitement des séries temporelles. Elle joue un rôle très important dans la prévision des processus stationnaires. Dans cette section, nous rappelons quelques propriétés de cette classe de processus.

Définition 1.2.1. Soit $(\varepsilon_t)_{t \in \mathbb{Z}}$ un processus stochastique. $(\varepsilon_t)_{t \in \mathbb{Z}}$ est dit bruit blanc fort si les variables aléatoires $(\varepsilon_t)_t$ sont i.i.d centrées avec $E(\varepsilon_t^2) = \sigma^2 < \infty$.

Définition 1.2.2. : Un processus $(Y_t)_{t \in \mathbb{Z}}$ est dit processus autorégressif d'ordre p , noté $AR(p)$, s'il satisfait pour tout $t \in \mathbb{Z}$ la relation

$$Y_t = \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + c + \varepsilon_t, \quad (1.1)$$

où $(\varepsilon_t)_{t \in \mathbb{Z}}$ est un bruit blanc fort de variance σ^2 , ϕ_i $i = 1, 2, \dots, p$ et c sont des constantes réelles telles que $\phi_p \neq 0$.

Dans une représentation autorégressive, chaque observation est composée par une combinaison linéaire de ses p valeurs passées, perturbée par un bruit blanc $(\varepsilon_t)_{t \in \mathbb{Z}}$.

Si le processus $(Y_t)_{t \in \mathbb{Z}}$ défini par (1.1) est faiblement stationnaire avec $\mu = E(Y_t)$ alors on a

$$\mu = \frac{c}{1 - \sum_{i=1}^p \phi_i}.$$

Avec le changement de variables $Z_t = Y_t - \mu$, on obtient

$$Z_t = \sum_{i=1}^p \phi_i Z_{t-i} + \varepsilon_t.$$

Ainsi $(Z_t)_{t \in \mathbb{Z}}$ est un $AR(p)$ centré.

Nous pouvons donc supposer, sans perte de généralité que $\mu = 0$ et donc $c = 0$.

Définition 1.2.3. *Le polynôme caractéristique associé au processus $AR(p)$ donné par l'équation 1.1 est défini par*

$$\Phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p.$$

Définition 1.2.4. *On définit l'opérateur "retard", noté L , comme l'opérateur linéaire qui associe à un $(X_t)_{t \in \mathbb{Z}}$ le processus $(Y_t)_{t \in \mathbb{Z}}$ tel que*

$$Y_t = LX_t = X_{t-1}, \quad t \in \mathbb{Z}.$$

L'opérateur L associe donc à toute valeur courante du processus sa valeur précédente.

A l'aide de l'opérateur retard, le modèle $AR(p)$ donné par la relation (1.1) s'écrit

$$\Phi(L)Y_t = \varepsilon_t, \quad t \in \mathbb{Z}.$$

Le théorème suivant donne une condition suffisante pour que l'équation (1.1) admette une solution strictement stationnaire.

Théorème 1.2.1. [18] :

Si le polynôme caractéristique ϕ vérifie $\Phi(z) \neq 0$ pour tout $z \in \mathbb{Z}$ tel que $|z| \leq 1$, alors la représentation autorégressive donnée par l'équation (1.1) possède une unique solution faiblement stationnaire.

Dans le cas d'un processus $AR(1)$, nous avons

Proposition 1.2.1. *Soit Y_t un $AR(1)$ donné par*

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

avec $|\phi| < 1$, alors

1. $Var(Y_t) = \gamma(0) \frac{\sigma}{1 - \phi^2}$.
2. $cov(Y_t, Y_{t-h}) = \gamma(h) \phi^h \frac{\sigma}{1 - \phi}$.

1.2.1 Estimateur de l'espérance, des auto-covariances et des autocorrélations

Soit $(Y_t)_{t \in \mathbb{Z}}$ un processus autorégressif stationnaire d'ordre p satisfaisant la relation (1.1) de moyenne μ . Pour estimer μ , la fonction d'autocovariance γ et la fonction d'autocorrélation ρ , nous considérons les estimateurs empiriques :

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n Y_i, \quad (1.2)$$

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{i=1}^{n-|h|} (Y_i - \hat{\mu})(Y_{i+|h|} - \hat{\mu}), \quad (1.3)$$

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}. \quad (1.4)$$

Si nous disposons de n observations, Y_1, Y_2, \dots, Y_n , alors il n'existe pas de paires d'observations séparées de plus de $n - 1$ intervalles de temps et donc l'expression (1.3) ne permet pas d'estimer les valeurs de $\gamma(h)$ pour $|h| \geq n$. De plus, lorsque $|h|$ est proche de n , il est clair que l'estimateur (1.3) de la covariance n'est pas fiable, dans la mesure où on ne dispose que de peu de paires d'observations $(Y_t, Y_{t+|h|})$.

Théorème 1.2.2. [18]

Les estimateurs définis dans (1.2), (1.3) et (1.4) convergent presque sûrement vers les valeurs théoriques μ , $\gamma(h)$ et $\rho(h)$.

Soit $(Y_t)_{t \in \mathbb{Z}}$ un processus autorégressif d'ordre p vérifiant (1.1).
Considérons

$$\mathbf{X}_n = \begin{pmatrix} Y_0 - \mu & Y_{-1} - \mu & \dots & Y_{-p+1} - \mu \\ Y_1 - \mu & Y_0 - \mu & \dots & Y_{-p+2} - \mu \\ \vdots & \vdots & & \vdots \\ Y_{n-1} - \mu & Y_{n-2} - \mu & \dots & Y_{n-p} - \mu \end{pmatrix}$$

Si on suppose que $(Y_t)_{t \in \mathbb{Z}}$ est stationnaire et soit Γ la matrice de covariance de $(Y_1 - \mu, \dots, Y_n - \mu)^t$, alors le Théorème 1.2.2 donne .

Théorème 1.2.3.

$$\frac{1}{n} \mathbf{X}_n^t \mathbf{X}_n \xrightarrow{p.s.} \Gamma$$

et en particulier pour $j, s = \{1, 2, \dots, p\}$ tels que $s \leq j$:

$$\frac{1}{n} \sum_{i=1}^n (Y_{i-j} - \mu)(Y_{i-s} - \mu) \xrightarrow{p.s.} \gamma(j-s)$$

Où γ est la fonction de covariance

1.3 La loi du logarithme itéré

Soit une suite variable aléatoire $(X_n)_{n \in \mathbb{N}}$ iid telle que :

$$E[X_n] = \mu, \quad \text{var}[X_n] = \sigma^2.$$

alors d'après la loi forte des grands nombres

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{p.s.} \mu$$

la loi du logarithme itéré nous donne la vitesse de convergence presque sûre assurée par la loi forte des grands nombres.

elle a été introduite par Alexandre Khintchine en (1924) pour des variables suivant une loi de Rademacher. elle a été ensuite démontrée pour des variables i.i.d et carré intégrale par P. Hartman et A. Wintner en (1940)

Théorème 1.3.1. (P. Hartman et A. Wintner) [9]

Soit une suite variable aléatoire $(X_n)_{n \in \mathbb{N}}$ iid telle que :

$$E[X_n] = \mu, \quad \text{var}[X_n] = \sigma^2.$$

et soit : $S_n = X_1 + \dots + X_n$ pour $n \geq 1$.

Alors

$$\liminf_{n \rightarrow \infty} \frac{S_n - n\mu}{\sqrt{2n\sigma^2 \log \log n}} = -1, \quad \limsup_{n \rightarrow \infty} \frac{S_n - n\mu}{\sqrt{2n\sigma^2 \log \log n}} = 1 \quad p.s.$$

en particulier :

$$\left| \frac{1}{n} \sum_{i=1}^n X_i - \mu \right| = O((n^{-1} \log \log n)^{\frac{1}{2}})$$

1.4 Mouvement Brownien

Le Mouvement Brownien joue un rôle fondamental dans la théorie des processus stochastiques à temps continu, aussi bien pour l'étude des trajectoires que pour la théorie de l'intégration stochastique.

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité.

Définition 1.4.1. :

Un processus stochastique $\{B(t) : t \geq 0\}$ défini sur $(\Omega, \mathcal{A}, \mathbb{P})$ et à valeurs réelles est appelé mouvement Brownien issu de 0 si :

1. $B(0) = 0$ p.s
2. le processus est à accroissements indépendants, c'est-à-dire pour tous les temps $0 \leq t_1 \leq t_2 \leq \dots \leq t_n$ $B(t_n) - B(t_{n-1}), B(t_{n-1}) - B(t_{n-2}), \dots, B(t_2) - B(t_1)$ sont indépendantes.
3. Pour tout $t \geq 0, h > 0$, les accroissements $B(t+h) - B(t)$ sont normalement distribués avec espérance 0 et variance h .
4. La fonction $t \mapsto B(t)$ est continue. p.s

Nous avons la propriété suivante

Proposition 1.4.1. Le processus B est un processus gaussien, continu, sa loi est caractérisée par sa moyenne nulle et sa covariance $\text{Cov}(B_s, B_t) = s \wedge t$.

La proposition suivante donne quelques propriétés du mouvement Brownien

Proposition 1.4.2. Soit $(B_t, t \geq 0)$ un mouvement Brownien, alors

- i) Le processus \hat{B} défini par $\hat{B}_t = -B_t$ est un mouvement Brownien.
- ii) Le processus \tilde{B} défini par $\tilde{B}_t = \frac{1}{c} B_{c^2 t}$ est un mouvement Brownien.
- iii) Le processus \bar{B}_t défini par $\bar{B}_t = t B_{\frac{1}{t}}$ pour $t > 0$ et $\bar{B}_0 = 0$ est un mouvement Brownien.

Définition 1.4.2. Un processus stochastique $\{B(t) : t \geq 0\}$ dans \mathbb{R}^n est appelé un mouvement brownien à d dimensions issu de 0 à si :

$$B(t) = (B_1(t), B_2(t), \dots, B_n(t))$$

où B_1, \dots, B_n sont des mouvements browniens linéaires indépendants issu de 0 respectivement.

Le théorème suivant donne la loi du logarithme itéré pour le mouvement brownien.

Théorème 1.4.1. [12]

$(B_t, t \geq 0)$ un mouvement Brownien, alors :

$$\limsup_{t \rightarrow +\infty} \frac{B(t)}{\sqrt{2t \ln(\ln(t))}} = 1 \text{ p.s.}$$

et

$$\liminf_{t \rightarrow +\infty} \frac{B(t)}{\sqrt{2t \ln(\ln(t))}} = -1 \text{ p.s.}$$

Lemme 1.4.1. [5] et [4] :

Soit $B(\cdot)$ un mouvement brownien standard alors

$$\max_{0 \leq s < t \leq 1} (B(t) - B(s)) \stackrel{\text{lois}}{=} \max_{0 \leq t \leq 1} |B(t)|$$

et pour $x \in \mathbb{R}$

$$P(\max_{0 \leq t \leq 1} |B(t)| \leq x) = \frac{4}{\pi} \sum_{n=0}^{\infty} \frac{(-1)^n}{2n+1} \exp\left\{-\frac{\pi^2(2n+1)^2}{8x^2}\right\} \quad (1.5)$$

Le mouvement Brownien peut être utilisé pour approximer les sommes partielles d'une suite de variables aléatoires sous certaines hypothèses.

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables définies sur un espace de probabilité (Ω, \mathcal{A}, P) et $(S_n)_{n \in \mathbb{N}}$ la suite de leurs sommes partielles

$$S_n = X_1 + X_2 + \dots + X_n.$$

On définit le processus somme partielle en temps continu $\{S(t) : t \geq 0\}$ comme suit

$$S(0) = 0, \quad S(t) = S_{[t]} \quad (t \in \mathbb{R}_+)$$

où $[.]$ est la partie entière.

Dans le cas d'une suite $(X_n)_{n \in \mathbb{N}}$ *i.i.d.*, nous avons

Théorème 1.4.2. *Komlós–Major–Tusnády [10] et [11] :*

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires indépendantes et identiquement distribuées sur (Ω, \mathcal{A}, P) avec $EX_1 = 0$ et $\text{Var } X_1 = 1$. Si

$$E |X_1|^\nu < \infty$$

pour certains $\nu > 2$, il existe un processus de Wiener standard $\{B(t) : t \geq 0\}$ tel que

$$|S(t) - B(t)| = o\left(t^{\frac{1}{\nu}}\right) \quad p.s.$$

pour $t \rightarrow \infty$.

Si les variables (X_n) sont dépendantes, nous avons

Théorème 1.4.3. *(Eberlein) [6] :*

Soit $\{X_n\}_{n \in \mathbb{N}}$ une suite de variables aléatoires à valeurs réelles et soit

$$S_n(m) = X_{m+1} + \dots + X_{m+n} \quad (m \in \mathbb{N}^*, n \in \mathbb{N}),$$

si $\{X_n\}_{n \in \mathbb{N}}$ vérifie les hypothèses suivantes :

a) $EX_n = 0$ pour tous $n \in \mathbb{N}$,

b) $\|E(S_n(m) | \mathcal{F}_m)\|_1 = \mathcal{O}\left(n^{\frac{1}{2}-\theta}\right)$, $m \in \mathbb{N}^$ pour un certain $\theta \in (0, \frac{1}{2})$,*

o $\mathcal{F}_m = \sigma(X_i, i \leq m)$ (la tribu engendrée par les X_i $i \leq m$)

c) il existe une constante σ_W telle que pour $m \in \mathbb{N}^$,*

$$ES_n^2(m) - \sigma_W = \mathcal{O}(n^{-\rho})$$

quand $n \rightarrow \infty$ pour un certain $\rho > 0$,

d) il existe un $\gamma > 0$ tel que pour $m \in \mathbb{N}^$,*

$$\|E(S_n^2(m) | \mathcal{F}_m) - ES_n^2(m)\|_1 = \mathcal{O}(n^{1-\gamma}) \quad p.s.$$

quand $n \rightarrow \infty$,

e) il existe une constante $M < \infty$ et $\kappa > 2$, telle que $E|X_n|^\kappa < M$ pour tout $n \in \mathbb{N}$.

Alors, il existe un processus de Wiener $\{B(t) : t \geq 0\}$, tel que :

$$\left| \sum_{n=1}^{[t]} X_n - \sigma_W B(t) \right| = \mathcal{O}\left(t^{\frac{1}{\nu}}\right) \quad a.s.$$

quand $t \rightarrow \infty$ pour un $\nu > 2$.

Les Théorèmes 1.4.2 et 1.4.3. vont nous servir pour démontrer les approximations de la statistique test dans le chapitre 4.

1.5 Test binomial

Le test binomial est utilisé lorsqu'une expérience a deux résultats possibles (c'est-à-dire succès/échec) avec des informations sur la probabilité de succès.

Nous allons introduire l'exemple suivant pour bien éclaircir les étapes du test.

Exemple 1.5.1. : *(l'influence de l'étude des mathématiques sur le niveau d'anglais)*

Un test qui a pour but de tester le niveau d'anglais pour les étudiants de la faculté des sciences a montré qu'une proportion de 45% d'étudiants ont une note supérieure à 10. Le même test est donné à 25 étudiants du département de mathématiques. Lors de ce test, 13 d'entre eux ont une note supérieure à 10.

On veut savoir si les étudiants du département de mathématiques ont un niveau d'anglais supérieur à celui de la population générale (étudiants de la faculté des sciences).

On introduit un paramètre p qui désigne la proportion d'étudiants du département de mathématiques qui ont une note supérieure à 10 au test d'anglais. Nous avons à tester les deux hypothèses :

$$H_0 : p = p_0 = 0.45 \quad \text{contre} \quad H_1 : p > 0.45$$

Soit T la statistique égale au nombre d'étudiants du département de mathématiques qui ont une note supérieure à 10.

Le nombre d'étudiants du département de mathématiques est $n = 25$ et T suit une loi binomiale $\mathcal{B}(25, p)$.

sous H_0 , T suit une lois $\mathcal{B}(25, 0.45)$

On fixe le risque α de première espèce, dont on rappelle que plus il diminue plus le test est significatif. Puis on cherche la région critique que l'on va noter R en posant la question à partir de quelle valeur T avec un risque α peut on accepter H_1

Dans notre exemple :

On se fixe à un niveau $\alpha = 5\%$, sur 25 étudiants de math à partir de combien d'étudiants qui ont eu note supérieure à 10 on rejette H_0 .

On rejette H_0 (i.e on accepte H_1) si le nombre des étudiants de mathématique qui ont eu une note supérieure à 10 est assez grand et donc la région critique est l'ensemble $R = \{T \geq k\}$, $k > 0$. Maintenant on cherche k de

telle sorte que

$$\mathbb{P}_{H_0}(T \geq k) \leq 0.05$$

Le tableau suivant donne, sous H_0 , (i.e quand T suit une loi $\mathcal{B}(25, 0.45)$) les probabilités $\mathbb{P}(T \geq k)$

k	...	18	17	16	15	14	13	12	11
$\mathbb{P}(T \geq k)$		0.005	0.017	0.043	0.059	0.182	0.306	0.457	0.615

On a $\mathbb{P}_{H_0}(T \geq 15) = 0.095$ et $\mathbb{P}_{H_0}(T \geq 16) = 0.043 \leq \alpha$ d'où $k = 16$. la région critique pour le risque $\alpha = 5\%$ est $R = \{T \geq 16\}$ et donc on accepte H_0 et on rejette H_1 c'est à dire que le niveau d'anglais des étudiant du département de mathématiques est comme celui de la population générale.

Remarque :

Il existe une autre méthode pour trancher entre H_0 et H_1 sans chercher la région critique.c'est celui de calculer la " p-valeur" .

Définition 1.5.1. (*p-valeur*) :

Soit T une statistique test et t sa valeur observée. La p-valeur est le risque minimal de rejeter l'hypothèse H_0 à partir de la valeur observée t . Elle fourni le niveau signification le plus petit pour lequel on rejette l'hypothèse H_0 . Plus la p-valeur est petite plus le test est en faveur de l'hypothèse alternative H_1 .

Si la p-valeur est inférieure ou égale à α , on rejette H_0 .

Dans l'exemple précédent, puisque la région critique est $R = \{T \geq k\}$ où $k > 0$ et la valeur observée de la statistique est $T = 13$ alors

$$p - \text{valeur} = \mathbb{P}_{H_0}(T \geq 13) = 0.306 > \alpha$$

Donc on accepte H_0 .

Définition 1.5.2. (*Puissance du test*) :

La puissance du test, notée η , est la probabilité de rejeter H_0 sous H_1 : $\eta = P_{H_1}(H_0)$.La puissance du test sert à mesurer la qualité du test.

Dans notre exemple :

Puisque notre test est un test unilatéral et la statistique T sous l'hypothèse H_1 suit une loi $\mathcal{B}(25, p)$ avec $p > 0.45$, alors la puissance du test est une fonction du paramètre p :

$$\eta = P_{H_1}(T \geq 16)$$

Par exemple si on veut tester que la proportion d'étudiants du département de mathématiques qui ont une note supérieure à 10 est de 50% au lieu de 45%, alors l'hypothèse alternative devient :

$$H_1 : p = p_1 = 0.5$$

Sous H_1 , T suit une loi $\mathcal{B}(25, 0.5)$. Par conséquent

$$\eta = \mathbb{P}_{H_1}(T \geq 16) = 0.114.$$

On remarque bien que la puissance est faible, donc le test n'est pas adapté pour tester $p_1 = 0.5$ contre $p_0 = 0.45$ qui sont assez proches. Dans ce cas la, si on veut mettre en évidence que $p = 0.5$ au lieu de 0.45, la taille de l'échantillon $n = 25$ n'est pas suffisante.

Chapitre 2

Test du changement de la moyenne à un point inconnu

Soit (Ω, \mathcal{A}, P) un espace de probabilité.

2.1 Problème

Soit X_1, X_2, \dots, X_n un échantillon de variables aléatoires réelles définies sur (Ω, \mathcal{A}, P) indépendants de fonction de distribution $f(x; \theta)$ où θ est un paramètre (la moyenne, la proportion...) de valeur initiale θ_0 connue. Dans ce chapitre on va étudier le cas où θ est la moyenne.

On veut tester l'hypothèse nulle H_0 selon laquelle toutes les observations sont tirées de la même population avec la fonction de distribution $f(x; \theta_0)$ contre l'hypothèse alternative :

$H_1 : X_1, \dots, X_k$ sont tirées de $f(x; \theta_0)$ et X_{k+1}, \dots, X_n de $f(x; \theta_1)$ $\theta_0 \neq \theta_1$.

On suppose que k inconnu (si k est connu ça serait un simple teste de comparaison entre deux échantillon).

Le problème de la détection du changement d'un paramètre est très important dans beaucoup domaine telle que le contrôle de qualité.

2.2 Test proposé (Test d'hypothèses séquentiel)

Le test d'hypothèse séquentiel est un test statistique où la taille de l'échantillon n'est pas fixée à l'avance contrairement au test classique. La statistique

test est évaluée séquentiellement après chaque nouvelle observation et le test s'arrête si on rejette H_0 . La taille de l'échantillon utilisée sera donc une variable aléatoire (un temps d'arrêt). ou bien le teste s'arrête et on accepte H_0 . si le test est tronqué

Définition 2.2.1. *Si on fixe la taille maximale de l'échantillon qu'on doit utiliser par une constante, alors le test est dit séquentiel tronqué et la constante est appelée le point de troncature.*

Considérons un échantillon $(X_1, X_2..X_n)$ de variables aléatoires réelles et indépendantes

Un schéma d'inspection du processus pour détecter un changement unilatéral, en particulier le cas unilatérale supérieur, a été donné par Page ([13]et[14]). Le paramètre considéré est la moyenne de la distribution des X_i , dont la valeur initiale $\theta = \theta_0$ est connue.

Le schéma consiste à enregistrer les sommes cumulées

$$S_0^X = 0, \quad S_r^X = \sum_{i=1}^r (X_i - \theta_0),$$

et de détecter un éventuelle changement si :

$$S_r^X - \min_{0 \leq i < r} S_i^X \geq h, \quad h > 0$$

Sur un graphe, comme le montre la figure 2.1, on peut voir que s'il n'y a pas de changement dans la moyenne ($\theta = \theta_0$), la courbe des sommes cumulées est horizontale, tandis que si une augmentation de la moyenne se produit, la courbe a une pente positive.

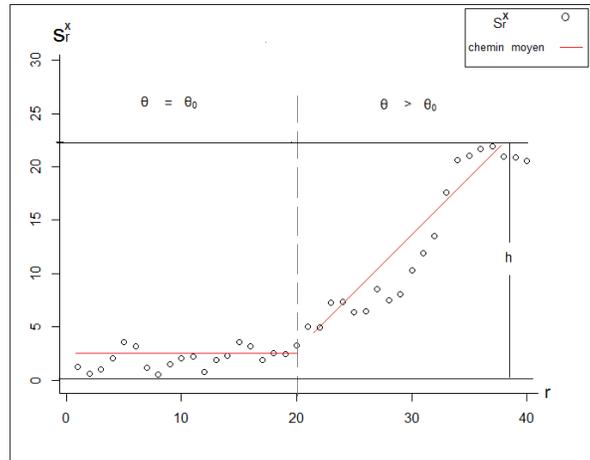


FIGURE 2.1 – Simulation des sommes S_r^X sur un échantillon de lois $N(5.1)$ et ça moyenne vaut 6 a l’instant 21.

D’où l’intérêt d’introduire la statistique

$$M^X = \max_{0 \leq r \leq n} \left(S_r^X - \min_{0 \leq i \leq r} S_i^X \right).$$

Page considère le test séquentiel tronqué pour le cas unilatéral supérieur ($\theta_1 > \theta_0$).

Test 1

H₀ : θ est constamment θ_0 .

H₁ : $\theta = \theta_0$ pour X_1, \dots, X_k et $\theta = \theta_1$ pour $X_{k+1}, X_{k+2}, \dots, X_n$.

A cause des difficultés techniques liées à l’étude de la statistique M^X , l’auteur introduit les variables aléatoires

$$Y_i = \begin{cases} a & \text{si } X_i \geq \theta_0 \\ -b & \text{sinon} \end{cases}$$

où a, b sont les nombres réels strictement positif vérifiant $E(Y_i | \theta_0) = 0$ ($i = [0, 1, \dots, n]$).

Soit

$$S_0 = 0, \quad S_r = \sum_{i=1}^r Y_i,$$

$$m_r = S_r - \min_{0 \leq i \leq r} S_i$$

et

$$M = \max_{0 \leq r \leq n} m_r$$

La région critique est $R = \{M \geq h\}$ avec $h > 0$.

2.3 Propriété du test

Pour simplifier on considère que les variable X_i sont symétriques par rapport a la moyenne initial θ_0 , c'est à dire $\mathbb{P}(X_i < \theta_0) = \mathbb{P}(X_i \geq \theta_0)$ de sorte que on peut prendre $a = b = 1$. Ainsi $Y_i = \text{sgn}(X_i - \theta)$ avec $\text{sgn}(x)$ représente le signe de x .

Posons pour $i = 0, 1, \dots, h-1$,

$$p_{r,i} = \mathbb{P}(m_r = i, \bigcap_{s=1}^{r-1} m_s < h).$$

Il est clair que $p_{0,0} = 1$ et $p_{0,i} = 0$ pour $i \neq 0$.

Proposition 2.3.1. *La probabilité de rejeter H_0 est :*

$$\mathbb{P}(R) = 1 - \sum_{i=0}^{h-1} p_{n,i} \tag{2.1}$$

Preuve :

$$\begin{aligned} \mathbb{P}(R) &= \mathbb{P}(M \geq h) \\ &= \mathbb{P}(\max_{0 \leq r \leq n} M_r \geq h) \\ &= 1 - \mathbb{P}(\max_{0 \leq r \leq n} m_r < h) \\ &= 1 - \mathbb{P}(m_1 < h, m_2 < h \dots m_n < h) \\ &= 1 - \mathbb{P}(m_n < h, \bigcap_{s=1}^{n-1} m_s < h) \\ &= 1 - \sum_{i=0}^{h-1} \mathbb{P}(m_n = i, \bigcap_{s=1}^{n-1} m_s < h) \\ &= 1 - \sum_{i=1}^{h-1} p_{n,i}. \end{aligned}$$

Supposons maintenant que $\mathbb{P}(Y_i = 1) = p = 1 - q$.

Proposition 2.3.2. *Sous les hypothèse précédentes et pour $h \geq 2$ nous avons*

$$\begin{cases} p_{r+1,0} = q (p_{r,0} + p_{r,1}) \\ p_{r+1,i} = p p_{r,i-1} + q p_{r,i+1} \quad (1 \leq i < h - 1) \\ p_{r+1,h-1} = p p_{r,h-2} \end{cases} \quad (2.2)$$

Preuve :

Posons $B_r = \bigcap_{s=1}^r \{m_s < h\}$ et $A_{r,i} = \{m_r = i, B_{r-1}\}$.

i) On a $p_{r+1,0} = \mathbb{P}(A_{r+1,0})$. Or

$$A_{r+1,0} = \{m_{r+1} = 0, B_r\} = \{S_{r+1} = \min_{0 \leq i \leq r+1} S_i, B_r\}.$$

Mais $S_{r+1} = \min_{0 \leq i \leq r+1} S_i$ est équivalent à $S_{r+1} < S_r$, soit à $Y_{r+1} = -1$.

On peut écrire donc

$$\begin{aligned} A_{r+1,0} &= \{S_{r+1} = \min_{0 \leq i \leq r+1} S_i, Y_{r+1} = -1, B_r\} \\ &= \left\{ S_r = \min_{0 \leq i \leq r} S_i, Y_{r+1} = -1, B_r \right\} \\ &\quad \cup \left\{ S_r = \min_{0 \leq i \leq r} S_i + 1, Y_{r+1} = -1, B_r \right\} \\ &= \{m_r = 0, Y_{r+1} = -1, B_r\} \cup \{m_{r+1} = +1, Y_{r+1} = -1, B_r\} \\ &= \{m_r = 0, Y_{r+1} = -1, B_{r-1}\} \\ &\quad \cup \{M_{r+1} = +1, Y_{r+1} = -1, B_{r-1}\} \\ &= [\{m_r = 0, , B_{r-1}\} \cup \{m_r = +1, B_{r-1}\}] \cap \{Y_{r+1} = -1\}. \end{aligned}$$

D'où

$$A_{r+1,0} = [A_{r,0} \cup A_{r,1}] \cap \{Y_{r+1} = -1\}$$

finalemt par passage au probabilité et en utilisant le faite que les variable aléatoires X_i sont indépendantes on obtient

$$p_{r+1,0} = q (p_{r,0} + p_{r,1}).$$

ii) On a

$$\begin{aligned}
A_{r+1,i} &= \{m_{r+1} = i, B_r\} \\
&= \left\{ S_{r+1} - \min_{0 \leq i \leq r+1} S_i = i, B_r \right\} \\
&= \left\{ S_r - \min_{0 \leq i \leq r+1} S_i = i - 1, Y_{r+1} = 1, B_r \right\} \\
&\quad \cup \left\{ S_r - \min_{0 \leq i \leq r+1} S_i = i + 1, Y_{r+1} = -1, B_r \right\} \\
&=: A \cup B
\end{aligned}$$

avec

$$A = \left\{ S_r - \min_{0 \leq i \leq r+1} S_i = i - 1, Y_{r+1} = 1, B_r \right\}$$

et

$$B = \left\{ S_r - \min_{0 \leq i \leq r+1} S_i = i + 1, Y_{r+1} = -1, B_r \right\}.$$

Pour l'ensemble A , puisque $Y_{r+1} = 1$ alors $S_{r+1} > \min_{0 \leq i \leq r+1} S_i$ et $\min_{0 \leq i \leq r+1} S_i = \min_{0 \leq i \leq r} S_i$. Ceci donne

$$A = \{m_r = i - 1, Y_{r+1} = 1, B_r\}.$$

Comme $i - 1 \leq h - 3$ alors

$$A = \{m_r = i - 1, Y_{r+1} = 1, B_{r-1}\}.$$

Pour l'ensemble B , comme $i \geq 1$ alors

$$S_r - \min_{0 \leq i \leq r+1} S_i = i + 1 \geq 2$$

$Y_{r+1} = -1$ entraine

$$S_{r+1} - \min_{0 \leq i \leq r+1} S_i \geq 1,$$

par suite $S_{r+1} > \min_{0 \leq i \leq r+1} S_i$ et donc $\min_{0 \leq i \leq r+1} S_i = \min_{0 \leq i \leq r} S_i$. en déduit que

$$B = \left\{ S_r - \min_{0 \leq i \leq r} S_i = i + 1, Y_{r+1} = -1, B_r \right\}.$$

Puisque $i + 1 \leq h - 1$ alors

$$B = \{m_r = i + 1, Y_{r+1} = -1, B_{r-1}\}.$$

On obtient donc

$$\begin{aligned} A_{r+1,i} &= [\{m_r = i - 1, B_{r-1}\} \cap \{Y_{r+1} = 1\}] \\ &\quad \cup [\{m_r = i + 1, B_{r-1}\} \cup \{Y_{r+1} = -1\}] \\ &= [A_{r,i-1} \cap \{Y_{r+1} = 1\}] \cup [A_{r,i+1} \cup \{Y_{r+1} = -1\}]. \end{aligned}$$

Ce qui montre que

$$p_{r+1,i} = p p_{r,i-1} + q p_{r,i+1}.$$

iii) Par le même raisonnement que *ii*) on obtient

$$A_{r+1,h-1} = \{m_r = h - 2, Y_{r+1} = 1, B_r\} \cup \{m_r = h, Y_{r+1} = -1, B_r\}.$$

Or

$$\{m_r = h, Y_{r+1} = -1, B_r\} = \{m_r = h, m_r < h, Y_{r+1} = -1, B_{r-1}\} = \emptyset.$$

Il en découle que

$$A_{r+1,h-1} = \{m_r = h - 2, Y_{r+1} = 1, B_r\} = \{Y_{r+1} = 1\} \cap A_{r,h-2}.$$

Par suite

$$p_{r+1,h-1} = p p_{r,h-2}.$$

Introduisons les notations suivantes :

$$\mathbf{p}_r = \begin{pmatrix} p_{r,0} \\ p_{r,1} \\ p_{r,2} \\ \vdots \\ p_{r,h-1} \end{pmatrix}$$

et

$$P = \begin{pmatrix} q & q & 0 & 0 & \dots & 0 & 0 \\ p & 0 & q & 0 & \dots & 0 & 0 \\ 0 & p & 0 & q & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & p & 0 \end{pmatrix}$$

Les relations 2.2 possèdent la représentation matricielle donnée par la proposition suivante :

Proposition 2.3.3. *Nous avons*

$$(a) \mathbf{p}_{r+1} = \mathbf{P} \cdot \mathbf{p}_r. \quad (2.3)$$

$$(b) \mathbf{p}_r = \mathbf{P}^r \cdot \mathbf{p}_0. \quad (2.4)$$

Preuve Elle repose sur les égalités

$$\left\{ \begin{array}{l} p_{r+1,0} = q (p_{r,0} + p_{r,1}) \\ p_{r+1,1} = p p_{r,0} + q p_{r,2} \\ p_{r+1,2} = p p_{r,1} + q p_{r,3} \\ \cdot \\ \cdot \\ p_{r+1,h-1} = p p_{r,h-2} \end{array} \right.$$

2.3.1 L'erreur de type 1

TABLE 2.1 – les valeur de n et h

		$\alpha = 0.05$		$\alpha = 0.01$	
n	h	n	h	n	h
21	10	75	19	20	12
26	11	83	20	27	14
31	12	91	21	35	16
36	13	100	22	43	18
41	14	119	24	53	20
47	15	139	26	64	22
54	16	161	28	76	24
60	17	185	30	89	26
67	18			103	28
				118	30

Sous l'hypothèse nulle H_0 , $p = \frac{1}{2}$ constamment car les variables X_i sont symétriques par rapport a la moyenne θ_0 . (i.e la probabilité que $X_i \geq \theta_0$ est égale a celle de $X_i \leq \theta_0$).

Par conséquent, on peut calculer l'erreur de première espèce (le rejet de H_0 à tort) par (2.1) et la relation(2.4) en posant $r = n$ (la taille de l'échantillon) ce qui donne :

$$\mathbf{p}_n = \mathbf{P}_{\frac{1}{2}}^n \cdot \mathbf{p}_0$$

où $\mathbf{P}_{\frac{1}{2}} = \mathbf{P}$ avec $p = \frac{1}{2}$.

La table (2.1) montre les valeurs de h et la taille d'échantillon n pour lesquelles les probabilités d'erreur de première espèce sont au plus α avec $\alpha = 0.05$ et $\alpha = 0.01$. Afin de s'assurer que les erreurs de première espèce sont au plus α pour les valeurs non tabulaires de n , la plus grande valeur de h doit être prise.

2.3.2 Puissance du test

On rappelle que la puissance du test est la probabilité de rejeter H_0 sous l'hypothèse H_1 .

Sous l'hypothèse H_1 , on distingue deux cas.

Si la probabilité que Y_i égale à 1 est constamment p pour tout $0 \leq i \leq n$, on considère que le changement s'est produit immédiatement et donc on peut calculer la probabilité du rejet de H_0 à l'aide de (2.1) et (2.4) en remplaçant r par la taille de l'échantillon n dans (2.4).

Si le changement se produit après la k ème observation la valeur de \mathbf{P}_n , utilisée pour calculer la probabilité de rejet de H_0 avec (2.1) est donné par la proposition :

Proposition 2.3.4. *Nous avons*

$$\mathbf{p}_n = \mathbf{P}^{n-k} \cdot \mathbf{P}_{\frac{1}{2}}^k \cdot \mathbf{p}_0 \quad (2.5)$$

Preuve. Comme le changement s'est produit après la k -ème observations alors la probabilité que $Y_i = 1$ pour les $n - k$ dernières observation ($i.e. i > k$) est égale à p et pour les k premières est égale à $\frac{1}{2}$. de (2.3) on a :

$$\mathbf{p}_n = \mathbf{P} \cdot \mathbf{p}_{n-1} = \mathbf{P}^2 \cdot \mathbf{p}_{n-2} = \dots = \mathbf{P}^{n-k} \cdot \mathbf{p}_{n-(n-k)} = \mathbf{P}^{n-k} \cdot \mathbf{p}_k.$$

à l'aide de (2.4) et comme la probabilité que $Y_i = 1$ pour les k premières observations est égale à $\frac{1}{2}$ alors on a

$$\mathbf{p}_k = \mathbf{P}_{\frac{1}{2}}^k \cdot \mathbf{p}_0,$$

D'où

$$\mathbf{p}_n = \mathbf{P}^{n-k} \cdot \mathbf{P}_{\frac{1}{2}}^k \cdot \mathbf{p}_0,$$

2.3.3 Simulation :

Dans la table 2.2, la puissance du **test 1** est comparée à la puissance du test binomial avec approximativement la même probabilité d'erreurs de type 1 pour un échantillon de taille 40.

D'après la table 2.1 Pour que le **test 1** ait une erreur de première espèce légèrement inférieures à 0.05, nous avons besoin de $h = 14$. On voit bien que la perte de puissance d'utiliser le **test 1** au lieu du test binomial est faible. Le test binomial correspondant est de rejeter l'hypothèse nulle H_0 si plus de 25 des Y_i sont positifs.

Remarque 1 :

L'hypothèse H_1 avec laquelle on a calculé la puissance du **test 1** est celle du cas 1.

Les deux mêmes tests sont mis en contraste dans la table (2.3), mais cette fois leurs puissances sont indiquées pour différentes positions du changement de $p = 0.5$ à $p = 0.75$ (et donc on est sous le cas 2 de l'hypothèse H_1).

On remarque bien que si le changement se produit vers le milieu de l'ensemble d'observations le **test 1** a une puissance sensiblement plus grande que le test binomial.

Également La table 2.3 indique les puissances du test binomial sur les $40 - k$ dernières observations, c'est à dire lorsque le changement s'est produit. Les différences entre la puissance de ce test et celui du **test 1** nous donne une indication sur la perte de puissance par ignorance de la position du changement.

TABLE 2.2 – Puissances des tests pour différents p .

p	Test 1	test binomial
0.50	0.044	0.040
0.55	0.132	0.133
0.60	0.304	0.317
0.65	0.543	0.572
0.70	0.778	0.807
0.75	0.929	0.945
0.80	0.987	0.992

Remarque 2 :

Dans le cas de la Table 2.3, la statistique test du test binomial sur tous l'échantillon $T =$ " le nombre des Y_i positives" sous l'hypothèse H_1 est définie comme suit :

$$T = T_0 + T_1,$$

avec

$T_0 =$ "le nombre des Y_i positives sur l'échantillon avec une probabilité de succès 0.50,
 $T_1 =$ "le nombre de Y_i positives sur l'échantillon avec une probabilité de succès 0.75.

Donc

$$T_0 \sim \mathcal{B}(k, 0.5) \quad \text{et} \quad T_1 \sim \mathcal{B}(n - k, 0.75).$$

La puissance est

$$\mathbb{P}_{H_1}(T_0 + T_1 \geq 26).$$

On calcule la loi de la somme par le produit de convolution :

$$\begin{aligned} \mathbb{P}_{H_1}(T_0 + T_1 = s) &= \sum_{i=0}^s \mathbb{P}_{H_1}(T_1 = i) P_{H_1}(T_0 = s - i) \\ &= \sum_{i=0}^s \binom{n-k}{i} (0.75)^i (0.25)^{n-k-i} \binom{k}{s-i} 0.5^k \end{aligned}$$

où

$$\binom{n}{k} = C_n^k = \frac{n!}{k!(n-k)!}$$

Table 2.3 Puissance du **test 1** pour différentes positions du changement

k	Test 1	Test binomial	Test d'binomial, k connu
0	0.929	0.945	0.945
10	0.814	0.762	0.894
20	0.509	0.438	0.461
25	0.308	0.282	0.618
30	0.157	0.163	0.244
40	0.044	0.040	0.040

2.3.4 Exemple de simulation du test

Soit un échantillon d'observations de taille 100 suivant une loi normale $N(4, 1)$ pour les 59 premières et de moyenne 5 pour les 41 dernières. La figure (2.2) montre les résultats de la simulation du test pour $\alpha = 0.05$ e donc par (2.1) $h = 22$.

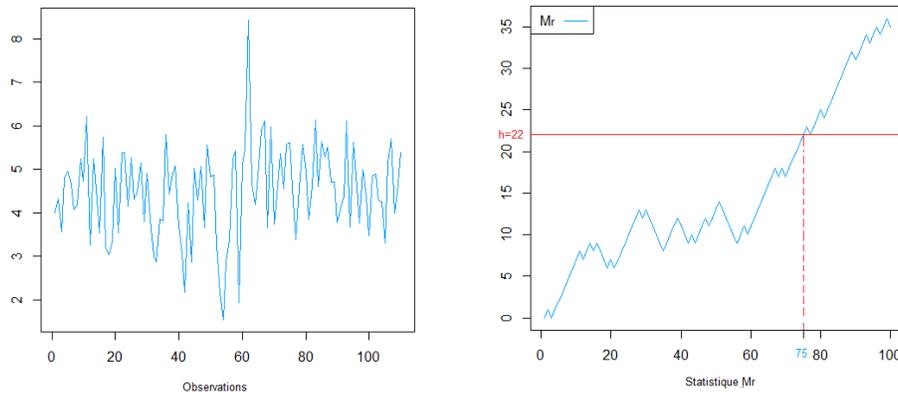


figure 2.2

on voit de la figure qu'on détecte un changement à l'instant 75.

Chapitre 3

Méthodes statistique de détection de ruptures

Dans ce chapitre, nous allons voir l'importance du logarithme du rapport de vraisemblance dans la construction du test statistique CUSUM très utilisé pour la détection des changements de paramètres.

3.1 Introduction :

Soit une suite d'observation $(X_n)_{n \geq 1}$ suivant une distribution de paramètre θ initialement donné par θ_0 . Le but des Méthodes statistiques de détection est de détecter le passage du paramètre θ de θ_0 à θ_1 ($\theta_1 \neq \theta_0$).

Comme on l'a déjà mentionné dans le contrôle de qualité ou bien dans d'autres domaines, le changement est le résultat d'un éventuel problème. Le paramètre θ est égal à θ_0 avant l'instant du changement que on va noter dans la suite t_p et après l'instant du changement θ prend la valeur θ_1 .

En général, les meilleur test pour un tel problème sont les test séquentiels où on détecte le changement à l'aide d'observation qui arrive en continu et donc on applique un test d'hypothèses à chaque nouvelle observation pour trancher entre deux hypothèses

$$H_0 : \theta = \theta_0 \quad \text{et} \quad H_1 : \theta = \theta_1.$$

Le test s'arrête si on rejette H_0 à un pas d'instant qu'on vas noter t_a , sinon le test continu.

Dans ce chapitre, nous allons présenter quelques méthodes de détection séquentielles.

3.2 La règle de Shewart et CUSUM

3.2.1 Le rapport de vraisemblance

Soit $(X_n)_{n \geq 1}$ un échantillon de variables aléatoires indépendantes de densité $f_\theta(\cdot)$ dépendant de θ qui vaut avant l'instant du problème t_p , la valeur $\theta = \theta_0$ et après $\theta = \theta_1$.

On suppose que θ_0 et θ_1 sont connus. On définit maintenant le logarithme du rapport de vraisemblance qui est l'origine d'une grande partie de règles de décision en statistique :

$$A(X_i) = \log \frac{f_{\theta_1}(X_i)}{f_{\theta_0}(X_i)}.$$

On note par $E_\theta(X_i)$ la moyenne de la variable aléatoire X sous la distribution f_θ .

Proposition 3.2.1. *Nous avons*

$$E_{\theta_0}(A(X)) < 0 \quad \text{et} \quad E_{\theta_1}(A(X)) > 0 \quad (3.1)$$

preuve.

La fonction log est strictement concave donc

$$\begin{aligned} E_{\theta_0}(A(X)) &< \log E_{\theta_0} \left(\frac{f_{\theta_1}(X)}{f_{\theta_0}(X)} \right) \\ &= \log \left(\int f_{\theta_0}(x) \frac{f_{\theta_1}(x)}{f_{\theta_0}(x)} dx \right) \\ &= \log \left(\int f_{\theta_1}(x) dx \right) \\ &= 0 \end{aligned}$$

Même chose pour la moyenne sous θ_1 , il suffit de montrer :

$$E_{\theta_1} \left[\log \frac{f_{\theta_0}(X_i)}{f_{\theta_1}(X_i)} \right] < 0.$$

Cela implique que

$$E_{\theta_1} \left[\log \frac{f_{\theta_1}(X_i)}{f_{\theta_0}(X_i)} \right] > 0.$$

et donc si le paramètre θ change le signe de la moyenne du logarithme de vraisemblance va changer.

Soit (X_1, X_2, \dots, X_n) un échantillon de variables aléatoires indépendantes de densité f_θ où θ est un paramètre. On veut tester l'hypothèse :

$$H_0 : \theta = \theta_0 \quad \text{contre} \quad H_1 : \theta = \theta_1$$

On définit le rapport de vraisemblance entre H_0 et H_1 par :

$$S_1^n = \sum_{i=1}^n A_i \quad A_i = \log \frac{f_{\theta_1}(X_i)}{f_{\theta_0}(X_i)}$$

D'après la propriété (3.1), les incréments de S_1^n sous H_0 sont en moyenne négatifs et donc la statistique S_1^n est en moyenne négative et décroissante tandis que sous H_1 elle est en moyenne positive croissante. D'où la règle de décision du rapport de vraisemblance entre les deux hypothèses H_0 et H_1 suivante :

$$\text{on rejette } H_0 \text{ si } S_1^n \geq h.$$

Notation : on note par S_j^k le rapport de vraisemblance sur l'échantillon $(X_j, X_{j+1}, \dots, X_n)$

3.2.2 La règle de Shewhart (1931) :

Cette méthode a été construite par Shewhart [15] pour détecter un changement en ligne son principe est le suivant.

Soit $(X_n)_{n \geq 1}$ une suite d'observations indépendantes. Comme c'est un test séquentiel, les observations arrivent en continu. On prend les N premières observations et donc on aura un échantillon X_1, \dots, X_N de taille N fixée.

Considérons les deux hypothèses : $H_0 : \theta = \theta_0$ $H_1 : \theta = \theta_1$

Si on rejette l'hypothèse H_0 , le test s'arrête pour rectifier le problème sinon on refait le test sur les N prochaines observations c'est à dire (X_{N+1}, \dots, X_{2N}) et ainsi de suite.

Pour les N premières observations, la règle de décision est :

$$\text{on rejette } H_0 \text{ si } S_1^N \geq h$$

où h le seuil fixé par l'expérimentateur. Le temps d'arrêt t_a est donc

$$t_a = N \cdot \min \{ K : S_{N(K-1)+1}^{NK} \geq h, K = 1, 2, \dots \}$$

..

3.2.3 La règle du CUSUM

La règle CUSUM est un algorithme qui a été utilisé en premier lieu pour détecter un changement de paramètre dans le cas où les observations sont indépendantes. Il a été généralisé ensuite au cas dépendant. Cette règle a été introduite par Page [14] en 1954.

Observations indépendantes

Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes de densité f_θ où θ est un paramètre avec :

$X_1, X_2, \dots, X_{t_p-1}$ sont de densité f_{θ_0} et $X_{t_p}, X_{t_p+1}, \dots$ sont de densité f_{θ_1} .

Si t_p tend vers l'infini, alors il n'y a pas de changement de paramètre. On note P_{θ_0} la loi de probabilité de toutes les variables $(X_n)_{n \geq 1}$ et E_{θ_0} l'espérance associée. Dans le cas contraire on les note $P^{(t_p)}$ et $E^{(t_p)}$.

Sous P_{θ_0} :

$$X_n \sim f_{\theta_0}(\cdot) \quad \forall n \geq 1,$$

et sous $P^{(t_p)}$:

$$\begin{aligned} X_n &\sim f_{\theta_0}(\cdot) \quad \forall n < t_p \\ X_n &\sim f_{\theta_1}(\cdot) \quad \forall n \geq t_p \end{aligned}$$

On suppose que θ_0 et θ_1 sont connus.

Construction du test : à chaque instant n , on test si il' y a changement de paramètre ou non à l'instant j tel que $1 \leq j \leq n$. Considérons donc les hypothèses :

H_0 : $\theta = \theta_0$ pour $i = 1, \dots, n$.

H_1 : $\theta = \theta_0$ pour $i < j$ et $\theta = \theta_1$ pour $i \geq j$.

Pour cela, on fait un rapport de vraisemblance entre ces deux hypothèses qui est définie comme suit :

$$\begin{aligned} \log \frac{f_{\theta_0, \theta_1, j}(X_1, \dots, X_n)}{f_{\theta_0}(X_1, \dots, X_n)} &= \log \frac{\prod_{i=1}^{j-1} f_{\theta_0}(X_i) \cdot \prod_{i=j}^n f_{\theta_1}(X_i)}{\prod_{i=1}^n f_{\theta_0}(X_i)} \\ &= \log \frac{\prod_{i=j}^n f_{\theta_1}(X_i)}{\prod_{i=j}^n f_{\theta_0}(X_i)} \\ &= \sum_{i=j}^n \log \frac{f_{\theta_1}(X_i)}{f_{\theta_0}(X_i)} \\ &= S_j^n. \end{aligned}$$

Ceci revient à faire le test de rapport de vraisemblance sur les observations $X_1 \dots X_n$ entre les hypothèses $\theta = \theta_0$ ou $\theta = \theta_1$. on fait ce rapport pour chaque j entre 1 et n , on obtient ainsi la statistique du test CUSUM.

Définition 3.2.1. *La statistique du test CUSUM est définie par*

$$g_n = \max_{1 \leq j \leq n} S_j^n = \max_{1 \leq j \leq n} \sum_{i=j}^n \log \frac{f_{\theta_1}(X_i)}{f_{\theta_0}(X_i)}. \quad (3.2)$$

D'après la propriété (3.1), la règle de décision est comme suit :

on rejette H_0 pour les n premières observations si $g_n \geq h$,

et le test s'arrête sinon le texte continu

Définition 3.2.2. *Le temps d'arrêt du test CUSUM est*

$$t_{CUSUM} = \inf \{n : g_n \geq h\}.$$

La figure (3.2) montre une simulation de la règle du CUSUM sur des observations suivant une loi normale $N(1, 1)$ avant l'instant du changement $t_p=115$ et après t_p la moyenne prend la valeur 2 avec un seuil $h=4.5$

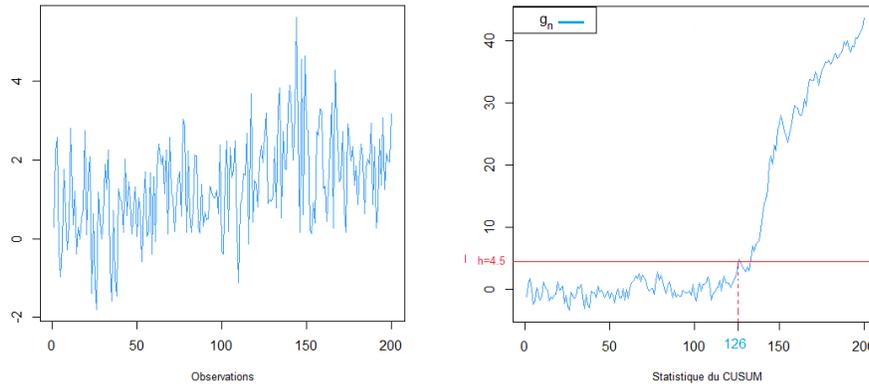


Figure 3.2 :Simulation du test du CUSUM sur un échantillons de lois $N(1,1)$, $h=4.5$, $t_p = 115$

on voit bien que d'après la figure que le temps de la détection $t_{cusum}=126$

Observations dépendantes

Jusqu'à maintenant, on a supposé que les variables sont indépendantes, néanmoins la plus part des systèmes qu'on surveille sont modélisés avec des variables dépendantes par exemple un modèle autorégressif. Il est nécessaire donc de trouver des méthodes pour détecter un changement de paramètre dans le cas dépendant.

L'idée consiste à considérer les densités conditionnelles.

sous P_{θ_0} :

$$X_n | X_1, \dots, X_{n-1} \sim f_{\theta_0}(\cdot | X_1, \dots, X_{n-1}) \quad \forall n \geq 1$$

$$\text{et sous } P^{(t_p)} : \begin{cases} X_n | X_1, \dots, X_{n-1} \sim f_{\theta_0}(\cdot | X_1, \dots, X_{n-1}) & \forall n < t_p \\ X_n | X_1, \dots, X_{n-1} \sim f_{\theta_1}(\cdot | X_1, \dots, X_{n-1}) & \forall n \geq t_p \end{cases}$$

La densité conjointe de tout l'échantillon avec $1 \leq t_p \leq n$ est :

$$f_{\theta_0, \theta_1, t_p}(X_1 \dots X_n) = f_{\theta_0}(X_1) \left[\prod_{i=2}^{t_p-1} f_{\theta_0}(X_i | X_1 \dots X_{i-1}) \right] \left[\prod_{i=t_p}^n f_{\theta_1}(X_i | X_1 \dots X_{i-1}) \right]$$

La généralisation de la règle du CUSUM aux données dépendantes, consiste simplement à calculer le logarithme du rapport de vraisemblance à l'aide des densités conditionnelles. Rappelons les hypothèses à tester :

$H_0 : \theta = \theta_0$ pour $i = 1, \dots, n$.

$H_1 : \theta = \theta_0$ pour $i < j$ et $\theta = \theta_1$ pour $i \geq j$.

Le rapport de vraisemblance entre ces deux hypothèses est :

$$\begin{aligned} \log \frac{f_{\theta_0, \theta_1, j}(X_1, \dots, X_n)}{f_{\theta_0}(X_1, \dots, X_n)} &= \log \frac{\prod_{i=j}^n f_{\theta_1}(X_i | X_1 \dots X_{i-1})}{\prod_{i=j}^n f_{\theta_0}(X_i | X_1 \dots X_{i-1})} \\ &= \sum_{i=j}^n \log \frac{f_{\theta_1}(X_i | X_1 \dots X_{i-1})}{f_{\theta_0}(X_i | X_1 \dots X_{i-1})} \end{aligned}$$

Définition 3.2.3. *La statistique du test CUSUM généralisé est*

$$g_n = \max_{1 \leq j \leq n} \sum_{i=j}^n K_i$$

avec

$$K_i = \log \frac{f_{\theta_1}(X_i | X_1 \dots X_{i-1})}{f_{\theta_0}(X_i | X_1 \dots X_{i-1})}.$$

Proposition 3.2.2. *Les incréments k_i calculés par les densités conditionnelles vérifient la propriété (3.1)*

Preuve : Soit j un instant de changement et $i \geq j$. On a

$$\begin{aligned}
& E_{\theta_0} \left[\log \frac{f_{\theta_1}(X_i | X_1 \dots X_{i-1})}{f_{\theta_0}(X_i | X_1 \dots X_{i-1})} \right] \\
& < \log \left[E_{\theta_0} \left(\frac{f_{\theta_1}(X_i | X_1 \dots X_{i-1} \dots X_{i-1})}{f_{\theta_0}(X_i | X_1 \dots X_{i-1})} \right) \right] \\
& < \log \left[E_{\theta_0} \left(\frac{f_{\theta_1}(X_i | X_1 \dots X_{i-1}) f_{\theta_0}(X_1 \dots X_{i-1})}{f_{\theta_0}(X_1 \dots X_i)} \right) \right] \\
& < \log \left(\int f_{\theta_0}(x_1 \dots x_i) \frac{f_{\theta_1}(x_i | x_1 \dots x_{i-1}) f_{\theta_0}(x_1 \dots x_{i-1})}{f_{\theta_0}(x_1 \dots x_i)} dx_1 \dots dx_i \right) \\
& < \log \left(\int \left(\int f_{\theta_1}(x_i | X_1 \dots X_{i-1}) dx_i \right) f_{\theta_0}(x_1 \dots x_{i-1}) dx_1 \dots dx_{i-1} \right) \\
& < 0.
\end{aligned}$$

Et enfin à l'aide de cette Proposition et avec le même raisonnement que pour le cas des données indépendantes (3.2), le temps d'arrêt de la règle du CUSUM généralisé est comme suit :

$$t_{CUSUM} = \inf \{n : g_n \geq h\}.$$

Simulation du test sur un AR(1) :

Soit $(X_n)_{n \geq 1}$ une suite d'observation suivant un modèle AR(1) avec :

$$X_n = 0.3X_{n-1} + \theta + \epsilon_n \quad \epsilon_n \sim N(0, 1)$$

Le changement de paramètre considéré est le suivant :

sous $H_0 : \theta = \theta_0 = 0$

sous $H_1 : \theta = \theta_1 = 0.5$

et donc les densités conditionnelles sont comme suit :

sous $H_0, X_n | X_1, \dots, X_{n-1}$ a pour densité

$$f_{\theta_0}(x_n | X_1, \dots, X_{n-1}) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x_n - 0.3X_{n-1})^2}{2}\right)$$

sous $H_1, X_n | X_1, \dots, X_{n-1}$ a pour densité

$$f_{\theta_1}(x_n | X_1 \dots X_{n-1}) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x_n - 0.3X_{n-1} - 0.5)^2}{2}\right)$$

La figure (3.3) montre les résultats de la simulation pour $h=3.5$. et $t_p = 120$

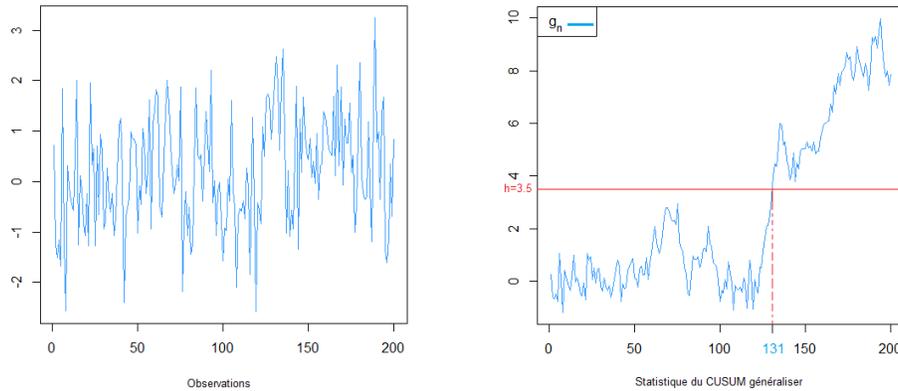


Figure 3.3

on voit bien de la figure qu'on a détecté un changement l'instant 131

Simulation du test sur un AR(2)

Soit $(X_n)_{n \geq 1}$ une suite d'observation suivant un modèle $AR(1)$ avec :

$$X_n = 0.3X_{n-1} + \theta X_{n-2} + \epsilon_n \quad \epsilon_n \sim N(0, 1)$$

Le changement de paramètre considéré est le suivant :

sous $H_0 : \theta = \theta_0 = 0.2$

sous $H_1 : \theta = \theta_1 = 0.6$

et donc les densités conditionnelles sont comme suit :

sous $H_0, X_n | X_1, \dots, X_{n-1}$ a pour densité

$$f_{\theta_0}(x_n | X_1, \dots, X_{n-1}) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x_n - 0.3X_{n-1} - 0.2X_{n-2})^2}{2}\right)$$

sous $H_1, X_n | X_1, \dots, X_{n-1}$ a pour densité

$$f_{\theta_1}(x_n | X_1 \dots X_{n-1}) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x_n - 0.3X_{n-1} - 0.6X_{n-2})^2}{2}\right)$$

Pour un seuil $h=3.5$. et $t_p = 120$ La figure (3.4) montre les résultat de la simulation .

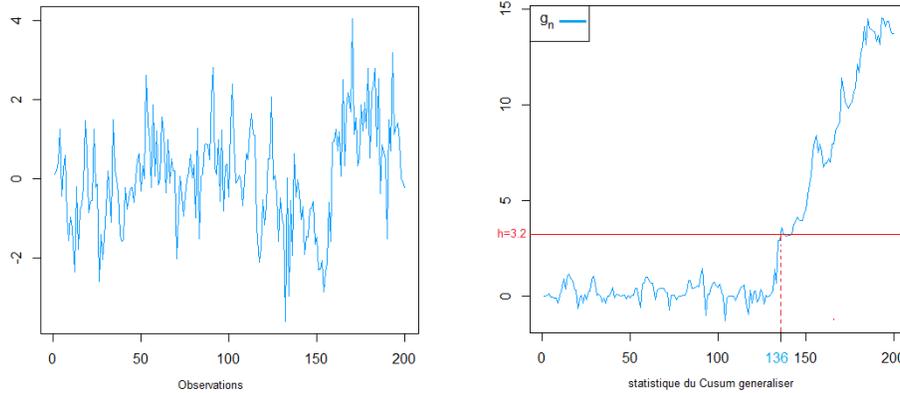


Figure 3.4

d'après la figure la détection a lieu à l'instant $t_a = 136$

Remarques générales : Le test CUSUM est séquentiel et donc la taille de l'échantillon est aléatoire. Généralement, le seuil h pour que l'erreur de première espèce ne dépasse pas un niveau de signification α , dépend à la fois de ce dernier et de la taille de l'échantillon. Donc on ne peut pas calculer le seuil par les méthodes classiques. Par contre, si le test est tronqué (c'est à dire on choisit à l'avance une taille n_0 qui ne doit pas être dépassée par le statisticien, alors l'approche couramment utilisée est basée sur la relation

$$\mathbb{P}(g_n \geq h) \leq \mathbb{P}\left(\max_{0 \leq n \leq n_0} g_n \geq h\right) \leq \alpha$$

Chapitre 4

Détection des changements des paramètres dans un modèle AR(p)

Dans ce chapitre, nous développons les résultats obtenus par E.Gombay et D.Serban dans [8]. Les auteurs s'intéressent à la détection des changements des paramètres d'un processus autorégressif d'ordre p et proposent des tests séquentiels. Les propriétés de ces tests reposent sur l'approximation des vecteurs scores associés aux échantillons en utilisant certains théorèmes limites notamment le principe d'invariance fort et la loi du logarithme itérée.

4.1 Introduction

Définition 4.1.1. (*Paramètre de nuisance*) :

Dans un test statistique, tout paramètre qui n'a pas un intérêt immédiat mais qu'on doit prendre en compte, est appelé paramètre de nuisance. Par exemple, si on veut tester la moyenne d'une loi normale alors la variance est un paramètre de nuisance.

Définition 4.1.2. (*Le vecteur de score*) :

Soit θ le vecteur des paramètres de l'échantillon $X=(X_1, X_2..X_n)$ et $L(X, \theta)$ la fonction de vraisemblance de l'échantillon.

Le vecteur de score est le gradient de la fonction log-vraisemblance c'est à dire

$$\nabla_{\theta} l(X, \theta),$$

où $l(X, \theta) = \log(L(X, \theta))$ est la fonction log-vraisemblance.

Soit $\{Y_i\}_{i \geq 1}$ un processus autorégressif d'ordre p défini par la relation :

$$Y_i - \mu = \phi_1 (Y_{i-1} - \mu) + \dots + \phi_p (Y_{i-p} - \mu) + \varepsilon_i, \quad i \geq 1 \quad (4.1)$$

où $\mu = E(Y_i)$, $\phi_i \in \mathbb{R}$, $i = 1, 2, \dots, p$ et $(\varepsilon_i)_{i \geq 1}$ est un bruit blanc fort.

Soit $\xi = (\mu, \sigma^2, \phi_1, \dots, \phi_p)^t$ le vecteur des paramètres.

On supposera tout au long de ce chapitre que :

$$- \text{le bruit blanc satisfait } \varepsilon_i \sim N(0, \sigma^2). \quad (C)$$

$$- \text{le processus } (Y_i)_{i > 0} \text{ est stationnaire.}$$

Nous considérons le problème de la détection d'un changement de paramètre à partir de la valeur initiale $\theta = \theta_0$ qui peut être n'importe quelle composante (ou ensemble de composantes) du vecteur des paramètres ξ . et où tous les autres paramètres sont inconnus, c'est-à-dire des paramètres de nuisance.

L'objectif est donc de tester les hypothèses suivantes :

$$H_0 : \theta = \theta_0 \text{ pour tous } Y_i, \quad i > 0.$$

$$H_1 : \theta = \theta_0 \text{ pour } Y_1, Y_2, \dots, Y_{t_p-1} \text{ et } \theta = \theta_1 \text{ pour tous } Y_i, i \geq t_p,$$

où $\theta_1 > \theta_0$. La méthode la plus connue pour le test séquentiel de la détection d'un paramètre est celle des sommes cumulatives (CUSUM) introduite par Page [13]

Soit (Y_1, Y_2, \dots, Y_n) un échantillon suivant le modèle (4.1) avec $1 \leq n \leq n_0$ où n_0 représente le point de troncature, c'est à dire la taille maximale de l'échantillon choisi par l'utilisateur (i.e Le test est séquentiel tronqué).

Si θ est le paramètre d -dimensionnel d'intérêt alors le vecteur de paramètre s'écrit $\xi = (\theta, \eta)$ où η représente le vecteur des paramètres de nuisance de dimension $p+2-d$ que nous estimons à l'instant n par $\hat{\eta}_n$ par l'estimateur du maximum de vraisemblance avec la contrainte $\theta = \theta_0$.

4.2 Construction du test

Définition 4.2.1. *Le vecteur score à l'instant n associé à l'échantillon Y_{j+1}, \dots, Y_n où $j+1$ avec $0 < j < n-1$ est supposé l'instant du changement est*

$$V_n(\xi) - V_j(\xi) = \nabla_{\xi} \sum_{i=1}^n \log f(Y_i; \xi) - \nabla_{\xi} \sum_{i=1}^j \log f(Y_i; \xi) = \nabla_{\xi} \sum_{i=j+1}^n \log f(Y_i; \xi).$$

avec f est la fonction de densité de ε_i

Soit $I(\xi)$ la matrice d'information de Fisher associée à une seule observation. On note $H_n(\theta_0, \hat{\eta}_n)$ le vecteur score $V_n(\theta_0, \hat{\eta}_n)$ normalisé par $I(\theta_0, \hat{\eta}_n)^{-\frac{1}{2}}$ c'est à dire

$$H_n(\theta_0, \hat{\eta}_n) = I(\theta_0, \hat{\eta}_n)^{-\frac{1}{2}} V_n(\theta_0, \hat{\eta}_n).$$

Pour $d = 1$, on note par $W_n(\theta_0, \hat{\eta}_n) - W_j(\theta_0, \hat{\eta}_n)$ la composante du vecteur $H_n - H_j$ issu de la dérivation de la fonction de log-vraisemblance par rapport à θ .

Règle de décision : La règle de décision du test de score est comme suit : On rejette l'hypothèse H_0 pour X_{j+1}, \dots, X_n si

$$W_n(\theta_0, \hat{\eta}_n) - W_j(\theta_0, \hat{\eta}_n) \geq h$$

avec $h > 0$. On refait ce test pour chaque $0 < j < n$.

D'où la règle de décision pour le test séquentiel de détection du changement :

On rejette l'hypothèse H_0 pour les n premières observations et on arrête le test si

$$\max_{0 < j < n} n_0^{-\frac{1}{2}} (W_n(\theta_0, \hat{\eta}_n) - W_j(\theta_0, \hat{\eta}_n)) \geq c(\alpha) \quad (4.2)$$

sinon le test continu ou bien le test s'arrête a l'instant n_0 et on accepte H_0 où α est le niveau de signification et $c(\alpha)$ est le seuil qui sera calculé à l'aide du Théorème 4.3.2 que nous allons annoncé dans la suite .

4.3 Résultats d'approximation :

Nous rappelons d'abord que $\varepsilon_i \sim N(0, \sigma^2)$.

Soit La représentation matricielle suivante de l'échantillon

$$Z_n = X_n \phi + \varepsilon$$

où

$$X_n = \begin{pmatrix} Y_0 - \mu & Y_{-1} - \mu & \dots & Y_{-p+1} - \mu \\ Y_1 - \mu & Y_0 - \mu & \dots & Y_{-p+2} - \mu \\ \vdots & \vdots & & \vdots \\ Y_{n-1} - \mu & Y_{n-2} - \mu & \dots & Y_{n-p} - \mu \end{pmatrix},$$

$Z_n = (Y_1 - \mu, \dots, Y_n - \mu)^t$, $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^t$ et $\phi = (\phi_1, \dots, \phi_p)^t$.
 et soit Γ la matrice de covariance du processus $(Y_1 - \mu, \dots, Y_p - \mu)^t$.
 Alors on a la proposition suivante.

Proposition 4.3.1. 1. Les estimateurs du maximum de vraisemblance du paramètre $\xi = (\mu, \sigma^2, \phi_1, \dots, \phi_p)^t$ sont

$$(a) \bar{\mu}_n = \frac{1}{n(1 - \sum_{j=1}^p \phi_j)} \sum_{i=1}^n \left(Y_i - \sum_{j=1}^p \phi_j Y_{i-j} \right)$$

$$(b) \hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n \left[(Y_i - \mu) - \sum_{j=1}^p \phi_j (Y_{i-j} - \mu) \right]^2$$

$$(c) \hat{\phi}_n = (\hat{\phi}_{n1}, \dots, \hat{\phi}_{np})^t = (X_n^t X_n)^{-1} X_n^t Z_n$$

2. La matrice de information de Fisher associée à une seule observation est :

$$\mathbf{I}(\mu, \sigma^2, \phi_1, \dots, \phi_p) = \begin{pmatrix} \frac{(1 - \phi_1 - \dots - \phi_p)^2}{\sigma^2} & 0 & 0 & \dots \\ 0 & \frac{1}{2\sigma^4} & 0 & \dots \\ 0 & 0 & \frac{1}{\sigma^2} \Gamma & \dots \end{pmatrix}$$

Preuve.

1. La fonction de vraisemblance associée à n observations est

$$\begin{aligned} L(\xi) &= \prod_{i=1}^n f(\varepsilon_i, \theta) \\ &= \left(\frac{1}{\sqrt{2\pi\sigma}} \right)^n e^{-\frac{1}{2\sigma} \sum_{i=1}^n (\varepsilon_i)^2} \\ &= \left(\frac{1}{\sqrt{2\pi\sigma}} \right)^n e^{-\frac{1}{2\sigma} \sum_{i=1}^n \left(Y_i - \mu - \sum_{j=1}^p \phi_j (Y_{i-j} - \mu) \right)^2}. \end{aligned}$$

Donc la fonction log vraisemblance est :

$$l_n(\mu, \sigma^2, \phi_1, \dots, \phi_p) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n \left[(Y_i - \mu) - \sum_{j=1}^p \phi_j (Y_{i-j} - \mu) \right]^2.$$

Par conséquent, pour $s=1, \dots, p$, les composantes du vecteur score sont :

$$\frac{\partial l_n}{\partial \mu} = \frac{(1 - \phi_1 - \dots - \phi_p)}{\sigma^2} \sum_{i=1}^n \left[(Y_i - \mu) - \sum_{j=1}^p \phi_j (Y_{i-j} - \mu) \right] \quad (1)$$

$$\frac{\partial l_n}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n \left[(Y_i - \mu) - \sum_{j=1}^p \phi_j (Y_{i-j} - \mu) \right]^2 \quad (2)$$

$$\frac{\partial l_n}{\partial \phi_s} = \frac{1}{\sigma^2} \sum_{i=1}^n (Y_{i-s} - \mu) \left[(Y_i - \mu) - \sum_{j=1}^p \phi_j (Y_{i-j} - \mu) \right] \quad (3)$$

(a) Par (1) on a $\frac{\partial l_n}{\partial \mu} = 0$ est équivalent à

$$\frac{(1 - \phi_1 - \dots - \phi_p)}{\sigma^2} \sum_{i=1}^n \left[(Y_i - \mu) - \sum_{j=1}^p \phi_j (Y_{i-j} - \mu) \right] = 0$$

Puisque le processus est supposé stationnaire alors

$$1 - \phi_1 - \dots - \phi_p \neq 0.$$

On obtient donc

$$\sum_{i=1}^n \left[(Y_i - \mu) - \sum_{j=1}^p \phi_j (Y_{i-j} - \mu) \right] = 0,$$

soit

$$\sum_{i=1}^n Y_i - n\mu - \sum_{i=1}^n \sum_{j=1}^p \phi_j Y_{i-j} + n\mu \sum_{j=1}^p \phi_j = 0.$$

D'où l'estimateur *e.m.v* de la moyenne μ est :

$$\bar{\mu}_n = \frac{1}{n \left(1 - \sum_{j=1}^p \phi_j \right)} \sum_{i=1}^n \left(Y_i - \sum_{j=1}^p \phi_j Y_{i-j} \right).$$

(b) Pour l'estimateur de la variance, en utilisant (2) $\frac{\partial l_n}{\partial \sigma^2} = 0$ donne

$$-\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n \left[(Y_i - \mu) - \sum_{j=1}^p \phi_j (Y_{i-j} - \mu) \right]^2 = 0.$$

Par suite

$$-n + \frac{1}{\sigma^2} \sum_{i=1}^n \left[(Y_i - \mu) - \sum_{j=1}^p \phi_j (Y_{i-j} - \mu) \right]^2 = 0.$$

On obtient l'estimateur *e.m.v* de σ^2 :

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n \left[(Y_i - \mu) - \sum_{j=1}^p \phi_j (Y_{i-j} - \mu) \right]^2$$

(c) La représentation matricielle de l'échantillon est :

$$Z_n = X_n \phi + \varepsilon$$

On peut écrire

$$l_n(\mu, \sigma^2, \phi_1, \dots, \phi_p) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \|Z_n - X_n \phi\|^2$$

et donc maximiser $l_n(\mu, \sigma^2, \phi_1, \dots, \phi_p)$ revient à minimiser

$$S(\phi) = \|Z_n - X_n \phi\|^2.$$

On a

$$\begin{aligned} S(\phi) &= \|Z_n - X_n \phi\|^2 \\ &= (Z_n - X_n \phi)^t (Z_n - X_n \phi) \\ &= Z_n^t Z_n - (X_n^t Z_n)^t \phi - \phi^t X_n^t Z_n + \phi^t X_n^t X_n \phi. \end{aligned}$$

Par une dérivation matricielles on obtient

$$\begin{aligned} \frac{\partial S(\phi)}{\partial \phi} &= -X_n^t Z_n - X_n^t Z_n + (X_n^t X_n + X_n^t X_n) \phi \\ &= -2X_n^t Z_n + 2X_n^t X_n \phi. \end{aligned}$$

L'estimateur *e.m.v* du vecteur ϕ_n est donc

$$\hat{\phi}_n = \left(\hat{\phi}_{n1}, \dots, \hat{\phi}_{np} \right)^t = (X_n^t X_n)^{-1} X_n^t Z_n$$

2. La matrice d'information de Fisher est $I(\xi) = (I_{i,j})_{\substack{1 \leq i \leq p+2 \\ 1 \leq j \leq p+2}}$ avec

$$I_{i,j} = -E \left[\left(\frac{\partial^2}{\partial \xi_i \partial \xi_j} \log f(\varepsilon_i, \xi) \right) \right].$$

On obtient

$$\mathbf{I}(\mu, \sigma^2, \phi_1, \dots, \phi_p) = \begin{pmatrix} \frac{(1 - \phi_1 - \dots - \phi_p)^2}{\sigma^2} & 0 & 0 & \dots \\ 0 & \frac{1}{2\sigma^4} & 0 & \dots \\ 0 & 0 & \frac{1}{\sigma^2} \mathbf{\Gamma} & \dots \end{pmatrix}$$

Dans le cas de la moyenne, on note l'estimateur du maximum de vraisemblance par $\bar{\mu}_n$ et on va noter par $\hat{\mu}$ la moyenne arithmétique :

$$\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^n Y_i$$

Remarques :

- (a) Dans la statistique test l'estimateur $\bar{\mu}$ du maximum de vraisemblance de la moyenne sera remplacée par la moyenne arithmétique $\hat{\mu}_n$ avec une erreur négligeable. La preuve sera donnée dans le Théorème 4.3.1
- (b) Dans une première étape, nous considérons le teste du changement dans la moyenne du processus et donc on suppose que la variances et les coefficients d'autorégression ϕ_i sont inconnus

Par (1) on a

$$W_n(\mu_0, \hat{\sigma}_n^2, \hat{\phi}_n) = (I_{1,1})^{-\frac{1}{2}} \frac{\partial l_n(\mu_0, \hat{\sigma}_n^2, \hat{\phi}_n)}{\partial \mu}$$

par suite

$$W_n(\mu_0, \hat{\sigma}_n^2, \hat{\phi}_n) = \frac{1}{\hat{\sigma}_n} \sum_{i=1}^n \left[(Y_i - \mu_0) - \sum_{j=1}^p \hat{\phi}_{nj} (Y_{i-j} - \mu_0) \right] \quad (4.3)$$

Si on veut tester le changement de la variance par rapport à la valeur initiale σ_0 alors on suppose que la moyenne et les coefficients d'autorégression sont inconnus.

Par (2) et en prenant $2^{-1/2}\sigma_0^{-2}$ en facteur :

$$W_n \left(\sigma_0^2, \hat{\mu}_n, \hat{\phi}_n \right) = (I_{2,2})^{-\frac{1}{2}} \frac{\partial l_n(\sigma_0^2, \hat{\mu}_n, \hat{\phi}_n)}{\partial \sigma^2}$$

et donc

$$W_n \left(\sigma_0^2, \hat{\mu}_n, \hat{\phi}_n \right) = 2^{-1/2} \sigma_0^{-2} \sum_{i=1}^n \left\{ \left[(Y_i - \hat{\mu}_n) - \sum_{j=1}^p \hat{\phi}_{nj} (Y_{i-j} - \hat{\mu}_n) \right]^2 - \sigma_0^2 \right\} \quad (4.4)$$

Enfin, si on veut tester le changement du vecteur des paramètres $\phi = (\phi_1, \dots, \phi_p)^t$, alors la moyenne et la variance seront supposés inconnus.

En utilisant (3) :

$$W_n \left(\phi_0, \hat{\mu}_n, \hat{\sigma}_n^2 \right) = \frac{1}{\hat{\sigma}_n} \mathbf{\Gamma}^{-1/2} \left(\phi_0, \hat{\mu}_n, \hat{\sigma}_n^2 \right) \nabla_{\phi} l \left(\phi_0, \hat{\mu}_n, \hat{\sigma}_n^2 \right) \quad (4.5)$$

Et comme les théorèmes sont basés sur des grands échantillons d'approximation, on supposera que n est assez grand

Théorème 4.3.1. *Supposons qu'aucun des paramètres ne change.*

i) *Sous les hypothèses $\phi = \phi_0$ ou $\sigma^2 = \sigma_0^2$ et la condition (C)*

$$|\hat{\mu}_n - \mu| = O \left((n^{-1} \log \log n)^{1/2} \right) \quad p.s.$$

ii) *Sous les hypothèses $\mu = \mu_0$ ou $\sigma^2 = \sigma_0^2$ et la condition (C)*

$$\left\| \hat{\phi}_n - \phi \right\| = O \left((n^{-1} \log \log n)^{1/2} \right) \quad p.s.$$

iii) *Sous les hypothèses $\mu = \mu_0$ ou $\phi = \phi_0$ et la condition (C)*

$$|\hat{\sigma}_n^2 - \sigma^2| = O \left((n^{-1} \log \log n)^{1/2} \right) \quad p.s.$$

Preuve :

(i). On a

$$\begin{aligned}
\bar{\mu}_n &= \frac{1}{n \left(1 - \sum_{j=1}^p \phi_j\right)} \sum_{i=1}^n \left(Y_i - \sum_{j=1}^p \phi_j Y_{i-j} \right) \\
&= \frac{1}{n \left(1 - \sum_{j=1}^p \phi_j\right)} \sum_{i=1}^n \left((Y_i - \mu + \mu) - \sum_{j=1}^p \phi_j (Y_{i-j} - \mu + \mu) \right) \\
&= \frac{1}{n \left(1 - \sum_{j=1}^p \phi_j\right)} \sum_{i=1}^n \left((Y_i - \mu) - \sum_{j=1}^p \phi_j (Y_{i-j} - \mu) + \mu \left(1 - \sum_{j=1}^p \phi_j\right) \right) \\
&= \mu + \frac{1}{n \left(1 - \sum_{j=1}^p \phi_j\right)} \sum_{i=1}^n \varepsilon_i
\end{aligned}$$

et donc

$$|\bar{\mu}_n - \mu| = \left| \frac{1}{\left(1 - \sum_{j=1}^p \phi_j\right)} \right| \left| \frac{\sum_{i=1}^n \varepsilon_i}{n} \right|.$$

La suite $(\varepsilon_i)_i$ étant *i.i.d.*, en appliquant la loi du logarithme itéré (Théorème 1.3.1), on obtient

$$|\bar{\mu}_n - \mu| \leq \left| \frac{1}{\left(1 - \sum_{j=1}^p \phi_j\right)} \right| \cdot O(n^{-1} \log \log n)^{\frac{1}{2}} \quad p.s$$

par suit :

$$|\bar{\mu}_n - \mu| = O\left((n^{-1} \log \log n)^{1/2}\right) \quad p.s$$

D'autre part par le Théorème 1.4.3 il existe un processus de Wiener $\{B(t) : t \geq 0\}$, $v > 2$ et une constante $\sigma_W > 0$, telles que :

$$\left| \sum_i^n (Y_i - \mu) - \sigma_W B(n) \right| = O\left(t^{\frac{1}{v}}\right) \quad p.s.$$

Ainsi

$$\left| \sum_{i=1}^n (Y_i - \mu) \right| = \left| \sum_{i=1}^n (Y_i - \mu) - \sigma_W B(n) + \sigma_W B(n) \right|$$

$$\leq \left| \sum_{i=1}^n (Y_i - \mu) - \sigma_W B(n) \right| + |\sigma_W B(n)|$$

Par la loi du logarithme itéré pour le mouvement brownien (Théorème 1.4.1), on obtient que

$$\left| \sum_{i=1}^n (Y_i - \mu) \right| = O \left((n \log \log n)^{1/2} \right) \quad p.s. \quad (4.6)$$

D'où

$$|\hat{\mu}_n - \mu| = O \left((n^{-1} \log \log n)^{1/2} \right) \quad p.s.$$

Et

$$|\hat{\mu}_n - \bar{\mu}_n| = O \left((n^{-1} \log \log n)^{1/2} \right) \quad p.s$$

(ii).

1) **Si μ est connu** : $Z_n = X_n \phi + \varepsilon$ donne

$$\phi = (X_n^t X_n)^{-1} X_n^t (Z_n - \varepsilon),$$

ceci implique que

$$\hat{\phi}_n = (X_n^t X_n)^{-1} X_n^t Z_n,$$

soit

$$\hat{\phi}_n - \phi = (X_n^t X_n)^{-1} X_n^t \varepsilon.$$

Posons $U = X_n^t \varepsilon$,

$$U = \begin{pmatrix} \sum_{i=1}^n (Y_{i-1} - \mu) \varepsilon_i \\ \sum_{i=1}^n (Y_{i-2} - \mu) \varepsilon_i \\ \dots \\ \sum_{i=1}^n (Y_{i-p} - \mu) \varepsilon_i \end{pmatrix}$$

et soit $U_s, s = 1, 2, \dots, p$ la s -ème composante du vecteur U . On a

$$\begin{aligned} \text{Var}(U_s) &= E \left[\left(\sum_{i=1}^n (Y_{i-s} - \mu) \varepsilon_i \right)^2 \right] \\ &= \sum_{i=1}^n E[(Y_{i-s} - \mu)^2 \varepsilon_i^2] + 2 \sum_{i < j}^n E[(Y_{i-s} - \mu)(Y_{j-s} - \mu) \varepsilon_i \varepsilon_j] \end{aligned}$$

En utilisant le fait que les ε_i et Y_{s-i} sont indépendantes pour $i > 0$, et que le processus est stationnaire $(Y_i)_{i \geq 1}$ on tire que

$$\text{Var}(U_s) = \sum_{i=1}^n E[(Y_{i-s} - \mu)^2] E(\varepsilon_i^2) = n \sigma^2 \gamma(0),$$

où γ est la fonction de covariance de $(Y_i)_{i \geq 1}$.

D'autre part, soit $s, k = \{1, \dots, p\}$ avec $s < k$. En utilisant encore une fois le fait que la variable ε_i est indépendante de Y_{i-s} et de Y_{i-k} pour $s, k > 0$ on peut écrire

$$\begin{aligned} \text{Cov}(U_s, U_k) &= \text{Cov}\left(\sum_{i=1}^n (Y_{i-s} - \mu)\varepsilon_i, \sum_{i=1}^n (Y_{i-k} - \mu)\varepsilon_i\right) \\ &= \sum_{i=1}^n \sum_{j=1}^n \text{Cov}((Y_{i-s} - \mu)\varepsilon_i, (Y_{j-k} - \mu)\varepsilon_j) \\ &= \sum_{i=1}^n \sum_{j=1}^n E[(Y_{i-s} - \mu)(Y_{j-k} - \mu)\varepsilon_i\varepsilon_j] \\ &= \sum_{i=1}^n E[(Y_{i-s} - \mu)(Y_{i-k} - \mu)\varepsilon_i^2] \\ &= n \sigma^2 \gamma(k - s). \end{aligned}$$

Par suite, si on note par $\mathbf{\Gamma}_2$ la matrice de covariance du vecteur U , alors

$$\mathbf{\Gamma}_2 = n \sigma^2 \mathbf{\Gamma}.$$

La matrice de covariance $C_{\mathbf{\Gamma}_2^{-\frac{1}{2}}U}$ du vecteur $\mathbf{\Gamma}_2^{-\frac{1}{2}}U$ vérifie

$$C_{\mathbf{\Gamma}_2^{-\frac{1}{2}}U} = n^{-1} \sigma^{-2} (\mathbf{\Gamma}^t)^{-\frac{1}{2}} \mathbf{\Gamma} \mathbf{\Gamma}^{-\frac{1}{2}} = n^{-1} \sigma^{-2} I_d$$

où I_d est la matrice identité de \mathbb{R}^d . Ceci assure que les composantes du vecteur $\mathbf{\Gamma}_2^{-\frac{1}{2}}U$ sont non corrélées. Le principe d'invariance [6] donne

$$\left\| \mathbf{\Gamma}_2^{-\frac{1}{2}}U \right\| = O\left((\log \log n)^{1/2}\right) \quad (4.7)$$

D'autre part on a

$$\begin{aligned}
\|\hat{\phi}_n - \phi\| &= \left\| (X_n^t X_n)^{-1} X_n^t \varepsilon \right\| \\
&= \left\| n^{-1/2} \left(\frac{1}{n} X_n^t X_n \right)^{-1} \sigma \mathbf{\Gamma}^{1/2} (n^{-1/2} \sigma^{-1} \mathbf{\Gamma}^{-1/2} X_n^t \varepsilon) \right\| \\
&= \left\| n^{-1/2} \left(\frac{1}{n} X_n^t X_n \right)^{-1} \sigma \mathbf{\Gamma}^{1/2} (\mathbf{\Gamma}_2^{-\frac{1}{2}} U) \right\|
\end{aligned}$$

Par (4.7) et en utilisant et le fait que $\frac{1}{n} X_n^t X_n \xrightarrow{p.s.} \mathbf{\Gamma}$, on obtient :

$$\|\hat{\phi}_n - \phi\| = O\left((n^{-1} \log \log n)^{1/2}\right) \quad p.s.$$

2) Si μ est inconnu : On va noter la statistique $T_{|\mu=\hat{\mu}_n}$ par \tilde{T} . On peut écrire

$$\begin{aligned}
\|\hat{\phi}_n - \phi\| &= \left\| (\tilde{X}_n^t \tilde{X}_n)^{-1} \tilde{X}_n^t \varepsilon \right\| \\
&= \left\| (\tilde{X}_n^t \tilde{X}_n)^{-1} (\tilde{X}_n^t \varepsilon - X_n^t \varepsilon + X_n^t \varepsilon) \right\|.
\end{aligned}$$

Or

$$\left\| \tilde{X}_n^t \varepsilon - X_n^t \varepsilon \right\| = \left| (\hat{\mu}_k - \mu) \sum_{i=1}^k \varepsilon_i \right| \|\mathbf{1}\|.$$

avec $\mathbf{1} = (1, 1, \dots, 1)^t$ un vecteur de dimension p . par (i) et la loi du logarithme itéré pour notre bruit blanc (Théorème 1.3.1)

$$\left| (\hat{\mu}_n - \mu) \sum_{i=1}^k \varepsilon_i \right| = O(\log \log n) \quad p.s$$

d'où

$$\|\tilde{X}_n^t \varepsilon - X_n^t \varepsilon\| = O(\log \log n) \quad p.s \quad (4.8)$$

de plus on a

$$\frac{1}{n} \tilde{X}_n^t \tilde{X}_n = \frac{1}{n} \left(\sum_{i=1}^n (Y_{i-j} - \hat{\mu}_n)(Y_{i-s} - \hat{\mu}_n) \right)_{\substack{1 \leq j \leq p \\ 1 \leq s \leq p}}$$

et on a pour tout $1 \leq j, s \leq p$:

$$\begin{aligned}
\sum_{i=1}^n (Y_{i-j} - \hat{\mu}_n)(Y_{i-s} - \hat{\mu}_n) &= \sum_{i=1}^n (Y_{i-j} - \mu + \mu - \hat{\mu}_n)(Y_{i-s} - \mu + \mu - \hat{\mu}_n) \\
&= \sum_{i=1}^n (Y_{i-j} - \mu)(Y_{i-s} - \mu) + (\mu - \hat{\mu}_n) \sum_{i=1}^n (Y_{i-j} - \mu) \\
&\quad + (\mu - \hat{\mu}_n) \sum_{i=1}^n (Y_{i-s} - \mu) + (\mu - \hat{\mu}_n)^2
\end{aligned}$$

Donc

$$\begin{aligned}
\sum_{i=1}^n (Y_{i-j} - \hat{\mu}_n)(Y_{i-s} - \hat{\mu}_n) - \sum_{i=1}^n (Y_{i-j} - \mu)(Y_{i-s} - \mu) &= (\mu - \hat{\mu}_n) \sum_{i=1}^n (Y_{i-j} - \mu) \\
&\quad + (\mu - \hat{\mu}_n) \sum_{i=1}^n (Y_{i-s} - \mu) + (\mu - \hat{\mu}_n)^2
\end{aligned}$$

par (i), on obtient

$$\left| \sum_{i=1}^n (Y_{i-j} - \hat{\mu}_n)(Y_{i-s} - \hat{\mu}_n) - \sum_{i=1}^n (Y_{i-j} - \mu)(Y_{i-s} - \mu) \right| = O(\log \log n) + O(n^{-1} \log \log n) \quad p.s.$$

Ainsi pour tout $1 \leq j, s \leq p$

$$\left| \sum_{i=1}^n (Y_{i-j} - \hat{\mu}_n)(Y_{i-s} - \hat{\mu}_n) - \sum_{i=1}^n (Y_{i-j} - \mu)(Y_{i-s} - \mu) \right| = O(\log \log n) \quad p.s.$$

Par conséquent

$$\left\| \frac{1}{n} \tilde{X}_n^t \tilde{X}_n - \frac{1}{n} X_n^t X_n \right\| = O(n^{-1} \log \log n) \quad p.s$$

Par suit

$$\begin{aligned}
\left\| \frac{1}{n} \tilde{X}_n^t \tilde{X}_n - \Gamma \right\| &= \left\| \frac{1}{n} \tilde{X}_n^t \tilde{X}_n - \frac{1}{n} X_n^t X_n + \frac{1}{n} X_n^t X_n - \Gamma \right\| \\
&\leq \left\| \frac{1}{n} \tilde{X}_n^t \tilde{X}_n - \frac{1}{n} X_n^t X_n \right\| + \left\| \frac{1}{n} X_n^t X_n - \Gamma \right\|
\end{aligned}$$

comme $\frac{1}{n}X_n^t X_n \xrightarrow{p.s} \mathbf{\Gamma}$ et $\left\| \frac{1}{n}\tilde{X}_n^t \tilde{X}_n - \frac{1}{n}X_n^t X_n \right\| = O(n^{-1} \log \log n)$ *p.s* alors :

$$\left\| \frac{1}{n}\tilde{X}_n^t \tilde{X}_n - \mathbf{\Gamma} \right\| \xrightarrow{p.s} 0,$$

ce qui montre que $\frac{1}{n}\tilde{X}_n^t \tilde{X}_n \xrightarrow{p.s} \mathbf{\Gamma}$. D'autre part on a

$$\begin{aligned} \|\hat{\phi}_n - \phi\| &= \left\| \left(\tilde{X}_n^t \tilde{X}_n \right)^{-1} (\tilde{X}_n^t \varepsilon - X_n^t \varepsilon + X_n^t \varepsilon) \right\| \\ &\leq \left\| n^{-1} \left(\frac{1}{n} \tilde{X}_n^t \tilde{X}_n \right)^{-1} (\tilde{X}_n^t \varepsilon - X_n^t \varepsilon) \right\| + \left\| n^{-1} \left(\frac{1}{n} \tilde{X}_n^t \tilde{X}_n \right)^{-1} (X_n^t \varepsilon) \right\|. \end{aligned}$$

sachant que $\frac{1}{n}\tilde{X}_n^t \tilde{X}_n \xrightarrow{p.s} \mathbf{\Gamma}$ et par (4.8), on conclu alors que :

$$\|\hat{\phi}_n - \phi\| = O\left((n^{-1} \log \log n)^{1/2}\right) \quad p.s.$$

(iii)

Si μ et inconnu et ϕ connu :

par (3.b)

$$\begin{aligned} \hat{\sigma}_n^2 &= \frac{1}{n} \sum_{i=1}^n \left[(Y_i - \hat{\mu}_n) - \sum_{j=1}^p \phi_j (Y_{i-j} - \hat{\mu}_\mu) \right]^2 \\ &= \frac{1}{n} \sum_{i=1}^n \left[(Y_i - \mu + \mu - \hat{\mu}_n) - \sum_{j=1}^p \phi_j (Y_{i-j} - \mu + \mu - \hat{\mu}_n) \right]^2 \\ &= \frac{1}{n} \sum_{i=1}^n \varepsilon_i^2 + (\mu - \hat{\mu}_n)^2 \left(1 - \sum_{j=1}^p \phi_j \right)^2 + 2(\mu - \hat{\mu}_n) \left(1 - \sum_{j=1}^p \phi_j \right) \frac{1}{n} \sum_{i=1}^n \varepsilon_i \end{aligned}$$

comme notre bruit blanc est gaussien alors $\text{var}(\varepsilon_i^2)$ existe et donc par la loi du logarithme itéré du Théorème 1.3.1 on obtient facilement que :

$$\left| \sigma^2 - \frac{1}{n} \sum_{i=1}^n \varepsilon_i^2 \right| = O\left((n^{-1} \log \log n)^{1/2}\right) \quad p.s \quad (4.9)$$

par (i) :

$$(\mu - \hat{\mu}_n)^2 \left(1 - \sum_{j=1}^p \phi_j \right)^2 = O(n^{-1} \log \log n) \quad p.s$$

et en utilisant la loi du logarithme itéré et (i) :

$$2(\mu - \hat{\mu}_n) \left(1 - \sum_{j=1}^p \phi_j\right) \frac{1}{n} \sum_{i=1}^n \varepsilon_i = O(n^{-1} \log \log n) \quad p.s$$

et donc finalement :

$$|\hat{\sigma}_n^2 - \sigma^2| = O\left((n^{-1} \log \log n)^{1/2}\right) \quad p.s$$

si μ est connu et ϕ est inconnu :

$$\begin{aligned} \hat{\sigma}_n^2 &= \frac{1}{n} \sum_{i=1}^n \left[(Y_i - \mu) - \sum_{j=1}^p \hat{\phi}_j (Y_{i-j} - \mu) \right]^2 \\ &= \frac{1}{n} \sum_{i=1}^n \varepsilon_i^2 - 2 \frac{1}{n} \sum_{i=1}^n \varepsilon_i \left[\sum_{j=1}^p (\hat{\phi}_{nj} - \phi_j) (Y_{i-j} - \mu) \right] + \frac{1}{n} \sum_{i=1}^n \left[\sum_{j=1}^p (\hat{\phi}_{nj} - \phi_j) (Y_{i-j} - \mu) \right]^2 \end{aligned}$$

le premier terme de la somme est régler par (4.9) le deuxième terme :

$$\left(\hat{\phi}_{n1} - \phi_1\right) \frac{1}{n} \sum_{i=1}^n \varepsilon_i (Y_{i-1} - \mu) + \cdots + \left(\hat{\phi}_{np} - \phi_p\right) \frac{1}{n} \sum_{i=1}^n \varepsilon_i (Y_{i-p} - \mu)$$

par (ii) :

$$\left(\hat{\phi}_{ns} - \phi_s\right) = O\left((n^{-1} \log \log n)^{1/2}\right) \quad p.s$$

et par le Théorème 1.4.3 et la loi du logarithme itéré pour le mouvement brownien pour le processus $\sum_{i=1}^n \varepsilon_i (Y_{i-s} - \mu)$ $s=\{1,2,..p\}$ (par le même raisonnement que (4.6)) nous donne :

$$\frac{1}{n} \sum_{i=1}^n \varepsilon_i (Y_{i-s} - \mu) = O\left((n^{-1} \log \log n)^{1/2}\right) \quad p.s$$

et donc le deuxième terme et d'ordre $O(n^{-1} \log \log n)$ p.s

pour le 3 ème terme on utilise (i) et le fait que $\frac{1}{n} X_n^t X_n \xrightarrow{p.s} \Gamma$

$$\frac{1}{n} \sum_{i=1}^n \left[\sum_{j=1}^p (\hat{\phi}_{nj} - \phi_j) (Y_{i-j} - \mu) \right]^2$$

$$= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p \left(\hat{\phi}_{nj} - \phi_j \right)^2 (Y_{i-j} - \mu)^2 + 2 \frac{1}{n} \sum_{i=1}^n \sum_{s < j}^p \left(\hat{\phi}_{nj} - \phi_j \right) \left(\hat{\phi}_{ns} - \phi_s \right) (Y_{i-j} - \mu) (Y_{i-s} - \mu)$$

le premier terme :

$$\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p \left(\hat{\phi}_{nj} - \phi_j \right)^2 (Y_{i-j} - \mu)^2 = \sum_{j=1}^p \left(\hat{\phi}_{nj} - \phi_j \right)^2 \frac{1}{n} \sum_{i=1}^n (Y_{i-j} - \mu)^2$$

par (ii) et $\frac{1}{n} \sum_{i=1}^n (Y_{i-j} - \mu)^2 \xrightarrow{p.s.} \gamma(0)$ on obtient :

$$\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p \left(\hat{\phi}_{nj} - \phi_j \right)^2 (Y_{i-j} - \mu)^2 = O(n^{-1} \log \log n) \quad p.s$$

même chose pour le 2 ème terme par (ii) et sachant que

$$\frac{1}{n} \sum_{i=1}^n (Y_{i-j} - \mu) (Y_{i-s} - \mu) \xrightarrow{p.s.} \gamma(j-s)$$

Alors :

$$\frac{1}{n} \sum_{i=1}^n \sum_{s < j}^p \left(\hat{\phi}_{nj} - \phi_j \right) \left(\hat{\phi}_{ns} - \phi_s \right) (Y_{i-j} - \mu) (Y_{i-s} - \mu) = O(n^{-1} \log \log n)$$

et donc finalement :

$$|\hat{\sigma}_n^2 - \sigma^2| = O\left((n^{-1} \log \log n)^{1/2}\right) \quad p.s$$

Théorème 4.3.2. :

(i) Sous l'hypothèse $H_0 : \mu = \mu_0$, où σ^2, ϕ supposé inconnu, et sous la condition (C), il existe un mouvement brownien standard $B(\cdot)$, tel que :

$$\left| W_n \left(\mu_0, \hat{\sigma}_n^2, \hat{\phi}_n \right) - B(n) \right| \stackrel{p.s.}{=} o(n^{1/v})$$

pour un $v > 2$.

(ii) sous l'hypothèse $H_0 : \phi = \phi_0$, où μ, σ^2 supposé inconnu, et sous la condition (C), il existe un processus $B^{(\cdot)} = (B_1, \dots, B_p)$, avec des composantes de mouvement brownien standard indépendantes, telles que :

$$\| W_n \left(\phi_0, \hat{\mu}_n, \hat{\sigma}_k^2 \right) - B(n) \| \stackrel{p.s.}{=} o(n^{1/v})$$

pour un $v > 2$.

(iii) Sous l'hypothèse $H_0 : \sigma^2 = \sigma_0^2$, où μ, ϕ sont supposés inconnu, et sous la condition (C), il existe un mouvement brownien standard $B(\cdot)$, tel que :

$$\left| W_n \left(\sigma_0^2, \hat{\mu}_n, \hat{\phi}_n \right) - B(n) \right| \stackrel{p.s.}{=} o(n^{1/v})$$

Ce théorème a été démontré dans [8] à l'aide du Théorème 4.3.1 et le Théorème (1.4.2).

calcul du seuil $c(\alpha)$:

On rappelle que $n \leq n_0$ le point de troncature, d'où on a

$$\max_{0 < j < n} n_0^{-\frac{1}{2}} (W_n(\theta_0, \hat{\eta}_n) - W_j(\theta_0, \hat{\eta}_n)) \leq \max_{0 < j < n \leq n_0} n_0^{-\frac{1}{2}} (W_n(\theta_0, \hat{\eta}_n) - W_j(\theta_0, \hat{\eta}_n)).$$

On peut calculer $c(\alpha)$ comme suit :

Le théorème 4.3.2 donne :

$$\max_{0 < j < n \leq n_0} n_0^{-\frac{1}{2}} (W_n(\theta_0, \hat{\eta}_n) - W_j(\theta_0, \hat{\eta}_n)) = \max_{0 < j < n \leq n_0} n_0^{-\frac{1}{2}} (B(n) - B(j) + o(n^{1/v}) + o(j^{1/v}))$$

pour n_0 assez grand on a :

$$\max_{0 < j < n \leq n_0} n_0^{-\frac{1}{2}} (B(n) - B(j)) \stackrel{\text{lois}}{=} \max_{0 \leq s < t \leq 1} (B(t) - B(s))$$

et donc on calcule le seuil $c(\alpha)$ par la contrainte suivante :

$$\mathbb{P} \left(\max_{0 \leq s < t \leq 1} (B(t) - B(s)) \geq c(\alpha) \right) \leq \alpha$$

par le lemme 1.4.1 on a .

$$\max_{0 \leq s < t \leq 1} (B(t) - B(s)) \stackrel{\text{lois}}{=} \max_{0 \leq t \leq 1} |B(t)|.$$

$$P(\max_{0 \leq t \leq 1} |B(t)| \leq c(\alpha)) = \frac{4}{\pi} \sum_{n=0}^{\infty} \frac{(-1)^n}{2n+1} \exp \left\{ -\frac{\pi^2 (2n+1)^2}{8 c(\alpha)^2} \right\} \quad (4.10)$$

En traçant la série (4.10), on peut trouver notre seuil $C(\alpha)$ par exemple $c(0.05)=2.14$ $c(0.01)=2.80$ voir Figure (4.1) :

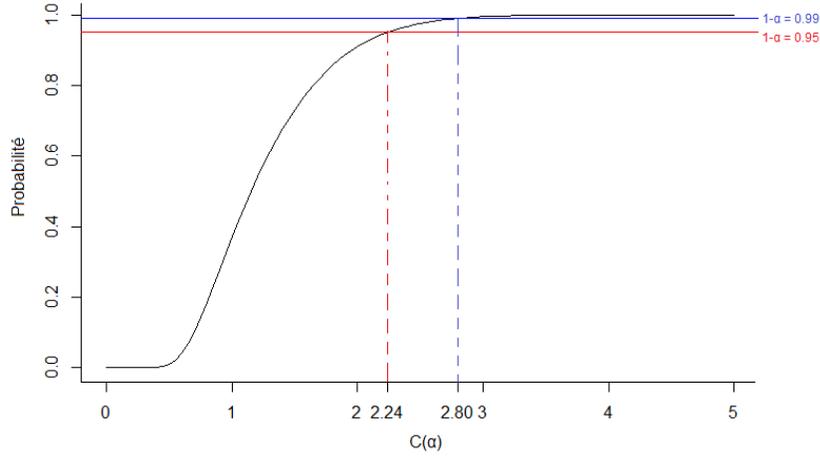


Figure 4.1 : Graphe de la fonction de répartition de $\max_{0 \leq t \leq 1} |B(t)|$

4.4 Simulation :

Dans cette section on va prendre des exemple sur notre test pour illustrer ce qu'on a fait et pour simplifier on va prendre un modèle AR(1). On supposera pour tous nos tests que $n_0 = 200$ et que l'instant du changement est $t_p = 100$. On notera par T_n la statistique test :

$$T_n = \max_{0 < j < n} n_0^{-\frac{1}{2}} (W_n(\theta_0, \hat{\eta}_n) - W_j(\theta_0, \hat{\eta}_n))$$

4.4.1 Détection du changement dans la moyenne :

Soit $\{Y_i\}_{i \geq 1}$ une suit d'observations suivant le model AR(1) :

$$Y_i - \mu = \phi(Y_{i-1} - \mu) + \varepsilon_i$$

On veut tester les hypothèses suivantes :

$H_0 : \mu = \mu_0$ pour tous Y_i , $i > 0$

$H_1 : \mu = \mu_0$ pour $Y_1, Y_2, \dots, Y_{t_p-1}$ et $\mu = \mu_1$ pour tous Y_i , $i \geq t_p$

et donc par (4.3) notre statistique sera comme suit :

$$W_n(\mu_0, \hat{\sigma}_n^2, \hat{\phi}_n) - W_j(\mu_0, \hat{\sigma}_n^2, \hat{\phi}_n) = \frac{1}{\hat{\sigma}_n} \sum_{i=j+1}^n [(Y_i - \mu_0) - \hat{\phi}_n(Y_{i-1} - \mu_0)]$$

La Figure (4.2) montre les résultats de la simulation du test pour des observation suivant le modele AR(1) avec $\phi = -0.5, \sigma^2 = 1$ et $\mu = \mu_0 = 0$ et a l'instant $t_p = 100$ $\mu = \mu_1 = 0.5$

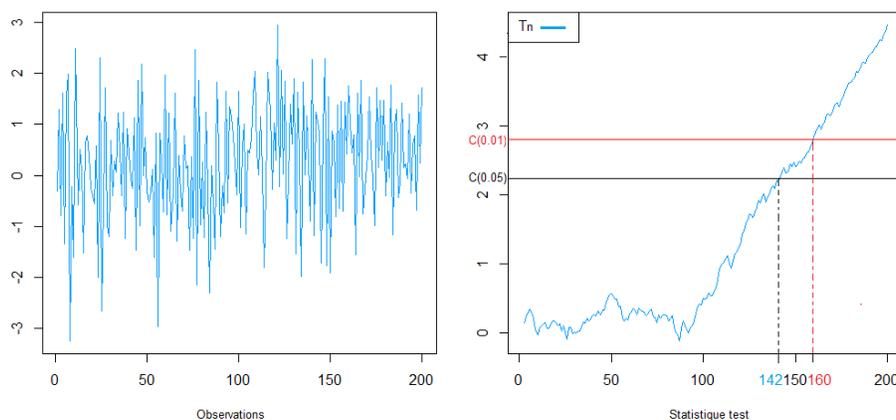


Figure 4.2 : Simulation du test sur la moyenne avec $\mu_1 = 0.5$

On voit bien de la figure que pour un niveau de signification de 1 % on détecte à l'instant $t_a = 160$ et on voit bien de la Figure que pour un niveau de signification de 5 % on détecte a l'instant $t_a = 142$.

Le même test est représenté dans la Figure (4.3) mais cette fois avec $\mu_1 = 1$

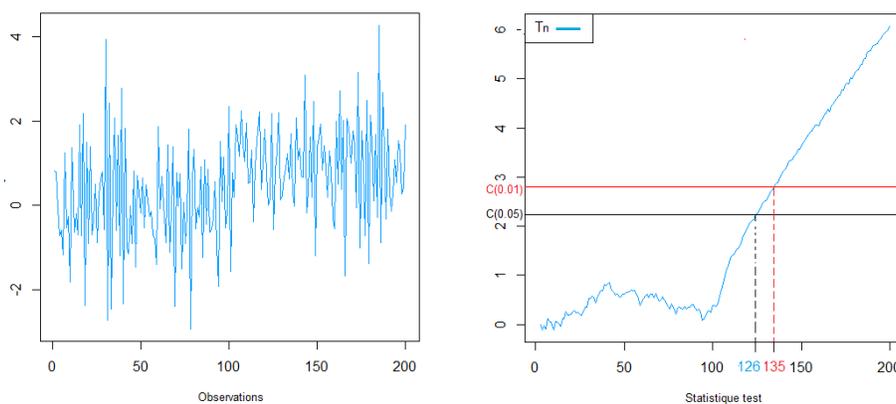


Figure 4.3 Simulation du test sur la moyenne avec $\mu_1 = 1$

On voit que pour $\alpha = 1\%$ on détecte le changement à l'instant 135 et pour $\alpha = 5\%$ on le détecte à l'instant 126.

4.4.2 Détection du changement dans la variance :

Soit $\{Y_i\}_{i \geq 1}$ une suite d'observations suivant le modèle $AR(1)$:

$$Y_i - \mu = \phi (Y_{i-1} - \mu) + \varepsilon_i$$

On veut tester les hypothèse suivant :

$$H_0 : \sigma = \sigma_0 \text{ pour tous } Y_i, i > 0$$

$$H_1 : \sigma = \sigma_0 \text{ pour } Y_1, Y_2, \dots, Y_{t_p-1} \text{ et } \sigma = \sigma_1 \text{ pour tous } Y_i, i \geq t_p$$

et donc par (4.4) notre statistique sera comme suite :

$$W_n \left(\sigma_0^2, \hat{\mu}_n, \hat{\phi}_n \right) - W_j \left(\sigma_0^2, \hat{\mu}_n, \hat{\phi}_n \right) = 2^{-1/2} \sigma_0^{-2} \sum_{i=j+1}^n \left\{ \left[(Y_i - \hat{\mu}_n) - \hat{\phi}_n (Y_{i-1} - \hat{\mu}_n) \right]^2 - \sigma_0^2 \right\}$$

La Figure (4.4) montre les résultats de la simulation du test sur la variance pour des observations suivant le modèle $AR(1)$ avec $\phi = -0.5, \mu = 0$ et $\sigma^2 = \sigma_0^2 = 1$ et à l'instant $t_p = 100$, $\sigma^2 = \sigma_1^2 = 1.2$

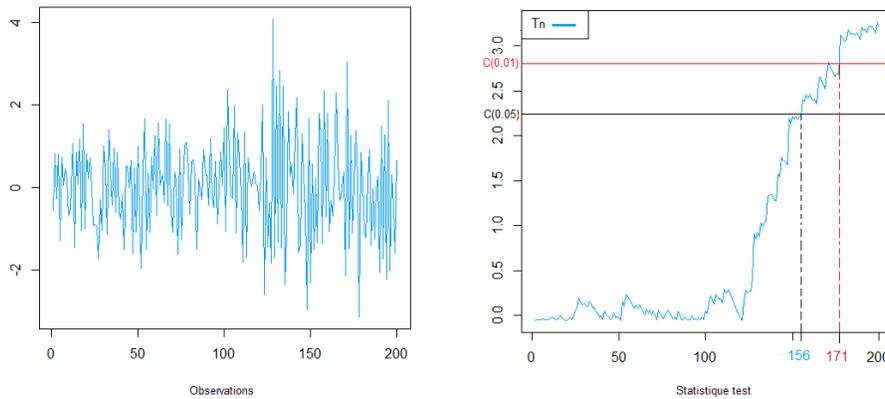


Figure 4.4 Simulation du test sur la moyenne avec $\sigma_1^2 = 1.2$

de la Figure (4.4) on détecte un changement à l'instant $t_a=171$ avec $\alpha = 0.01$ et à l'instant $t_a=156$ pour $\alpha = 0.05$.

la Figure (4.5) représente le même test pour $\sigma_1^2 = 1.8$

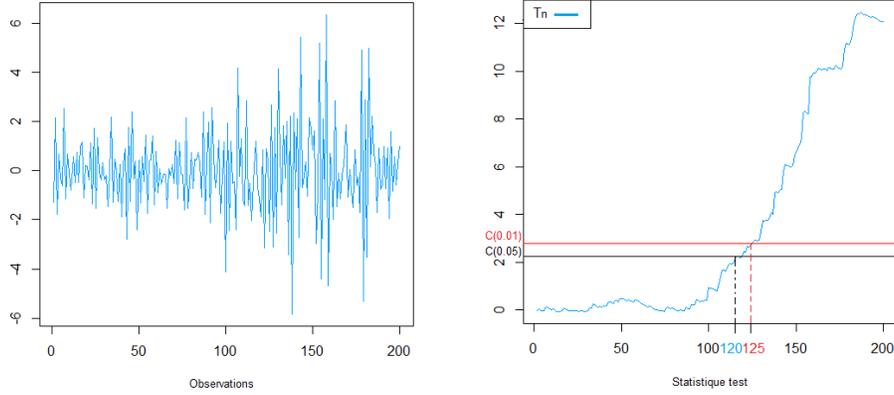


Figure 4.5 Simulation du test sur la moyenne avec $\sigma_1^2 = 1.8$

De la Figure on voit que la détection se fait à l'instant 125 pour $\alpha = 0.01$ et pour $\alpha = 0.05$ à l'instant 120.

4.4.3 Détection du changement des paramètres d'un modèle autorégressif :

Soit $\{Y_i\}_{i \geq 1}$ une suite d'observations suivant le modèle $AR(1)$:

$$Y_i - \mu = \phi_0 (Y_{i-1} - \mu) + \varepsilon_i$$

On veut tester les hypothèses suivantes :

$$H_0 : \phi = \phi_0 \text{ pour tous } Y_i, i > 0$$

$H_1 : \phi = \phi_0$ pour $Y_1, Y_2, \dots, Y_{t_p-1}$ et $\phi = \phi_1$ pour tous $Y_i, i \geq t_p$. Par (4.5) notre statistique est :

$$W_n(\phi_0, \hat{\mu}_n, \hat{\sigma}_n^2) - W_j(\phi_0, \hat{\mu}_n, \hat{\sigma}_n^2) = \frac{1}{\hat{\sigma}_n} \gamma(0)^{-1/2} \sum_{i=j+1}^n (Y_{i-1} - \hat{\mu}_n) [(Y_i - \hat{\mu}_n) - \phi_0 (Y_{i-1} - \hat{\mu}_n)]$$

On montre dans la Figure (4.6) les résultats de la simulation du test pour $\mu = 0, \sigma = 1$ et $\phi = \phi_0 = -0.5$ et à l'instant $t_p = 100$, $\phi = \phi_1 = -0.2$

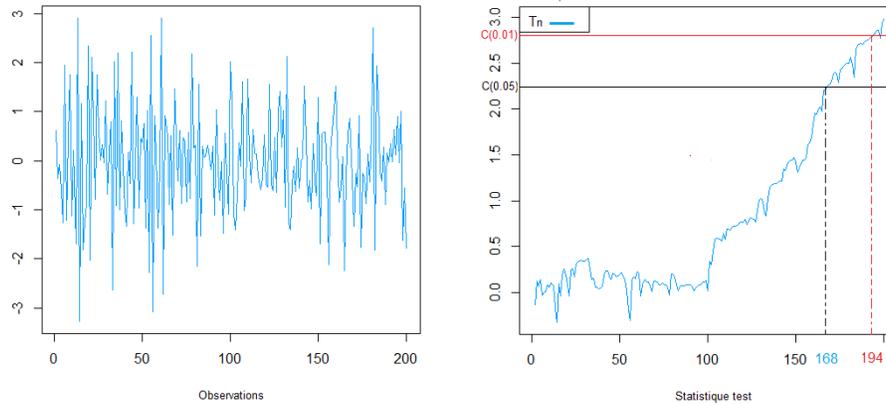


Figure 4.6 Simulation du test sur la moyenne avec $\phi_1 = -0.2$

De la Figure, on voit bien que qu'on détecte un changement pour $\alpha = 0.01$ à l'instant $t_a = 194$ et pour $\alpha = 0.05$ le changement est détecté à l'instant $t_a = 168$ et de même que le cas de la moyenne et la variance on refait le test pour un écart plus grand entre ϕ_0 et ϕ_1 .

La Figure (4.7) montre le résultat de la simulation pour $\phi_1 = 0.5$

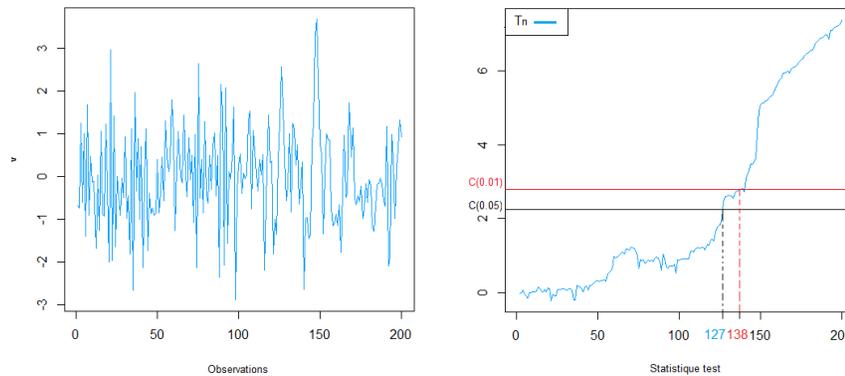


Figure 4.7 Simulation du test sur la moyenne avec $\phi_1 = 0.5$

et donc on détecte un changement à l'instant $t_a = 138$ pour un niveau significatif $\alpha = 0.01$, et pour $\alpha = 0.05$ le changement est détecté à l'instant $t_a = 127$.

Conclusion :

Le mémoire porte sur l'étude statistique des méthodes de détection des ruptures, en particulier les tests séquentiels qui sont très utilisés dans ce domaine. L'objectif principal étant d'étudier le test du CUSUM et ses propriétés principales. Nous avons présenté le schéma d'inspection proposé par Page [14] pour détecter un changement dans la moyenne et on l'a comparé avec le test binomial. Nous avons étudié aussi le test statistique du CUSUM basé sur le rapport de vraisemblance qui nécessite la distribution de probabilité de l'échantillon qui n'est pas toujours possible.

Enfin, nous avons développé les résultats établis dans [8] qui ont utilisé l'idée du test CUSUM de Page afin de construire un test séquentiel tronqué pour détecter les ruptures dans un modèle autorégressif d'ordre p . Les auteurs ont obtenu une approximation de la statistique test par un mouvement Brownien et ce grâce au comportement asymptotique des estimateurs des paramètres du modèle.

Bibliographie

- [1] A Aue. *Sequential change-point analysis based on invariance principles*. PhD thesis, Universität zu Köln, 2003.
- [2] J Bigot. *Séries chronologiques. Licence*, 3, 2017.
- [3] JP Boulay. *Statistique mathématique applications. Ellipses édition Marketing S.A .32.rue Bague 75740 paris cedex 15 ,2010*.
- [4] KL Chung. *A course in probability theory*, 2^e edition. *Academic Press, New york. Second Dean, P. and Bird, NF (1967). Monte Carlo estimates of critical percolation probabilities. Camb. Phil. Soc*, 63, 1974.
- [5] KL Chung and RJ Williams. *Introduction to stochastic integration*. Birhauser : Boston, 1983.
- [6] E Eberlein. *On strong invariance principles under dependence assumptions. The Annals of Probability*, pages 260–270, 1986.
- [7] E Gombay. *Change detection in autoregressive time series. Journal of Multivariate Analysis*, 99(3) :451–464, 2008.
- [8] E Gombay and D Serban. *Monitoring parameter change in ar (p) time series models. Journal of Multivariate Analysis*, 100 :715–725, 2009.
- [9] O Klesov. *Limit theorems for multi-indexed sums of random variables*, volume 71. Springer, 2014.
- [10] J Komlós, P Major, and G Tusnády. *An approximation of partial sums of independent RV's, and the sample DF. I. Wahrsch. verw. Gebiete*, 32 :111–131, 1975.
- [11] J Komlós, P Major, and G Tusnády. *An approximation of partial sums of independent RV's, and the sample DF. II. Wahrsch. verw. Gebiete*, 34 :33–58, 1976.

- [12] S Méléard. *Modèles aléatoires en Ecologie et Evolution*. Springer, 2016.
- [13] ES Page. Continuous inspection schemes. *Biometrika*, 41 :100–115, 1954.
- [14] ES Page. A test for a change in a parameter occurring at an unknown point. *Biometrika*, 42 :523–527, 1955.
- [15] A Shewhart. *Economic control of quality of manufactured product*. Macmillan And Co Ltd, London, 1931.
- [16] Y Velenik. Chapitres choisis de théorie des probabilités.
- [17] G Verdier. *Détection Statistique de Rupture de Modèle dans les Systèmes Dynamiques-Application à la Supervision de Procédés de Dépollution Biologique*. PhD thesis, Université Montpellier II-Sciences et Techniques du Languedoc, 2007.
- [18] MC Viano and A Philippe. Maîtrise d'économétrie : Cours de séries temporelles. *Université des Sciences et Technologies de Lille, UFR de Mathématiques Pures et Appliquées*, 1999.