

Ministère de L'Enseignement Supérieur Et de La Recherche Scientifique
Université Djillali Liabès de Sidi Bel-Abbès
Faculté des Sciences exactes
Département de Mathématiques

THÈSE

présentée par

BELGUERNA Abderrahmane

pour obtenir

Le grade de DOCTEUR EN SCIENCES

Spécialité : Mathématiques

Option : Probabilités et Statistique

Propriétés asymptotiques pour des estimateurs non paramétriques Des processus autorégressifs

soutenu le, devant le jury composé de

Mr GHERIBALLAH ABDELKADER	Professeur	UDL-S.B.A	Président
Mr GUNDOUZI TOUFIK	M.C.A	UTM-SAIDA	Examineur
Mr KANDOUZI ABDELJEBBAR	M.C.A	UTM-SAIDA	Examineur
Mr RABHI ABBES	M.C.A	UDL-S.B.A	Examineur
Mr MADANI FETHI	M.C.A	UTM-SAIDA	Examineur
Mr BEAISSA SAMIR	M.C.A	UDL-S.B.A	Directeur de thèse

A mes parents,

mes enfants Ayoub, Hind et Anès,

et à toute ma famille.

Remerciements

En premier lieu, je tiens à exprimer mon profond sentiment de respect et de reconnaissance à mon directeur de thèse, Monsieur **BENAISSA Samir**, pour ses conseils qui sont toujours judicieux et son encouragement durant la période de préparation et recherche en doctorat. Par la même occasion, qu'il trouve en ces mots l'expression de mon admiration pour ses grandes qualités scientifiques et humaines.

Je remercie sincèrement Monsieur **GHERIBALLAH Abdelkader** pour l'honneur qu'il me fait en présidant le jury de cette thèse.

Je remercie vivement Messieurs **GUNDOUZI Toufik**, **KANDOUZI Abdeljebbar**, **RABHI Abbes** et **MADANI Fethi** qui me font l'honneur de participer à mon jury. Je les remercie énormément pour l'attention qu'ils ont accordé à ce travail.

Il est important pour moi de remercier ma famille qui a toujours été une source inépuisable d'encouragements.

Merci à tous ceux qui ont contribué, de près ou de loin, à l'aboutissement de ce travail.

A toute ma famille.

Table des matières

Introduction générale	9
1 Introduction	13
1.0.1 Les opérateurs discrets	15
1.1 Équation aux différences linéaires à coefficients constants	16
1.1.1 Motivation	16
1.1.2 Équation aux différences linéaires du $p^{\text{ème}}$ ordre	17
1.2 Notions sur les processus stochastiques	21
1.2.1 Processus stationnaire	21
1.2.2 Propriétés de la fonction d'autocovariance	24
1.2.3 Processus linéaires et processus linéaires généraux	24
1.2.4 Processus autorégressif $AR(p)$	26
1.2.5 Équations de Yule-Walker	30
1.2.6 Estimation du processus autorégressif d'ordre un $AR(1)$	31
1.2.7 Estimation du processus autorégressif d'ordre p $AR(p)$	32
1.2.8 Les modèles $ARMA$	34
1.3 Simuler des processus	35
1.3.1 Simuler un échantillon aléatoire d'une variables continue	35
1.3.2 Simuler un processus autorégressif	37
1.3.3 Simuler un processus MA	37
1.3.4 AR (Auto-Regressive models)	43
2 Estimation non-paramétrique par la méthode de noyau	48
2.1 Modes de convergence :	52
2.1.1 Convergence en loi :	52
2.1.2 Convergence en probabilité :	53
2.1.3 Convergence Presque sûre :	55
2.2 Estimation de la densité de probabilité d'une variable aléatoire.	57

TABLE DES MATIÈRES

2.2.1	Estimation par histogramme	58
2.2.2	Estimateur à noyau	59
2.3	Estimation par la méthode du noyau	63
2.3.1	La fonction de répartition empirique	63
2.4	L'estimation de la fonction de répartition	67
2.4.1	L'estimation de la fonction de densité	69
3	Notions de dépendance et inégalités des moments	73
3.1	Notions de dépendance	73
3.1.1	Conditions de mélange	74
3.1.2	L'association	77
3.2	Inégalités des moments	78
4	La convergence asymptotique du paramètre d'un model autoregressif d'ordre 1 dans le cas des erreurs α-mélangeant	82
4.1	Convergence presque complète	83
4.2	Principaux résultats	84
4.3	Intervalle de confiance	88
5	Distributions asymptotiques des estimateurs de la densité d'erreur dans les modèles autorégressifs de premier ordre	90
5.1	Modèles, estimateurs, et hypothèses	91
5.2	Distribution asymptotique de \hat{f}_n en un point fixe	92
5.3	Distribution asymptotique de la mesure globale	97
	Conclusion et perspectives	104
	Bibliographie	105

Table des figures

1.1	ACF. MA(1). WN	38
1.2	ACF. MA(3). WN	39
1.3	ACF. AR(1)	45
1.4	PACF. ACF. AR(3)	46
1.5	Partial ACF. ACF. AR(3)	47

Liste des tableaux

3.1	Relations entre quelques conditions de dépendance	75
-----	---	----

Notation et Abréviation

Les notations et les abréviations suivantes sont utilisées :

ACF : Fonction d'auto-corrélation (Anglais : auto correlation function).

PACF : Fonction d'auto-corrélation partiel (Anglais : partiel auto correlation function).

AR : Auto régressif (Anglais : Auto regressive).

MA : Moyenne mobile (Anglais : Moving average).

ARMA : Auto régressif Moyenne mobile (Anglais : Auto regressive moving average).

ARIMA : Auto régressif Moyenne mobile intégré (Anglais : Auto regressive integrated moving average).

AIC : Critère d'information d'Akaikae (Anglais : Akaikae information criteria).

BIC : Critère d'information Bayésien.

Publications et communications

Publication Internationale : A. Belguerna and S. Benaissa. The asymptotic convergence of the parameter estimator in the first order autoregressive model $AR(1)$ under strong mixing conditions. International Journal of Mathematics and Statistics, Year 2014, Volume 15, Issue Number 3, 1-8.

Communication Internationale : A. Belguerna and S. Benaissa. Estimation d'une régression non paramétrique. International conference on Mathematics for the natural and life sciences (ICM-NLS14). Organisé par la faculté des sciences de la nature et de la vie et des sciences de la terre et de l'univers, université de Tlemcen. 9-11 Novembre 2014 à Tlemcen -Algérie-.

Introduction générale

Pour tenter de comprendre le monde autour de nous, des observations sont souvent fait étalée dans le temps. Valeurs à l'avenir (le futur) dépendent, avec les observations disponibles actuellement (le passé et le présent), généralement d'une manière stochastique. Une telle dépendance est utile de prévoir le futur de son passé. En effet, nous va décrire la dynamique de système à partir de laquelle les données observées sont générées et sera donc prévoir et éventuellement contrôler les événements futurs. Les objectifs de l'analyse des séries chronologiques sont diverses, selon le contexte des applications. Les statisticiens considèrent généralement une série temporelle comme une réalisation d'un processus stochastique. Une tâche fondamentale est de dévoiler la loi de probabilité qui régit la série chronologique observée. Avec une telle loi de probabilité, nous pouvons comprendre la dynamique prévus des événements futurs , et contrôler les événements futurs par intervention. Ce sont les trois principaux objectifs de l'analyse des séries chronologiques.

Il y a une infinité de processus stochastiques qui peuvent générer les même données observées , comme le nombre d'observations est toujours fini. Cependant, certains de ces processus sont plus plausible et admettent meilleure interprétation que d'autres. Sans contraintes supplémentaires sur le processus sous-jacent, il est impossible d'identifier le processus à partir d'un nombre fini d'observations. Une approche populaire est de confiner la loi de probabilité à une famille déterminée et puis choisit un membre de cette famille qui est la plus plausible. La première que l'on appelle la modélisation et le dernier est appelé estimation, ou, plus généralement, inférence statistique.

La classe la plus populaire des modèles de séries chronologiques linéaires se compose des modèles autorégressifs d'ordre p , ($AR(p)$), les modèles à moyennes mobiles d'ordre q ($MA(q)$) (En Anglais : moving average) et les modèles mixtes d'ordre (p, q) ($ARMA(p, q)$). Il existe aussi des modèles particuliers comme $ARIMA$, $SARIMA$

Un processus autorégressif est un modèle de régression dans lequel la variable de la série est expliquée par ses valeurs passées plutôt que par d'autres variables.

Les processus autorégressifs supposent que chaque point peut être prédit par la somme pondérée d'un ensemble de points précédents, plus un terme aléatoire d'erreur.

Ces processus forment une classe flexible de modèles pour de nombreux phénomènes observés. Ils sont construits à partir de l'idée que l'observation au temps t s'explique linéairement par les observations précédentes.

Les processus autorégressifs sont utilisés pour modéliser des séries chronologiques dans de nombreux domaines, en finance, en biologie, en climatologie, en médecine, en économétrie et en météorologie et dans bien d'autres domaines. Par exemple, en finance, on s'intéresse à modéliser le taux de change d'une devise. En météorologie, les scientifiques modélisent par exemple la température dans le dernier mois pour prédire la température qu'il fera demain. L'idée est de prendre un échantillon de données et de construire le meilleur modèle qui ajuste ces données. Ce modèle nous permet de tirer certaines conclusions sur la série.

Les premiers processus autorégressifs ont été introduits par George Udny Yule (1871-1951). Les recherches actuelles se concentrent autour de deux thèmes : l'étude des propriétés théoriques de ces processus (existence, moments et queue d'une loi stationnaire, comportement extrême) dont la connaissance permet de mieux choisir les modèles en fonction des données, et d'autre part l'estimation des paramètres des modèles pour pouvoir répondre aux problèmes pratiques et faire des prévisions. Les processus de moyenne mobile supposent que chaque point est fonction des erreurs entachant les points précédant, plus sa propre erreur. Les modèles ARMA (modèles autorégressifs et moyenne mobile), ou aussi modèle de Box-Jenkins, sont des principaux modèles des séries temporelles. Étant donné une série temporelle X_t , le modèle ARMA est un outil pour comprendre et prédire, éventuellement, les valeurs futures de cette série. Le modèle ARMA qui combine les deux modèles AR et MA est composé de deux parties : une part autorégressive (AR) et une part moyenne-mobile (MA). Le modèle est généralement noté $ARMA(p, q)$, où p est l'ordre de la partie AR et q l'ordre de la partie MA. En particulier, un modèle autorégressif $AR(p)$ est un $ARMA(p, 0)$. Un modèle moyenne mobile $MA(q)$ est un $ARMA(0, q)$.

Les modèles ARMA, une fois choisi les ordres p et q , peuvent être ajustés sur des données par la méthode des moindres carrés, on recherche les paramètres qui minimisent la somme des carrés des résidus. Prendre des valeurs de p et q les plus petites est généralement vu comme une bonne pratique (principe de parcimonie). Pour un modèle AR pur, les équations de Yule-Walker permettent de réaliser l'ajustement.

Un modèle ARIMA est étiqueté comme modèle $ARIMA(p, d, q)$, dans lequel :

p est le nombre de termes auto-régressifs,

d est le nombre de différences,

q est le nombre de moyennes mobiles.

Le modèle *ARMA* ne permet de traiter que les séries dites stationnaires (des moments du premier ordre qui sont invariants au cours du temps). Les modèles *ARIMA* permettent de traiter les séries non stationnaires après avoir déterminé le niveau d'intégration (le nombre de fois qu'il faut différencier la série avant de la rendre stationnaire).

Un modèle de bruit blanc faible est un processus stochastique $\varepsilon = (\varepsilon_t)_{t \in I}$ où l'ensemble I est un intervalle de temps qui peut être discret ou continu tel que pour tout t et t' dans I $\mathbb{E}(\varepsilon_t) = 0$ et $\mathbb{E}(\varepsilon_t \varepsilon_{t'}) = \sigma^2 I_{tt'}$.

On dit qu'il est fort lorsque de plus les variables ε_t sont indépendantes.

L'hypothèse d'observations indépendantes est souvent subjective, voir erronée, car elle ne reflète pas l'évolution exacte du phénomène aléatoire. En effet, les observations dépendantes sont plus ajustées à la réalité. Il existe, cependant, de nombreuses notions de dépendance. On s'intéresse ici à celles qui s'expriment en termes de coefficients de mélange entre les tribus engendrées par le passé et le futur de la suite de variables aléatoires $(X_n)_{n \geq 1}$. Les coefficients de α -mélange α_n , de β -mélange β_n , de ρ -mélange ρ_n , de φ -mélange φ_n et de ψ -mélange ψ_n . Les suites α -mélangeantes ont beaucoup d'intérêt, les processus linéaires sont sous certaines conditions α -mélangeants et leurs coefficients de mélange ont un ordre de grandeur explicite. Cependant, il est techniquement très difficile d'évaluer ces coefficients α_n des suites mélangeantes (voir Doukhan [42]) pour une étude complète de ce sujet).

Une autre notion de dépendance est celle appelée dépendance positive ou association, des variables aléatoires sont dites associées lorsque la covariance de toute paire de fonctions croissantes de ces variables est positive. Les coefficients qui interviennent en manipulant la dépendance positive sont les plus naturels possibles. Ces quantités sont plus faciles à évaluer que les coefficients de mélange.

Une autre notion de dépendance faible a été définie récemment par Doukhan et Louhichi qui permet notamment de traiter le mélange et l'association dans une approche unifiée. Dans cette notion générale de dépendance, au lieu de mesurer l'écart entre les tribus engendrées par le passé et le futur des variables aléatoires, on mesure plutôt l'écart entre des fonctions convenables du passé et du futur de la suite.

Ce travail sera organisé comme suit : dans le premier chapitre préliminaires nous rappelons les notions fondamentales sur les processus stochastique et en particulier les processus autorégressifs, nous abordons les travaux essentiels (Ciesielski) d'étude de stationnarité, la causalité et d'autres propriétés asymptotique de ces derniers processus. nous donnons des principaux résultats de convergence. Ces résultats seront utilisés ensuite dans les chapitres deux, quatre et cinq. Dans le chapitre deux, on aborde l'estimation non-paramétrique par la méthode à noyau. Dans le chapitre trois nous entamons les différentes notions de dépendance et des résultats générales concernant ces notions et les processus autorégressifs. Des

inégalités des moments sont nécessaires pour l'analyse du comportement asymptotique des estimateurs des processus autorégressifs dans les principaux résultats du chapitre quatre et pour l'étude de l'estimation de la fonction de densité dans le chapitre cinq. Nos résultats sur la convergence presque complète de l'estimateur des moindres carrés du paramètre d'un processus autorégressif d'ordre sous erreur dépendante, se trouvent dans le chapitre quatre. Le chapitre cinq est consacré à l'estimation de la fonction de densité.

Chapitre 1

Introduction

L'étude des séries temporelles, ou séries chronologiques, correspond à l'analyse statistique d'observations régulièrement espacées dans le temps. Elle ont été utilisées en astronomie ('on the periodicity of sunspots', 1906), en météorologie ('time-series regression of sea level on weather', 1968), en théorie du signal ('Noise in FM receivers', 1963), en biologie ('the autocorrelation curves of schizophrenic brain waves and the power spectrum', 1960), en économie ('time-series analysis of imports, exports and other economic variables', 1971)...etc.

De façon générale, quand on parle de séries stationnaires, on a en tête une représentation de forme X_t , où $t \in \mathbb{Z}$, représentant les observations (potentielles) du processus, dont on peut définir un ensemble d'autocovariance $\gamma(t, s) = \mathbb{E}[(X_t - \mu)(X_s - \mu)]$ qui ne dépend que la distance entre t et s , $\gamma(t, s) = \gamma(t + h, s + h)$ pour tout $h \in \mathbb{Z}$ (notion 'faible' de stationnarité). On demande généralement à cette autocovariance $\gamma(t, s)$ de tendre vers 0 quand la différence entre t et s tend vers l'infini : la covariance entre des éléments très éloignés dans la série tend vers 0.

Cette approche, basé sur l'utilisation des corrélations, correspond à l'analyse de type 'temporelle' : elle consiste à l'étudier les corrélation croisées de fonctions de la série (X_t). Ces méthodes sont généralement paramétriques de type moyenne-mobiles (moving average MA) ou autorégressives (AR)- voire les deux (ARMA). Toutes ces méthodes consistants à estimer des paramètres peuvent généralement être vus comme des généralisations de la régression linéaire.

Historiquement, ce sont les astronomes qui les premiers ont travaillé sur des séries chronologiques.

Soit un processus scalaire $AR(1)$, noté $(X_t)_{t \in \mathbb{Z}}$ vérifiant :

$$X_t = \theta X_{t-1} + \zeta_t, \quad t \in \mathbb{Z}, \quad X_t \in \mathbb{R}$$

où $\theta \in \Theta \subset \mathbb{R}$; $|\theta| < 1$ et les processus $(\zeta_t)_{t \in \mathbb{Z}}$ souvent appelé processus bruit. L'estimateur des moindres carrés de θ s'écrit sous la forme

$$\hat{\theta}_n = \left(\frac{1}{n} \sum_{t=1}^n X_{t-1}^2 \right)^{-1} \left(\frac{1}{n} \sum_{t=1}^n X_t X_{t-1} \right).$$

De plus, le cas où les variables aléatoires ζ_t sont indépendantes et identiquement distribuées (i.i.d.) et de moment d'ordre 4 fini. $\sqrt{n}(\hat{\theta}_n - \theta)$ converge en loi vers une variable aléatoire gaussienne centrée de variance $\sigma_\theta^2 = 1 - \theta^2$. Voir [73] où l'on peut aussi trouver des résultats asymptotiques relatifs au processus $X_t = \sum_k \theta_k \zeta_{t-k}$. Voir [78], Davis et Resnick montrent, à propos de ce dernier modèle, qu'il est possible d'obtenir des résultats de convergence en loi non standard pour la fonction d'autocorrélation $\gamma(h) = (\mathbb{E}X_0^2)^{-1} \mathbb{E}(X_0 X_h)$ (estimée par $(\sum_{t=1}^n X_t X_t)^{-1} \sum_{t=1}^n X_t X_{t+h}$). Deux articles en 1927 ont ouvert une étude sur les processus autorégressifs et les moyennes mobiles : l'article de **Yule**[86] et celui de **Slutsky**[84] .

Yule a introduit dans la littérature les modèles autorégressifs, en considérant des modèles de la forme

$$X_t = \theta_1 X_{t-1} + \theta_2 X_{t-2}.$$

Étant données deux valeurs initiales, cette suite présente un comportement saisonnier, fonction des paramètres θ_1 et θ_2 . **Yule** remarque qu'en fait, le comportement dépend des racines (complexes) de l'équation $z^2 - \theta_1 z - \theta_2 = 0$, et plus particulièrement de leur position par rapport au disque unité. Si leur module est inférieur à 1, alors on observe un comportement sinusoïdal amorti. En fait, la forme générale des solutions sera

$$X_t = A \rho^t \cos(\omega t - \alpha), \text{ lorsque } 0 < \rho < 1.$$

Le modèle autorégressif proposé par Yule est le suivant

$$X_t = \theta_1 X_{t-1} + \theta_2 X_{t-2} + \zeta_t \tag{1.1}$$

où (ζ_t) correspond à un bruit blanc.

Exemple La série définie par $X_t = 1.8X_{t-1} - 0.8X_{t-2}$ peut être écrite également

$$X_t = -2(0.9)^t \cos(4t - 1/2)$$

Slutsky a introduit les moyennes mobiles la même année que **Yule** a introduit les processus autorégressifs... mais son article, écrit en 1927 en russe n'a été traduit qu'en 1937 en anglais. Pour cela, il a utilisé des nombres générés par la loterie officielle, et a réussi à générer une série décrivant le cycle économique en Angleterre, de 1855 à 1877. La forme générale était la suivante,

$$X_t = \alpha_0 \zeta_t + \alpha_1 \zeta_{t-1} + \dots + \alpha_q \zeta_{t-q}, \tag{1.2}$$

où (ζ_t) constituent un bruit blanc. Cette écriture a suggéré d'élargir la relation (1.1) sous une forme proche de (1.2), à savoir

$$\theta_0 X_t + \theta_1 X_{t-1} + \theta_2 X_{t-2} = \zeta_t. \quad (1.3)$$

Les processus introduits par **Yule** deviendront les processus $AR(p)$ et ceux introduits par **Slutsky** les processus $MA(q)$. L'analogie entre les deux processus sera même poussée plus loin lorsqu'il sera montré que les processus $AR(p)$ et $MA(q)$ sont respectivement des processus $MA(\infty)$ et $AR(\infty)$, sous certaines conditions. Plus généralement on peut montrer que tout $AR(p)$ peut avoir une représentation $MA(\infty)$ et de manière duale, on peut exprimer aussi toute $MA(q)$ comme un $AR(\infty)$.

1.0.1 Les opérateurs discrets

Nous considérons l'ensemble des suites numériques et sur cet ensemble nous définirons les opérateurs de retard L , d'avance F , de différence ∇ et de sommation S . (Nous conservons les symboles anglais L pour Backward et F pour forward).

i) L'opérateur de retard L

Il est défini par

$$LY_t = Y_{t-1}, \forall t$$

la linéarité est évidente : $L(\alpha Y_t + \beta Z_t) = \alpha Y_{t-1} + \beta Z_{t-1} = \alpha LY_t + \beta LZ_t$ on posera $L^0 Y_t = 1.Y_t$ (opérateur identité).

L'opérateur L^n est défini par $L(L^{n-1} Y_t) = Y_{t-n}$.

La somme d'opérateurs défini par :

$$(\alpha_1 L^{n_1} + \dots + \alpha_p L^{n_p}) Y_t = (\alpha_1 L^{n_1} Y_t + \dots + \alpha_p L^{n_p} Y_t) = \alpha_1 Y_{t-n_1} + \dots + \alpha_p Y_{t-n_p}.$$

Ainsi un polynôme en L est bien défini.

ii) L'opérateur d'avance F

Il est défini par :

$$FY_t = Y_{t+1}, \forall t$$

toutes les définitions précédentes bien entendu s'applique à F et la linéarité est aussi évidente.

iii) L'opérateur de différence ∇

défini par :

$$\nabla Y_t = Y_t - Y_{t-1}, \forall t$$

1.1 Équation aux différences linéaires à coefficients constants

on a aussi $\nabla = 1 - L$, il est possible de reprendre les définitions *i*).

Itérations de ∇ :

$$\begin{aligned}\nabla^2 Y_t &= (1 - L)^2 Y_t \\ &= (1 - 2L + L^2) Y_t \\ \nabla^d Y_t &= \sum_{k=0}^d (-1)^k C_d^k Y_{t-k}.\end{aligned}$$

iv) **L'opérateur de sommation S**

Il est défini par :

$$SY_t = \sum_{h=-\infty}^t Y_h, \forall t$$

où $SY_t = (1 + L + L^2 + \dots)Y_t$.

On a ainsi (au moins formellement)

$$\nabla^{-1} = \frac{1}{1 - L} = 1 + L + L^2 + \dots \quad \nabla^{-1} = S.$$

S est bien sûre linéaire et les définitions *i*) sont applicables pour les itérations on a :

$$SY_t = \sum_{h=k+1}^t Y_h + SY_k,$$

pour $t \geq k + 1$

$$S^2 Y_t = \sum_{i=k+1}^t \sum_{h=k+1}^i Y_h + S^2 Y_k + (t - k)SY_k.$$

1.1 Équation aux différences linéaires à coefficients constants

1.1.1 Motivation

On suppose donné une suite de temps (année, trimestre, mois, semaine, jour,...). Soit Y_t la valeur d'une variable au temps t et w_t une autre quantité connue en t . Supposons que Y_t dépende de Y_{t-1} et de w_t par le modèle suivant :

$$Y_t = \theta Y_{t-1} + w_t \tag{1.4}$$

(1.4) définit une équation aux différences du 1^{er} ordre, linéaire à coefficients constants. On veut analyser la dynamique d'un tel système.

1.1.2 Équation aux différences linéaires du $p^{\text{ème}}$ ordre

Généralisations (1.4) en faisant dépendre Y_t de p valeurs passées :

$$Y_t = \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \dots + \theta_p Y_{t-p} + w_t. \quad (1.5)$$

Première méthode (matricielle) on écrit une équation vectorielle du 1^{er} ordre :

$$\begin{aligned} \widetilde{Y}_t = \begin{pmatrix} Y_t \\ Y_{t-1} \\ \vdots \\ Y_{t-(p-2)} \\ Y_{t-(p-1)} \end{pmatrix} &= \begin{pmatrix} \theta_1 & \theta_2 & \cdots & \theta_{p-1} & \theta_p \\ 1 & 0 & \cdots & 0 & 0 \\ \vdots & \ddots & & \vdots & \vdots \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} Y_{t-1} \\ Y_{t-2} \\ \vdots \\ Y_{t-(p-1)} \\ Y_{t-p} \end{pmatrix} + \begin{pmatrix} w_t \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix} \\ \Leftrightarrow \widetilde{Y}_t &= \widetilde{\theta} \widetilde{Y}_{t-1} + \widetilde{w}_t. \end{aligned} \quad (1.6)$$

On voit donc que l'on peut trouver une forme analogue à (1.4) pour une dépendance d'ordre p en exprimant la relation (1.5) sous la forme matricielle (1.6). En itérant t fois, on a :

$$\widetilde{Y}_t = \widetilde{\theta}^{t+1} \widetilde{Y}_1 + \widetilde{\theta}^t \widetilde{w}_0 + \widetilde{\theta}^{t-1} \widetilde{w}_1 + \dots + \widetilde{\theta} \widetilde{w}_{t-1} + \widetilde{w}_t \quad (1.7)$$

\Rightarrow

$$\widetilde{Y}_{t+j} = \widetilde{\theta}^{j+1} \widetilde{Y}_t + \widetilde{\theta}^j \widetilde{w}_t + \widetilde{\theta}^{j-1} \widetilde{w}_{t+1} + \dots + \widetilde{\theta} \widetilde{w}_{t+j-1} + \widetilde{w}_{t+j}. \quad (1.8)$$

Remarque que seule la première ligne des systèmes (1.7) et (1.8) nous intéresse.

La multiplicateur dynamique du \widetilde{w}_t est ici l'élément (1,1) de la matrice $\widetilde{\theta}^j$. Donc on peut écrire :

$$\widetilde{Y}_{t+j} = \sum_{i=1}^p (\widetilde{\theta}^{j+1})_{1i} \widetilde{Y}_{t-j} + \sum_{i=0}^j (\widetilde{\theta}^{j-i})_{11} \widetilde{w}_{t+i}. \quad (1.9)$$

Tout le problème est donc de calculer la 1^{ère} ligne des puissances de $\widetilde{\theta}$.

Proposition 1.1 *Les valeurs propres de $\widetilde{\theta}$ sont les solutions de l'équation caractéristique*

$$\lambda^p - \theta_1 \lambda^{p-1} - \dots - \theta_{p-1} \lambda - \theta_p = 0.$$

Preuve : On développe $\det(\widetilde{\theta} - \lambda I)$ par rapport à la dernière colonne.

Supposons pour simplifier que les p valeurs sont distinctes. Rangeons les sur la diagonale d'une matrice D . Alors il existe une matrice Λ non singulière, telle que

$$\widetilde{\theta} = \Lambda D \Lambda^{-1} \Leftrightarrow \widetilde{\theta}^j = \Lambda D^j \Lambda^{-1},$$

1.1 Équation aux différences linéaires à coefficients constants

$$(\tilde{\theta}^j)_{11} = \sum_{k=1}^p \underbrace{\Lambda_{1k}(\Lambda_{k1}^{-1})}_{c_k} \lambda_k^j. \quad (1.10)$$

On voit que

$$\sum_{k=1}^p c_k = (\Lambda\Lambda^{-1})_{11} = 1.$$

Proposition 1.2

$$c_k = \frac{\lambda_k^{p-1}}{\prod_{\substack{i=1 \\ i \neq k}}^p (\lambda_k - \lambda_i)}.$$

Preuve : Comme le $k^{\text{ième}}$ vecteur propre de $\tilde{\theta}$ est $(\lambda_k^{p+1-i})_{i=1, \dots, p} = (\Lambda_{ik})_{i=1, \dots, p}$, on voit que $(\Lambda^{-1})_{ik} = \frac{1}{\prod_{\substack{i=1 \\ i \neq k}}^p (\lambda_k - \lambda_i)}$.

Le multiplicateur dynamique donnant l'effet sur Y_{t+j} de l'augmentation de w_t d'une unité est donc :

$$(\theta^j)_{11} = \sum_{k=1}^p \frac{\lambda_k^{p+j-1}}{\prod_{\substack{i=1 \\ i \neq k}}^p (\lambda_k - \lambda_i)} = \sum_{k=1}^p c_k \lambda_k^p = \psi_j. \quad (1.11)$$

Remarque 1.1 Supposons que toutes les valeurs propres de $\tilde{\theta}$ soient < 1 (en module). Alors $\tilde{\theta}^{j+1} \rightarrow 0$ si $j \rightarrow +\infty$ et $\sum_{j=0}^{+\infty} \tilde{\theta}^j = (I - \tilde{\theta})^{-1}$ est bornée. Si tous les Y_t et tous les w_t sont bornées et connues depuis un lointain passé, on peut négliger la première somme dans (1.9) et écrire

$$Y_t = w_t + \psi_1 w_{t-1} + \psi_2 w_{t-2} + \dots$$

Sous forme matricielle,

$$\tilde{Y}_t = \sum_{j=0}^{+\infty} \tilde{\theta}^j \tilde{w}_{t-j}.$$

Si

$$\tilde{w}_{t-j} \rightarrow \tilde{w}_{t-j} + \begin{pmatrix} c \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \tilde{w}_{t-j} + ce_1, \quad j = 0, 1, \dots$$

\Rightarrow

$$\tilde{Y}_t \rightarrow \tilde{Y}_t + \underbrace{\left(\sum_{j=0}^{+\infty} \tilde{\theta}^j \right)}_{(I - \tilde{\theta})^{-1}} ce_1.$$

1.1 Équation aux différences linéaires à coefficients constants

On démontre que l'élément

$$\left((I - \tilde{\theta})^{-1}\right)_{11} = (1 - \theta_1 - \dots - \theta_p)^{-1}.$$

En conclusion, l'effet à long terme d'une augmentation de c de tous les w_{t-j} passés est une augmentation de Y_t de $(1 - \theta_1 - \dots - \theta_p)^{-1}c$.

Deuxième méthode : opérateurs de décalage

Posons

$$LY_t = Y_{t-1},$$

L est un opérateur qui associe à un élément de la suite $(Y_t, t \in \mathbb{Z})$ l'élément précédent l'équation (1.4) peut s'écrire sous la forme

$$Y_t = \theta LY_t + w_t,$$

ou encore

$$(1 - \theta L)Y_t = w_t. \tag{1.12}$$

Appliquons aux deux membres de (1.12) l'opérateur $1 + \theta L + \dots + \theta^j L^j$:

$$(1 - \theta^{j+1} L^{j+1})Y_t = (1 + \theta L + \dots + \theta^j L^j)w_t$$

\Leftrightarrow

$$Y_t = \theta^{j+1} Y_{t-j-1} + w_t + \theta w_{t-1} + \dots + \theta^j w_{t-j}.$$

En remplaçant t par $t + j$, on retrouve

$$Y_t = \theta^{t+1} Y_{-1} + w_t + \theta^t w_0 + \theta^{t-1} w_1 + \dots + \theta w_{t-1} + w_t.$$

De même l'équation (1.5) peut s'écrire

$$(1 - \theta_1 L - \dots - \theta_p L^p)Y_t = w_t. \tag{1.13}$$

Proposition 1.3 Les solutions $\lambda_1, \lambda_2, \dots, \lambda_p$ de l'équation caractéristique de la proposition 1.1 sont les inverses des racines du polynôme $\varphi(z)$ de variable (complexe) z où

$$\varphi(z) = 1 - \theta_1 z - \dots - \theta_p z^p.$$

Les racines de $\varphi(z)$ soit donc λ_i^{-1} , $i = 1, \dots, p$. L'équation (1.13), qui peut s'écrire,

$$(1 - \lambda_1 L)(1 - \lambda_2 L) \dots (1 - \lambda_p L)Y_t = w_t$$

1.1 Équation aux différences linéaires à coefficients constants

est stable (i.e. Y_t reste borné, si w_t est borné), si $|\lambda_i| < 1$, donc $|\lambda_i^{-1}| > 1$, c'est-à-dire si les racines de $\varphi(z)$ sont en module > 1 ("à l'extérieur du disque-unité"), ou, de manière équivalente, si les valeurs propres de $\tilde{\theta}$ sont en module < 1 .

Supposons (comme avant) que les λ_i soient distinctes et que $|\lambda_i| < 1$, $i = 1, \dots, p$. Invertissons la relation (1.13) formellement.

$$\varphi(z)Y_t = w_t \Leftrightarrow Y_t = \varphi^{-1}(z)w_t.$$

Peut-on donner un sens à cette expression ?

On a

$$\varphi^{-1}(z) = \frac{1}{(1 - \lambda_1 z)(1 - \lambda_2 z) \dots (1 - \lambda_p z)} = \psi(z).$$

Par décomposition en éléments simples, on a :

$$\psi(z) = \frac{c_1}{(1 - \lambda_1 z)} + \frac{c_2}{(1 - \lambda_2 z)} + \dots + \frac{c_p}{(1 - \lambda_p z)} \quad (1.14)$$

comme $\psi(0) = 1$ on a $\sum_{k=1}^p c_k = 1$. De plus,

$$\begin{aligned} \lim_{z \rightarrow \lambda_k^{-1}} \psi(z)(1 - \lambda_k z) = c_k &= \left[\prod_{\substack{i=1 \\ i \neq k}}^p (1 - \lambda_i \lambda_k^{-1}) \right]^{-1} \\ &= \frac{\lambda_k^{p-1}}{\prod_{\substack{i=1 \\ i \neq k}}^p (\lambda_k - \lambda_i)}. \end{aligned} \quad (1.15)$$

Comme, dans l'expression suivante, chaque série entre parenthèses converge :

$$\begin{aligned} \psi(z) &= \sum_{k=1}^p c_k (1 + \lambda_k z + \lambda_k^2 z^2 + \lambda_k^3 z^3 + \dots) \\ &= \underbrace{\sum_{k=1}^p c_k}_{=1} + \left(\sum_{k=1}^p c_k \lambda_k \right) z + \left(\sum_{k=1}^p c_k \lambda_k^2 \right) z^2 + \dots \\ &= 1 + \sum_{j=1}^{+\infty} \underbrace{\left(\sum_{k=1}^p c_k \lambda_k^j \right)}_{\psi_j} z^j \\ &= 1 + \sum_{j=1}^{+\infty} \left[\sum_{k=1}^p \frac{\lambda_k^{p-1+j}}{\prod_{\substack{i=1 \\ i \neq k}}^p (\lambda_k - \lambda_i)} \right] z^j. \end{aligned} \quad (1.16)$$

1.2 Notions sur les processus stochastiques

Et on trouve le résultat de la proposition 1.2. On peut alors écrire explicitement Y_t en fonction du passé de la série $(w_t, t \in \mathbb{Z})$:

$$Y_t = w_t + \sum_{j=1}^{+\infty} \psi_j w_{t-j} \simeq w_t + \sum_{j=1}^{\nu} \psi_j w_{t-j}$$

où ν est choisi assez grand pour qu'on puisse négliger le reste.

1.2 Notions sur les processus stochastiques

Les processus stochastiques sont des outils de modélisation, sont utilisés dans un grand nombre de domaines scientifiques.

Définition 1.1 *Un processus stochastique (ou aléatoire) est une famille de variables aléatoires $\{X_t; t \in T\}$ toutes définies sur le même ensemble fondamental Ω . Les variables aléatoires peuvent être discrètes ou continues et l'ensemble d'indices peut être :*

1. $T = \{0, 1, 2, \dots\}$ ou $T = \mathbb{Z}$. Dans ce cas on parle d'un processus en temps discret.
2. $T = [0, +\infty)$ ou $T = \mathbb{R}$. Le processus est alors en temps continu.

Il y a deux niveaux auxquels on peut observer l'évolution d'un processus :

- Fixons un $\omega \in \Omega$ (une réalisation de l'expérience aléatoire sous-jacente). L'ensemble $\{X_t; t \in T\}$ est une réalisation du processus stochastique, aussi appelée trajectoire.

On peut étudier les propriétés des trajectoires.

- Fixons n instants (t_1, t_2, \dots, t_n) , avec $t_1 < t_2 < \dots < t_n$ où n est un entier positif arbitraire.

Considérons le vecteur aléatoire $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$, extrait du processus, et notons sa fonction de répartition conjointe :

$$F_{t_1, t_2, \dots, t_n}(x_1, x_2, \dots, x_n) = \mathbb{P}(X_{t_1}, X_{t_2}, \dots, X_{t_n}). \quad (1.17)$$

Les propriétés distributionnelles du processus sont données par la collection de toutes les fonctions de répartition F_{t_1, t_2, \dots, t_n} pour tout $n \geq 1$ fini et tout n -uple (t_1, t_2, \dots, t_n) de T . Pour obtenir des processus utiles, il faut particulariser.

1.2.1 Processus stationnaire

Définition 1.2 *Le processus $\{X_t; t \in T\}$ est dit fortement (ou strictement) stationnaire, si pour*

1.2 Notions sur les processus stochastiques

tout $n \geq 1$ fini;

tout $t_1 < t_2 < \dots < t_n \in T$;

tout h tel que $t_i + h \in T$, $i = 1, 2, \dots, n$:

$$F_{t_1, t_2, \dots, t_n}(x_1, x_2, \dots, x_n) = F_{t_1+h, t_2+h, \dots, t_n+h}(x_1, x_2, \dots, x_n). \quad (1.18)$$

C'est à dire si les fonctions de répartition du processus sont toutes invariantes par décalage dans le temps. Nous nous concentrons ici sur les processus en temps discret.

Exemples en temps discret

1. une suite $\{X_t; t \in T\}$, $T = \{0, 1, 2, \dots\}$, de variables aléatoires indépendantes de même loi de Bernoulli est un processus en temps discret strictement stationnaire.

2. La suite $\{Y_t; t \in T\}$ définie par $Y_t = (-1)^t X_t$ où $\{X_t\}$ est le processus ci-dessus, n'est pas strictement stationnaire.

3. Soit $\{\zeta_t; t \in T\}$ une suite de variables aléatoires indépendantes de même loi normale $\mathcal{N}(0, \sigma^2)$. Elles forment un processus strictement stationnaire appelé **bruit blanc gaussien**. Si les moments d'ordre 1 et 2 existent, on se contente souvent d'une notion de stationnarité plus faible que la précédente.

La suite de variables aléatoires $\{X_t; t \in \mathbb{Z}\}$ est appelée **processus stochastique**. Les observations (données) empiriques X_1, \dots, X_T sont interprétées comme des réalisations particulières de ce processus, tandis que $\{X_t; t \in \mathbb{Z}\}$, lorsqu'on ajoute des conditions sur sa structure probabiliste (ses moyennes, ses variances et covariances, ...) sert à dénoter le modèle (paramétrique) théorique. Celui-ci aidera à mieux comprendre le comportement "en moyenne" et des propriétés statistiques des estimateurs des paramètres du modèle.

Ayant introduit la notion d'un processus stochastique servant à modéliser des séries chronologiques, on se concentre maintenant sur la classe importante des **processus stationnaires**. La définition suivante caractérise les suites qui sont stationnaires en moyenne et dont la structure de covariance reste elle aussi constante.

Définition 1.3 Un processus est dit faiblement stationnaire (ou stationnaire du seconde ordre), si

i) tous les moments d'ordre 1 et 2 existent (i.e. $\mathbb{E}(X_t) < \infty$ et $\mathbb{E}(X_t^2) < \infty$);

ii) les espérances $\mathbb{E}(X_t) = \mu$ sont identiques pour tout t ;

iii) les covariances $\text{Cov}(X_t, X_{t+h}) = \gamma(h)$ ne dépendent que de h pour tout t et $t+h \in T$.

Dans cette définition, la propriété i) exprime la stationnarité en moyenne de la suite, ii) assure que la variance de chaque variable reste finie, et iii) précise ce qu'on entend par

"invariance". Par cette propriété, on peut introduire que

$$\text{Cov}(X_t, X_{t+h}) = \text{Cov}(X_{t-1}, X_{t-1+h}) = \dots = \text{Cov}(X_0, X_{t+h}) \quad \forall t, h \in \mathbb{N}.$$

Définition 1.4 $\gamma(h)$ s'appelle la fonction d'autocovariance (fac) du processus. En particulier

$$\gamma(0) = \text{Var}(X_t).$$

La fonction d'autocorrélation $\rho(h)$ est donné par

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

Remarques

1. Un processus fortement stationnaire du second ordre, si les moments d'ordre 1 et 2 n'existent pas.
2. Un processus stationnaire du second ordre n'est pas nécessairement fortement stationnaire : posons

$$X_t = Z_1 \cos \psi t + Z_2 \sin \psi t$$

où ψ est une constante et Z_1 et Z_2 sont des variables aléatoires non corrélées avec $\mathbb{E}(Z_1) = \mathbb{E}(Z_2) = 0$ et $\text{Var}(Z_1) = \text{Var}(Z_2) = 1$. Alors

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

$$\text{Cov}(X_t, X_{t+h}) = \mathbb{E}[(Z_1 \cos \psi t + Z_2 \sin \psi t)(Z_1 \cos \psi(t+h) + Z_2 \sin \psi(t+h))] = \cos \psi h$$

Sont indépendantes de t . $\{X_t; t \in T\}$ est stationnaire du second ordre. Il est évident que X_{t_1} et X_{t_2} n'ont pas nécessairement la même loi. Le processus n'est donc pas fortement stationnaire.

Définition 1.5 La suite de variables aléatoires $\{\zeta_t\}_{t=0,1,\dots}$ constitue un bruit blanc faible (respectivement fort) si les trois propriétés suivantes sont respectées.

- i) $\mathbb{E}(\zeta_t) = 0 \forall t \in Z$;
- ii) $\mathbb{E}(\zeta_t^2) = \sigma^2$ est constante et strictement positive ;
- iii) $\text{Cov}(\zeta_t, \zeta_s) = 0$ si $t \neq s$ (respectivement, les ζ_t sont i.i.d.).

Si la suite $\{\zeta_t\}_{t=0,1,\dots}$ est un bruit blanc faible, on notera $\{\zeta_t\} \sim WN(0, \sigma^2)$. Si $\{\zeta_t\}$ constituent un bruit blanc fort, on notera $\{\zeta_t\} \sim IID(0, \sigma^2)$.

Le bruit blanc, même non portant de la structure de corrélation non triviale ("processus purement aléatoire"), jouera un rôle prépondérant dans notre modélisation : Il intervient

1.2 Notions sur les processus stochastiques

comme module de base pour construire des processus stationnaires portant des corrélations intéressantes ainsi qu'il, servira comme modèle pour les "résidus" d'une modélisation, la composante qui reste après avoir modélisé toute structure de corrélation intéressante et qui est supposée de ne plus contribuer à cette structure non triviale. En effet, tester si ces résidus sont compatibles avec un bruit blanc sera un des moyens principaux pour une diagnostique du modèle ajusté aux données stationnaires mais autocorrélées.

1.2.2 Propriétés de la fonction d'autocovariance

Soit $\gamma(\cdot)$ la fac (fonction d'autocovariance) d'un processus en temps discret faiblement stationnaire $\{X_t; t \in \mathbb{Z}\}$. Alors, pour tout h :

- i) $\gamma(0) \geq 0$;
- ii) $|\gamma(h)| \leq \gamma(0)$;
- iii) $\gamma(-h) = \gamma(h)$.

Ces conditions ne sont que nécessaires. Il faut encore que la fac soit définie non négative.

Remarque

1. On constate qu'on peut associer un processus gaussien à chaque fac.
2. une fonction d'autocorrélation possède toutes les propriétés d'une fac, avec en plus $\rho(0) = 1$.

1.2.3 Processus linéaires et processus linéaires généraux

Un processus linéaire est un processus stochastique X_t formé par une combinaison linéaire (non nécessairement finie) de bruits blancs forts. On définit également la classe des processus linéaires généraux, qui sont constitués de combinaisons linéaires de bruits blancs faibles. Introduisons formellement ces deux types importants de processus.

Définition 1.6 $\{X_t\}_{t \in \mathbb{Z}}$ est un processus linéaire (respectivement un processus linéaire général) de moyenne μ s'il peut être écrit sous la forme :

$$X_t = \mu + \sum_{k=-\infty}^{+\infty} b_k \zeta_{t-k} \quad (1.19)$$

où $\{\zeta_t\}_{t \in \mathbb{Z}}$ est un bruit blanc fort (respectivement faible), avec variance σ_ζ^2 , et où la suite des coefficients b_k est supposée telle que

$$\sum_{k=-\infty}^{+\infty} b_k^2 < \infty. \quad (1.20)$$

1.2 Notions sur les processus stochastiques

Dans cette dernière définition, nous devons préciser le sens qui est donné à la série (1.19). Pour ce faire, considérons une suite de nombres $m_n \rightarrow \infty$ et formons le processus linéaire tronqué

$$\mu + \sum_{k=-m_n}^{m_n} b_k \zeta_{t-k},$$

qui est bien défini puisque le nombre de termes dans la somme est fini. La convergence (1.19) signifie que la moyenne quadratique

$$\mathbb{E}[(\mu + \sum_{k=-m_n}^{m_n} b_k \zeta_{t-k} - X_t)^2] \leq \sum_{|k| \geq m_n} b_k^2 \mathbb{E}|\zeta_{t-k}|^2 = \sigma_\zeta^2 \sum_{|k| \geq m_n} b_k^2$$

tend vers 0 lorsque $n \rightarrow \infty$ (et donc lorsque $m_n \rightarrow \infty$).

La restriction (1.20) faite sur les coefficients b_k est usuelle, et signifie que, lorsque k grandit, les coefficients $|b_k|$ décroissent rapidement (autrement, la somme infinie (1.20) ne saurait être finie). La condition (1.20) se dit également " b_k est de carré sommable", ou encore $b_k \in l^2(\mathbb{Z})$. En terme de modélisation, cette condition signifie que l'influence de la série $\{\zeta_t\}$ est limitée dans le temps : X_t sera principalement déterminé par le bruit blanc aux instants proches de t . La condition (1.20) est également nécessaire pour que la série (1.19) converge au sens de la moyenne quadratique, comme nos considérations ci-dessus le montrent.

Le résultat suivant montre que les processus linéaires que nous avons définis sont des processus stationnaires

Proposition 1.4 *Si $\{X_t\}$ est un processus linéaire général défini par (1.19) avec $\zeta_t \sim WN(0, \sigma_\zeta^2)$, alors $\{X_t\}$ est stationnaire et on a :*

$$\gamma(0) = \text{Var}(X_t) = \sigma_\zeta^2 \sum_{-\infty}^{\infty} b_k^2,$$

$$\gamma(s) = \sigma_\zeta^2 \sum_{-\infty}^{\infty} b_k b_{k+s} \quad \forall s \in \mathbb{Z}.$$

Preuve On a déjà remarqué ci-dessus que $\mathbb{E}(X_t) = \mu$ (pour tout $t \in \mathbb{Z}$). Pour vérifier les deux autres conditions, on calcule la covariance entre X_t et X_{t+h}

$$\begin{aligned} \text{Cov}(X_t, X_{t+h}) &= \mathbb{E}(X_t X_{t+h}) - \mathbb{E}(X_t) \mathbb{E}(X_{t+h}) \\ &= \mathbb{E}(\mu^2 + \mu \sum_{l=-\infty}^{\infty} b_l \zeta_{t+h-l} + \mu \sum_{k=-\infty}^{\infty} b_k \zeta_{t-k}) \end{aligned}$$

$$\begin{aligned}
& + \sum_{k,l=-\infty}^{\infty} b_k b_l \zeta_{t-k} \zeta_{t+h-l} - \mu^2 \\
& = \mu \sum_{l=-\infty}^{\infty} b_l \mathbb{E}(\zeta_{t+h-l}) + \mu \sum_{k=-\infty}^{\infty} b_k \mathbb{E}(\zeta_{t-k}) \\
& + \sum_{k,l=-\infty}^{\infty} b_k b_l \mathbb{E}(\zeta_{t-k} \zeta_{t+h-l})
\end{aligned}$$

où le passage de l'espérance dans la somme infini est justifiée par le théorème de la convergence monotone et le fait que $\mathbb{E}(X_t) < \infty$ pour tout t . Puisque $\zeta_t \sim WN(0, \sigma_\zeta^2)$, on a donc

$$\begin{aligned}
Cov(X_t, X_{t+h}) & = \sum_{k,l=-\infty}^{\infty} b_k b_l Cov(\zeta_{t-k} \zeta_{t+h-l}) \\
& = \sigma_\zeta^2 \sum_{k,l=-\infty}^{\infty} b_k b_l \delta_{t-k, t+h-l} \\
& = \sigma_\zeta^2 \sum_{k,l=-\infty}^{\infty} b_k b_{k+h}.
\end{aligned}$$

Donc $Cov(X_t, X_{t+h})$ ne dépend pas de t et on a les résultats annoncés.

1.2.4 Processus autorégressif $AR(p)$

Définition 1.7 Un processus faiblement stationnaire $\{X_t; t \in \mathbb{Z}\}$ est dit autorégressif d'ordre p , noté $AR(p)$, s'il vérifie :

$$X_t = \theta_1 X_{t-1} + \theta_2 X_{t-2} + \dots + \theta_p X_{t-p} + \zeta_t \quad (1.21)$$

$\mathbb{E}(X_t) = 0$, où $\{\zeta_t\} \sim WN(0, \sigma^2)$. En isolant ζ_t on trouve :

$$\begin{aligned}
(1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_p L^p) X_t & = \zeta_t \\
\Gamma(L) X_t & = \zeta_t.
\end{aligned} \quad (1.22)$$

Formellement, les deux processus $\{X_t; t \in \mathbb{Z}\}$ et $\{\zeta_t; t \in \mathbb{Z}\}$ sont liés par l'opérateur $\Gamma(L)$ qui est un polynôme de degré p dans l'opérateur de décalage. On généralise facilement (1.22) à un processus d'espérance μ :

$$\Gamma(L)(X_t - \mu) = \zeta_t$$

Nous pouvons donc, sans perdre de généralité, supposer que $\mathbb{E}(X_t) = 0$.

Causalité : Prenons un processus $AR(1)$ et itérons :

$$X_t = \theta X_{t-1} + \zeta_t$$

$$\begin{aligned}
 &= \theta(\theta X_{t-2} + \zeta_{t-1}) + \zeta_t \\
 &= \theta^2 X_{t-2} + \theta \zeta_{t-1} + \zeta_t \\
 &\vdots \\
 &= \theta^n X_{t-n} + \theta^{n-1} \zeta_{t-(n-1)} + \dots + \zeta_t \\
 &= \theta^n X_{t-n} + \sum_{i=0}^{n-1} \theta^i \zeta_{t-i} \\
 &= \theta^n X_{t-n} + \left(\sum_{i=0}^{n-1} \theta^i L^i \right) \zeta_t.
 \end{aligned}$$

i) Si $|\theta| < 1$, et si n très grand, la contribution de $\theta^n X_{t-n}$ tend vers 0, et comme $\sum_{i=0}^{+\infty} \theta^i = (1 - \theta)^{-1}$ converge, X_t converge presque sûrement et en moyenne quadratique :

$$\begin{aligned}
 X_t \longrightarrow \sum_{i=0}^{+\infty} \theta^i \zeta_{t-i} &= \left(\sum_{i=0}^{+\infty} \theta^i L^i \right) \zeta_t \\
 &= (1 - \theta L)^{-1} \zeta_t
 \end{aligned}$$

On dit que $\{X_t; t \in \mathbb{Z}\}$ possède une représentation causale en fonction du passé de bruit blanc $\{\zeta_t; t \in \mathbb{Z}\}$ car dans ce cas, X_t est complètement déterminé par le passé du processus $\{\zeta_t; t \in \mathbb{Z}\}$;

ii) Si $|\theta| > 1$, la représentation de $\{X_t; t \in \mathbb{Z}\}$ en fonction de $\{\zeta_t; t \in \mathbb{Z}\}$ est non causale. On peut néanmoins représenter $\{X_t; t \in \mathbb{Z}\}$ en fonction des valeurs **futurs** de $\{\zeta_t; t \in \mathbb{Z}\}$. Dans la pratique, ce n'est pas très utile et on cherche toujours (et on pourra toujours trouver) un bruit blanc pour lequel $\{X_t; t \in \mathbb{Z}\}$ a une représentation causale. La représentation canonique est alors

$$X_t - \frac{1}{\theta} X_{t-1} = \eta_t$$

où

$$\eta_t = (1 - \theta F)(1 - \theta L)^{-1} \zeta_t = -\theta \zeta_{t+1} + (1 - \theta^2) \sum_{i=0}^{\infty} \theta^i \zeta_{t-i};$$

iii) Si $|\theta| = 1$, il n'existe pas une représentation stationnaire de $\{X_t; t \in \mathbb{Z}\}$ en fonction de $\{\zeta_t; t \in \mathbb{Z}\}$. Par exemple, pour $\theta = 1$, $X_t = X_{t-1} + \zeta_t$, qui peut s'écrire $X_t - X_{t-1} = \zeta_t + \zeta_{t-1} + \dots + \zeta_{t-h+1}$ et donc $\mathbb{E}(X_t - X_{t-h}) = h\sigma^2$. Or pour un processus

1.2 Notions sur les processus stochastiques

stationnaire, il est possible de montrer que $\mathbb{E}(X_t - X_{t-h})^2 \leq 4\text{Var}(X_t)$. Puisqu'il est impossible que pour tout h , $h\sigma^2 \leq 4\text{Var}(X_t)$, le processus n'est pas stationnaire. Donc une marche aléatoire n'est pas stationnaire car la suite des covariances dépend de h . En fait, on observe en plus que sa variance n'est pas constante mais augmente en temps h .

L'évolution de X_t est aléatoire, et on ne peut pas faire de prévision car il manque une structure adéquate.

Pour illustrer ce processus non stationnaire, on peut penser aux prix des actifs d'aujourd'hui donnant une information sur les prix du lendemain, mais dont la différence est quand même aléatoire, avec un certain contrôle de la variabilité ($\sigma_\zeta^2 < \infty$). Ici non plus, on ne peut pas faire de prévision, car $\{\zeta_t\}$ ne possède pas une structure des corrélations.

Fac d'un processus causale : un processus $AR(1)$ causale est stationnaire.

En effet

$$\begin{aligned}
 \mathbb{E}(X_t) &= \mathbb{E}\left(\lim_{n \rightarrow +\infty} \sum_{i=0}^n \theta^i \zeta_{t-i}\right) \\
 &= \lim_{n \rightarrow +\infty} \sum_{i=0}^n \theta^i \mathbb{E}(\zeta_{t-i}) = 0 \\
 \text{Cov}(X_t, X_{t+h}) &= \mathbb{E}\left(\lim_{n \rightarrow +\infty} \sum_{i=0}^n \theta^i \zeta_{t-i} \sum_{j=0}^n \theta^j \zeta_{t+h-j}\right) \\
 &= \mathbb{E}\left(\lim_{n \rightarrow +\infty} \sum_{j=0}^n \theta^{|h|+2j} \zeta_{t+h-j}^2\right) \\
 &= \lim_{n \rightarrow +\infty} \sum_{j=0}^n \theta^{|h|+2j} \underbrace{\mathbb{E}(\zeta_{t+h-j}^2)}_{\sigma^2} \\
 &= \theta^{|h|} \sum_{i=0}^{+\infty} \theta^{2i} \sigma^2 = \frac{\theta^{|h|}}{1 - \theta^2} \sigma^2 \\
 &= \gamma(h).
 \end{aligned}$$

En particulier :

$$\text{Var}(X_t) = \gamma(0) = \frac{1}{1 - \theta^2} \sigma^2.$$

Les moments d'ordre 1 et 2 ne dépendent donc pas de t et le processus est bien stationnaire d'ordre 2.

1.2 Notions sur les processus stochastiques

Pour que ces formules aient un sens, on voit qu'il faut que $|\theta| < 1$, sinon, par exemple, la variance serait négative ou infinie.

La propriété de **causalité** peut aussi s'exprimer à l'aide de la racine du polynôme (d'ordre 1) $\Gamma(z) = 1 - \theta z$, qui vaut θ^{-1} : le processus $AR(1)$ est causale, si $|\theta^{-1}| > 1$.
 A retenir : un processus $AR(1) : X_t = \theta X_{t-1} + \zeta_t$ sera auto-corrélé positivement si $0 < \theta < 1$, et auto-corrélé négativement si $-1 < \theta < 0$. Cette série va osciller autour de 0, en s'en écartant suivant la valeur ζ_t du processus d'innovation (si $-1 < \theta < 1$). Si $\theta = +1$, on obtient un marche aléatoire, et si $\theta > +1$ ou $\theta < -1$ le processus n'est pas stationnaire, et on obtient un modèle qui explosera (à moyen terme). La valeur θ , dans le cas où le processus est stationnaire, est la corrélation entre deux dates consécutives $\theta = corr(X_t, X_{t-1})$.

$$AR(1) = \begin{cases} \text{Fonction d'autocorrélation} & \begin{cases} \theta > 0 & \text{décroissance exponentielle.} \\ \theta < 0 & \text{sinusoïde amortie.} \end{cases} \\ \text{Fonction d'autocorrélation partielle} & \begin{cases} \text{première non nulle.} \\ \text{toutes nulles après.} \end{cases} \end{cases}$$

Le processus $AR(2)$ Ces processus sont également appelés modèles de Yule, dont la forme générale est

$$(1 - \theta_1 L - \theta_2 L^2) X_t = \zeta_t$$

où les racines du polynôme caractéristique $\varphi(z) = 1 - \theta_1 z - \theta_2 z^2$ sont supposées à l'intérieur du disque unité (de telle sorte que le processus ζ_t corresponde à l'innovation). Cette condition s'écrit

$$\begin{cases} 1 - \theta_1 + \theta_2 > 0 \\ 1 - \theta_1 - \theta_2 > 0 \\ \theta_1^2 + 4\theta_2 > 0 \end{cases}$$

C'est à dire le couple (θ_1, θ_2) doit se trouver dans 4 parties.

A retenir : Le comportement d'un processus $AR(2) : X_t = \theta_1 X_{t-1} + \theta_2 X_{t-2} + \zeta_t$ dépendra fortement des racines de son équation caractéristique $1 - \theta_1 z - \theta_2 z^2 = 0$. Le cas le plus intéressant est celui où l'équation caractéristique a deux racines complexes conjuguées $r \exp(\pm i\theta)$ pour $r < 1$: le processus est alors stationnaire (et oscille alors autour de 0, sans exploser, de la même façon que les processus $AR(1)$ dans le cas où $|\theta| < 1$).

Le processus est alors quasi-cyclique, de fréquence θ , avec un bruit aléatoire.

$$AR(2) = \begin{cases} \text{Fonction d'autocorrélation} & \begin{cases} \text{décroissance exponentielle.} \\ \text{où sinusoïde amortie.} \end{cases} \\ \text{Fonction d'autocorrélation partielle} & \begin{cases} \text{deux première non nulles.} \\ \text{toutes nulles après.} \end{cases} \end{cases}$$

Plus généralement, le processus $AR(p)$ est causal, si toutes les racines du polynôme $\Gamma(z)$ sont de module > 1 . On peut alors exprimer X_t en fonction du passé du bruit blanc $\{\zeta_t; t \in \mathbb{Z}\}$. Supposons (pour simplifier) que toutes ces racines soient distinctes. Par (15), on peut alors écrire

$$X_t = \sum_{j=0}^{+\infty} \psi_j \zeta_{t-j}$$

où ψ_j peut s'écrire explicitement en fonction des racines de $\Gamma(z)$. Sa fac est facilement exprimé à l'aide des ψ_j : on a

$$\gamma(h) = \sigma^2 \sum_{j=0}^{+\infty} \psi_j \psi_{j+|h|}. \quad (1.23)$$

En particulier, la variance de $\{X_t; t \in \mathbb{Z}\}$ s'écrit :

$$\gamma(0) = \sigma^2 \sum_{j=0}^{+\infty} \psi_j^2.$$

On voit, par exemple, que si $\psi_j = \theta^j$ (processus $AR(1)$), on retrouver la formule donnée plus haut :

$$\sum_{j=0}^{+\infty} \psi_j^2 = \sum_{j=0}^{+\infty} (\theta^2)^j = (1 - \theta^2)^{-1}.$$

1.2.5 Équations de Yule-Walker

Prenons un processus autorégressif causal. Multiplions les deux membres de (1.21) par X_{t-j} et prenons l'espérance

$$\mathbb{E}(X_t X_{t-j}) = \sum_{i=1}^p \theta_i \mathbb{E}(X_{t-i} X_{t-j}) + \mathbb{E}(\zeta_t X_{t-j})$$

En utilisant la représentation causale,

$$\begin{aligned} \mathbb{E}(\zeta_t X_{t-j}) &= \sum_{k=0}^{+\infty} \psi_k \mathbb{E}(\zeta_t \zeta_{t-j-k}) \\ &= \begin{cases} \sigma^2 & \text{si } j = 0 \\ 0 & \text{si } j > 0 \end{cases} \end{aligned}$$

donc

$$\gamma(j) = \sum_{i=1}^p \theta_i \gamma(i-j) \quad j = 1, \dots, p \quad (1.24)$$

$$\gamma(0) = \sum_{i=1}^p \theta_i \gamma(i) + \sigma^2 \Rightarrow \sigma^2 = \gamma(0) - \sum_{i=1}^p \theta_i \gamma(i). \quad (1.25)$$

En regroupant, (1.24) sous forme matricielle, les p équations (1.24) pour $1 \leq j \leq p$, on obtient

$$\begin{pmatrix} \gamma(0) & \gamma(1) \cdots & \cdots \gamma(p-2) & \gamma(p-1) \\ \gamma(1) & \gamma(0) \cdots & & \gamma(p-2) \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ \gamma(p-1) & \gamma(p-2) \cdots & & \gamma(0) \end{pmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_{p-1} \\ \theta_p \end{pmatrix} = \begin{pmatrix} \gamma(1) \\ \gamma(2) \\ \vdots \\ \gamma(p-1) \\ \gamma(p) \end{pmatrix} \quad (1.26)$$

Les équations (1.25) et (1.26) sont appelées équations de Yule-Walker.

Ces équations permettent également de déterminer les valeurs des paramètres du modèle à partir d'estimation de la fonction d'autocovariance.

1.2.6 Estimation du processus autorégressif d'ordre un AR(1)

Considérons, sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, un processus autorégressif d'ordre un AR(1) $(X_t)_{t \in \mathbb{Z}}$, avec de fonction de covariance $\gamma(\cdot)$, défini par

$$\underbrace{X_t - \theta X_{t-1}}_{\text{partie autorégressif}} = \underbrace{\zeta_t}_{\text{bruit blanc}}$$

où la suite $(\zeta_t)_{t \in \mathbb{Z}}$ est un processus stationnaire, étant donné un ensemble d'observations $\{X_t, 1 \leq t \leq n\}$ nous souhaitons estimer le coefficient (paramètre inconnu) θ du modèle AR(1). Pour cela on utilise l'estimateur des **moindres carrés**.

Notons $\hat{\theta}_n$ l'estimateur des moindres carrés de θ obtenu par minimisation de la quantité suivante

$$Q(\theta) = \sum_{t=1}^n (X_t - \theta X_{t-1})^2.$$

Cette minimisation revient à résoudre l'équations linéaire suivant

$$\frac{\partial Q(\theta)}{\partial \theta} = \sum_{t=1}^n X_{t-1} (X_t - \theta X_{t-1}) = 0.$$

En introduisant la quantité suivante :

$$\hat{\theta}_n = \frac{\sum_{t=1}^n X_{t-1} X_t}{\sum_{t=1}^n X_{t-1}^2}. \quad (1.27)$$

1.2.7 Estimation du processus autorégressif d'ordre p $AR(p)$

Dans cette section nous nous intéressons aux problèmes de l'estimation des paramètres d'un processus $AR(p)$ à partir d'une suite de n observations. Nous supposons que les données ont été préalablement traitées de façon à supprimer d'éventuelles **tendances affine** et/ou **saisonnière**. L'estimation des paramètres d'un processus $AR(p)$ comprend aussi, en principe, l'estimation des ordres ne sera pas traité dans cette thèse. Nous supposons donc que p est connu et nous nous intéressons uniquement à l'estimation des paramètres $\{\theta_j, 1 \leq j \leq p\}$, et σ^2 intervenant dans l'équation récurrente définissant le processus $X_t = \theta_1 X_{t-1} + \theta_2 X_{t-2} + \dots + \theta_p X_{t-p} + \zeta_t$.

On verra que, pour obtenir de bons estimateurs de $\{\theta_j, 1 \leq j \leq p\}$ et de σ^2 , il suffit de partir des $(p+1)$ premiers coefficients d'autocovariance empirique et de résoudre les équations de **Yule-Walker**. Cela signifie que, quel que soit n les observations n'interviennent, dans l'expression de l'estimateur, que par un nombre fixé, égal à $p+1$, de valeur de la covariance empirique

$$\hat{\gamma}_n(h) = \frac{1}{n} \sum_{t=1}^{n-h} (X_{t+h} - \hat{\mu}_n)(X_t - \hat{\mu}_n) \quad (1.28)$$

où $0 \leq h \leq p$ et $\hat{\mu}_n = n^{-1} \sum_{t=1}^n X_t$. La construction de bons estimateurs ne peut se faire avec un nombre fixé (indépendant de n) de valeurs de la suite des covariances empiriques. Il existe de nombreuses méthodes. Nous avons établi, une relation simple (équations (1.24) de Yule-Walker) entre les $(p+1)$ coefficients du modèle et les $(p+1)$ premiers coefficients d'autocovariance d'un processus $AR(p)$ **causal** défini par l'équation récurrente : $X_t = \theta_1 X_{t-1} + \theta_2 X_{t-2} + \dots + \theta_p X_{t-p} + \zeta_t$. En posant $\Theta = (\theta, \theta_2, \dots, \theta_p)$, $\gamma_p = (\gamma(1), \gamma(2), \dots, \gamma(p))^T$ et

$$\tilde{\gamma}_p = \begin{pmatrix} \gamma(0) & \gamma(1) \cdots & \cdots \gamma(p-2) & \gamma(p-1) \\ \gamma(1) & \gamma(0) \cdots & & \gamma(p-2) \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ \gamma(p-1) & \gamma(p-2) \cdots & & \gamma(0) \end{pmatrix}.$$

Les équations de **Yule-Walker** ont pour expression matricielle :

$$\tilde{\gamma}_p \Theta = \gamma_p, \quad (1.29)$$

$$\sigma^2 = \gamma(0) - \Theta^T \gamma_p. \quad (1.30)$$

En substituant, dans ces relations, les covariances $\gamma(h)$ par les covariances empirique $\hat{\gamma}(h)$, on obtient un système linéaire qui fournit les estimateurs $\hat{\Theta}_n$ et $\hat{\sigma}_n^2$ comme solution de :

$$\hat{\tilde{\gamma}}_p \hat{\Theta}_n = \hat{\gamma}_p, \quad (1.31)$$

$$\hat{\sigma}_n^2 = \hat{\gamma}(0) - \hat{\Theta}_n^T \hat{\gamma}_p. \quad (1.32)$$

Si $\hat{\gamma}_p > 0$, alors $\hat{\gamma}_p$ est de rang plein. En divisant alors les deux membres de $\hat{\gamma}_p^T \hat{\Theta}_n = \hat{\gamma}_p$ par $\hat{\gamma}(0)$ et en introduisant l'autocorrélation empirique $\hat{\rho}(h) = \hat{\gamma}(h)/\hat{\gamma}(0)$, on aboutit aux deux équations :

$$\hat{\Theta}_n = \hat{C}_{-1}^p \hat{\Upsilon}_p \quad (1.33)$$

$$\hat{\sigma}_n^2 = \hat{\gamma}(0)(1 - \hat{\Upsilon}_p^T \hat{C}_p^{-1} \hat{\Upsilon}_p) \quad (1.34)$$

où

$$\hat{\Upsilon}_p = (\hat{\rho}(1), \hat{\rho}(2), \dots, \hat{\rho}(p))$$

et

$$\hat{C}_p^{-1} = \begin{pmatrix} \hat{\rho}(1) & \hat{\rho}(1) & \cdots & \cdots & \hat{\rho}(p) \\ \hat{\rho}(1) & \hat{\rho}(0) & \cdots & \cdots & \hat{\rho}(p-1) \\ \vdots & \ddots & & \vdots & \vdots \\ \vdots & & \ddots & \vdots & \vdots \\ \hat{\rho}(p) & \hat{\rho}(p-1) & \cdots & \cdots & \hat{\rho}(0) \end{pmatrix}.$$

Les théorèmes suivants précisent le comportement asymptotique de la suite Θ et permettent alors de construire des intervalles de confiance ou de fournir des tests d'hypothèse.

Théorème 1.1 *Soit X_t un processus $AR(p)$ causal où $\zeta_t \sim IID(0, \sigma^2)$ et soit un échantillon $\{X_1, X_2, \dots, X_n\}$ de taille n . On note $\hat{\Theta}_n = \hat{C}_p^{-1} \hat{\Upsilon}_p$ et $\hat{\sigma}_n^2 = \hat{\gamma}(0)(1 - \hat{\Upsilon}_p^T \hat{C}_p^{-1} \hat{\Upsilon}_p)$.*

Alors, quand $n \rightarrow \infty$, on a :

$$\begin{cases} \hat{\sigma}_n & \xrightarrow{\mathbb{P}} \sigma^2 \\ \sqrt{n}(\hat{\Theta}_n - \Theta) & \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2 \tilde{\gamma}_p^{-1}) \end{cases}$$

Théorème 1.2 *Soit X_t un processus $AR(p)$ causal où $\zeta_t \sim IID(0, \sigma^2)$ et soit échantillon $\{X_1, X_2, \dots, X_n\}$ de taille n . On note $\hat{\Theta}_n = \hat{C}_m^{-1} \hat{\Upsilon}_m$ où $m > p$. Alors, quand $n \rightarrow \infty$, on a :*

$$\sqrt{n}(\hat{\Theta}_n - \Theta_m) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2 \tilde{\gamma}_m^{-1}) \quad (1.35)$$

où $\Theta_m = \{\theta_1, \dots, \theta_p, 0, \dots, 0\}$ est la suite du meilleur prédicteur linéaire de X_t en fonction de $\{X_{t-1}, \dots, X_{t-m}\}$.

Preuve du théorème 1.1 et 1.2 : Voir [75]

1.2.8 Les modèles ARMA

Les modèles ARMA sont un mélange des modèles AR et MA proposés par Yule et Slutsky. Un processus $(X_t)_{t \in \mathbb{Z}}$ est un processus ARMA(p, q) s'il existe un bruit blanc $(\zeta_t)_{t \in \mathbb{Z}}$ (c'est à dire un processus stationnaire tel que ζ_t et ζ_{t-k} soient indépendants, pour tout k , pour tout t) tel que

$$X_t = \theta_1 X_{t-1} + \dots + \theta_p X_{t-p} + \zeta_t + \delta_1 \zeta_{t-1} + \dots + \delta_q \zeta_{t-q} \text{ pour tout } t. \quad (1.36)$$

Sous certaines conditions, ces processus sont stationnaires. Comme nous le verrons par la suite, ces processus peuvent s'écrire sous forme

$$\Gamma(L)X_t = \Upsilon(L)\zeta_t \text{ où } \Gamma(L) = \mathbb{1} - \theta_1 L - \dots - \theta_p L^p \text{ et } \Upsilon(L) = \mathbb{1} + \delta_1 L + \dots + \delta_q L^q$$

Les différentes conditions sur les paramètres θ_k et δ_k signifient que dans un ARMA(p, q), tous les paramètres du polynôme $\Gamma(L)$ peuvent être nuls à l'exception du paramètre correspondant au $p^{\text{ème}}$ retard et que tous les paramètres du polynôme $\Upsilon(L)$ peuvent être nuls à l'exception du paramètre correspondant au $q^{\text{ème}}$ retard. En effet, si $\theta_p = 0$, alors le processus $(X_t)_{t \in \mathbb{Z}}$ satisfait une représentation ARMA($p - 1, q$). De la même façon, si $\delta_q = 0$, alors le processus $(X_t)_{t \in \mathbb{Z}}$ satisfait une représentation ARMA($p, q - 1$).

Ainsi, on constate que les processus AR et MA ne sont que des cas particuliers des processus ARMA. Un AR(p) correspond à un ARMA($p, 0$), de la même façon un MA(q) correspond à un ARMA($0, q$).

Proposition 1.5 *Un processus autorégressif d'ordre p , possède une densité spectrale définie par*

$$f(\lambda) = \frac{\sigma^2}{2\pi |A(e^{-i\lambda})|^2} \quad (1.37)$$

(σ^2 est la variance de ζ), et le polynôme A ne peut donc posséder de zéros de module 1. Réciproquement si un processus centre stationnaire au second ordre X possède une densité spectrale de la forme (1.37) pour un nombre reel $\sigma > 0$ et un polynôme unitaire A de degré p (nécessairement sans zéro de module 1) ce processus est autorégressif d'ordre p les $(a_j)_{1 \leq j \leq p}$ étant données par $A(t) = 1 + \sum_{j=0}^p a_j t^j$.

Proposition 1.6 *La densité spectrale d'un processus AR(p) s'écrit d'une seule manière sous la forme*

$$f(\lambda) = \frac{\sigma^2}{2\pi |A(e^{-i\lambda})|^2} \quad (1.38)$$

pour un réel $\sigma > 0$, et un polynôme unitaire A de degré p dont toutes les racines sont de module > 1 . La représentation par $\sum_{k=0}^p a_k X_{t-k} = \zeta_t$ du processus X est alors appelée représentation canonique de X .

Proposition 1.7 (représentation canonique) Soit (X_t) un processus A.R.M.A stationnaire au second ordre, Sa densité spectrale est donnée par :

$$f(\lambda) = \frac{\sigma^2 C(e^{-i\lambda})|^2}{2\pi |A(e^{-i\lambda})|^2} \quad (1.39)$$

1.3 Simuler des processus

Dans cette section, nous expliquons comment simuler en \mathbf{R} des processus simples tels qu'un bruit blanc ou un processus autorégressif. Nous allons également décrire la fonction `arima.sim` qui permet de générer des processus **ARIMA**. Cependant, avant de simuler ces processus, on se demandera comment un logiciel peut créer un échantillon aléatoire. La réponse à cette question passe par la notion de générateur de nombres aléatoires par le logiciel. Nous allons voir que ce générateur permet de simuler un échantillon aléatoire provenant d'une variable uniforme sur $(0; 1)$. A partir de la simulation d'une variable uniforme, il sera alors possible de générer des échantillons pour d'autres variables aléatoires continues (exponentielle, normale,...).

1.3.1 Simuler un échantillon aléatoire d'une variables continue

Pour obtenir une suite de nombres aléatoires, \mathbf{R} , comme d'ailleurs un grand nombre de logiciels, possède une fonction prédéfinie. Cette fonction se présente en réalité sous la forme d'une grande suite de nombres qui, pratiquement, est semblable à un échantillon issu d'une distribution uniforme sur $(0; 1)$. Cependant, bien qu'elle soit très longue, la suite obtenue est forcément finie, et la fonction génératrice apparaît comme une fonction cyclique qui "boucle" cette série. La fonction génératrice de nombres aléatoires est donc une fonction déterministe et cyclique, et les nombres générés sont dits pseudo-aléatoires.

Le générateur aléatoire de nombres simule donc une variable aléatoire uniforme sur $(0; 1)$ à partir de laquelle il est possible de simuler des variables aléatoires continues. Nous allons à présent décrire certaines de ces méthodes.

1.3 Simuler des processus

La méthode de la transformation inverse est basée sur le théorème de la transformation inverse. Ce théorème peut être énoncé comme suit :

Théorème 1.3 *Si U est une variable uniforme sur l'intervalle $(0;1)$ et F est une fonction de répartition continue quelconque, alors la variable aléatoire X définie par $X = F^{-1}(U)$ a pour fonction de répartition F .*

Grâce à ce résultat, nous pouvons simuler une variable aléatoire X de fonction de répartition continue F à partir d'une variable uniforme U en posant $X = F^{-1}(U)$. Par exemple, pour simuler une variable aléatoire exponentielle de moyenne 1, il suffit de prendre $F(x) = 1 - e^{-x}$ et un simple calcul montre que la variable

$$F^{-1}(U) = -\ln(1 - U)$$

est de distribution exponentielle d'espérance 1. Mais, si U est uniforme sur $(0;1)$, il en est de même pour $1 - U$ et on a donc que

$$X = -\ln U$$

est une variable aléatoire exponentielle d'espérance 1.

Pour simuler un échantillon aléatoire provenant d'une variable exponentielle avec \mathbf{R} , il faut utiliser la fonction `rexp()` avec deux paramètres : n , qui est la taille de l'échantillon requis, et `rate`, qui est le taux de hasard de l'exponentielle. Dans \mathbf{R} , simulons un tel échantillon de taille 100 et de paramètre 2 :

`x < -rexp(100, 2)` \mathbf{R} affiche alors à l'écran les 100 valeurs de la simulation.

Une autre possibilité pour simuler des variables aléatoires continues est d'utiliser la méthode de rejet. Celle-ci permet de simuler une variable aléatoire continue ayant une fonction de densité f à partir d'une autre variable simulée de densité g selon la procédure suivante :

Etape 1 Simuler une réalisation y provenant de la variable aléatoire Y dont la fonction de densité est g . Simuler indépendamment de Y une réalisation u provenant de variable aléatoire U uniforme sur $(0;1)$;

Etape 2 Si $u \leq f(y)/cg(y)$ pour une certaine constante c , alors on garde la réalisation y et on pose $x = y$. Sinon, on ne génère aucune valeur.

Cette procédure est répétée autant de fois que nécessaire pour arriver à la taille requise pour l'échantillon, et on montre que la variable aléatoire X générée par cette méthode a la fonction de densité f (voir Ross (1994)).

1.3.2 Simuler un processus autorégressif

Pour simuler simplement T réalisations d'un processus autorégressif d'ordre p ayant les paramètres a_1, \dots, a_p et dont la variance des innovations est σ_ε^2 , on peut utiliser l'algorithme suivant : Etape 1 Fixer p valeurs initiales réelles arbitraires x_0, \dots, x_{p-1} ;

Etape 2 Générer des innovations i.i.d. $\{\varepsilon_t\}_{t=1, \dots, T+T_0}$;

Etape 3 Calculer récursivement les valeurs

$$X_t = \sum_{k=1}^p a_k X_{t-k} + \varepsilon_t,$$

Etape 4 éliminer les T_0 premières valeurs ainsi générées.

1.3.3 Simuler un processus MA

(calculer la moyenne mobile avec différents coefficients.)

```

n < -200
x < -rnorm(n)
y < -(x[2 : n] + x[2 : n - 1])/2
op < -par(mfrow = c(3, 1), mar = c(2, 4, 2, 2) + .1)
plot(ts(x), xlab = "", ylab = "whitenoise")
plot(ts(y), xlab = "", ylab = "MA(1)")acf(y, main = "")
par(op)
n < -200
x < -rnorm(n)
y < -(x[1 : (n - 3)] + x[2 : (n - 2)] + x[3 : (n - 1)] + x[4 : n])/4
op < -par(mfrow = c(3, 1), mar = c(2, 4, 2, 2)+.1)plot(ts(x), xlab = "", ylab = "whitenoise")
plot(ts(y), xlab = "", ylab = "MA(3)")
acf(y, main = "")
par(op)

```

Au lieu de calculer la moyenne mobile à la main, tu peux employer la fonction de "filtre".

```

n < -200
x < -rnorm(n)
y < -x[2 : n] - x[1 : (n - 1)]
op < -par(mfrow = c(3, 1), mar = c(2, 4, 2, 2) + .1)

```

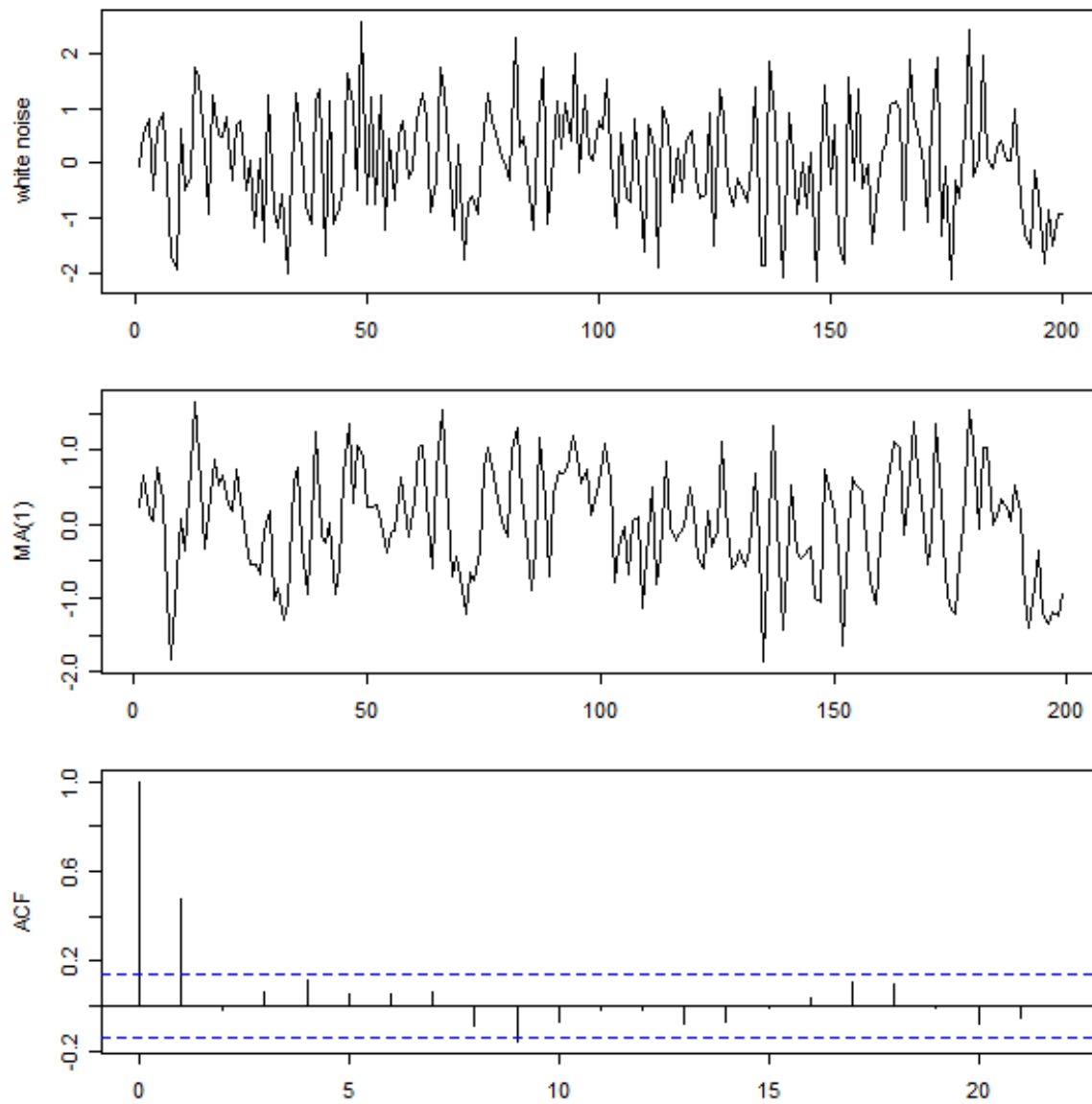


FIGURE 1.1 – ACF. MA(1). WN

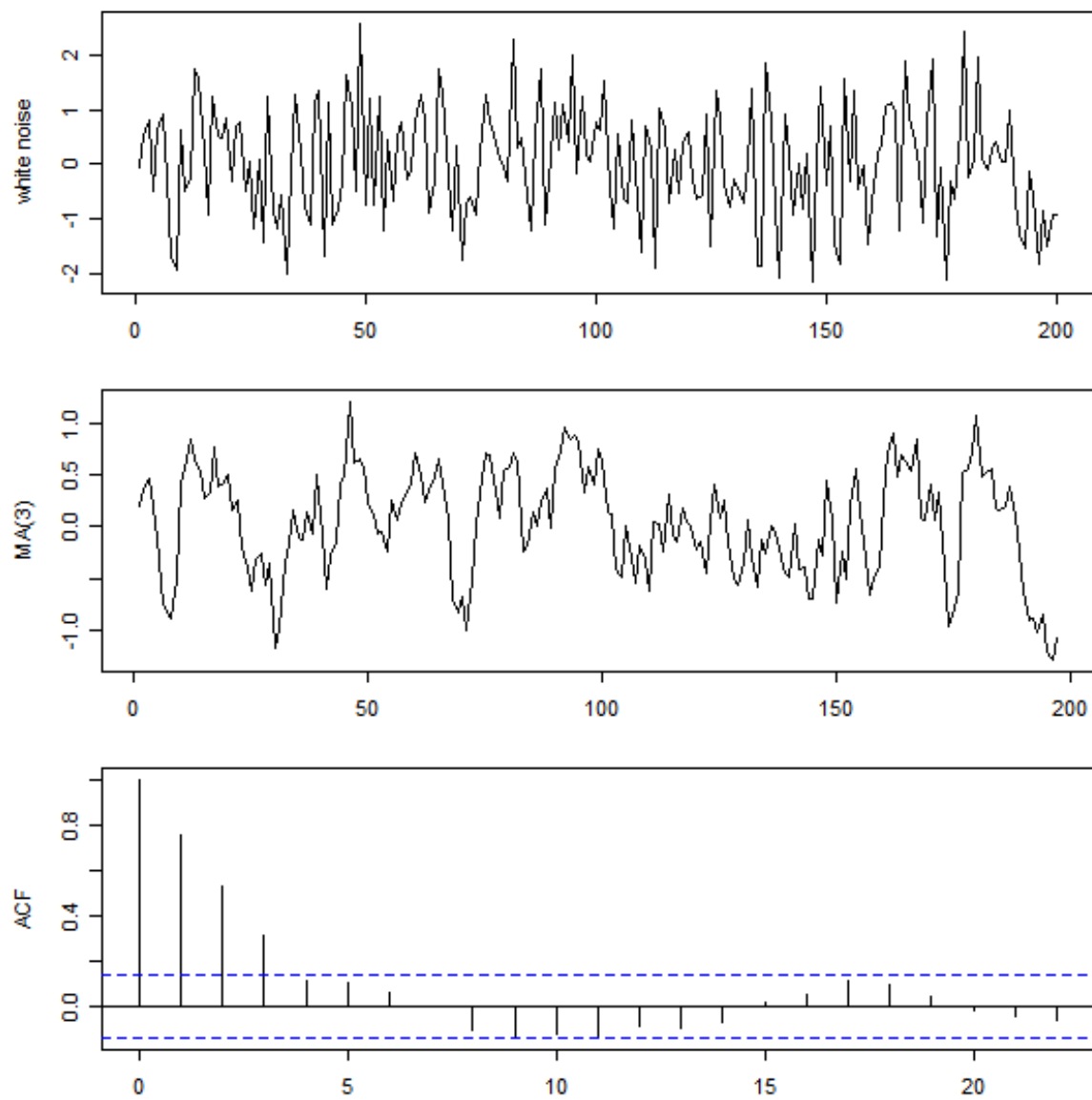
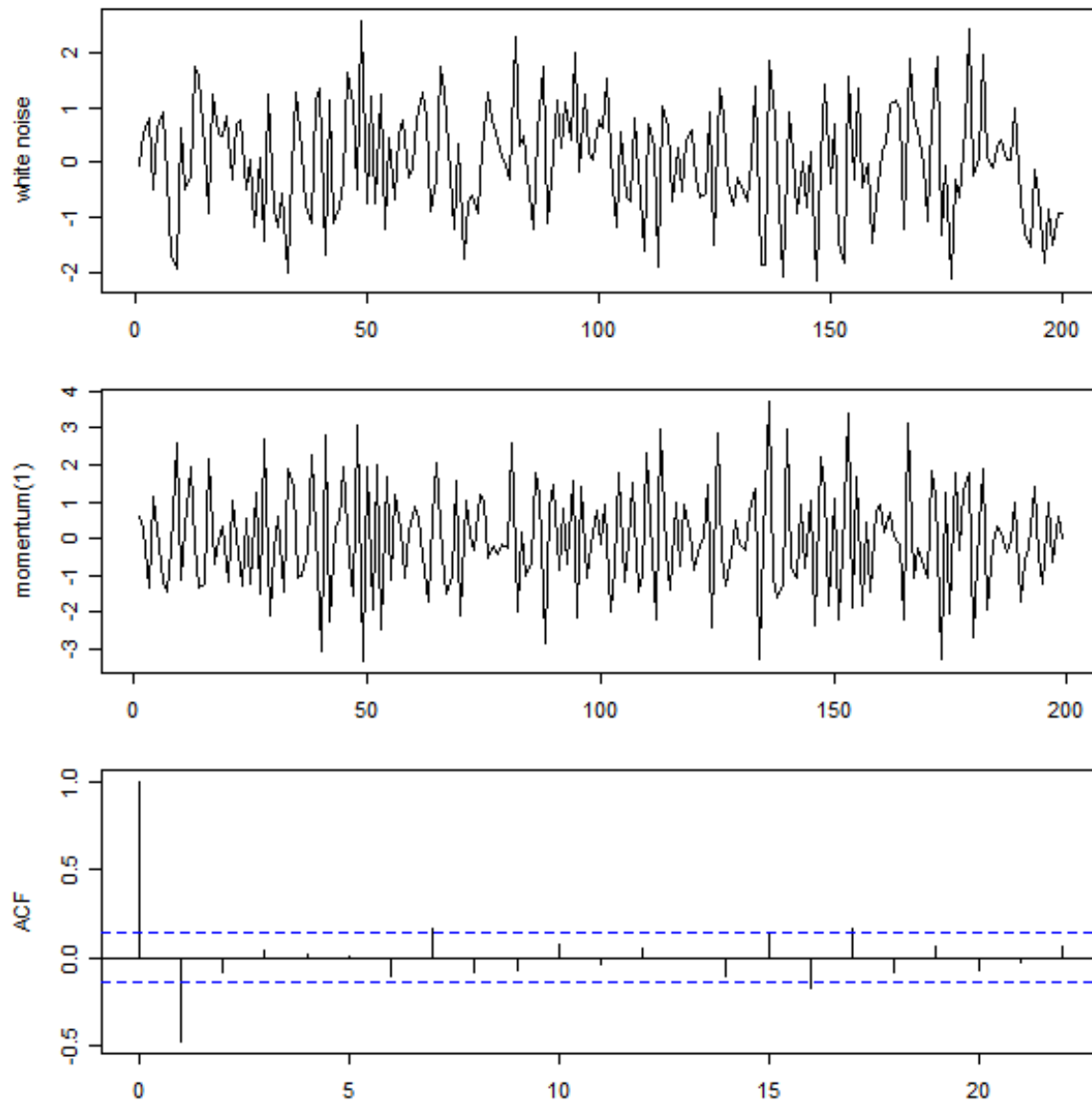


FIGURE 1.2 – ACF. MA(3). WN

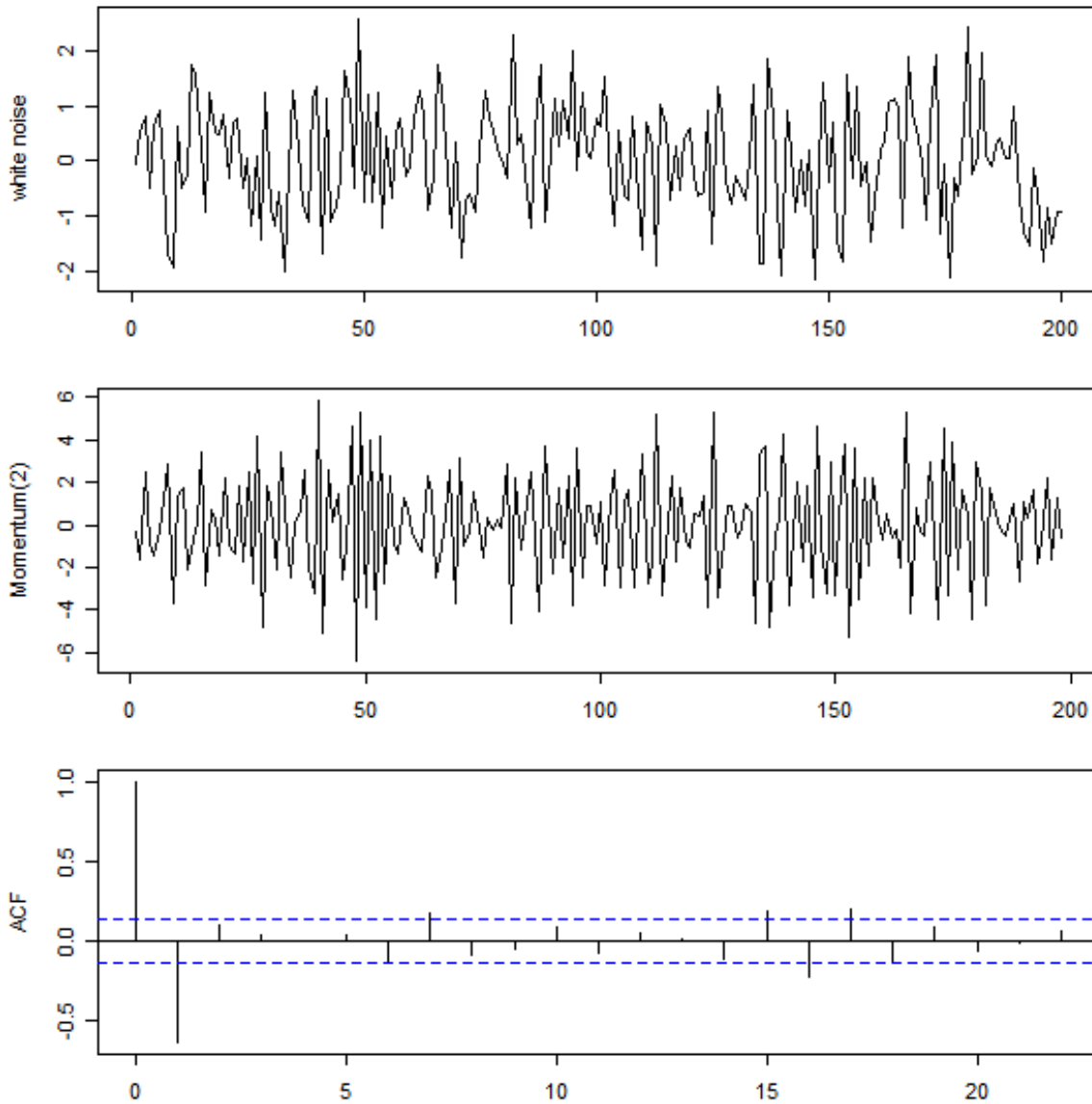
1.3 Simuler des processus

```
plot(ts(x), xlab = "", ylab = "whitenoise")
plot(ts(y), xlab = "", ylab = "momentum(1)")
acf(y, main = "")
par(op) n < -200x < -rnorm(n)
y < -x[3 : n] - 2 * x[2 : (n - 1)] + x[1 : (n - 2)]
op < -par(mfrow = c(3, 1), mar = c(2, 4, 2, 2) + .1)
plot(ts(x), xlab = "", ylab = "whitenoise")
plot(ts(y), xlab = "", ylab = "Momentum(2)")
acf(y, main = "")
par(op)
```

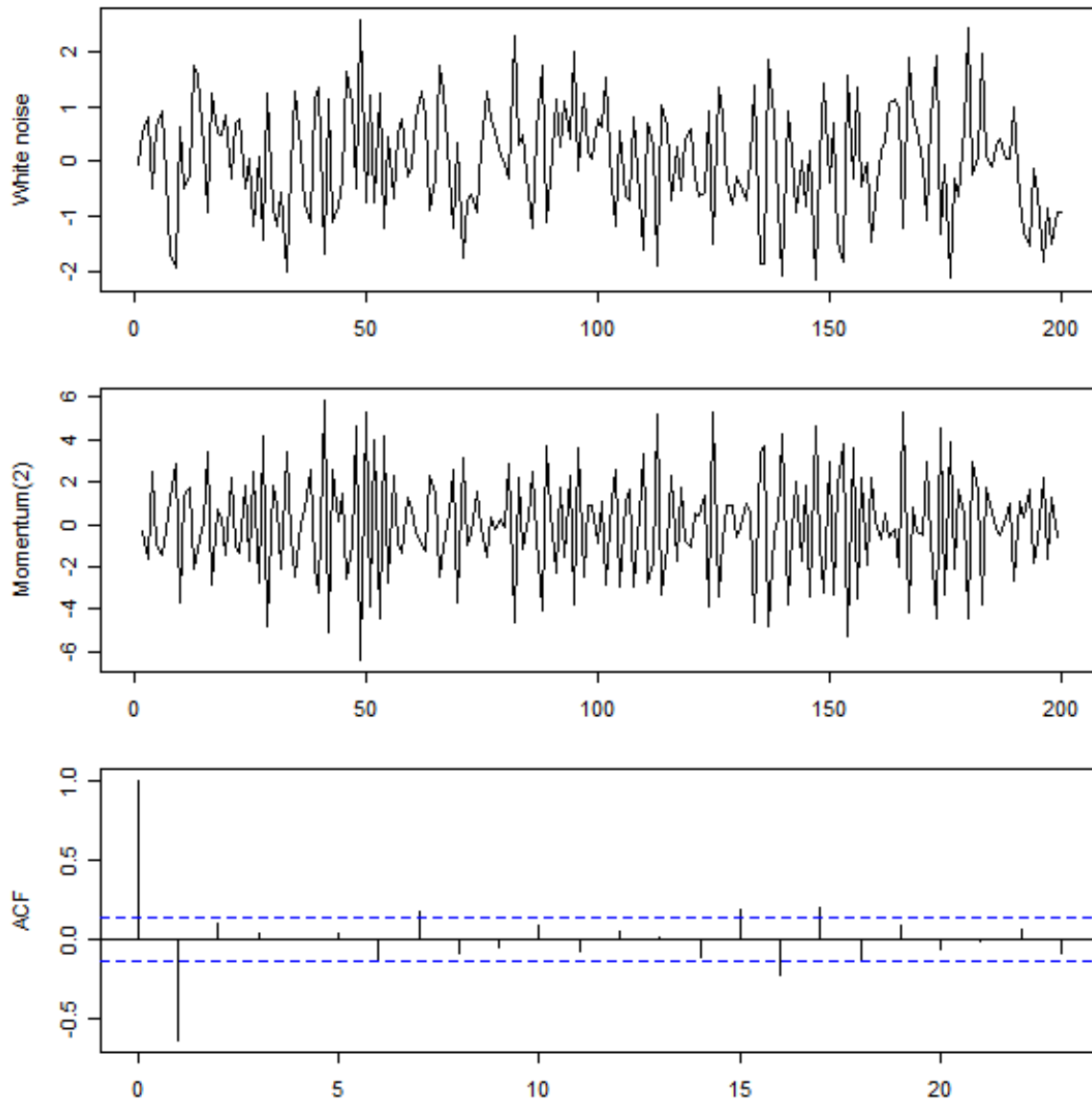
1.3 Simuler des processus



1.3 Simuler des processus



```
n < -200
x < -rnorm(n)
y < -filter(x, c(1, -2, 1))
op < -par(mfrow = c(3, 1), mar = c(2, 4, 2, 2) + .1)
plot(ts(x), xlab = "", ylab = "Whitenoise")
plot(ts(y), xlab = "", ylab = "Momentum(2)")
acf(y, na.action = na.pass, main = "")
par(op)
```



1.3.4 AR (Auto-Regressive models)

Des autres moyens d'établir une série chronologique doit calculer chaque limite en ajoutant le bruit au limite précédente : ceci s'appelle une promenade aléatoire. Par exemple,

$n < -200$

$x < -rep(0, n)$

$for(iin2 : n)$

$x[i] < -x[i - 1] + rnorm(1)$

1.3 Simuler des processus

Ceci peut être écrit, plus simplement, avec la fonction de "cumsum".

```
n < -200
x < -rnorm(n)
y < -cumsum(x)
op < -par(mfrow = c(3, 1), mar = c(2, 4, 2, 2) + .1)
plot(ts(x), xlab = "", ylab = "")
plot(ts(y), xlab = "", ylab = "AR(1)")
acf(y, main = "")
par(op)
```

Plus généralement, un processus d'AR(q) est un processus dans du lequel chaque limite est une combinaison linéaire les limites précédentes de q et un bruit blanc (avec des coefficients fixes).

```
n < -200
x < -rep(0, n)
for(i in 4 : n)x[i] < -.3 * x[i - 1] - .7 * x[i - 2] + .5 * x[i - 3] + rnorm(1)
op < -par(mfrow = c(3, 1), mar = c(2, 4, 2, 2) + .1)
plot(ts(x), xlab = "", ylab = "AR(3)")acf(x, main = "", xlab = "")
pacf(x, main = "", xlab = "")
par(op)
n < -200x < -arima.sim(list(ar = c(.3, -.7, .5)), n)op < -par(mfrow = c(3, 1), mar =
c(2, 4, 2, 2)+.1)plot(ts(x), xlab = "", ylab = "AR(3)")acf(x, xlab = "", main = "")pacf(x, xlab =
"", main = "")par(op)
```

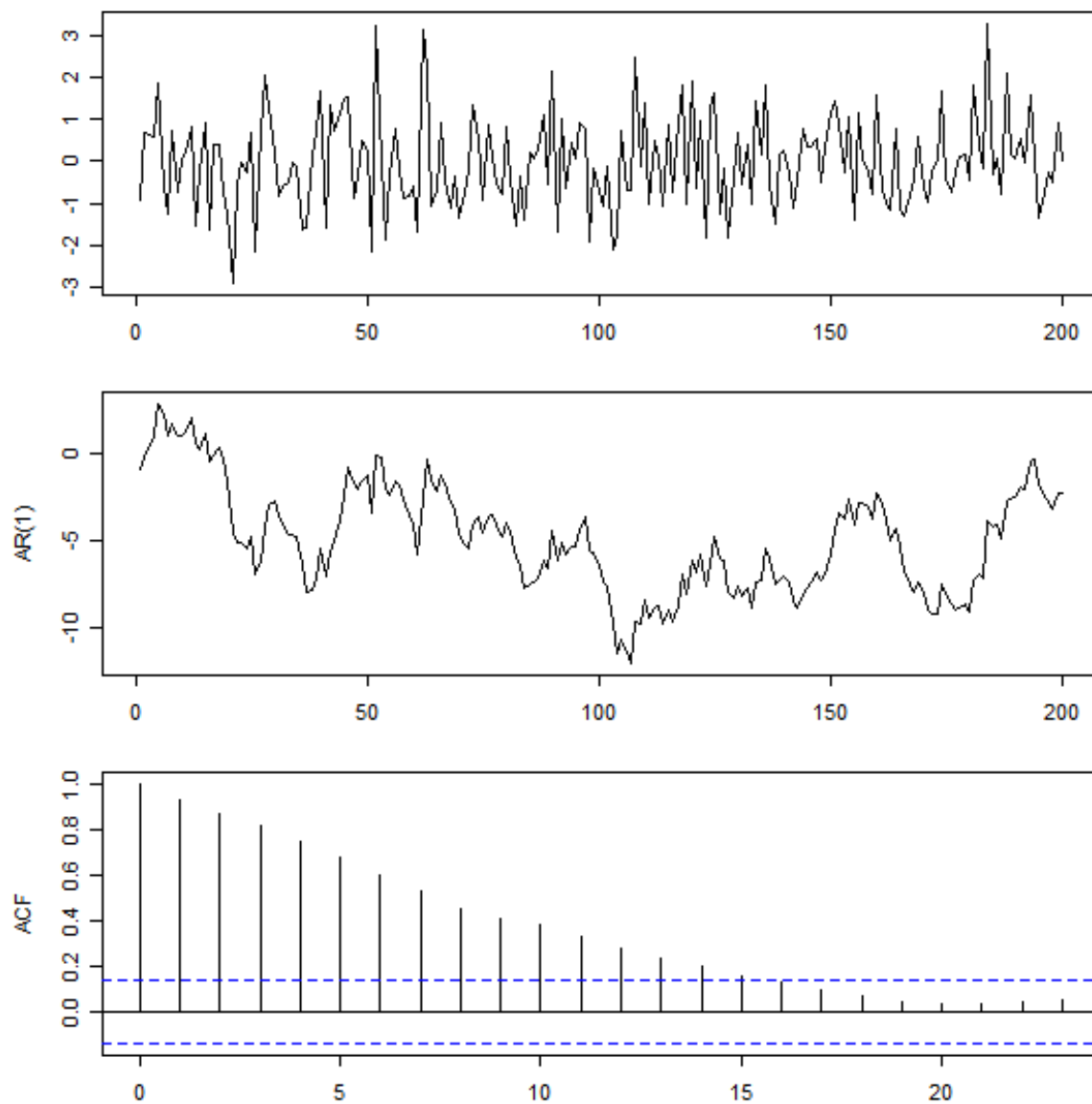


FIGURE 1.3 – ACF. AR(1)

1.3 Simuler des processus

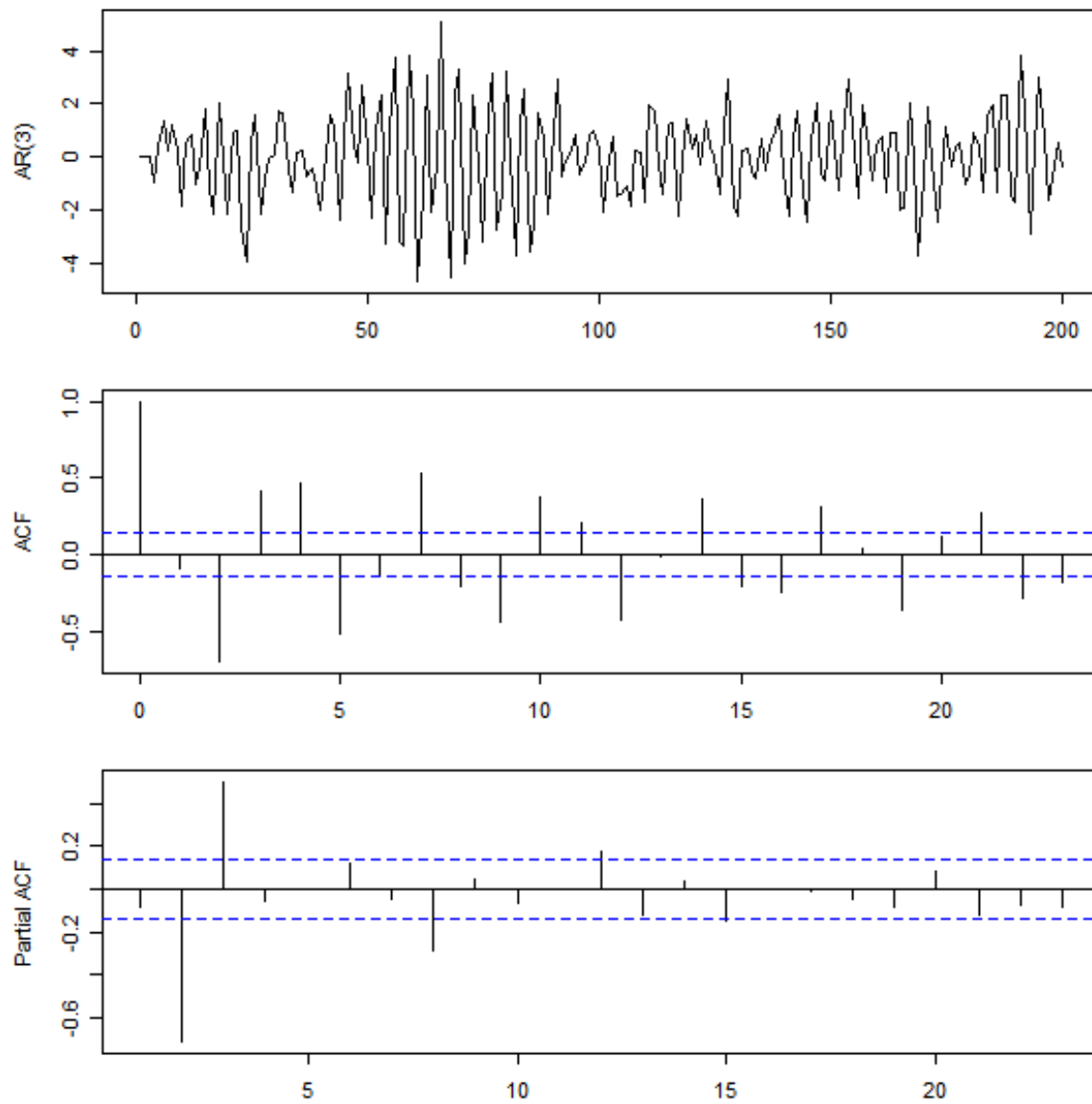


FIGURE 1.4 – PACF. ACF. AR(3)

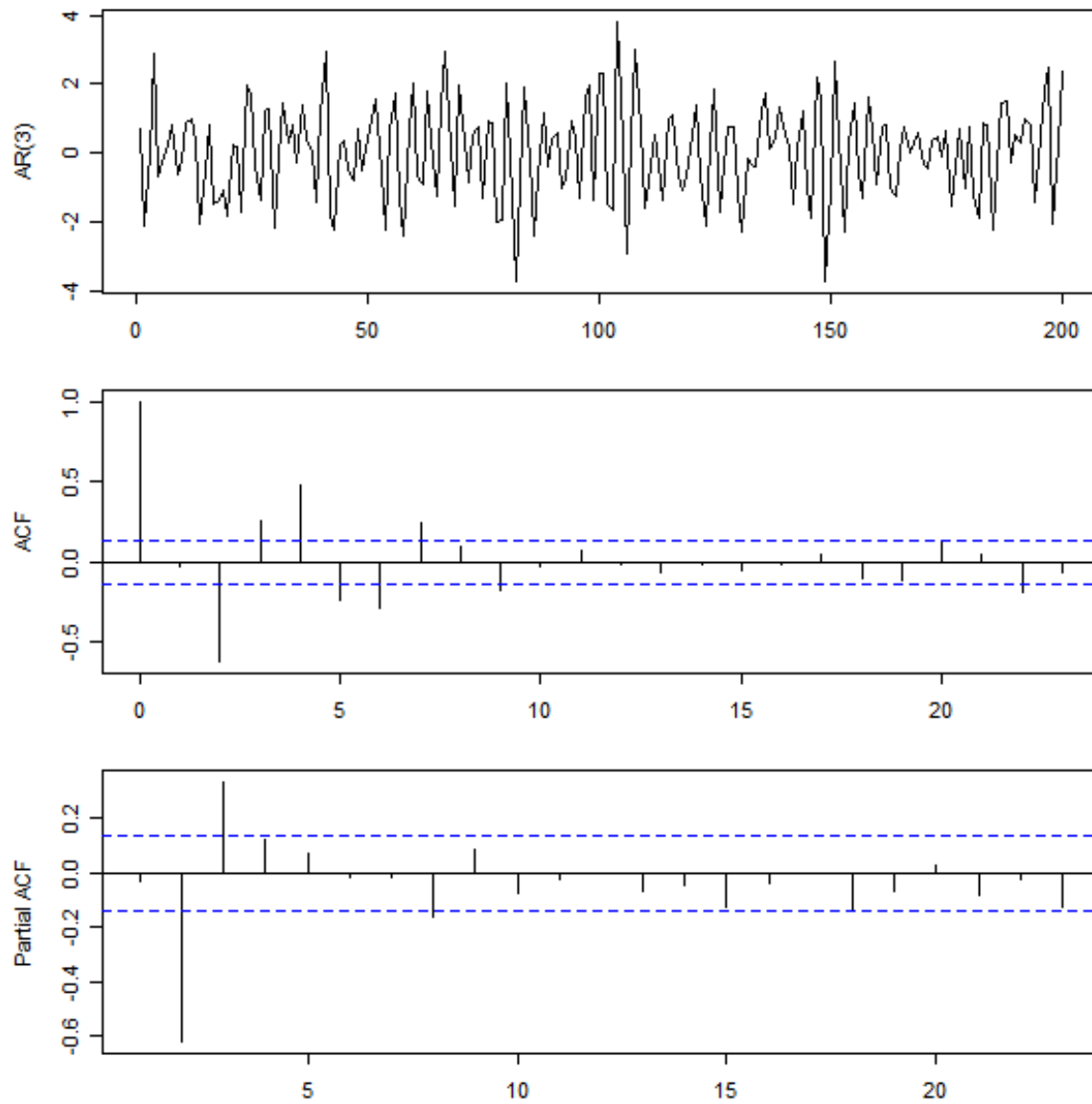


FIGURE 1.5 – Partial ACF. ACF. AR(3)

Chapitre 2

Estimation non-paramétrique par la méthode de noyau

La particularité de la statistique non-paramétrique est que le paramètre inconnu qu'on cherche à détecter, à estimer ou à classifier n'est pas supposé d'appartenir à une famille indiquée par un petit nombre de paramètres réels. En général, dans la théorie non-paramétrique on suppose que le nombre de paramètres qui décrivent la loi des observations est une fonction croissant du nombre d'observations, ou encore que le nombre de paramètres est infini. Pour donner un exemple concret, considérons le modèle linéaire multiple. C'est un modèle très populaire dans le milieu des praticiens et profondément étudié par les théoriciens. Il s'agit d'une expérience qui résulte en l'observation des couples (X_i, Y_i) , $i = 1, \dots, n$ où, en général, X_i est un vecteur p -dimensionnel et Y_i est une valeur réelle. On suppose que pour un vecteur $\beta \in \mathbb{R}^p$ et un réel α , la dépendance de Y_i en X_i est expliquée par la fonction affine

$$\alpha + \beta^T x$$

à une erreur aléatoire près, c'est-à-dire

$$Y_i = \alpha + \beta^T X_i + \zeta_i \tag{2.1}$$

Rappelons les résultats suivants vu leurs utilités dans la suite.

Théorème de Bochner

Théorème 2.1 *Soit $K : (\mathbb{R}^m; B^m) \rightarrow (R; B)$ une fonction mesurable, où B^m est la tribu borélienne de \mathbb{R}^m , vérifiant :*

$$\exists M \text{ constante} / \forall z \in \mathbb{R}^m, |K(z)| \leq M,$$

$$\int_{\mathbb{R}^m} |K(z)| dz < \infty$$

et

$$\|z\|^m |K(z)| \rightarrow 0 \text{ quand } \|z\| \rightarrow \infty.$$

Par ailleurs, soit $g : (\mathbb{R}^m; B^m) \rightarrow (R; B)$ une fonction mesurable telle que

$$\int_{\mathbb{R}^m} |g(z)| dz < \infty.$$

On définit :

$$g_n(x) = \frac{1}{h_n^m} \int_{\mathbb{R}^m} |K(z/h_n)| g(x-z) dz,$$

où $0 < h_n \rightarrow 0$ quand $n \rightarrow \infty$. Si g est continue, alors

$$\lim_{n \rightarrow \infty} g_n(x) = g(x) \int_{\mathbb{R}^m} |K(z)| dz$$

Si g est uniformément continue alors la convergence ci dessus est uniforme.

Inégalité de type Bernstein Il existe plusieurs versions d'inégalités de ce type. Nous nous contentons de rappeler dans le lemme ci dessous une version simplifiée, qui nous servira dans ce travail et dont la preuve est donnée dans l'article de Hoeffding (1963).

Lemme 2.1 Soit $\Delta_1, \Delta_2, \dots, \Delta_n$ des v.a.r. centrées, indépendantes et de même loi, telles qu'il existe deux réels positifs d et δ^2 vérifiant :

$$|\Delta_1| \leq d \text{ et } \mathbb{E}(\Delta_1^2) \leq \delta^2.$$

Alors, pour tout $\varepsilon \in]0, \frac{\delta^2}{d}[$ on a

$$\mathbb{P} \left(n^{-1} \left| \sum_{i=1}^n \Delta_i \right| > \varepsilon \right) \leq 2 \exp \left\{ -\frac{n\varepsilon^2}{4\delta^2} \right\}.$$

Convergence du maximum de variables aléatoires i.i.d. Soit Z_1, Z_2, \dots, Z_n n v.a. i.i.d., de fonction de répartition F , posons

$$T_{k_n} = \max_{1 \leq i \leq n} \{Z_1, Z_2, \dots, Z_n\}$$

et

$$T_k = \sup\{t : F(t) < 1\}.$$

En supposant que $T_k < \infty$, on montre que $T_{k_n} \rightarrow T_k (n \rightarrow \infty)$ p.s.

En effet, on a d'une part

$$\mathbb{P}(T_{k_n}) = \mathbb{P}(Z_1 \leq t, Z_2 \leq t, \dots, Z_n \leq t) = F^n(t) = 1 \text{ si } F(t) = 1 \text{ et } F^n(t) = 0 \text{ si } F(t) < 1,$$

qui est la fonction de la répartition de la v.a. presque sûrement égale à T_k ; donc $T_{k_n} \rightarrow T_k$ en loi et comme T_k est constante alors $T_{k_n} \rightarrow T_k$ en probabilité. D'autre part, on a T_{k_n} est croissante et presque sûrement majorée par T_k , elle converge donc, presque sûrement, vers X . Mais

$$T_{k_n} \rightarrow T_k \text{ en probabilité}$$

et

$$T_{k_n} \rightarrow X \text{ en probabilité}$$

alors,

$$T_k = X \text{ p.s.}$$

Lemme 2.2 (lemme de Borel-Centelli voir [7]) Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité, et $\{A_n\}$ une suite d'événements.

(a) Si $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$ alors

$$\mathbb{P}\left(\limsup_n A_n\right) = 0.$$

(b) Si les événements A_1, A_2, \dots , sont indépendants et $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$, alors

$$\mathbb{P}\left(\limsup_n A_n\right) = 1.$$

Corollaire 2.1 (voir [7]) Soit $\{A_n\}$ une suite des événements indépendants. Alors

$$\mathbb{P}\left(\limsup_n A_n\right) = 0 \iff \sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty.$$

Preuve. Nous avons déjà vu que, même sans indépendance :

$$\sum_{n=1}^{+\infty} \mathbb{P}(A_n) < +\infty \Rightarrow \mathbb{P}(\limsup_{n \rightarrow +\infty} A_n) = 0.$$

Mais, d'autre part, on a aussi :

$$\sum_{n=1}^{+\infty} \mathbb{P}(A_n) = +\infty \Rightarrow \mathbb{P}(\limsup_{n \rightarrow +\infty} A_n) = 1;$$

en effet, comme les $A_n^c, n \geq 1$, sont indépendants, on a :

$$\mathbb{P}(\cap_{k=n}^N A_k^c) = \prod_{k=n}^N \mathbb{P}(A_k^c) = \prod_{k=n}^N (1 - \mathbb{P}(A_k)) \leq \prod_{k=n}^N \exp(-\mathbb{P}(A_k))$$

c'est-à-dire

$$\mathbb{P}(\cap_{k=n}^N A_k^c) \leq \exp(-\sum_{k=n}^N \mathbb{P}(A_k)) \rightarrow 0 \text{ quand } n \rightarrow +\infty; \quad (2.2)$$

donc

$$\mathbb{P}(\cap_{k=n}^{\infty} A_k^c) = 0,$$

et donc :

$$\mathbb{P}(\liminf_{n \rightarrow \infty} A_k^c) = \mathbb{P}(\cap_{k=n}^{\infty} A_k^c).$$

Exemple. Si l'o joue à "pile ou face", et que A_n est l'évènement "obtenir face au n^{me} lancer", comme $\mathbb{P}(A_n) = 1/2$ pour tout $n \geq 1$, on a $\sum_{n=1}^{+\infty} \mathbb{P}(A_n) = +\infty$, donc $\mathbb{P}(\limsup_{n \rightarrow +\infty} A_n) = 1$: presque sûrement, on aura une infinité de fois "face". De même, on aura une infinité de fois de "pile".

Remarque. L'indépendance est essentielle. Par exemple si $\Omega = [0, 1]$ et $A_n = [0, 1/n]$, on a $\sum_{n=1}^{+\infty} \mathbb{P}(A_n) = \sum_{n=1}^{+\infty} \frac{1}{n} = +\infty$, mais $\limsup_{n \rightarrow +\infty} A_n = \{0\}$.

Définition 2.1 Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires définie sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. On dit que X_n **converge presque sûre** (p.s) vers zéro, si pour tout $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P}(|X_n| > \varepsilon \text{ pour certains } n \geq m) = 0.$$

Définition 2.2 (Hsu et Robbins [8]) Une suite $\{X_n, n \in \mathbb{N}\}$ de variables aléatoires converge presque complètement vers zéro, si pour tout $\varepsilon > 0$,

$$\sum_{n=1}^{\infty} \mathbb{P}(|X_n| > \varepsilon) < \infty.$$

Selon la définition de ce qui précède, on remarque que la convergence presque complète implique la convergence presque sûre, Cependant la réciproque a lieu si les variables aléatoires $\{X_n\}$ sont indépendantes.

Exemple :

Soit $\omega \in \Omega = (0, 1)$, et \mathbb{P} est la mesure de Lebesgue ,

$$X_n(\omega) = n \mathbb{1}_{(0, \frac{1}{n})}(\omega).$$

2.1 Modes de convergence :

pour tout $\varepsilon > 0$ puisque $\mathbb{P}(|X_n| \geq \varepsilon \text{ pour certains } n \geq m) \leq \frac{1}{m}$, la suite $\{X_n\}_{n=1}^\infty$ converge vers 0 p.s.. Mais pour $N > \varepsilon$, on a $\sum_{n=1}^\infty \mathbb{P}(|X_n| > \varepsilon) \geq \sum_{n=N}^\infty \frac{1}{n} = \infty$. La suite $\{X_n\}_{n=1}^\infty$ ne converge pas presque complètement vers 0.

2.1 Modes de convergence :

Il existe quatre types de convergence pour des suites de variables aléatoires : la convergence en distribution, la convergence en probabilité, la convergence presque sûre et la convergence en moyenne quadratique. Ces quatre types de convergence sont rangés là par type croissant d'implication. La convergence en moyenne quadratique et la convergence presque sûre impliquent la convergence en probabilité qui à son tour implique la convergence en loi.

2.1.1 Convergence en loi :

Parmi les notions de convergence, la convergence en loi est la convergence la plus faible. La définition suivante est valable pour les v.a. réelles et vectorielles.

On va considérer une suite de variables aléatoires X_n indexées par n et de loi de probabilité cumulative (ou distribution) associée qui est définie par :

$$F_n(x) = P(X_n < x). \quad (2.3)$$

Dans cette partie nous donnons quelques outils. La définition suivante est valable pour les v.a. réelles et vectorielles.

Définition 2.3 *on dit qu'une suite de v.a. $(X_n; n \in \mathbb{N}^*)$ converge en loi vers une v.a. X si et seulement si pour tout fonction g à valeurs réelles, bornée et continue, on a :*

$$\lim_{n \rightarrow \infty} \mathbb{E}[g(X_n)] = \mathbb{E}[g(X)]$$

et l'on écrit :

$$g(X_n) \xrightarrow{\mathcal{L}} g(X)$$

Proposition 2.1 *Pour qu'une suite X_n converge en loi vers X , il suffit que cette suite converge en probabilité vers X .*

2.1 Modes de convergence :

Exemple 1 Soit X_n de loi uniforme sur $\left\{0; \frac{1}{n}; \dots; \frac{n-1}{n}\right\}$. Alors la suite $(X_n; n \in \mathbb{N}^*)$ converge en loi vers la loi uniforme sur $[0, 1]$. En effet, si g est continue bornée, on déduit de la convergence des sommes de Riemann que :

$$\mathbb{E}[g(X_n)] = \frac{1}{n} \sum_{k=0}^{n-1} g(k/n) \xrightarrow{n \rightarrow \infty} \int_{[0,1]} g(x) dx$$

2.1.2 Convergence en probabilité :

On présente dans ce paragraphe la convergence en probabilité

Définition 2.4 On dit qu'une suite X_n de variables aléatoires converge en probabilité vers la variable aléatoire X . si quelque soit le réel $\varepsilon > 0$, on a :

$$\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$$

ou encore :

$$\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| \leq \varepsilon) = 1.$$

Donc :

$$X_n \xrightarrow{\mathbb{P}} X \Leftrightarrow \forall \varepsilon > 0, \lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0.$$

Cette convergence implique la convergence en loi.

Cette notion de convergence est basée sur la loi faible des grands nombres et en constitue une définition alternative. Elle est très intéressante pour ses propriétés. En particulier nous avons le théorème suivant appelé théorème de **Slutsky** :

Théorème 2.2 soit X_n une suite de variables aléatoires avec $X_n \xrightarrow{\mathbb{P}} x^*$. Si $Y_n = f(X_n)$ et si f est continue en x^* alors :

$$Y_n = f(X_n) \xrightarrow{\mathbb{P}} f(X_n \xrightarrow{\mathbb{P}} x^*) = f(x^*)$$

La convergence en probabilité et la convergence en loi se marient bien du fait que la convergence en probabilité est plus forte que la convergence en loi. C'est sur cette association qu'est basé le théorème suivant qui est connu sous le nom de théorème de transformation linéaire de Cramer :

Théorème 2.3 Soient X_n une suite de variables aléatoires vectorielles et A_n une suite de matrices aléatoires de dimensions conformes. Sous les conditions :

$$X_n \xrightarrow{\mathcal{L}} X \text{ et } A_n \xrightarrow{\mathbb{P}} A$$

2.1 Modes de convergence :

on a

$$A_n X_n \xrightarrow{\mathcal{L}} AX$$

Ce théorème manipule des transformations linéaires. Il peut se généraliser à des transformations générales continues.

La convergence en probabilité n'implique pas la convergence des moments et en particulier celle des espérances mathématiques. On peut donc définir une notion de convergence plus forte qui est la convergence en moyenne quadratique :

Définition 2.5 La suite de variables aléatoires X_n converge en moyenne quadratique vers la constante x^* si :

$$\lim_{n \rightarrow \infty} \mathbb{E}(X_n - x^*)^2 = 0.$$

Exemple 2 Théorème de Bernoulli : On effectue n expériences successives indépendantes où on s'intéresse à chaque fois à la réalisation d'un certain événement A . On associe alors à chaque expérience i , une variable de Bernoulli.

$$\left\{ \begin{array}{l} \mathbb{P}(X_i = 1) = p \\ \mathbb{P}(X_i = 0) = q = 1 - p \end{array} \right\}.$$

Notons $f_n = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}_n$, on a

$$\begin{aligned} \mathbb{E}(f_n) &= p \\ \text{Var}(f_n) &= \frac{p \cdot q}{n}. \end{aligned}$$

En appliquant l'inégalité de Tchebychev,

$$\forall \varepsilon > 0 : \mathbb{P}(|f_n - p| > \varepsilon) \leq \frac{p \cdot q}{n \varepsilon^2} \rightarrow 0$$

donc

$$f_n \rightarrow 0, \text{ lorsque } n \rightarrow +\infty$$

Proposition 2.2 On suppose que les espérances et les variances des X_n et de X existent. Pour que la suite X_n converge en probabilité vers X , il suffit que l'espérance et la variance de $Y_n = X_n - X$, convergent toutes les deux vers 0 lorsque n tend vers $+\infty$.

2.1.3 Convergence Presque sûre :

On présente dans ce paragraphe la convergence presque sûre (p.s.) et les conditions qui permettent de permuter convergence p.s. et espérance.

Définition 2.6 Une suite (X_n) de variable aléatoire définie sur $(\Omega, \mathcal{A}, \mathbb{P})$ converge presque sûrement vers la variable aléatoire X si et seulement si :

$$\mathbb{P}\left(\omega \in \Omega / \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\right) = 1.$$

Exemple 3 Soit $(\Omega, \mathcal{A} = \mathbb{P}(\otimes), \mathbb{P})$ un espace de probabilité, tel que $\Omega = \{\omega_1, \omega_2, \omega_3\}$; et

$$\mathbb{P}\{\omega_1\} = \frac{1}{2}, \mathbb{P}\{\omega_2\} = \frac{1}{4}, \mathbb{P}\{\omega_3\} = \frac{1}{4}$$

Définissons la variable aléatoire X_n par

$$X_n(\omega_1) = \frac{\alpha}{n}, X_n(\omega_2) = -1 + \frac{\alpha}{n}, X_n(\omega_3) = 1 - \frac{\alpha}{n}$$

où α est une constante positive. Montrons que $(X_n \xrightarrow{P.S.} X)$ avec X la variable aléatoire définie par

$$X(\omega_1) = 0, X(\omega_2) = -1, X(\omega_3) = 1.$$

En effet, posons $D_f = \{x \in \mathbb{R} / x \geq 0\}$, alors :

$$\begin{aligned} \mathbb{P}\{\omega \in \Omega / \lim_{n \rightarrow \infty} X_n(\omega_1) = 0, \lim_{n \rightarrow \infty} X_n(\omega_2) = -1, \lim_{n \rightarrow \infty} X_n(\omega_3) = 1\} &= \mathbb{P}\{\omega_1, \omega_2, \omega_3\} \\ &= \mathbb{P}\{\Omega\} = 1. \end{aligned}$$

Théorème 2.4 Théorème de delta

Soit $r : \mathbb{N} \rightarrow \mathbb{R}^+$ telle que $\lim_{n \rightarrow +\infty} r(n) = +\infty$, a une constante réelle, et $\{X_n, n \geq 1\}$, une suite de v.a.r. vérifiant :

$$r(n)(X_n - a) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma).$$

soit $g : \mathbb{R} \rightarrow \mathbb{R}$, dérivable.

Alors :

$$r(n)(g(X_n) - g(a)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma | g'(a) |).$$

2.1 Modes de convergence :

Preuve :

Considérons le développement de g à l'ordre 1 au voisinage de a :

$$g(x) - g(a) = (x - a)g'(a) + (x - a)R(x),$$

$$\lim_{x \rightarrow a} R(x) = 0,$$

où :

c'est-à-dire :

$$\forall \varepsilon > 0, \exists \alpha > 0 : |x - a| < \alpha \Rightarrow |R(x)| < \varepsilon,$$

d'où :

$$\{|X_n - a| < \alpha\} \subset \{|R(X_n)| < \varepsilon\},$$

$$\mathbb{P}\{|X_n - a| < \alpha\} \leq \mathbb{P}\{|R(X_n)| < \varepsilon\}.$$

Or, d'après **2**, et donc :

$$\lim_{n \rightarrow +\infty} \mathbb{P}\{|X_n - a| < \alpha\} = 1 \Rightarrow \lim_{n \rightarrow +\infty} \mathbb{P}\{|R(X_n)| < \varepsilon\} = 1,$$

c'est-à-dire

$$R(X_n) \xrightarrow{\mathbb{P}} 0.$$

Il s'ensuit :

$$r(n)(g(X_n) - g(a)) = r(n)(X_n - a)g'(a) + (X_n - a)r(n)R(X_n).$$

D'après **2**, $r(n)(X_n - a)R(X_n) \xrightarrow{\mathbb{P}} 0$, et $r(n)(X_n - a)g'(a) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma |g'(a)|)$, d'où le résultat recherché. Très utile en statistique mathématique lors de l'étude de comportements asymptotiques, ce théorème est aisément généralisable au cas multi dimensionnel : **Théorème**

Soit $r : \mathbb{N} \rightarrow \mathbb{R}^+$ telle que $\lim_{n \rightarrow +\infty} r(n) = +\infty$, a un vecteur de \mathbb{R}^p . $\{X_n, n \geq 1\}$, une suite de vecteurs de \mathbb{R}^p . tels que :

$$r(n)(X_n - a) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \Sigma\right).$$

soit $g : \mathbb{R}^p \rightarrow \mathbb{R}^q$ différentiable, G sa matrice (q, p) des dérivées premières :

2.2 Estimation de la densité de probabilité d'une variable aléatoire.

$$G = \begin{pmatrix} \frac{\delta g_1}{\delta u_1} & \cdots & \frac{\delta g_1}{\delta u_p} \\ \vdots & & \\ \frac{\delta g_q}{\delta u_1} & \cdots & \frac{\delta g_q}{\delta u_p} \end{pmatrix}$$

Alors :

$$r(n) (g(X_n) - g(a)) \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, G(a) \sum G^t(a) \right).$$

Exemple 4 Soit X une v.a.r. de loi de Poisson $\mathcal{P}(\lambda)$, $\{X_n, n \geq 1\}$, une suite d'observations indépendantes de X . Le paramètre d'intérêt est non pas $\lambda = \mathbb{E}(X)$, mais $e^{-\lambda} = \mathbb{P}(X = 0)$. On sait que : $\bar{X}_n \xrightarrow{\mathbb{P}} \lambda$ (loi faible des grands nombres).

Par le théorème central limite :

$$\sqrt{n} (\bar{X}_n - \lambda) \xrightarrow{\mathcal{L}} \mathcal{N} (0, \sqrt{\lambda})$$

soit $g : \mathbb{R}_+^* \rightarrow \mathbb{R}_+^*$ définie par $u \rightarrow g(u) = e^{-u}$. Par le théorème 2

$$e^{-\bar{X}_n} \xrightarrow{\mathbb{P}} e^{-\lambda}.$$

En outre :

$$\sqrt{n} (g(\bar{X}_n) - g(\lambda)) \xrightarrow{\mathcal{L}} \mathcal{N} (0, \sqrt{\lambda} | -e^{-\lambda} |).$$

$$\sqrt{n} (e^{-\bar{X}_n} - e^{-\lambda}) \xrightarrow{\mathcal{L}} \mathcal{N} (0, \sqrt{\lambda} e^{-\lambda}).$$

2.2 Estimation de la densité de probabilité d'une variable aléatoire.

Problème : Comment faire une inférence statistique quand on ne connaît pas la loi des observations X_1, \dots, X_T ?

2.2.1 Estimation par histogramme

Une première solution est de reproduire la représentation en escalier de la fonction de répartition empirique F_n pour f

$$\hat{f}(x) = \sum_{i=1}^k \omega_i \mathbb{1}_{[a_i, a_{i+1}[}(x), \quad a_1 < \dots < a_{k+1}. \quad (2.4)$$

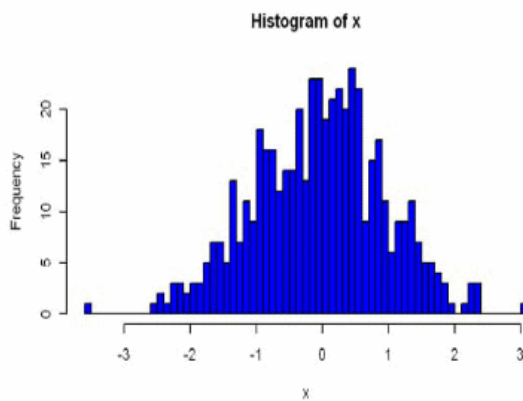
En choisissant les ω_i telles que $\sum_{i=1}^k \omega_i (a_i - a_{i+1}) = 1$ et $\omega_i (a_i - a_{i+1}) = \widehat{\mathbb{P}}_F(X \in [a_i, a_{i+1}[)$,

Par exemple,

$$\omega_i (a_i - a_{i+1}) = \frac{1}{T} \sum_{j=1}^T \mathbb{1}_{[a_i, a_{i+1}[}(X_j)$$

est un estimateur convergent de $\widehat{\mathbb{P}}_F(X \in [a_i, a_{i+1}[)$.

le code : `hist(x)density` donne les valeurs des a_i et `hist(x)breaks` les valeurs des a_i . le code : `x < -rnorm(500), hist(x, breaks = 50, col = "blue")` donne



Histogramme d'une suite normale avec $k = 50, T = 500$.

2.2 Estimation de la densité de probabilité d'une variable aléatoire.

Inconvénients : i) L'estimateur dépend du choix de la partition (a_i) , souvent construite en fonction des données (comme dans R)

ii) Problème des extrémités a_1 et a_{k+1} : ils ne peuvent pas être infinis mais doivent suffisamment approcher le support de f .

iii) k et (a_i) doivent dépendre de T pour que \hat{f} converge vers f mais... $a_{i+1} - a_i$ ne doit pas décroître trop vite vers 0 pour que l'estimation soit convergente : il faut suffisamment d'observations par intervalle $[a_i, a_{i+1}[$

iv) L'histogramme est une fonction discontinue

2.2.2 Estimateur à noyau

Au lieu de considérer une approximation uniforme autour de chaque X_i , on peut utiliser une fonction plus lisse :

$$\hat{f}_T(x) = \frac{1}{Th} \sum_{i=1}^T K\left(\frac{x - X_i}{h}\right) \quad (2.5)$$

où K est un noyau (par exemple une densité de probabilité et h un facteur d'échelle. Sur R on peut utiliser les noyaux suivants

1) Le noyau normal [kernel="gaussian" ou "g"]

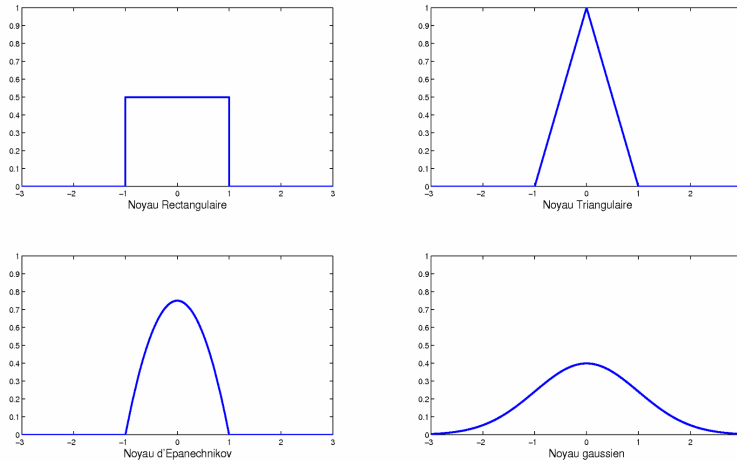
2) Le noyau d'Epanechnikov [kernel="epanechnikov" ou "e"] $K(y) = C(1 - y^2)\mathbb{1}_{[-1,1]}(y)$

3) Le noyau triangulaire [kernel="triangular" ou "t"] $K(y) = (1 + y)\mathbb{1}_{[-1,0]}(y) + (1 - y)\mathbb{1}_{[0,1]}(y)$.

4) Les noyaux "rectangular", "biweight", "cosine", "optcosine". L'estimation est très influencée par le choix du noyau

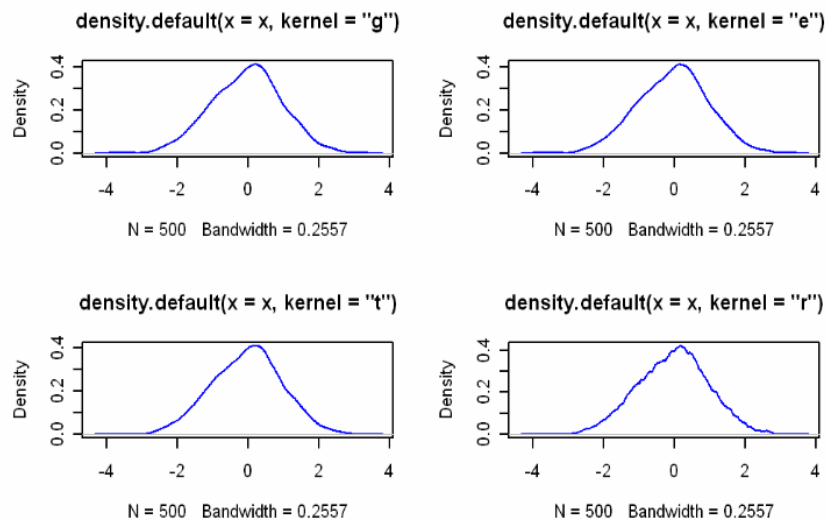
2.2 Estimation de la densité de probabilité d'une variable aléatoire.

Les courbes de ces noyaux sont présentées ci-dessous :



Exemple :

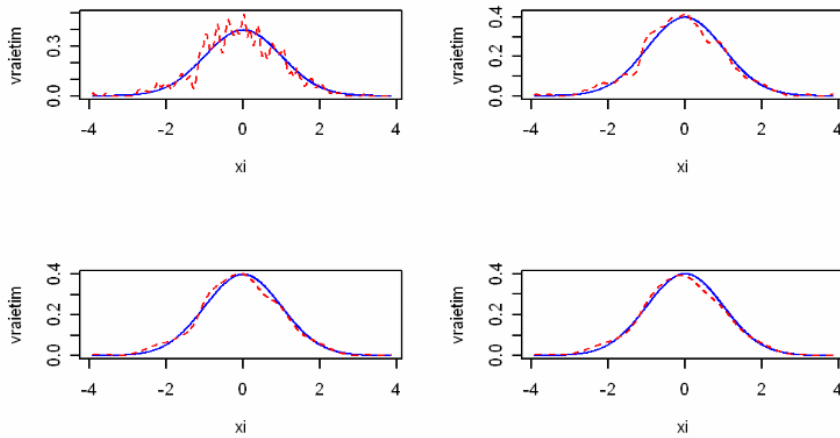
Pour la série simulée précédemment on obtient



Le noyau rectangulaire donne un estimateur moins lisse. Le choix de la fenêtre h est crucial, Si h grand, un grand nombre des X_i contribuent à l'estimation de $f(x)$, on obtient un esti-

2.2 Estimation de la densité de probabilité d'une variable aléatoire.

mateur avec un biais très grand et une variance très petite. Si h petit, peu de X_i contribuent à l'estimation de $f(x)$, on obtient un estimateur avec un biais très petit et une variance très grande



Estimateurs à noyau en fonction du choix de la fenêtre h_T . La vraie densité en bleu et l'estimateur en rouge. Quatre choix pour h_T : $T^{-1/2}, T^{-1/3}, T^{-1/4}$ et $T^{-1/5}$

Le choix $h_T = T^{-1/5}$ (en bas à droite) semble être meilleur Fenêtre optimale En étudiant l'erreur moyenne intégrée

$$d(f, \hat{f}) = \mathbb{E} \int (f(x) - \hat{f}(x))^2 dx,$$

on peut trouver un choix optimal pour la fenêtre h . De la décomposition

$$d(f, \hat{f}) = \int (f(x) - \mathbb{E}(\hat{f}(x)))^2 dx + \int \text{Var} \hat{f}(x) dx \text{ (biais}^2 + \text{Var)}$$

et les approximations

$$f(x) - \mathbb{E}(\hat{f}(x)) \simeq \frac{f''(x)}{2} h_T^2$$

$$\mathbb{E} \left[\exp \left(-\frac{(X_i - x)^2}{2h_T^2} \right) \right] \simeq f(x) \sqrt{2\pi} h_T,$$

on en déduit que le biais est de l'ordre de

$$\int \left(\frac{f''(x)}{2} \right)^2 dx h_T^4,$$

2.2 Estimation de la densité de probabilité d'une variable aléatoire.

et que le terme de variance est approximativement

$$\frac{1}{Th_T\sqrt{2\Pi}}$$

Par conséquent, l'erreur tend vers 0 quand T tend vers ∞ si i) h_T tend vers 0 et ii) Th_T tend vers l'infini. La fenêtre optimale est donnée par

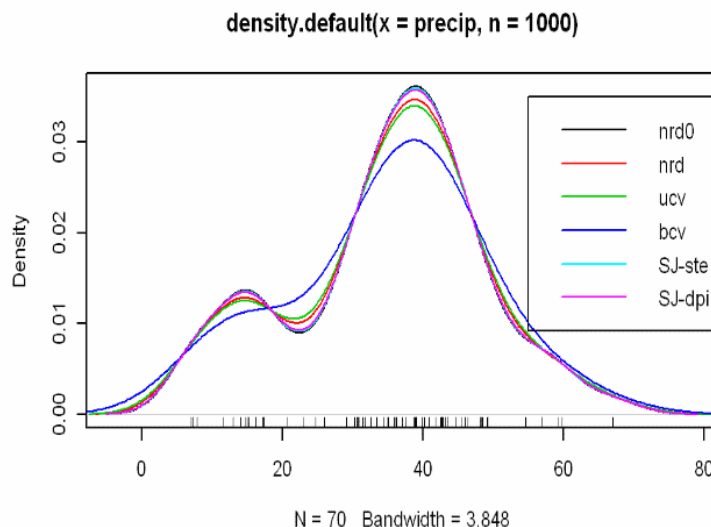
$$h_{opt} = \left(T\sqrt{2\Pi} \int (f''(x))^2 dx \right)^{1/5}$$

h_{opt} dépend de la dérivée seconde qui elle même inconnue. Pour résoudre ce problème il existe plusieurs méthodes : La fenêtre optimale basée sur "rule of thumb", avec le noyau de Gauss, a la forme

$$h_{opt} = \frac{0.9 \min(\hat{\sigma}, \hat{q}_{75} - \hat{q}_{25})}{1.34T^{1/5}},$$

ou σ est l'écart-type estimé et \hat{q}_{25} et \hat{q}_{75} sont les quantiles estimés (Silverman (1986, page 48, eqn (3.31)). Sur R il y a les méthodes suivantes :

- bw.nrd0* implémente la "rule-of-thumb" de Silverman.
- bw.nrd* une variation de la précédente Scott (1992), utilisant le 1.06.
- bw.ucv* et *bw.bcv* utilisent la validation croisée non biaisée et biaisée respectivement.
- bw.SJ* implémente la méthode de Sheather et Jones (1991). Exemple : On considère la série des précipitations annuelles "precip", la densité estimée avec les 6 choix possibles est donnée par



2.3 Estimation par la méthode du noyau

2.3.1 La fonction de répartition empirique

Définition 2.7 Soient X_1, X_2, \dots, X_n un n -échantillon de X de fonction de répartition F , on appelle fonction de répartition empirique la fonction F_n définie par :

$$F_n(x) : \mathbb{R} \rightarrow [0, 1]$$

$$x \rightarrow F_n(x) = \frac{1}{n} \sum_{i=1}^n I_{]-\infty, x]}(x_i)$$

Proposition 2.3 Soient X_1, X_2, \dots, X_n un n -échantillon de X de fonction de répartition F , alors F_n est un estimateur sans biais pour $F(x)$.

Preuve

Notre objectif est de vérifier que $E[F_n(x)] \rightarrow F(x)$ quand $n \rightarrow \infty$.

Par définition de la fonction de répartition :

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I_{[x_i \leq x]}$$

. Ceci implique

$$\begin{aligned} E[F_n(x)] &= \frac{1}{n} \sum_{i=1}^n E[I_{x_i \leq x}] \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{P}(X_i \leq x) \\ &= \frac{n}{n} F(x) \\ &= F(x). \end{aligned}$$

Ainsi $F_n(x)$ est asymptotiquement sans biais.

Proposition 2.4 Soient X_1, X_2, \dots, X_n un n -échantillon de X de fonction de répartition F , alors F_n converge presque complètement vers F .

Preuve

L'objectif est de vérifier que :

$$\forall \epsilon > 0, \sum_{n=1}^{\infty} \mathbb{P}[|F_n(x) - F(x)| > \epsilon] < \infty$$

2.3 Estimation par la méthode du noyau

. En effet,

$$\begin{aligned} \mathbb{P}[|F_n(x) - F(x)| > \epsilon] &= \mathbb{P}[|F_n(x) - E[F_n(x)] + E[F_n(x)] - F(x)| > \epsilon] \\ &\leq \mathbb{P}[|F_n(x) - E[F_n(x)]| > \frac{\epsilon}{2}] + \mathbb{P}[|E[F_n(x)] - F(x)| > \frac{\epsilon}{2}]. \end{aligned}$$

Pour que

$$\forall \epsilon > 0, \sum_{n=1}^{\infty} \mathbb{P}[|F_n(x) - F(x)| > \epsilon] < +\infty$$

. Il suffit de vérifier que

$$\sum_{n=1}^{\infty} \mathbb{P}[|F_n(x) - E[F_n(x)]| > \frac{\epsilon}{2}] < +\infty$$

et

$$\sum_{n=1}^{\infty} \mathbb{P}[|E[F_n(x)] - F(x)| > \frac{\epsilon}{2}] < +\infty$$

D'après la proposition précédente $E[F_n(x)] = F(x)$ ce qui est équivalent à

$$\forall \epsilon > 0, \sum_{n=1}^{\infty} \mathbb{P}[|E[F_n(x)] - F(x)| > \frac{\epsilon}{2}] = 0$$

. Maintenant il reste de montrer que

$$\sum_{n=1}^{\infty} \mathbb{P}[|F_n(x) - E[F_n(x)]| > \frac{\epsilon}{2}] < +\infty$$

En effet; on considère la suite

$$x_n = \sqrt{\frac{\log n}{n}} \rightarrow 0$$

quand

$$n \rightarrow \infty$$

et on démontre qu'il existe ϵ_0 tel que

$$\sum_{n=1}^{\infty} \mathbb{P} \left[|F_n(x) - E[F_n(x)]| > \epsilon_0 \sqrt{\frac{\log n}{n}} \right] < +\infty$$

Comme la suite $x_n \rightarrow 0$

$$\forall \epsilon, \exists N$$

tel que

$$\forall n > N, \epsilon_0 |x_n| < \frac{\epsilon}{2}$$

2.3 Estimation par la méthode du noyau

Si $n > N$

$$\mathbb{P}[|F_n(x) - E[F_n(x)]| > \frac{\epsilon}{2}] \leq \mathbb{P}[|F_n(x) - E[F_n(x)]| > \epsilon_0 | x_n]$$

. Par contre si $n \leq N$

$$\sum_{n=1}^N \mathbb{P}[|F_n(x) - E[F_n(x)]| > \frac{\epsilon}{2}] < +\infty$$

Maintenant il reste de montrer que

$$\sum_{n=N}^{\infty} \mathbb{P}[|F_n(x) - E[F_n(x)]| > \frac{\epsilon}{2}] < +\infty$$

$$\sum_{n=N}^{\infty} \mathbb{P}[|F_n(x) - E[F_n(x)]| > \frac{\epsilon}{2}] \leq \sum_{n=N}^{\infty} \mathbb{P}[|F_n(x) - E[F_n(x)]| > \epsilon_0 \sqrt{\frac{\log n}{n}}]$$

. Ainsi il suffit de montrer qu'il existe ϵ_0 tel que

$$\sum_{n=N}^{\infty} \mathbb{P} \left[|F_n(x) - E[F_n(x)]| > \epsilon_0 \sqrt{\frac{\log n}{n}} \right]$$

. En effet ;

$$\begin{aligned} \sum_{n=N}^{\infty} \mathbb{P} \left[\frac{1}{n} \sum_{i=1}^n I_{]-\infty, x]}(x_i) - E \left[\frac{1}{n} \sum_{i=1}^n I_{]-\infty, x]}(x_i) \right] \geq \epsilon_0 \sqrt{\frac{\log n}{n}} \right] = \\ \sum_{n=N}^{\infty} \mathbb{P} \left[\sum_{i=1}^n I_{]-\infty, x]}(x_i) - E \left[\sum_{i=1}^n I_{]-\infty, x]}(x_i) \right] \geq \epsilon_0 \sqrt{n \log n} \right] \end{aligned}$$

On applique l'inégalité de Bernstein :

$$x_i \longleftrightarrow I_{]-\infty, x]}(x_i)$$

$$\alpha_i = 0$$

$$\beta_i = 1$$

$$t = \epsilon_0 \sqrt{n \log n}$$

Donc

$$\begin{aligned} \sum_{n=N}^{\infty} \mathbb{P}[|F_n(x) - E[F_n(x)]| > \epsilon_0 \sqrt{\frac{\log n}{n}}] &\leq 2 \sum_{n=N}^{\infty} \exp \frac{-2\epsilon_0^2 n \log n}{n} \\ &\leq 2 \sum_{n=N}^{\infty} n^{-2\epsilon_0^2} \end{aligned}$$

A condition que $2\epsilon_0^2 > 1$,

$$\sum_{n=N}^{\infty} \mathbb{P}[|F_n(x) - E[F_n(x)]| > \frac{\epsilon}{2}] < \infty$$

. Par conséquent

$$\sum_{n=1}^{\infty} \mathbb{P}[|F_n(x) - E[F_n(x)]| > \frac{\epsilon}{2}] < \infty$$

. Ainsi $F_n(x)$ est un estimateur qui converge presque complètement.

Proposition 2.5 Soient X_1, X_2, \dots, X_n un n -échantillon de X de fonction de répartition F , alors F_n converge en moyenne quadratique vers F .

Preuve

Pour montrer que F_n converge en moyenne quadratique vers F , il suffit de montrer que

$$\mathbb{E}[F_n(x) - F(x)]^2 \rightarrow 0$$

$$\text{Var}[F_n(x) - F(x)] = \mathbb{E}[F_n(x) - F(x)]^2 - [\mathbb{E}[F_n(x) - F(x)]]^2$$

Pour montrer que $\mathbb{E}[F_n(x) - F(x)]^2 \rightarrow 0$, il suffit de montrer que $\text{Var}[F_n(x)] \rightarrow 0$ car

$$[\mathbb{E}[F_n(x) - F(x)]]^2 = 0$$

et

$$\text{Var}[F_n(x) - F(x)] = \text{Var}[F_n(x)]$$

On réordonne les observations :

$$X_1, \dots, X_n = X_{(1)}, \dots, X_{(n)}$$

On obtient donc la définition suivante de $F_n(x)$:

Si $x \leq X_i$, alors $nF_n(x) = 0$.

Si $X_i \leq x \leq X_{i+1}$, alors $nF_n(x) = i$.

Si $x \geq X_n$, alors $nF_n(x) = n$.

Les valeurs possibles pour $nF_n(x)$ sont $\{0, 1, \dots, n\}$.

$$\begin{aligned} \mathbb{P}[nF_n(x) = 0] &= \mathbb{P}[\forall i, X_i > x] \\ &= \mathbb{P}\left[\bigcap_{i=1}^n X_i > x\right] \\ &= \prod_{i=1}^n \mathbb{P}[X_i > x] \\ &= \prod_{i=1}^n \{1 - \mathbb{P}(X_i \leq x)\} \\ &= (1 - F(x))^n \end{aligned}$$

$$\mathbb{P}[nF_n(x) = i] = 0$$

D'où F_n est un estimateur qui converge en moyenne quadratique.

2.4 L'estimation de la fonction de répartition

Définition 2.8 Soient X_1, X_2, \dots, X_n un n -échantillon de X de fonction de répartition F , on appelle estimateur à noyau H l'estimateur défini par :

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n H\left(\frac{x - X_i}{h_n}\right)$$

Où $(h_n)_n$ est une suite convergente vers 0 quand n tend vers ∞ et H est une fonction de répartition. h_n est appelé paramètre de lissage.

Théorème 2.5 Soient X_1, X_2, \dots, X_n un n -échantillon de X de fonction de répartition F , si

$$\int_{-\infty}^{+\infty} tH'(t)dt < +\infty$$

alors l'estimateur à noyau F_n est asymptotiquement sans biais.

Preuve

Notre objectif est de vérifier que $\mathbb{E}[F_n(x)] \rightarrow F(x)$ quand $n \rightarrow \infty$.

$$\begin{aligned} \mathbb{E}[F_n(x)] &= \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n H\left(\frac{x - X_i}{h_n}\right)\right] \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}\left[H\left(\frac{x - X_i}{h_n}\right)\right] \\ &= \mathbb{E}\left[H\left(\frac{x - X_i}{h_n}\right)\right] \\ &= \int_{-\infty}^{+\infty} H\left(\frac{x - z}{h_n}\right) f(z) dz \end{aligned}$$

Il suffit de prendre $2\epsilon_0^2 > 1$ pour conclure que

$$\sum_{n=1}^{\infty} \mathbb{P}[|F_n(x) - E[F_n(x)]| > \frac{\epsilon}{2}] < +\infty$$

. Pour la deuxième inégalité, on a d'après le théorème précédent $\mathbb{E}[F_n(x)] - F(x)$ tend vers 0 quand n tend vers $+\infty$ ce qui est équivalent à

$$\forall \epsilon > 0, \exists N \in \mathbb{N} / \forall n, n \geq N \Rightarrow |\mathbb{E}[F_n(x)] - F(x)| < \epsilon$$

donc :

$$\begin{aligned} \sum_{n=1}^{\infty} \mathbb{P}[|\mathbb{E}[F_n(x)] - F(x)| > \frac{\epsilon}{2}] &= \sum_{n=1}^N \mathbb{P}[|\mathbb{E}[F_n(x)] - F(x)| > \frac{\epsilon}{2}] + \sum_{n=N}^{\infty} \mathbb{P}[|\mathbb{E}[F_n(x)] - F(x)| > \frac{\epsilon}{2}] \\ &= \sum_{n=1}^N \mathbb{P}[|\mathbb{E}[F_n(x)] - F(x)| > \frac{\epsilon}{2}] < \infty \end{aligned}$$

2.4 L'estimation de la fonction de répartition

Ainsi $F_n(x)$ est un estimateur qui converge presque complètement.

Proposition 2.6 Soient X_1, X_2, \dots, X_n un n -échantillon de X de fonction de répartition F , alors sous les mêmes conditions du théorème précédent l'estimateur à noyau F_n converge en moyenne quadratique vers F .

Preuve

Pour montrer que F_n converge en moyenne quadratique vers F il faut montrer que

$$\mathbb{E}([F_n(x) - F(x)]^2) \rightarrow 0$$

quand $n \rightarrow \infty$ On a

$$\begin{aligned} \text{Var}[F_n(x)] &= \text{Var}[F_n(x) - F(x)] \\ &= \mathbb{E}([F_n(x) - F(x)]^2) - (\mathbb{E}[F_n(x) - F(x)])^2 \end{aligned}$$

Donc il suffit de montrer que :

$$\text{Var}[F_n(x)] \rightarrow 0$$

quand $n \rightarrow \infty$. Puisque le deuxième terme tend vers 0. En effet ;

$$\begin{aligned} \text{Var}[F_n(x)] &= \text{Var}\left[\frac{1}{n} \sum_{i=1}^n H\left(\frac{x - X_i}{h_n}\right)\right] \\ &= \frac{1}{n^2} \sum_{i=1}^n \text{Var}\left[H\left(\frac{x - X_i}{h_n}\right)\right] \\ &= \frac{1}{n} \text{Var}\left[H\left(\frac{x - X_i}{h_n}\right)\right] \end{aligned}$$

et

$$\text{Var}\left[H\left(\frac{x - X_i}{h_n}\right)\right] = \mathbb{E}\left[H^2\left(\frac{x - X_i}{h_n}\right)\right] - (\mathbb{E}\left[H\left(\frac{x - X_i}{h_n}\right)\right])^2$$

D'une part, une intégration par parties et un changement de variables donnent

$$\begin{aligned} \mathbb{E}\left[H^2\left(\frac{x - X_i}{h_n}\right)\right] &= \int_{-\infty}^{+\infty} H^2\left(\frac{x - z}{h_n}\right) f(z) dz \\ &= 2 \int_{-\infty}^{+\infty} H(t) H'(t) F(x - h_n t) dt \end{aligned}$$

D'autre part le développement de Taylor à l'ordre 1 nous permet d'écrire

$$\mathbb{E}\left[H^2\left(\frac{x - X_i}{h_n}\right)\right] = 2F(x) \int_{-\infty}^{+\infty} H(t) H'(t) dt - 2h_n \int_{-\infty}^{+\infty} t H(t) H'(t) f(\xi) dt$$

2.4 L'estimation de la fonction de répartition

, $\xi \in]x - h_n t, x[$. Au passage à la limite quand n tend vers ∞ , $\mathbb{E}[H^2(\frac{x - X_i}{h_n})]$ tend vers $2F(x) \int_{-\infty}^{+\infty} H(t)H'(t)dt < \infty$. Or, $\mathbb{E}[H(\frac{x - X_i}{h_n})]$ tend vers $F(x)$. Ainsi,

$$\text{Var}[F_n(x)] = \frac{1}{n} \text{Var}[H(\frac{x - X_i}{h_n})]$$

tend vers 0 quand n tend vers ∞ .

D'où la convergence en moyenne quadratique de l'estimateur à noyau F_n vers F . Ceci achève la démonstration de la proposition.

2.4.1 L'estimation de la fonction de densité

Définition 2.9 Soient X_1, X_2, \dots, X_n un n -échantillon de X de densité f , On appelle un estimateur à noyau pour la densité f la variable aléatoire définie par :

$$f_n(x) = \frac{1}{nh_n} \sum_{i=1}^n K(x - X_i h_n)$$

. Où K est un noyau borné intégrable et $(h_n)_n$ est une suite de réels avec h_n (resp. nh_n tend vers 0 (resp. $+\infty$) quand n tend vers ∞ .

Théorème 2.6 Si K est borné tel que :

$$\int_{-\infty}^{+\infty} tK(t)dt < \infty$$

alors f_n est asymptotiquement sans biais.

Preuve

$$\begin{aligned} \mathbb{E}[f_n(x)] &= \mathbb{E}\left[\frac{1}{nh_n} \sum_{i=1}^n K(x - X_i h_n)\right] \\ &= \frac{1}{nh_n} \sum_{i=1}^n \mathbb{E}[K(x - X_i h_n)] \\ &= \frac{1}{h_n} \mathbb{E}[K(x - X_i h_n)] \\ &= \frac{1}{h_n} \int_{-\infty}^{+\infty} K\left(\frac{x - X_i}{h_n}\right) f(z) dz \end{aligned}$$

2.4 L'estimation de la fonction de répartition

On considère maintenant le changement de variable suivant

$$t = \frac{x - z}{h_n} \Rightarrow dt = -\frac{1}{h_n} dz$$

$$\begin{aligned} \mathbb{E}[f_n(x)] &= \frac{1}{h_n} \int_{-\infty}^{+\infty} K(t) f(x - h_n t) h_n dt \\ &= \int_{-\infty}^{+\infty} K(t) f(x - h_n t) dt \end{aligned}$$

Maintenant on applique le développement de Taylor à l'ordre 1, on trouve

$$\begin{aligned} \mathbb{E}[f_n(x)] &= \int_{-\infty}^{+\infty} K(t) (f(x) - h_n t f'(\xi)) dt \\ &= f(x) \int_{-\infty}^{+\infty} K(t) dt - h_n \int_{-\infty}^{+\infty} t K(t) f'(\xi) dt \\ &= f(x) - h_n \int_{-\infty}^{+\infty} t K(t) f'(\xi) dt \end{aligned}$$

où $\xi \in]x - h_n t, x[$. Comme h_n tend vers 0 quand n tend vers $+\infty$ alors $\mathbb{E}[f_n(x)] = f(x)$. D'où f_n est asymptotiquement sans biais.

Proposition 2.7 Soient X_1, X_2, \dots, X_n un n -échantillon de X de densité f , alors sous les mêmes conditions du théorème précédent f_n converge en moyenne quadratique vers f .

Preuve

Comme dans la proposition 2.5, il suffit de prouver que $\text{Var}[f_n(x)]$ tend vers 0 quand n tend vers ∞ pour démontrer la convergence en moyenne quadratique de l'estimateur f_n de f . À cet effet; on a :

$$\begin{aligned} \text{Var}[f_n(x)] &= \text{Var}\left[\frac{1}{nh_n} \sum_{i=1}^n K(x - X_i h_n)\right] \\ &= \frac{1}{n^2 (h_n)^2} \sum_{i=1}^n \text{Var}[K(x - X_i h_n)] \\ &= \frac{1}{n (h_n)^2} \text{Var}[K(x - X_i h_n)] \\ &= \frac{1}{n (h_n)^2} [\mathbb{E}[K^2(x - X_i h_n)] - (\mathbb{E}[K(x - X_i h_n)])^2] \\ &= \frac{1}{nh_n} \left[\frac{1}{h_n} \mathbb{E}[K^2(x - X_i h_n)] - \frac{1}{n} \left[\frac{1}{h_n} (\mathbb{E}[K(x - X_i h_n)]) \right]^2 \right] \end{aligned}$$

2.4 L'estimation de la fonction de répartition

On pose

$$A = \frac{1}{nh_n} \left[\frac{1}{h_n} \mathbb{E}[K^2(x - X_i h_n)] \right]$$

et

$$B = \frac{1}{n} \left[\frac{1}{h_n} (\mathbb{E}[K(x - X_i h_n)])^2 \right]$$

D'une part, on a

$$\frac{1}{h_n} E[K^2(x - X_i h_n)] = \frac{1}{h_n} \int_{-\infty}^{+\infty} K^2(x - zh_n) f(z) dz$$

. On considère le changement de variable suivant :

$$t = \frac{x - z}{h_n} \Rightarrow dt = \frac{-1}{h_n} dz$$

.

$$\frac{1}{h_n} \mathbb{E}[K^2(x - X_i h_n)] = \int_{-\infty}^{+\infty} K^2(t) f(x - h_n t) dt$$

. L'application du développement de Taylor à l'ordre 1 donne :

$$f(x) = f(x_0) + (x - x_0) f'(\xi)$$

où $\xi \in]x, x_0[$. ou encore

$$\begin{aligned} f(x - h_n t) &= f(x) + (x - h_n t - x) f'(\xi) \\ &= f(x) - h_n t f'(\xi) \end{aligned}$$

Il résulte que

$$\begin{aligned} \frac{1}{h_n} \mathbb{E}[K^2(x - X_i h_n)] &= \int_{-\infty}^{+\infty} K^2(t) (f(x) - h_n t f'(\xi)) dt \\ &= f(x) \int_{-\infty}^{+\infty} K^2(t) dt - h_n \int_{-\infty}^{+\infty} t K^2(t) f'(\xi) dt \end{aligned}$$

On voit clairement que $\frac{1}{h_n} \mathbb{E}[K^2(\frac{x - X_i}{h_n})]$ tend vers $f(x) \int_{-\infty}^{+\infty} K^2(t) < \infty$ quand n tend vers ∞ par suite la quantité A tend vers 0 puisque nh_n tend vers ∞ quand n tend vers ∞ . D'autre part

$$\begin{aligned} \frac{1}{h_n} E[K(x - X_i h_n)] &= \frac{1}{h_n} \int_{-\infty}^{+\infty} K(x - zh_n) f(z) dz \\ &= f(x) \int_{-\infty}^{+\infty} K(t) dt - h_n \int_{-\infty}^{+\infty} t K(t) f'(\xi) dt \end{aligned}$$

2.4 L'estimation de la fonction de répartition

D'où $\frac{1}{h_n} E[K(x - X_i h_n)]$ tend vers $f(x)$ quand n tend vers ∞ . Finalement la quantité B tend vers 0 quand n tend vers ∞ .

Alors $\text{Var}[f_n(x)]$ tend vers 0 quand n tend vers ∞ . Ceci affirme la convergence en moyenne quadratique de f_n vers f .

Chapitre 3

Notions de dépendance et inégalités des moments

Ce chapitre se divise en deux sections. Dans la première section nous étudions les Différentes types de dépendance et la relation entre eux. La deuxième section sera consacrée aux quelques inégalités des moments sous des conditions de dépendance, ces inégalités sont utiles pour la suite de ce travail.

3.1 Notions de dépendance

*La théorie asymptotique classique dans la statistique est construite autour du théorème de la limite centrale CLT et la loi des grands nombres pour des suites des variables aléatoires indépendantes. Dans l'étude des propriétés asymptotiques des modèles linéaires qui sont considérés comme des suites dépendantes des variables aléatoires, l'approche populaire consiste à exprimer un processus linéaire en termes de processus autorégressifs ou à moyenne mobile MA où le bruit blanc $\{\varepsilon_t\}$ est supposé être *i.i.d.* Malheureusement, la représentation MA n'est plus pertinente dans le contexte des processus non linéaires, où plus de structures de dépendance complexes seront rencontrées. Nous devons imposer certaine indépendance asymptotique pour apprécier les propriétés des grands échantillons des processus non linéaires inférentiels. Un processus mélangeant peut être considéré comme une suite de variables aléatoires dont le passé et le futur sont asymptotiquement indépendant.*

Une généralisation utile d'une suite de variables aléatoires indépendantes, identiquement distribuées est un processus aléatoire mélangeant, en général, les variables aléatoires loin dans le temps les uns des autres sont presque indépendant (faiblement dépendant).

Pour des suites mélangeants, à la fois la loi des grands nombres (c'est à dire, le théorème

3.1 Notions de dépendance

ergodique) et le théorème de la limite centrale peut être établie. Dans cette section, nous introduisons les différentes conditions de mélange. Le α -mélange est le plus faible parmi les conditions de mélange les plus utilisées, nous citons certains théorèmes limites et les inégalités de probabilités pour des processus α -mélange. elles jouent des rôles très importantes dans le développement de la théorie asymptotique pour les processus linéaire et non linéaire. Pour une discussion plus détaillée sur les conditions de mélange, nous renvoyons le lecteur à Bradley (1986)[20] et Doukhan (1994) [42]. Enfin, nous présentons une forme générique du théorème de la limite centrale qui est constamment rencontré dans la régression non paramétrique à noyau.

3.1.1 Conditions de mélange

Pour simplifier la notation, nous présentons les conditions de mélange pour des processus strictement stationnaires (du fait que le processus de mélange n'est pas nécessairement stationnaire). L'idée est de définir des coefficients de mélange pour mesurer la force (de différentes façons) de dépendance entre deux segments d'un processus, qui sont éloignés l'un de l'autre dans le temps. Soit $\{X_t, t = 0, 1, 2, \dots\}$ un processus strictement stationnaire. Pour $n = 1, 2, \dots$, on définit

$$\alpha(n) = \sup_n \{|\mathbb{P}(A)\mathbb{P}(B) - \mathbb{P}(AB)|, A \in \mathcal{F}_{-\infty}^m, B \in \mathcal{F}_{m+n}^\infty\}, \quad (3.1)$$

$$\beta(n) = \mathbb{E} \left[\sup_{B \in \mathcal{F}_n^\infty} |\mathbb{P}(B) - \mathbb{P}(B|X_0, X_{-1}, X_{-2}, \dots)| \right], \quad (3.2)$$

$$\rho(n) = \sup_{X \in \mathcal{L}^2(\mathcal{F}_{-\infty}^0), Y \in \mathcal{L}^2(\mathcal{F}_n^\infty)} |\text{Corr}(X, Y)|, \quad (3.3)$$

$$\varphi(n) = \sup_{A \in \mathcal{F}_{-\infty}^0, B \in \mathcal{F}_n^\infty, \mathbb{P}(A) > 0} |\mathbb{P}(B) - \mathbb{P}(B|A)|, \quad (3.4)$$

$$\psi(n) = \sup_{A \in \mathcal{F}_{-\infty}^0, B \in \mathcal{F}_n^\infty, \mathbb{P}(A)\mathbb{P}(B) > 0} |1 - \mathbb{P}(B|A)/\mathbb{P}(B)|, \quad (3.5)$$

où \mathcal{F}_i^j est la σ -algèbre engendrée par $\{X_t, i \leq t \leq j\}$, et $\mathcal{L}^2(\mathcal{F}_i^j)$ se compose des variables aléatoires \mathcal{F}_i^j -mesurable de second moment fini. (Les lecteurs sont renvoyés au chapitre 1 de Chow et Teicher (1997) [44] pour la définition d'une σ -algèbre et les fonctions mesurables). Intuitivement, \mathcal{F}_i^j assemble toutes les informations sur les processus recueillies entre le temps i et le temps j . Quand au moins un des coefficients de mélange converge vers 0 quand n tend vers ∞ , nous pouvons dire que le processus $\{X_t\}$ est asymptotiquement indépendant. Noter

3.1 Notions de dépendance

que $\mathcal{F}_{n+1}^\infty \subset \mathcal{F}_n^\infty$ pour tout $n \geq 1$. Ainsi, l'ensemble des coefficients de mélange définies ci-dessus sont monotones non croissante.

Définition 3.1 Le processus $\{X_t\}$ est dite :

1. α -mélangeant (mélange fort) si $\lim_{n \rightarrow \infty} \alpha(n) = 0$, cette définition a été introduite par Rosenblatt (1956),
2. β -mélangeant (régularité absolue) si $\lim_{n \rightarrow \infty} \beta(n) = 0$,
3. ρ -mélangeant (régularité complète) si $\lim_{n \rightarrow \infty} \rho(n) = 0$, cette définition a été introduite par Kolmogorov et Rosanov (1960),
4. φ -mélangeant (mélange uniforme) si $\lim_{n \rightarrow \infty} \varphi(n) = 0$, et
5. ψ -mélangeant si $\lim_{n \rightarrow \infty} \psi(n) = 0$.

On peut classé, quelques facteurs de mélange maintenant comme suit :

1. Le diagramme suivant illustre les relations entre les cinq conditions de mélange :

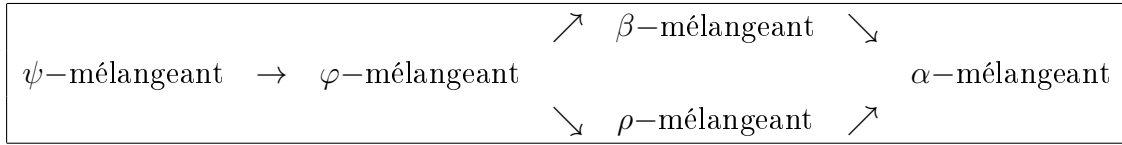


TABLE 3.1 – Relations entre quelques conditions de dépendance

En plus, il est a maintenu que

$$\alpha(n) \leq \frac{1}{4}\rho(n) \leq \frac{1}{2}\varphi^{1/2}(n).$$

Pour la démonstration, voir R.C.Bradley (1986) [20]. Le α -mélangeant, aussi dite mélange fort, ce n'est pas comme son non l'indique, c'est le plus faible parmi les cinq conditions cité dessus, il est clair qui est impliqué par n'importe quel des autres quatres conditions. Par contre le ψ -mélangeant est le plus fort. En général, β -mélangeant (dite aussi régularité absolu) et le ρ -mélangeant aucun n'implique l'autre. Mais pour des processus gaussien, ρ -mélangeant est équivalent à α -mélangeant, il est donc plus faible que β -mélangeant. Voir Doukhan (1994) [42].

2. Les propriétés de mélange sont héréditaire dans le sens que, pour tout fonction mesurable $m(\cdot)$, le processus $\{m(X_t)\}$ possède les propriété de mélange de $\{X_t\}$. Voir Billingsley (1968)[24].

3.1 Notions de dépendance

3. On considère le processus $MA(\infty)$

$$X_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j},$$

où $a_j \rightarrow 0$ rapidement (on note qu'un processus causale $ARMA(p, q)$ vérifié cette condition), et ε_t est une suite *i.i.d.* Si la fonction de densité de probabilité de ε_t existe (comme les distributions Normale, Cauchy, Exponentiel et Uniforme), alors $\{X_t\}$ est β -mélangeant avec $\beta(n) \rightarrow 0$ exponentiellement rapide (Voir Pham et Tran (1985) [52]). Mais ce résultat c'est pas toujours vérifié quand ε_t est discrète. Par exemple $X_{t+1} = 0,5X_t + \varepsilon_t$ n'est pas α -mélangeant si ε_t suit une distribution binômiale (Voir Andrews 1984 [53])

4. Si $\{X_t\}$ est une chaîne de Markov strictement stationnaire, les coefficients de mélange sont effectivement définies dans $(\mathcal{F}_{-\infty}^0, \mathcal{F}_n^{\infty})$ remplacé par $(\sigma(X_0), \sigma(X_n))$, où $\sigma(X_t)$ désigne la σ -algèbre engendrée seulement par la variable aléatoire X_t (Voir le théorème 4.1 de Bradley (1986) [20]). En outre, les coefficients de mélange décroissent vers 0 exponentiellement rapide si $\{X_t\}$ est ρ -, φ -, ou ψ -mélange (théorème 4.2 de Bradley 1986 [20]).

5. Il résulte de (4) ci-dessus et le lemme 1.3 dans Bosq (1998) [54] que si $\{X_t\}$ est une chaîne de Markov strictement stationnaire et α -mélange, le coefficient de mélange est majoré par

$$\alpha(n) \leq \frac{1}{2} \int |f_{0,n}(x, y) - f(x)f(y)| dx dy,$$

où f est la fonction de densité marginal de X_t , et $f_{0,n}$ est la densité conjointe de (X_0, X_n) .

6. Davydov (1973) [56] montre que, pour une chaîne de Markov strictement stationnaire $\{X_t\}$,

$$\beta = \int \|F_n(\cdot|x) - F(\cdot)\| F(dx), \quad (3.6)$$

où F est la distribution marginale de X_t , $F_n(\cdot|x)$ est la distribution conditionnelle de X_n avec $X_0 = x$, et $\|\cdot\|$ désigne la variation totale. Si $\{X_t\}$ est géométriquement ergodique satisfaisant à la condition supplémentaire

$$\|F_n(\cdot|x) - F(\cdot)\| \leq A(x)\eta^n, \quad (3.7)$$

presque sûrement en respectant la distribution de $F(\cdot)$, où $\eta \in (0, 1)$ est un constant et $\int A(x)F(dx) < \infty$, il résulte immédiatement de (2.2) que $\{X_t\}$ est β -mélange avec des coefficients décroissent exponentiellement. Nummelin et Tuominen (1982) [57] ont fourni des conditions suffisantes si (2.3) est maintenue; voir également Doukhan (1994) [42].

3.1 Notions de dépendance

7. Nous pouvons définir un processus "d'un côté infini" $\{X_t, t \geq 1\}$ soit, par exemple, α -mélange si $\alpha(n) \rightarrow 0$, où $\alpha(n)$ est défini comme en (2.1)

3.1.2 L'association

Définition 3.2 Une suite de variables aléatoires X_1, X_2, \dots, X_n est dite associée si pour toute paire de fonctions croissantes (par coordonnées) $f, g : \mathbb{R}^n \rightarrow \mathbb{R}$,

$$\text{Cov}(f(X_1, X_2, \dots, X_n), g(X_1, X_2, \dots, X_n)) \geq 0,$$

bien évidemment lorsque cette variance existe.

Cette définition a été introduite par Esary, Proschan et Walkup (1967) [58], dont l'objectif était de trouver des applications en fiabilité et en statistique. La notion d'association appelée aussi dépendance positive a été développée dans les travaux de Harris (1960) [59] et de Fortuin, Kastelyn et Ginibre (1971) [60], leurs travaux sont appliqués à la théorie de la mécanique statistique.

Remarque 1 Dans la définition de l'association, le paire des fonctions croissantes peut être remplacé par un paire de fonctions décroissantes, car $\text{Cov}(f(X), g(X)) = \text{Cov}(-f(X), -g(X))$.

Remarque 2 Une famille infinie de variables aléatoires est dite associée si toute sous famille finie est associée.

On présente dans cette section quelques propriétés basiques de l'association (Esary, Proschan et Walkup) [58].

(P1) Si deux ensembles de variables associées sont indépendants l'un de l'autre, leur union est un ensemble associé.

(P2) Tout singleton formé d'une variable aléatoire réelle est associé.

(P3) Si $X = (X_1, \dots, X_n)$ est associé et si f_1, \dots, f_k sont des fonctions toutes croissantes de \mathbb{R}_n dans \mathbb{R} (ou toutes décroissantes), alors le vecteur $Y = (f_1(X), \dots, f_k(X))$ est associé.

(P4) Si $X^{(k)} = (X_1^{(k)}, \dots, X_n^{(k)})$ est associé pour tout k et si $X^{(k)}$ converge en loi vers $X = (X_1, \dots, X_n)$ lorsque k tend vers $+\infty$, alors X est associé.

(P5) Tout vecteur aléatoire à composantes indépendantes est associé.

3.2 Inégalités des moments

Théorème 3.1 (voir Yu.A.Davydov [61]) Soient X et Y deux variables aléatoires centrées et de variances finies.

$\forall p, q, r$ avec $r > 1$ et $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} = 1$ on a

$$|Cov(X, Y)| \leq 8\alpha(X, Y)^{\frac{1}{p}} (\mathbb{E}(|X|^q))^{\frac{1}{q}} (\mathbb{E}(|Y|^r))^{\frac{1}{r}}$$

Théorème 3.2 (voir Q.M.Shao et H.Yu [41]) Soient $2 < p < r < \infty$, $2 < v < r$ et $(X_n)_{n \geq 1}$ une suite α -mélangeante de variables aléatoires telles que pour tout n , $\mathbb{E}(X_n) = 0$ et $(\mathbb{E}(|X_n|^r))^{\frac{1}{r}} < \infty$, On suppose que

$$\alpha_n \leq Cn^{-\theta} \quad \text{pour } C > 0 \quad \text{et } \theta > 0.$$

Alors $\forall \epsilon > 0$, il existe $K = K(\epsilon, r, p, v, \theta, C) < \infty$ tel que

$$\mathbb{E}(|S_n|^p) \leq K[(nC_n)^{\frac{p}{2}} \max_{i \leq n} (\mathbb{E}(|X_i|^v))^{\frac{p}{v}} + n^{(p - \frac{(r-p)\theta}{r}) \vee (1+\epsilon)} \max_{i \leq n} (\mathbb{E}(|X_i|^r))^{\frac{p}{r}}],$$

où

$$S_n = \sum_{i=1}^n X_i, \quad C_n = \left(\sum_{i=1}^n (i+1)^{\frac{2}{v-2}} \alpha_i \right)^{\frac{v-2}{v}} \quad \text{et } a \vee b = \max(a, b),$$

En particulier, pour tout $\epsilon > 0$,

1. Si $\theta > \frac{v}{v-2}$ et $\theta \geq \frac{(p-1)r}{r-p}$, on a

$$\mathbb{E}(|S_n|^p) \leq K[n^{\frac{p}{2}} \max_{i \leq n} (\mathbb{E}(|X_i|^v))^{\frac{p}{v}} + n^{(1+\epsilon)} \max_{i \leq n} (\mathbb{E}(|X_i|^r))^{\frac{p}{r}}],$$

2. Si $\theta \geq \frac{pr}{2(r-p)}$, on a

$$\mathbb{E}(|S_n|^p) \leq Kn^{\frac{p}{2}} \max_{i \leq n} (\mathbb{E}(|X_i|^r))^{\frac{p}{r}}.$$

Théorème 3.3 (voir Q.M.Shao [40]) Soit $(X_n)_{n \geq 1}$ une suite ρ -mélangeante de variables aléatoires, telles que $(\mathbb{E}(|X|^q))^{\frac{1}{q}} < \infty$ pour certain $q \geq 2$.

Alors il existe $K = K(q, \rho(\cdot)) < \infty$ tel que

$$\begin{aligned} \mathbb{E}(\max_{1 \leq i \leq n} |S_k(i)|^q) &\leq K \left[n^{\frac{q}{2}} \exp \left(K \sum_{i=0}^{\lfloor \log 2n \rfloor} \rho(2^i) \right) \max_{k \leq i \leq k+n} (\mathbb{E}(|X_i|^2))^{\frac{q}{2}} \right. \\ &\quad \left. + n \exp \left(K \sum_{i=0}^{\lfloor \log 2n \rfloor} \rho^{\frac{2}{q}}(2^i) \right) \max_{k \leq i \leq k+n} (\mathbb{E}(|X_i|^q)) \right]. \end{aligned}$$

où $S_k(i) = \sum_{j=k+1}^{k+1} X_j$.

3.2 Inégalités des moments

Théorème 3.4 (Théorème de la limite centrale pour des suites α -mélangeantes)

Soit X_t une suite stationnaire avec $\mathbb{E}(X_t) = 0$, On suppose que X est α -mélangeante et que pour un $\delta > 0$

$$\mathbb{E}(|X_t|^{2+\delta}) < \infty,$$

et

$$\sum_{n=0}^{\infty} \alpha^{2+\delta}(n) < \infty.$$

Alors

$$\lim_{n \rightarrow \infty} \frac{\sigma_n^2}{n} = \mathbb{E}(|X_1|^2) + 2 \sum_{k=1}^{\infty} \mathbb{E}(X_1 X_k) \equiv \sigma^2,$$

où

$$S_n = \sum_{k=1}^n X_k$$

et

$$\sigma_n^2 = \text{Var}(S_n).$$

Pour la preuve voir P.Doukhan, P. Massart et E. Rio (1994) [43] et Durrett (1991) [62].

Lemme 3.1 (voir R. Yokoyama [19]) Soit Y_i une suite α -mélange d'une variable aléatoire réelle sur l'espace probabilisé $(\Omega, \beta, \mathbb{P})$, avec $\mathbb{E}(Y_1) = 0$ et $\mathbb{E}(|Y_i|^{m+\delta}) < \infty$ pour tout entier $m > 2$ et $\delta > 0$. Si

$$\sum_{i=1}^{\infty} (i+1)^{\frac{m}{2}-1} [\alpha(i)]^{\frac{\delta}{m+\delta}} < \infty,$$

alors il existe un constant C tel que

$$\mathbb{E}(|\sum_{i=1}^n Y_i|^m) \leq C n^{\frac{m}{2}}, \quad n \geq 1. \quad (3.8)$$

Lemme 3.2 (K. Yoshihara [12]) Soit ε_i une suite α -mélange de coefficient $\alpha(n)$ supposons que pour un tel $\delta > 0$ et pour tout $m \geq 2$

1. $\mathbb{E}(\varepsilon_i) = 0$ et $\mathbb{E}(|\varepsilon_i|^{m+\delta}) \leq M < \infty$, ($i=1,2,\dots$), et
2. $\sum_{i=1}^{\infty} (i+1)^{\frac{m}{2}-1} (\alpha(i))^{\frac{\delta}{m+\delta}} < \infty$.

Alors, pour toute suite $\{a_k\}$ et pour tout entier n , on a

$$\mathbb{E} \left(\sum_{i=b+1}^{b+n} a_i \varepsilon_i \right)^m \leq C'_m A_{b,n}^m \quad (\forall b \geq 0, n \geq 1), \quad (3.9)$$

avec C'_m est un constant ne dépend que de m et $A_{b,n}^m = \sum_{i=b+1}^{b+n} a_i^m$, ($A_{0,n}^2 = A_n^2$)

3.2 Inégalités des moments

Lemme 3.3 (voir [50]) Soit (ε_i) une suite de variables aléatoires α -mélange, s'il existe les nombres positifs c_1^* , c_2^* et c_3^* tel que

$$c_1^*h \leq \mathbb{E} \left[K \left(\frac{t - \varepsilon_i}{h} \right) \right], \quad \mathbb{E} \left[K \left(\frac{t - \varepsilon_i}{h} \right) \right]^2 \leq c_2^*h \quad \forall i \geq 1,$$

et

$$\mathbb{E} \left[K \left(\frac{t - \varepsilon_i}{h} \right) \right] K \left(\frac{t - \varepsilon_j}{h} \right) \leq c_3^*h^2 \quad \forall i \neq j.$$

Alors on a

$$\text{Var} \left[\sum_{i=1}^n K \left(\frac{t - \varepsilon_i}{h} \right) \right] = nh\sigma^2 + o(nh),$$

où $\sigma^2 = \text{var}[\varepsilon_i]$

Ce chapitre fait l'objet d'une publication [70] paru dans le journal "International Journal of Mathematics and Statistics".

[http ://www.ceser.in/ceserp/index.php/ijms/issue/view/267](http://www.ceser.in/ceserp/index.php/ijms/issue/view/267)

A. Belguerna and S. Benaissa. The asymptotic convergence of the parameter estimator in the first order autoregressive model $AR(1)$ under strong mixing conditions. International Journal of Mathematics and Statistics, Year 2014, Volume 15, Issue Number 3, 1-8.

Chapitre 4

La convergence asymptotique du paramètre d'un modèle autoregressif d'ordre 1 dans le cas des erreurs α -mélangeant

Dans la première et la deuxième section du chapitre quatre nous étudions la convergence asymptotique de l'estimateur du paramètre d'un processus autorégressifs d'ordre 1. La troisième section sera consacrée à trouver un intervalle de confiance pour l'estimateur de ce paramètre pour un seuil de confiance donné.

Soit le processus autoregressif d'ordre 1 stationnaire suivant :

$$AR(1) : X_i = \varphi X_{i-1} + \varepsilon_i, \quad (4.1)$$

où φ est le paramètre du processus (4.1), et ε_i est l'erreur. Le but principale de ce travail est d'étudier la convergence presque complète de :

$$\sqrt{n}(\hat{\varphi}_n - \varphi),$$

dans le cas des erreurs α -mélangeant,

i.e :

$$\sum_{n=1}^{\infty} \mathbb{P}(|\sqrt{n}(\hat{\varphi}_n - \varphi)| > K) < +\infty$$

où encore

$$\sum_{n=1}^{\infty} \mathbb{P} \left(\left| \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^n X_{i-1} \varepsilon_i}{\frac{1}{n} \sum_{i=1}^n X_{i-1}^2} \right| > K \right) < +\infty$$

4.1 Convergence presque complète

Définition 4.1 (Voir **R.C. Bradley** [20] et [21]) Supposons que $(\varepsilon_i, i \in \mathbb{N}^*)$ est une variable aléatoire réelle sur l'espace probabilisé $(\Omega, \beta, \mathbb{P})$, et soit \mathbb{F}_l^n la σ -algèbre engendrée par $(\varepsilon_i, l \leq i \leq n)$ et soit :

$$\alpha(n) = \sup \left\{ |\mathbb{P}(AB) - \mathbb{P}(A)\mathbb{P}(B)| : A \in \mathbb{F}_{-\infty}^l, B \in \mathbb{F}_{l+n}^\infty \right\}.$$

La suite (ε_i) est dite α -mélange si :

$$\alpha(n) \rightarrow 0 \quad \text{lorsque} \quad n \rightarrow \infty.$$

Remarque 3 (Voir **P. Billingsley** [24]) En général, si $h: \mathbb{R} \rightarrow \mathbb{R}$ est une fonction mesurable la suite $X := (X_i)_i$ est α -mélangeant, alors $Y := (h(X_i))_i$ l'est aussi. En plus, si l et n sont des entiers fixes, alors

$$\sigma(h(X_j), j \leq l) \subset \sigma(X_j, j \leq l) \quad \text{et} \quad \sigma(h(X_j), j \geq l+n) \subset \sigma(X_j, j \geq l+n),$$

qui donne $\alpha_Y(n) \leq \alpha_X(n)$ pour tout entier n .

Définition 4.2 (Hsu et Robbins [8]) Une suite $\{X_n, n \in \mathbb{N}^*\}$ de variables aléatoires converge presque complètement vers 0, si $\forall \epsilon > 0$,

$$\sum_{n=1}^{\infty} \mathbb{P}(|X_n| > \epsilon) < \infty$$

Lemme 4.1 (Lemme de Borel-Cantelli voir [7]) Soit $(\Omega, \beta, \mathbb{P})$ un espace probabilisé, $\{A_n\}$ une suite d'évènements.

1. Si $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$ alors

$$\mathbb{P}(\limsup_n A_n) = 0$$

2. Si les évènements A_1, A_2, \dots sont indépendants et $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$, alors

$$\mathbb{P}(\limsup_n A_n) = 1$$

Corollaire 4.1 (voir [7]) Soit $\{A_n\}$ une suite d'évènements indépendants. Alors

$$\mathbb{P}(\limsup_n A_n) = 0 \Leftrightarrow \sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$$

4.2 Principaux résultats

revenons maintenant sur le processus autorégressif d'ordre un $AR(1)$ défini dans (4.1), où φ est un paramètre déterministe du processus (4.1), et ε_i est une suite α -mélange d'une variable aléatoire réelle sur l'espace probabilisé $(\Omega, \beta, \mathbb{P})$ identiquement distribuées de moyenne nulle, $\mathbb{E}(|\varepsilon_i|^m) < \infty$ pour $(m \geq 1)$ et $\sigma_i^2 = \mathbb{E}(\varepsilon_i^2) < \infty$, pour la simplicité on va supposer que notre processus (4.1) est initialisé par $X_0 = 0$, le paramètre du model φ qui est inconnu peut être estimé par exemple par le principe des moindres carrés comme suit :

$$\hat{\varphi}_n = \frac{\sum_{i=1}^n X_i X_{i-1}}{\sum_{i=1}^n X_{i-1}^2}, \quad n \geq 2 \quad (4.2)$$

les observations X_1, \dots, X_n étant données et $\hat{\varphi}_n$ est l'estimateur de φ ([2]).

4.2 Principaux résultats

Théorème 4.1 Pour tout $0 < \epsilon < \sqrt{T_1}$ et $m > 2$. Si

$\sum_{i=1}^{\infty} (i+1)^{\frac{m}{2}-1} [\alpha(i)]^{\frac{\delta}{m+\delta}} < \infty, \forall \delta > 0$ alors il existe un constant C'_m tel que

$$\mathbb{P}(\sqrt{n}|\hat{\varphi}_n - \varphi| > K) \leq \frac{2^{m-1}}{(K\epsilon)^{2m}} \left(\frac{B_n^m}{n^m} + C'_m n^{-\frac{m}{2}} \right) + \exp\left(-\frac{n(\epsilon^2 - T_1)^2}{2T_2}\right) \quad (4.3)$$

où $T_1 = \mathbb{E}(X_i^2) < \infty, T_2 = \mathbb{E}(X_i^4) < \infty, i \geq 1, B_n = \sum_{i=1}^n \sigma_i^2 = O(n^\alpha), 0 < \alpha < 2$ et C'_m est un constant ne dépend que de m .

Pour démontrer ce théorème, en introduisant les lemmes suivants :

Lemme 4.2 (voir [9] et [10]) $\forall A, B, m > 0$ on a

$$(A + B)^m \leq \max(1, 2^{m-1})(A^m + B^m)$$

Lemme 4.3 (voir [19])[Yokoyama (1980)] Soit Y_i une suite α -mélange d'une variable aléatoire réelle sur l'espace probabilisé $(\Omega, \beta, \mathbb{P})$, avec $\mathbb{E}(Y_1) = 0$ et $\mathbb{E}(Y_i^{m+\delta}) < \infty$ pour tout entier $m > 2$ et $\delta > 0$. Si $\sum_{i=1}^{\infty} (i+1)^{\frac{m}{2}-1} [\alpha(i)]^{\frac{\delta}{m+\delta}} < \infty$, alors il existe un constant C tel que

$$\mathbb{E}\left(\left|\sum_{i=1}^n Y_i\right|^m\right) \leq Cn^{\frac{m}{2}}, \quad n \geq 1. \quad (4.4)$$

Preuve du théorème 4.1

Par un simple calcul, on a

$$\varphi = \frac{\sum_{i=1}^n X_i X_{i-1} - \sum_{i=1}^n X_{i-1} \varepsilon_i}{\sum_{i=1}^n X_{i-1}^2}$$

et

$$\widehat{\varphi}_n - \varphi = \frac{\sum_{i=1}^n X_{i-1} \varepsilon_i}{\sum_{i=1}^n X_{i-1}^2} \quad (4.5)$$

et par suite on a

$$\mathbb{P}(\sqrt{n}|\widehat{\varphi}_n - \varphi| > K) = \mathbb{P}\left(\left|\frac{\frac{1}{\sqrt{n}} \sum_{i=1}^n X_{i-1} \varepsilon_i}{\frac{1}{n} \sum_{i=1}^n X_{i-1}^2}\right| > K\right)$$

pour facilité l'écriture, posons

$$I = \left| \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^n X_{i-1} \varepsilon_i}{\frac{1}{n} \sum_{i=1}^n X_{i-1}^2} \right| \quad (4.6)$$

Successivement par l'inégalité de Hölder, on obtient

$$\begin{aligned} I &\leq \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^n |X_{i-1}| |\varepsilon_i|}{\frac{1}{n} \sum_{i=1}^n X_{i-1}^2} \\ &\leq \frac{\frac{1}{\sqrt{n}} (\sum_{i=1}^n X_{i-1}^2)^{\frac{1}{2}} (\sum_{i=1}^n \varepsilon_i^2)^{\frac{1}{2}}}{\frac{1}{n} \sum_{i=1}^n X_{i-1}^2} \\ &\leq \frac{\frac{1}{\sqrt{n}} (\sum_{i=1}^n \varepsilon_i^2)^{\frac{1}{2}}}{\frac{1}{n} (\sum_{i=1}^n X_{i-1}^2)^{\frac{1}{2}}}. \end{aligned}$$

Et comme $K < I$ donc

$$K < \frac{\frac{1}{\sqrt{n}} (\sum_{i=1}^n \varepsilon_i^2)^{\frac{1}{2}}}{\frac{1}{n} (\sum_{i=1}^n X_{i-1}^2)^{\frac{1}{2}}}$$

on arrive à la majoration suivante

$$\mathbb{P}(I > K) \leq \mathbb{P}\left(\frac{\frac{1}{\sqrt{n}} (\sum_{i=1}^n \varepsilon_i^2)^{\frac{1}{2}}}{\frac{1}{n} (\sum_{i=1}^n X_{i-1}^2)^{\frac{1}{2}}} > K\right). \quad (4.7)$$

Puisque \mathbb{P} est une probabilité, pour tout $\epsilon > 0$ on a

$$\begin{aligned} \mathbb{P}(I > K) &\leq \mathbb{P}\left(\frac{1}{\sqrt{n}} (\sum_{i=1}^n \varepsilon_i^2)^{\frac{1}{2}} > \frac{1}{n} (\sum_{i=1}^n X_{i-1}^2)^{\frac{1}{2}} K\right) \\ &\leq \mathbb{P}\left(\frac{1}{\sqrt{n}} (\sum_{i=1}^n \varepsilon_i^2)^{\frac{1}{2}} > K\epsilon\right) + \mathbb{P}\left(\frac{1}{n} (\sum_{i=1}^n X_{i-1}^2)^{\frac{1}{2}} \leq \epsilon\right) \\ &\leq \pi_1 + \pi_2 \end{aligned}$$

posons

$$\pi_1 = \mathbb{P}\left(\frac{1}{\sqrt{n}} (\sum_{i=1}^n \varepsilon_i^2)^{\frac{1}{2}} > K\epsilon\right)$$

4.2 Principaux résultats

et

$$\pi_2 = \mathbb{P} \left(\frac{1}{n} \left(\sum_{i=1}^n X_{i-1}^2 \right)^{\frac{1}{2}} \leq \epsilon \right)$$

maintenant il suffit de majorer π_1 et π_2 , d'une part le résultat suivant se tiennent également, puisque ε_i et ε_i^2 sont des suites des variables aléatoires α -mélange. Soit $Y_i = \varepsilon_i^2 - \mathbb{E}(\varepsilon_i^2)$, $\mathbb{E}(Y_i) = 0$, sous l'hypothèse $\sum_{i=1}^{\infty} (i+1)^{\frac{m}{2}-1} [\alpha(i)]^{\frac{1}{m+1}} < \infty$, appliquant le lemme (4.3) précédent, pour tout $m > 2$, $\mathbb{E}(|Y_i|^{m+1}) < \infty$ et $\sigma_i^2 = \mathbb{E}(\varepsilon_i^2) < \infty$, alors

$$\mathbb{E} \left(\left| \sum_{i=1}^n Y_i \right|^m \right) \leq C'_m n^{\frac{m}{2}}$$

d'où

$$\mathbb{E} \left(\left| \sum_{i=1}^n \frac{1}{n} Y_i \right|^m \right) \leq C'_m n^{-\frac{m}{2}} \quad (4.8)$$

On obtient donc

$$\mathbb{E} \left(\left| \frac{1}{n} \sum_{i=1}^n (\varepsilon_i^2 - \sigma_i^2) \right|^m \right) \leq C'_m n^{-\frac{m}{2}} \quad (4.9)$$

pour tout $m > 2$,

$$\left| \frac{1}{n} \sum_{i=1}^n (\varepsilon_i^2 - \sigma_i^2) \right| = \left| \frac{\sum_{i=1}^n \varepsilon_i^2}{n} - \frac{\sum_{i=1}^n \sigma_i^2}{n} \right|$$

on a

$$||A| - |B|| \leq |A - B|$$

ou encore

$$|A|^m \leq (|B| + |A - B|)^m$$

et grâce au lemme (4.2) on a

$$|A|^m \leq \max\{1, 2^{m-1}\} (|B| + |A - B|)^m.$$

On pose $A = \frac{1}{n} \sum_{i=1}^n \varepsilon_i^2$ et $B = \frac{1}{n} \sum_{i=1}^n \sigma_i^2$.

Et par conséquence

$$\begin{aligned} \mathbb{E}(|A|^m) &\leq 2^{m-1} (\mathbb{E}(|B|^m) + \mathbb{E}(|A - B|^m)) \\ &\leq 2^{m-1} \left(\frac{B_n^m}{n^m} + C'_m n^{-\frac{m}{2}} \right). \end{aligned}$$

D'où finalement,

$$\mathbb{E} \left| \frac{1}{n} \sum_{i=1}^n \varepsilon_i^2 \right|^m \leq 2^{m-1} \left(\frac{B_n^m}{n^m} + C'_m n^{-\frac{m}{2}} \right). \quad (4.10)$$

4.2 Principaux résultats

L'inégalité de Markov pour $m > 2$ et la relation (4.10) entraînent

$$\begin{aligned}\pi_1 &\leq \frac{\mathbb{E}\left|\frac{1}{n}\sum_{i=1}^n \varepsilon_i^2\right|^m}{(K\varepsilon)^{2m}} \\ &\leq \frac{2^{m-1}}{(K\varepsilon)^{2m}} \left(\frac{B_n^m}{n^m} + C'_m n^{-\frac{m}{2}}\right).\end{aligned}$$

Donc pour $m > 2$ on a une première majoration

$$\pi_1 \leq \frac{2^{m-1}}{(K\varepsilon)^{2m}} \left(\frac{B_n^m}{n^m} + C'_m n^{-\frac{m}{2}}\right). \quad (4.11)$$

D'autre part, nous obtenons successivement

$$\begin{aligned}\pi_2 &= \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n X_{i-1}^2 \leq \varepsilon^2\right) \\ &= \mathbb{P}\left(n\varepsilon^2 - \sum_{i=1}^n X_{i-1}^2 \geq 0\right) \\ &= \mathbb{E}\left(\mathbb{1}_{\{n\varepsilon^2 - \sum_{i=1}^n X_{i-1}^2 \geq 0\}}\right) \\ &\leq \mathbb{E}\left(\exp\lambda\left(n\varepsilon^2 - \sum_{i=1}^n X_{i-1}^2\right)\right) \quad (\lambda > 0) \\ &\leq e^{\lambda n\varepsilon^2} \mathbb{E}\left(\exp -\lambda \sum_{i=1}^n X_{i-1}^2\right) \\ &\leq e^{\lambda n\varepsilon^2} \prod_{i=1}^n \mathbb{E}\left(\exp -\lambda X_{i-1}^2\right).\end{aligned}$$

on trouve également

$$\pi_2 \leq e^{\lambda n\varepsilon^2} \prod_{i=1}^n \mathbb{E}\left(\exp -\lambda X_{i-1}^2\right).$$

on pose

$$X = \lambda X_{i-1}^2,$$

vérifions que pour $x \geq 0$ on a

$$e^{-x} \leq 1 - x + \frac{1}{2}x^2. \quad (4.12)$$

En effet

soit $\psi(x) = e^{-x}$ et $\phi(x) = 1 - x + \frac{1}{2}x^2$ ce qui implique que $\psi'(x) = -e^{-x}$

de plus

$$e^x \geq 1 + x \quad \forall x, \quad (4.13)$$

on déduit que $\psi'(x) = -e^{-x} \leq -1 + x = \phi'(x)$. Aussi $\psi(0) = 1 = \phi(0)$ il en découle que $\psi(x) \leq \phi(x)$ pour tout $x \geq 0$,

ce qui montre 4.12.

4.12 et 4.13 entraînent que pour tout $\lambda > 0$

$$\begin{aligned}e^{\lambda n\varepsilon^2} \prod_{i=1}^n \mathbb{E}\left(\exp\left(-\frac{1}{n^2}\lambda X_{i-1}^2\right)\right) &\leq e^{\lambda n\varepsilon^2} \left(1 - \lambda T_1 + \frac{\lambda^2}{2}T_2\right)^n \\ &\leq e^{\lambda n\varepsilon^2} \left(\exp\left(\lambda T_1 + \frac{\lambda^2}{2}T_2\right)\right)^n \\ &\leq e^{\lambda n\varepsilon^2} \exp\left(-n\lambda T_1 + \frac{\lambda^2}{2}nT_2\right)\end{aligned}$$

4.3 Intervalle de confiance

où $T_1 = \mathbb{E}(X_i^2) < \infty, T_2 = \mathbb{E}(X_i^4) < \infty$.

Et par conséquence

$$\pi_2 = \mathbb{P} \left(\sum_{i=1}^n X_{i-1}^2 \leq n\epsilon^2 \right) \leq \exp \left[\lambda (n\epsilon^2 - nT_1) + \frac{n\lambda^2 T_2}{2} \right]. \quad (4.14)$$

Considérons la fonction suivante $h(\lambda) = \lambda(n\epsilon^2 - nT_1) + \frac{n\lambda^2 T_2}{2}$ avec $\lambda > 0$, l'équation $h'(\lambda) = 0$ a pour unique solution $\lambda = -\frac{(\epsilon^2 - T_1)}{T_2}$, aussi que $h''(\lambda) > 0$ ce qui prouve que $\lambda = -\frac{(\epsilon^2 - T_1)}{T_2}$ est un minimum pour la fonction h , et d'après 4.14 il découle que

$$\pi_2 = \mathbb{P} \left(\sum_{i=1}^n X_{i-1}^2 \leq n\epsilon^2 \right) \leq \exp \left[-\frac{n(\epsilon^2 - T_1)^2}{2T_2} \right]. \quad (4.15)$$

4.7, 4.11 et 4.15 entraînent que pour tout $K > 0, T_1 < \infty, T_2 < \infty, 0 < \epsilon < \sqrt{T_1}, m > 2$ et $0 < \alpha < 2$

$$\mathbb{P}(\sqrt{n}|\hat{\varphi}_n - \varphi| > K) \leq \frac{2^{m-1}}{(K\epsilon)^{2m}} \left(\frac{B_n^m}{n^m} + C'_m n^{-\frac{m}{2}} \right) + \exp \left(-\frac{n(\epsilon^2 - T_1)^2}{2T_2} \right).$$

Ceci achève la preuve du théorème.

Corollaire 4.2 La suite $(\hat{\varphi}_n)_{n \in \mathbb{N}}$ définie dans le théorème 4.1 converge presque complètement vers le paramètre φ du processus autorégressif $AR(1)$.

Preuve du corollaire 4.2 La convergence presque complète suit l'inégalité 4.1.

En effet, on sait que la série $\{V_n, n \in \mathbb{N}\}$ de terme général réel et positif défini par

$$V_n = \frac{2^{m-1}}{(K\epsilon)^{2m}} \left(\frac{B_n^m}{n^m} + C'_m n^{-\frac{m}{2}} \right) + \exp \left(-\frac{n(\epsilon^2 - T_1)^2}{2T_2} \right), \quad (4.16)$$

par suite

$$\sum_{n=1}^{+\infty} \mathbb{P}(\sqrt{n}|\hat{\varphi}_n - \varphi| > K) \leq \sum_{n=1}^{+\infty} \frac{2^{m-1}}{(K\epsilon)^{2m}} \left(\frac{B_n^m}{n^m} + C'_m n^{-\frac{m}{2}} \right) + \exp \left(-\frac{n(\epsilon^2 - T_1)^2}{2T_2} \right) < \infty.$$

Ce qui montre le corollaire.

4.3 Intervalle de confiance

L'inégalité 4.1 permet la construction d'un intervalle de confiance pour le paramètre du processus autorégressif d'ordre un $AR(1)$. Pour K suffisamment grand, tel que $K = \epsilon\sqrt{n}$, on

4.3 Intervalle de confiance

a

$$\begin{aligned}
& \lim_{n \rightarrow +\infty} \frac{2^{m-1}}{(K\epsilon)^{2m}} \left(\frac{B_n^m}{n^m} + C'_m n^{-\frac{m}{2}} \right) + \exp \left(-\frac{n(\epsilon^2 - T_1)^2}{2T_2} \right) = \\
& \lim_{n \rightarrow +\infty} \frac{2^{m-1}}{n^m \epsilon^{4m}} \left(\frac{(O(n^\alpha))^m}{n^m} + C'_m n^{-\frac{m}{2}} \right) + \exp \left(-\frac{n(\epsilon^2 - T_1)^2}{2T_2} \right) = \\
& \lim_{n \rightarrow +\infty} \frac{2^{m-1}}{\epsilon^{4m}} \left(O \left(\frac{1}{n^{m(2-\alpha)}} \right) + C'_m n^{-\frac{3m}{2}} \right) + \exp \left(-\frac{n(\epsilon^2 - T_1)^2}{2T_2} \right) = 0.
\end{aligned}$$

Ce que signifie, pour η un niveau donné, on trouve un nombre naturel n_η tel que

$$\forall n \geq n_\eta, \text{ on a, } \frac{2^{m-1}}{\epsilon^{4m}} \left(O \left(\frac{1}{n^{m(2-\alpha)}} \right) + C'_m n^{-\frac{3m}{2}} \right) + \exp \left(-\frac{n(\epsilon^2 - T_1)^2}{2T_2} \right) \leq \eta. \quad (4.17)$$

Par conséquence,

$$\mathbb{P}(\sqrt{n}|\hat{\varphi}_{n_\eta} - \varphi| \leq \epsilon) \geq 1 - \eta. \quad (4.18)$$

Ce qui signifie que le paramètre du processus autorégressif d'ordre un appartient à l'intervalle de centre $\hat{\varphi}_{n_\eta}$ et de rayon ϵ avec une probabilité plus grande ou égale à $1 - \eta$.

Chapitre 5

Distributions asymptotiques des estimateurs de la densité d'erreur dans les modèles autorégressifs de premier ordre

Dans la régression paramétrique et modèles autorégressifs, plusieurs auteurs ont étudiés les propriétés des estimateurs pour les distributions des erreurs. La convergence faible des processus empiriques basées sur les résidus dans les modèles de régression paramétriques est abordée dans Koul (1970, 1977, 1992, 1996), Loynes (1980), Portony (1986) et Mammen (1996), tandis que Boldin (1982), Koul (1991) et koul et Ossiander (1994) pour discuter sur les paramètres des modèles autorégressifs. La convergence uniforme de la densité d'erreur dans ces modèles est discutée dans koul (1992). Lee et Na (2002) s'étendent le résultat asymptotique (sur la base du L_2 -norme) dans Bickel et Rosenblatt (1973) pour les estimateurs de la densité d'erreurs dans les modèles de premier ordre tout en autorégressifs, tandis que Horv'ath et Zitikis (2003) considèrent la distribution asymptotique de la L_p -norme des estimateurs. Dans le présent chapitre, nous obtenons d'abord la dirtribution asymptotique des estimateurs du noyau de densité des erreurs à un point fixe. Puis nous montrons la distribution asymptotique du maximum d'un écart convenablement normalisé de l'estimateur de la densité de l'attente de la densité d'erreur du noyau (basé sur l'erreur vraie), qui est la même que celle dans le cas de l'échantillon en créer un établi par Bickel et Rosenblatt (1973). Le document est organisé comme suit. Les modèles, les estimateurs et les hypothèses sont introduits dans la section suivante. l'article 3 énonce et démontre la distribution asymptotique de l'estimateur à un point fixe tandis que la section 4 montre la distribution asymptotique

de la mesure globale. Dans les sections suivantes, toutes les limites sont considérées comme la taille de l'échantillon n tend vers ∞ .

5.1 Modèles, estimateurs, et hypothèses

Dans cette section, nous introduisons le modèle et les estimateurs à noyau de densité d'erreur. Supposons que la suite $\{X_i\}$ satisfait le modèle autorégressif de premier ordre

$$X_i = \theta X_{i-1} + \zeta_i \quad (5.1)$$

où les ζ_i sont des variables aléatoires indépendantes et identiquement distribuées (iid) de moyenne 0, variance σ^2 , quatrième moment fini et densité f inconnue. Nous supposons aussi $|\theta| < 1$ et la séquence $\{X_i\}$ est stationnaire. Puis, nous avons la représentation

$$X_i = \sum_{j=0}^{\infty} \theta^j \zeta_{i-j}, \quad (5.2)$$

Par conséquent, en utilisant Théorème 7.1.2 dans Brockwell et Davis (1991), nous avons la distribution asymptotique suivante :

$$\frac{1}{\sqrt{n}} \sum_{i=0}^n X_i \rightarrow \mathcal{N} \left(0, \left(\frac{\sigma\theta}{1-\theta} \right)^2 \right) \quad (5.3)$$

Supposons que nous observons $X_0, X_1, X_2, \dots, X_n$. Soit un estimateur de θ avec la propriété suivante :

$$\sqrt{n}(\hat{\theta} - \theta) = O_{\mathbb{P}}(1). \quad (5.4)$$

la propriété (5.4) est naturelle puisque l'estimateur des moindres carrés la satisfait. Soit

$$\hat{\zeta}_i = X_i - \hat{\theta} X_{i-1} \quad (5.5)$$

désigne les résidus. Sur la base de ces résidus, on construit un estimateur de la densité d'erreur $f(t)$ comme suit :

$$\hat{f}_n(t) = \frac{1}{n} \sum_{i=1}^n K_{h_n}(t - \hat{\zeta}_i), \quad t \in \mathbb{R},$$

où $K_h(t) = \frac{1}{h} K\left(\frac{t}{h}\right)$, et h_n étant des nombres positifs (appelé généralement bande passante) tendant vers zéro, et K est la fonction de densité du noyau. Nous allons montrer que $\hat{f}_n(t)$ est asymptotiquement normale à un point fixe. Nous avons aussi besoin de définir la densité

5.2 Distribution asymptotique de \widehat{f}_n en un point fixe

d'erreur du noyau sur la base des véritables erreurs (que nous ne pouvons pas observer)

$\zeta_1, \zeta_2, \dots, \zeta_n$:

$$f_n(t) = \frac{1}{n} \sum_{i=1}^n K_{h_n}(t - \zeta_i), \quad t \in \mathbb{R}.$$

Nous allons considérer la distribution asymptotique de la distance maximale d'un écart convenablement normalisé à partir de $\widehat{f}_n(t)$ à $E(f_n(t)) = \int K(x)f(t - h_n x)dx$. Nous concluons cette section avec quelques hypothèses de base sur K et h_n qui sont utilisés par la suite. i)

(K_1) K existe sur la droite réelle et K'' est bornée ;

(K_2) K est bornée ;

(K_3) $\int |x|K(x)dx < \infty$;

(K_4) $\int |x|K^2(x)dx < \infty$;

(K_5) $\int |xK'(x)|dx < \infty$ et $\int |x||K'(x)|^2dx < \infty$;

ii)(H) : $h_n \rightarrow 0$.

Note. La densité normale satisfait toutes les hypothèses sur K .

5.2 Distribution asymptotique de \widehat{f}_n en un point fixe

Dans cette section, nous considérons la distribution asymptotique de $\widehat{f}_n(t)$. Nous avons le résultat de normalité asymptotique suivante

Théorème 5.1 Supposons que à un instant fixe $t \in \mathbb{R}$, il existe une constante $0 < c < \infty$ telle que f satisfait

$$|f(t) - f(t - y)| \leq c|y|, \quad \text{pour tout } y \in \mathbb{R}. \quad (5.6)$$

Supposons que h_n satisfait

$$nh_n^5 \rightarrow \infty. \quad (5.7)$$

Puis, sous les hypothèses (K_1) – (K_5), (H) et (5.4), nous avons ce que, dans la distribution,

$$\frac{1}{\sqrt{\text{Var}(f_n(t))}} \left[\widehat{f}_n - \int K(x)f(t - h_n x)dx \right] \rightarrow \mathcal{N}(0, 1). \quad (5.8)$$

avant de prouver le théorème 5.6, nous allons énoncer et prouver un lemme auxiliaire.

Lemme 5.1 Supposons que (5.4), (5.6) et nous allons maintenir les hypothèses (K_1), (K_5) et (H). Alors, nous avons

$$|\widehat{f}_n(t) - f_n(t)| = O_{\mathbb{P}}\left(\frac{1}{nh_n^3}\right) \quad (5.9)$$

5.2 Distribution asymptotique de \widehat{f}_n en un point fixe

Preuve. D'après (5.1) et (5.5), nous avons

$$\widehat{\zeta}_i - \zeta_i = -(\widehat{\theta} - \theta)X_{i-1}. \quad (5.10)$$

En utilisant (K_1) et (5.10), nous pouvons obtenir

$$\begin{aligned} |\widehat{f}_n(t) - f_n(t)| &= \left| \frac{1}{nh_n} \sum_{i=1}^n \left[K\left(\frac{t - \widehat{\zeta}_i}{h_n}\right) - K\left(\frac{t - \zeta_i}{h_n}\right) \right] \right| \\ &= \left| \frac{1}{nh_n} \sum_{i=1}^n \left[\frac{(\widehat{\theta} - \theta)X_{i-1}}{h_n} K'\left(\frac{t - \zeta_i}{h_n}\right) + \frac{(\widehat{\theta} - \theta)^2 X_{i-1}^2}{2h_n^2} K''(\xi_{it}) \right] \right| \\ &\leq \frac{|\widehat{\theta} - \theta|}{nh_n^2} \left| \sum_{i=1}^n X_{i-1} K''\left(\frac{t - \zeta_i}{h_n}\right) \right| + c_1 \frac{|\widehat{\theta} - \theta|^2}{2nh_n^3} \sum_{i=1}^n X_{i-1}^2, \\ &= J_{n1}(t) + J_{n2}(t), \text{ say,} \end{aligned} \quad (5.11)$$

où c_1 ($0 < c_1 < \infty$) est une borne supérieure pour $|K''|$, (i.e., $|K''(x)| \leq c_1 \forall x \in \mathbb{R}$), et ξ_{it} est un nombre compris entre $(t - \widehat{\zeta}_i)/h_n$ et $(t - \zeta_i)/h_n$. nous allons maintenant montrer que

$$J_{n1}(t) = O_{\mathbb{P}}\left(\sqrt{(n^2 h_n^3)^{-1} + (n^2 h_n^2)^{-1}}\right). \quad (5.12)$$

Nous commençons la preuve avec la borne

$$\begin{aligned} \left| \sum_{i=1}^n X_{i-1} \right|^2 &\leq 2 \left| \sum_{i=1}^n X_{i-1} \left[K'\left(\frac{t - \zeta_i}{h_n}\right) - \mathbb{E}\left(K'\left(\frac{t - \zeta_i}{h_n}\right)\right) \right] \right|^2 \\ &\quad + 2 \left| \sum_{i=1}^n X_{i-1} \right|^2 \left(\mathbb{E}\left(K'\left(\frac{t - \zeta_i}{h_n}\right)\right) \right)^2. \end{aligned} \quad (5.13)$$

Puisque X_i et $(\zeta_{i+1}, \zeta_{i+2}, \dots)$ sont indépendants d'après (5.2), nous avons l'égalité

$$\begin{aligned} &\mathbb{E}\left(\left| \sum_{i=1}^n X_{i-1} \left[K'\left(\frac{t - \zeta_i}{h_n}\right) - \mathbb{E}\left(K'\left(\frac{t - \zeta_i}{h_n}\right)\right) \right] \right|^2 \right) \\ &= \sum_{i=1}^n \mathbb{E}(X_{i-1}^2) \text{Var}\left(K'\left(\frac{t - \zeta_i}{h_n}\right)\right). \end{aligned} \quad (5.14)$$

Nous allons maintenant estimer la variance sur le côté droit de (5.14). D'après (5.6) et $\int |x| |K'(x)|^2 dx < \infty$, nous avons

$$\left| \int K'^2(x) f(t - xh_n) dx - f(t) \int K'^2(x) dx \right| \leq ch_n \int K'^2(x) dx = O(h_n).$$

5.2 Distribution asymptotique de \widehat{f}_n en un point fixe

Donc,

$$\text{Var} \left(K' \left(\frac{t - \zeta_i}{h_n} \right) \right) \leq \mathbb{E} \left(\left(K' \left(\frac{t - \zeta_i}{h_n} \right) \right)^2 \right) = h_n \int K'^2(x) f(t - xh_n) dx = O(h_n).$$

Combinons cette borne avec (5.14) et en utilisant $\mathbb{E}(X_i^2) = \mathbb{E}(X_0^2) < \infty$, nous montrons que

$$\mathbb{E} \left(\left| \sum_{i=1}^n X_{i-1} \left[K' \left(\frac{t - \zeta_i}{h_n} \right) - \mathbb{E} \left(K' \left(\frac{t - \zeta_i}{h_n} \right) \right) \right] \right|^2 \right) = O(nh_n).$$

Donc, par l'inégalité de Markov, on obtient

$$\left| \sum_{i=1}^n X_{i-1} \left[K' \left(\frac{t - \zeta_i}{h_n} \right) - \mathbb{E} \left(K' \left(\frac{t - \zeta_i}{h_n} \right) \right) \right] \right|^2 = O_{\mathbb{P}}(nh_n). \quad (5.15)$$

Nous allons maintenant estimer le second terme du côté droit de (5.13). D'après (5.6) et $\int |x| |K'(x)| dx < \infty$, nous avons

$$\left| \int K'(x) f(t - xh_n) dx - f(t) \int K'(x) dx \right| \leq ch_n \int K'(x) dx = O(h_n).$$

Donc,

$$\mathbb{E} \left(K' \left(\frac{t - \zeta_i}{h_n} \right) \right) = h_n \int K'(x) f(t - xh_n) dx = O(h_n). \quad (5.16)$$

De (5.4), il s'ensuit que $\widehat{\theta} - \theta = O_{\mathbb{P}}(n^{-1/2})$. A partir de (5.3), nous avons $\left| \sum_{i=1}^n X_{i-1} \right|^2 = O_{\mathbb{P}}(n)$. Combinons ces points avec (5.13), (5.15) et (5.16), il est facile de voir que (5.12) est vérifié. D'après (5.12) et l'hypothèse (H), nous obtenons

$$J_{n1} = O_{\mathbb{P}}((nh_n^3)^{-1}). \quad (5.17)$$

Nous allons maintenant prouver que

$$J_{n2} = O_{\mathbb{P}}((nh_n^3)^{-1}). \quad (5.18)$$

Puisque $\mathbb{E}(\sum_{i=1}^n X_{i-1}^2) = n\mathbb{E}(X_0^2) = O(n)$, en utilisant l'inégalité de Markov, nous obtenons

$\sum_{i=1}^n X_{i-1}^2 = O_{\mathbb{P}}(n)$. Ainsi, en utilisant la propriété $\widehat{\theta} - \theta = O_{\mathbb{P}}(n^{-1/2})$, nous prouvons (5.18).

Combinons (5.11), (5.17) et (5.18), la déclaration (5.9) suit.

Proof. du Théorème 5.1. Puisque

$$\mathbb{E}(f_n(t)) = \int K(x) f(t - xh_n) dx,$$

5.2 Distribution asymptotique de \widehat{f}_n en un point fixe

Nous pouvons écrire

$$\widehat{f}_n(t) - \int K(x)f(t - xh_n) dx = |\widehat{f}_n(t) - f_n(t)| + |f_n(t) - \mathbb{E}(f_n(t))|. \quad (5.19)$$

D'après (5.6), nous obtenons

$$\left| \int K(x)f(t - xh_n) dx - f(t) \int K(x) dx \right| \leq ch_n \int |x|K(x) dx,$$

et aussi la même liée par K remplacée par K^2 . En outre, combinons ces limites avec les hypothèses (K_3) et (K_4) , nous avons

$$\int K(x)f(t - xh_n) dx = f(t) \int K(x) dx + O(h_n),$$

et aussi la même expression avec K remplacé par K^2 . Par conséquent, nous avons

$$\begin{aligned} \text{Var}(f_n(t)) &= \frac{1}{nh_n^2} \text{var} \left(K \left(\frac{t - \zeta_1}{h_n} \right) \right) \\ &= \frac{1}{nh_n} \left[\int K^2(x)f(t - xh_n) dx - h_n \left(\int K(x)f(t - xh_n) dx \right)^2 \right] \\ &= \frac{1}{nh_n} [f(t) \int K^2(x) dx + O(h_n)]. \end{aligned} \quad (5.20)$$

Nous avons l'égalité

$$\frac{f_n(t) - \mathbb{E}(f_n(t))}{\sqrt{\text{Var}(f_n(t))}} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{K \left(\frac{t - \zeta_i}{h_n} \right) - \mathbb{E} \left(K \left(\frac{t - \zeta_1}{h_n} \right) \right)}{\sqrt{\text{Var} \left(K \left(\frac{t - \zeta_1}{h_n} \right) \right)}}.$$

Pour prouver la normalité asymptotique, nous allons utiliser le théorème de Lindeberg-Feller de la limite centrale. Nous vérifierons l'état du Lindeberg-Feller comme suit. Fixer tout $\epsilon > 0$. En utilisant le fait que K est borné, nous avons

$$\mathbb{E} \left[\frac{\left(K \left(\frac{t - \zeta_i}{h_n} \right) - \mathbb{E} \left(K \left(\frac{t - \zeta_1}{h_n} \right) \right) \right)^2}{\text{Var} \left(K \left(\frac{t - \zeta_1}{h_n} \right) \right)} \mathbb{1} \left\{ \left| \frac{K \left(\frac{t - \zeta_i}{h_n} \right) - \mathbb{E} \left(K \left(\frac{t - \zeta_1}{h_n} \right) \right)}{\sqrt{\text{Var} \left(K \left(\frac{t - \zeta_1}{h_n} \right) \right)}} \right| > \epsilon \sqrt{n} \right\} \right]$$

5.2 Distribution asymptotique de \widehat{f}_n en un point fixe

$$\begin{aligned}
&= O \left(\frac{1}{\text{Var} \left(K \left(\frac{t - \zeta_1}{h_n} \right) \right)} \mathbb{E} \left[\mathbb{1} \left\{ \left| \frac{K \left(\frac{t - \zeta_i}{h_n} \right) - \mathbb{E} \left(K \left(\frac{t - \zeta_1}{h_n} \right) \right)}{\sqrt{\text{Var} \left(K \left(\frac{t - \zeta_1}{h_n} \right) \right)}} \right| > \epsilon \sqrt{n} \right\} \right] \right) \\
&= O \left(\frac{\mathbb{P} \left(\left| K \left(\frac{t - \zeta_1}{h_n} \right) - \mathbb{E} \left(K \left(\frac{t - \zeta_1}{h_n} \right) \right) \right| > \epsilon \sqrt{n \text{Var} \left(K \left(\frac{t - \zeta_1}{h_n} \right) \right)} \right)}{\text{Var} \left(K \left(\frac{t - \zeta_1}{h_n} \right) \right)} \right)
\end{aligned}$$

De la probabilité dans la dernière expression, par l'inégalité de Tchebychev, nous avons

$$\mathbb{P} \left(\left| K \left(\frac{t - \zeta_1}{h_n} \right) - \mathbb{E} \left(K \left(\frac{t - \zeta_1}{h_n} \right) \right) \right| > \epsilon \sqrt{n \text{Var} \left(K \left(\frac{t - \zeta_1}{h_n} \right) \right)} \right) \leq \frac{1}{\epsilon^2 n}$$

Ainsi, nous obtenons que

$$\begin{aligned}
&\mathbb{E} \left[\frac{\left(K \left(\frac{t - \zeta_i}{h_n} \right) - \mathbb{E} \left(K \left(\frac{t - \zeta_1}{h_n} \right) \right) \right)^2}{\text{Var} \left(K \left(\frac{t - \zeta_1}{h_n} \right) \right)} \mathbb{1} \left\{ \left| \frac{K \left(\frac{t - \zeta_i}{h_n} \right) - \mathbb{E} \left(K \left(\frac{t - \zeta_1}{h_n} \right) \right)}{\sqrt{\text{Var} \left(K \left(\frac{t - \zeta_1}{h_n} \right) \right)}} \right| > \epsilon \sqrt{n} \right\} \right] \\
&= O \left(\frac{1}{\epsilon^2 n \text{Var} \left(K \left(\frac{t - \zeta_1}{h_n} \right) \right)} \right) \\
&= O \left(\frac{1}{\epsilon^2 n [f(t) \int K^2(x) dx + O(h_n)]} \right) \rightarrow 0,
\end{aligned}$$

ou nous utilisons également (5.20), $f(t) > 0$, et $nh_n \rightarrow \infty$ (garanti par (5.7) et l'hypothèse (H)). D'après le théorème de Lindeberg-Feller de la limite centrale (par exemple, le théorème

5.3 Distribution asymptotique de la mesure globale

(4.12) en Kallenberg (1997)), nous obtenons que, dans la distribution,

$$\frac{1}{\sqrt{\text{Var}(f_n(t))}}(f_n(t) - \mathbb{E}(f_n(t))) \rightarrow \mathcal{N}(0, 1).$$

combinons ce résultat avec (5.19), nous voyons que, afin de montrer (5.8), il suffit de prouver que

$$\frac{|\hat{f}_n(t) - f_n(t)|}{\sqrt{\text{Var}(f_n(t))}} = o_{\mathbb{P}}(1). \quad (5.21)$$

D'après (5.20), l'hypothèse (H) et $f(t) > 0$, il s'ensuit que $1/\sqrt{\text{Var}(f_n(t))} = O_{\mathbb{P}}(\sqrt{nh_n})$. Par conséquent, afin de montrer (5.21), il suffit de prouver que

$$\sqrt{nh_n}|\hat{f}_n(t) - f_n(t)| = o_{\mathbb{P}}(1).$$

Cela résulte de (5.9), i.e.,

$$\sqrt{nh_n}|\hat{f}_n(t) - f_n(t)| = O_{\mathbb{P}}\left(\frac{1}{nh_n^5}\right)$$

et l'hypothèse $nh_n^5 \rightarrow \infty$ dans (5.7).

5.3 Distribution asymptotique de la mesure globale

Dans cette section, nous étendons le théorème (3.1) de Bickel et Rosenblatt (1973) pour f_n . Nous montrons que la distribution asymptotique de la mesure globale de l'écart de $\hat{f}_n(t)$ à partir de $\mathbb{E}(f_n(t))$ est la même que celle dans le cas où nous pouvons observer l'erreur ϵ . Soit

$$\widehat{M}_n = \sup\{\sqrt{nh_n/f(t)}|\hat{f}_n(t) - \mathbb{E}(f_n(t))| : 0 \leq t \leq 1\},$$

$$M_n = \sup\{\sqrt{nh_n/f(t)}|f_n(t) - \mathbb{E}(f_n(t))| : 0 \leq t \leq 1\}.$$

(Il n'y a pas de perte lors de l'examen $[0, 1]$ plutôt que de tout autre intervalle sur lequel la densité est bornée loin de 0 et ∞ .) Le principal résultat est le suivant

Théorème 5.2 *Sous les hypothèses de lemmes 5.2 et 5.4, nous avons*

$$\mathbb{P}\left((2\delta \log n)^{1/2} \left(\frac{\widehat{M}_n}{\lambda(K)^{1/2}} - d_n\right) < y\right) \rightarrow e^{-2y}, \quad \forall y \in \mathbb{R},$$

où $\lambda(K)$ et d_n sont donnés dans le lemme.

5.3 Distribution asymptotique de la mesure globale

Nous aurons besoin de trois lemmes pour prouver le résultat principal. Le premier lemme est sur la distribution asymptotique de M_n donnée par le théorème (3.1) de Bickel et Rosenblatt (1973). Nous le réaffirmons ici.

Lemme 5.2 Supposons les hypothèses suivantes : A1. Soit K (a) s'annule à l'extérieur de l'intervalle $[-A, A]$ et est absolument continue sur $[-A, A]$ avec dérivé K' où (b) est absolument continue sur $(-\infty, \infty)$ avec dérivé K' de telle sorte que $\int |K'(x)|k dx < \infty$, $k = 1, 2$; A2. f est continue, positive et bornée sur $[0, 1]$; A3. $f^{1/2}$ est absolument continue et sa dérivé $1/2f'/f^{1/2}$ est délimitée en valeur absolue. En outre,

$$h_n = n^{-\delta} \text{ for some } \delta \in (0, 1/4) \quad (5.22)$$

et

$$\int_{|x|>3} |x|^{3/2} |K'(x)| (\log \log |x|)^{1/2} [|W'(x)| + |W(x)|] dx < \infty.$$

puis,

$$\mathbb{P} \left((2\delta \log n)^{1/2} \left(\frac{M_n}{\lambda(K)^{1/2}} - d_n \right) < y \right) \rightarrow e^{-2y}, \quad \forall y \in \mathbb{R},$$

où

$$\lambda(k) = \int k^2(x) dx \quad (5.23)$$

et d_n est défini comme suit.

Si $C_1(K) > 0$ où $C_1(K) = \frac{K^2(A)+K^2(-A)}{2} / \lambda(K)$, puis

$$d_n = (2\delta \log n)^{\frac{1}{2}} + \frac{1}{(2\delta \log n)^{\frac{1}{2}}} \left[\log \frac{C_1(K)}{\sqrt{\pi}} - \frac{1}{2}(\log \delta + \log \log n) \right].$$

Si $C_1(K) = 0$, puis

$$d_n = (2\delta \log n)^{\frac{1}{2}} + \frac{1}{(2\delta \log n)^{\frac{1}{2}}} \left[\log \frac{C_2(K)}{2\pi} \right],$$

où

$$C_2(K) = \frac{1}{2} \left[\int [K'(x)]^2 dx \right] / \lambda(K).$$

le second lemme est l'inégalité de type exponentiel de la différence de martingale.

Lemme 5.3 Si (Z_i) est une différence de martingale et

$$|Z_i| \leq k \text{ a.s.}$$

puis, pour tout $\epsilon > 0$,

$$\mathbb{P} \left(\left| \sum_{i=1}^n Z_i \right| > \epsilon \right) \leq 2 \exp \left\{ -\frac{\epsilon^2}{2nk^2} \right\}.$$

5.3 Distribution asymptotique de la mesure globale

Pour la démonstration du lemme 5.3, voir Azuma (1967). Maintenant nous utilisons le lemme 5.3 pour prouver le troisième lemme, qui est sur la proximité de \hat{f}_n et f_n .

Lemme 5.4 *Supposons que la première fonction dérivée K' est bornée et intégrable. Soit la fonction de densité f une fonction Lipschitzienne continue d'ordre 1, c'est à dire, il existe une constante $0 < l < \infty$ telle que*

$$|f(x) - f(y)| \leq l|x - y|, \forall x \in \mathbb{R} \text{ et } y \in \mathbb{R}. \quad (5.24)$$

En outre, la fenêtre h_n satisfait

$$\frac{\log n}{nh_n^5} \rightarrow 0, \quad (5.25)$$

et il existe une constante $3/2 < \gamma < \infty$ telle que

$$\frac{\sqrt{\log n}}{nh_n^{3\gamma+3/2}} \rightarrow 0. \quad (5.26)$$

Puis, sous (5.4), (K_1) et (H) , nous avons

$$\sup_{t \in [0,1]} |\hat{f}_n(t) - f_n(t)| = o_{\mathbb{P}} \left(\frac{1}{\sqrt{nh_n \log n}} \right). \quad (5.27)$$

Preuve. Sous l'hypothèse (K_1) , nous voyons que (5.11) est encore valable. Par conséquent, nous avons

$$\sup_{t \in [0,1]} |\hat{f}_n(t) - f_n(t)| \leq \sup_{t \in [0,1]} J_{n1} + J_{n2}. \quad (5.28)$$

Nous estimons la limite des deux termes dans le côté droit de (5.28). nous commençons avec le premier terme.

$$\begin{aligned} \sup_{t \in [0,1]} J_{n1} &= \frac{|\hat{\theta} - \theta|}{nh_n^2} \sup_{t \in [0,1]} \left| \sum_{i=1}^n X_{i-1} K' \left(\frac{t - \zeta_i}{h_n} \right) \right| \\ &\leq \frac{|\hat{\theta} - \theta|}{nh_n^2} \sup_{t \in [0,1]} \left| \sum_{i=1}^n X_{i-1} \left[K' \left(\frac{t - \zeta_i}{h_n} \right) - \mathbb{E} \left(K' \left(\frac{t - \zeta_1}{h_n} \right) \right) \right] \right| \\ &+ \frac{|\hat{\theta} - \theta|}{nh_n^2} \sup_{t \in [0,1]} \left| \sum_{i=1}^n X_{i-1} \left| \sup_{t \in [0,1]} \mathbb{E} \left(K' \left(\frac{t - \zeta_1}{h_n} \right) \right) \right| \right| \\ &= I_{n1} + I_{n2}, \text{ say.} \end{aligned} \quad (5.29)$$

Nous allons montrer que

$$I_{n1} = o_{\mathbb{P}} \left(\frac{1}{\sqrt{nh_n \log n}} \right). \quad (5.30)$$

5.3 Distribution asymptotique de la mesure globale

nous notons

$$W_i(t) = X_{i-1} \mathbb{1}_{\{|X_{i-1}| \leq n^{1/3} h_n^\gamma\}} \left[K' \left(\frac{t - \zeta_i}{h_n} \right) - \mathbb{E} \left(K' \left(\frac{t - \zeta_1}{h_n} \right) \right) \right].$$

Soit \mathcal{F}_n la σ -algèbre engendrée par les variables aléatoires $\{\zeta_i, i \leq n\}$. Puisque X_{i-1} et ζ_i sont indépendants, le processus (pour t fixe)

$$\left(\sum_{i=1}^n W_i(t), \mathcal{F}_n, 1 \leq n \leq n \right) \text{ est une martingale.}$$

Puisque K' est bornée, il existe une constante $0 < c_2 < \infty$ telle que $|K'(x)| \leq c_2, \forall x \in \mathbb{R}$. Ainsi, nous avons

$$\left| K' \left(\frac{t - \zeta_i}{h_n} \right) - \mathbb{E} \left(K' \left(\frac{t - \zeta_1}{h_n} \right) \right) \right| \leq 2c_2. \quad (5.31)$$

Par conséquent, il s'ensuit que $|W_i(t)| \leq 2c_2 n^{1/3} h_n^\gamma$. Donc, en appliquant le lemme 5.3 avec $k = 2c_2 n^{1/3} h_n^\gamma$ et $\varepsilon = n h_n^\gamma / \sqrt{\log n}$, nous obtenons que, pour tout $t \in \mathbb{R}$,

$$\mathbb{P} \left(\left| \sum_{i=1}^n W_i(t) \right| > n \frac{h_n^\gamma}{\sqrt{\log n}} \right) \leq 2e^{-\frac{n^{1/3}}{8c_2^2 \log n}} \quad (5.32)$$

Dévisons $[0, 1]$ en n^2 intervalles de longueur égale à $1/n^2$. Puis, nous avons

$$\sup_{t \in [0, 1]} \left| \sum_{i=1}^n W_i(t) \right| \leq \max_{1 \leq j \leq n^2} \left| \sum_{i=1}^n W_i(j/n^2) \right| \quad (5.33)$$

$$+ \max_{1 \leq j \leq n^2} \sup_{t \in [(j-1)/n^2, j/n^2]} \left| \sum_{i=1}^n (W_i(t) - W_i(j/n^2)) \right| \quad (5.34)$$

$$(5.35)$$

Pour le premier terme du côté droit de (5.33), en utilisant (5.32), nous avons

$$\begin{aligned} & \mathbb{P} \left(\max_{1 \leq j \leq n^2} \left| \sum_{i=1}^n W_i(j/n^2) \right| > n \frac{h_n^\gamma}{\sqrt{\log n}} \right) \\ & \leq \mathbb{P} \left(\sum_{i=1}^{n^2} \left| \sum_{i=1}^n W_i(j/n^2) \right| > n \frac{h_n^\gamma}{\sqrt{\log n}} \right) \leq 2n^2 e^{-\frac{n^{1/3}}{8c_2^2 \log n}} \rightarrow 0. \end{aligned}$$

Par conséquent, nous obtenons que

$$\max_{1 \leq j \leq n^2} \left| \sum_{i=1}^n W_i(j/n^2) \right| = O_{\mathbb{P}} \left(n \frac{h_n^\gamma}{\sqrt{\log n}} \right). \quad (5.36)$$

5.3 Distribution asymptotique de la mesure globale

Puisque K'' est bornée, en utilisant le développement de Taylor, nous obtenons que, pour tout $1 \leq j \leq n^2$ et $t \in [j/n^2, (j+1)/n^2]$,

$$\left| K' \left(\frac{t - \zeta_i}{h_n} \right) - K' \left(\frac{j/n^2 - \zeta_i}{h_n} \right) \right| \leq c_1 \frac{|t - j/n^2|}{h_n} \leq \frac{c_1}{n^2 h_n} \quad (5.37)$$

D'après (5.3), nous avons, pour tout $1 \leq j \leq n^2$ et $t \in [j/n^2, (j+1)/n^2]$,

$$\begin{aligned} & \left| \mathbb{E} \left(K' \left(\frac{t - \zeta_i}{h_n} \right) - K' \left(\frac{j/n^2 - \zeta_i}{h_n} \right) \right) \right| \\ &= h_n \left| \int K'(x) [f(t - h_n x) - f(j/n^2 - h_n x)] dx \right| \\ &\leq l h_n |t - j/n^2| \int |K'(x)| dx \leq \frac{l C h_n}{n^2}. \end{aligned}$$

en utilisant cette borne et (5.37), il s'ensuit que

$$\begin{aligned} |W_i(t) - W_i(j/n^2)| &\leq |X_{i-1}| \mathbb{1}_{\{|X_{i-1}| \leq n^{1/3} h_n^\gamma\}} \left(\frac{c_1}{n^2 h_n} + \frac{l C h_n}{n^2} \right) \\ &\leq n^{-5/3} h_n^\gamma \left(\frac{c_1}{h_n} + l C h_n \right). \end{aligned}$$

par conséquent, le second terme sur le côté droit de (5.33) a une limite supérieure $n^{-2/3} h_n^\gamma \left(\frac{c_1}{h_n} + l C h_n \right)$, qui est en $O(n^{-2/3} h_n^{\gamma-1})$ par $h_n h_n \rightarrow 0$. Combinons ceci avec (5.33) et (5.36), nous obtenons que

$$\sup_{t \in [0,1]} \left| \sum_{i=1}^n W_i(t) \right| = O_{\mathbb{P}} \left(\frac{n h_n^\gamma}{\sqrt{\log n}} + n^{-2/3} h_n^{\gamma-1} \right). \quad (5.38)$$

D'après les définitions de I_{n1} et $W_i(t)$, et en utilisant (5.31), nous obtenons que

$$I_{n1} \leq \frac{|\hat{\theta} - \theta|}{n h_n^2} \left[\sup_{t \in [0,1]} \left| \sum_{i=1}^n W_i(t) \right| + 2c_2 \sum_{i=1}^n |X_{i-1}| \mathbb{1}_{\left\{ \sum_{i=1}^n |X_{i-1}| > n^{1/3} h_n^\gamma \right\}} \right]. \quad (5.39)$$

remarquant que, pour tout i , nous avons

$$\mathbb{E} \left(X_i \mathbb{1}_{\left\{ \sum_{i=1}^n |X_i| > n^{1/3} h_n^\gamma \right\}} \right) \leq \frac{\mathbb{E}(|X_i|^4)}{n^{1/3} h_n^\gamma} = \frac{\mathbb{E}(|X_i|^4)}{n h_n^{3\gamma}},$$

5.3 Distribution asymptotique de la mesure globale

et d'après (5.2) et en utilisant l'inégalité de Holder, on obtient

$$\begin{aligned} \mathbb{E}(|X_i|^4) &\leq \mathbb{E} \left(\left(\sum_{j=0}^{\infty} |\theta|^{\frac{3j}{4}} |\theta^{\frac{j}{4}} \zeta_{i-j}| \right)^4 \right) \leq \left(\sum_{j=0}^{\infty} |\theta|^j \right)^3 \sum_{j=0}^{\infty} |\theta|^j \mathbb{E}(|\zeta_{i-j}|^4) \\ &= \frac{\mathbb{E}(|\zeta_0|^4)}{(1-|\theta|)^4}, \end{aligned}$$

ou nous avons utilisé $\mathbb{E}(|\zeta_0|^4) < \infty$ et $|\theta| < 1$. Par conséquent, il s'ensuit que

$$\mathbb{E} \left(\sum_{i=1}^n |X_{i-1}| \mathbb{1}_{\left\{ \left| \sum_{i=1}^n X_{i-1} \right| > n^{1/3} h_n^\gamma \right\}} \right) \leq \frac{\mathbb{E}(|\zeta_0|^4)}{(1-|\theta|)^4 h_n^{3\gamma}}.$$

par conséquent, en utilisant l'inégalité de Markov, nous avons

$$\sum_{i=1}^n |X_{i-1}| \mathbb{1}_{\left\{ \left| \sum_{i=1}^n X_{i-1} \right| > n^{1/3} h_n^\gamma \right\}} = O_{\mathbb{P}} \left(\frac{1}{n h_n^{3\gamma}} \right).$$

Combinons ceci avec (5.39) et (5.38), et en utilisant (5.4), (5.25) et (5.26), nous complétons la preuve de (5.30). Nous allons maintenant estimer I_{n2} . D'après (5.3), il s'ensuit que $\sum_{i=1}^n |X_{i-1}| = O_{\mathbb{P}}(\sqrt{n})$. Par conséquent, en utilisant (5.4), et le fait que K' est bornée, nous obtenons que

$$I_{n2} = O_{\mathbb{P}} \left(\frac{1}{n h_n^2} \right).$$

Notons

$$\frac{1}{n h_n^2} = \frac{1}{\sqrt{n h_n \log n}} \sqrt{\frac{\log n}{n h_n^3}} = o \left(\frac{1}{\sqrt{n h_n \log n}} \right),$$

D'après (5.25). Donc nous, obtenons que

$$I_{n2} = o_{\mathbb{P}} \left(\frac{1}{\sqrt{n h_n \log n}} \right).$$

Combinons ceci avec (5.29) et (5.30), il s'ensuit que

$$\sup_{t \in [0,1]} J_{n1} = o_{\mathbb{P}} \left(\frac{1}{\sqrt{n h_n \log n}} \right). \quad (5.40)$$

D'après (5.18), nous avons encore $J_{n2} = O_{\mathbb{P}}((n h_n^3)^{-1})$. Par conséquent, d'après (5.25), nous obtenons

$$J_{n2} = O_{\mathbb{P}} \left(\frac{1}{\sqrt{n h_n \log n}} \sqrt{\frac{\log n}{n h_n^3}} \right) = o_{\mathbb{P}} \left(\frac{1}{\sqrt{n h_n \log n}} \right).$$

5.3 Distribution asymptotique de la mesure globale

Combinons ceci avec (5.28) et (5.40), nous obtenons (5.27).

Note. En prenant $\delta \in (0, 1/6)$ (d'où $1/(3\delta) - 1/2 > 3/2$), nous pouvons vérifier $h_n = n^{-\delta}$ et tout $\gamma(3/2 < \gamma < 1/(3\delta) - 1/2)$ satisfait les hypothèses ci-dessus.

Enfin, nous donnons la preuve du résultat principal.

Preuve du théorème 5.2 Nous décomposons la différence $\widehat{f}_n(t) - \mathbb{E}(f_n(t))$ en deux parties :

$$\widehat{f}_n(t) - \mathbb{E}(f_n(t)) = [\widehat{f}_n(t) - f_n(t)] + [f_n(t) - \mathbb{E}(f_n(t))].$$

En vue du lemme 5.2, il suffit de prouver

$$\frac{1}{\sqrt{nh_n \log n}} \sup_{t \in [0,1]} \left| \frac{\widehat{f}_n(t) - f_n(t)}{\sqrt{f(t)}} \right| = o_{\mathbb{P}}(1). \quad (5.41)$$

Puisque f est continue et positive sur $[0, 1]$, il est facile d'obtenir que $1/f$ est bornée sur $[0, 1]$. Par conséquent, en utilisant le lemme 5.4, nous obtenons 5.4.

Conclusion et perspectives

Les processus autorégressifs sont des outils très puissants pour la modélisation des séries chronologiques dans de nombreux domaines, en finance, en biologie, en climatologie, en médecine, en économétrie et en météorologie et dans bien d'autres domaines. Les études sur les processus autorégressifs sont énormes et les perspectives aussi, autant de l'aspect théorie que l'application. En plus que nous avons vu, on cite les points suivants comme des perspectives :

- Prise en compte de la non stationnarité.*
- La mise en pratique des résultats, utilisant l'outils informatique.*
- Obtenir des résultats pour des processus autorégressifs d'ordre supérieur à 1.*
- Pour le dernier chapitre, l'étude de l'estimation de la densité par la méthode de noyau lorsque l'erreur est une suite des variables aléatoires dépendantes, comme par exemple erreur α -mélangeant, où d'autres conditions de dépendance comme le ρ -mélangeant, le φ -mélangeant et le ψ -mélangeant...*
- Réduction d'ordre des processus autorégressifs.*
- Sur d'autres type de données vectorielles et fonctionnelles.*
- Etudier les modèles ARIMA, SARIMA et GARCH.*
- Etude des résultats obtenus pour des variables aléatoires (erreur) β -mélangeant, sous d'autres conditions de dépendance comme le ρ -mélangeant, le φ -mélangeant et le ψ -mélangeant...*
- Est-il possible d'avoir la convergence presque complète dans le cas où l'erreur est ND.*

Bibliographie

- [1] Mann H.B and Wald A (1943). *On the statistical treatment of linear stochastic difference equations. Econometria* 11, 173-220.
- [2] Anderson T.W (1959). *On asymptotic distributions of estimates of parameters of stochastic difference equations, Ann. Math. Statist.* 30, 676-687.
- [3] Lai T.L and Wei C.Z (1982). *Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. Annals of statistics* 10, 154-166.
- [4] White J.S (1958). *The limiting distribution of the serial correlation coefficient in the explosive case. Annals of Mathematical Statistics* 29, 1188-1197.
- [5] Phillips P.C.B (1987). *Towards a unified asymptotic theory for autoregression. Biometrika* 74, 535-547.
- [6] Chan N.H and Wei C.Z (1988). *Limiting distribution of squares estimates of unstable autoregressive process, Annals of Statistics.* 16, 367-401.
- [7] Guikhman I.I and Skorokhod A.B (1971). *Theory of random process. Naouka, Tome1.*
- [8] . Hus P.L and Robbins H (1947). *Complete convergence and law of large numbers, Proc. Nat. Acad. Sci. U.S.A.* 33, 25-31.
- [9] Chow Y.S and Teicher H (1978). *Probability theory, Independence, interchangeability, martingales, Springer-Verlag, New York.*
- [10] Hong Jyy.I (2003). *On the strong law of large numbers for sums of random elements in Banach space.*
- [11] Devroye Luc. *A course in Density Estimation (Progress in Probability) 1st Edition. Birkhäuser Verlag, Boston, 1987.*
- [12] Yoshihara K (1978). *Moment inequalities for mixing sequences. Kodai Math. J,* 316-328.
- [13] Oodaira H and Yoshihara K (1971). *The law of the iterated logarithm for stationary proucess satisfying mixing conditions. Kodai Math. Sem. Rep. 23,* 311-334.

-
- [14] Guodong X, Schanchao Y and Aiwu C (2009). A maximal moment inequality for α -mixing sequences and its applications. *Statistics and probability letters (Elsivier)* 79, 1429-1437.
- [15] Jianqing F and Qiwei Y (2003). *Nonlinear Time Series, Nonparametric and Parametric Methods, Springer Series in Statistics, Springer-Verlag New York.*
- [16] Fuxia Cheng (2005). Asymptotic Distributions of Error Density Estimators in First-order Autoregressive Models, *The Indian Journal of Statistics, Volume 67, Part 3, 553-567.*
- [17] Peter J. Brockwell and Richard A. Davis. (1991). *Time Series : Theory and Methods. Springer Series in Statistics, Spring-Verlag, New York.*
- [18] Peter J. Brockwell and Richard A. Davis (2002). *Introduction to Time Series and Forecasting, Second Edition, Springer-Verlag New York, Inc.*
- [19] Ryoza Yokoyama, (1980). Moment Bounds for Stationary Mixing Sequences. *Z, Wahrscheinlichkeitstheorie verw.Gebiete by Springer-Verlag.* 52, 45-57.
- [20] R.C. Bradley., (1986). Basic properties of strong mixing conditions. In : *Dependence in Probability and Statistics, (E. Eberlein and M.S. Taqqu, eds.), pp. 165- 192. Birkh user, Boston, 1986.*
- [21] R.C. Bradley., (2005). *Introduction to Strong Mixing Conditions, Volume 2. Department of Mathematics, Indiana University, Bloomington. Custom Publishing of I.U., Bloomington, 2005.*
- [22] Azuma, K. (1967). weighted sums of certain dependent random variables. *Tohoku Math. J., 19, 357-367.*
- [23] Bickel, P.J. and Rosenblatt, M. (1973). On some global measures of the deviations of density function estimators. *Ann. Statist., 6, 1071-1095.*
- [24] Billingsley.P (1999). *Convergence of probability Measures 2nd edition, Wiley series in probability and statistics. Probability and statistics, Wiley, New York.*
- [25] Boldin, M.V. (1982). Estimation of the distribution of noise in an autoregression scheme. *Theory Probab. Appl., 27, 866-871.*
- [26] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* 58, 13-30.
- [27] Horv th, L. and Zitikis, R. (2003). Asymptotics of L_p -norms of density estimators in first-order autoregressive models. *Statist. Probab. Lett., 65, 331-342.*
- [28] Kallenberg, O. (1997). *Foundations of Modern Probability. Springer-Verlag, New York.*

-
- [29] Koul, H.L. (1970). Some convergence theorems for ranks and weighted empirical cumulatives. *Ann. Math. Statist.*, 41, 1768-1773.
- [30] Koul, H.L. (1977). Behavior of robust estimators in the regression model with dependent errors. *Ann. Statist.*, 5, 681-699.
- [31] Koul, H.L. (1991). A weak convergence result useful in robust autoregression. *J. Statist. Plann. Inference*, 29, 291-308.
- [32] Koul, H.L. (1992). *Weighted Empirical and Linear Models*. (1992). Lecture Notes-Monograph Series 21, Institute of Mathematical Statistics, Hayward, California.
- [33] Koul, H.L. (1996). Asymptotics of some estimators and sequential empiricals in nonlinear time series. *Ann. Statist.*, 24, 380-404.
- [34] Koul, H.L. and Osiander, M. (1994). Weak convergence of randomly weighted residuals empiricals with application to autoregression. *Ann. Statist.*, 22, 540-562.
- [35] Lee, S. and Na, S. (2002). On the Bickel-Rosenblatt test for first-order autoregressive models. *Statist. Probab. Lett.*, 56, 23-35.
- [36] Loynes, R.M. (1980). The empirical d.f. of residuals from generalized regression. *Ann. Statist.*, 8, 285-298.
- [37] Mammen, E. (1996). Empirical process of residuals for high-dimensional linear models. *Ann. Statist.*, 24, 307-335.
- [38] Portnoy, S. (1986). Asymptotic behavior of the empirical distribution of M-estimated residuals from a regression model with many parameters. *Ann. Statist.*, 14, 1152-1170.
- [39] Shao Q.M., Weak convergence multidimensional empirical processes for strong mixing sequences, *Chinese Ann. Math. Ser. A* 7 (1986), 547-552.
- [40] Shao Q.M., Maximal inequality for partial sums of ρ -mixing sequences, *Ann. Probab.* 23 (1995), 948-965.
- [41] Shao Q.M and Yu H., Weak convergence for weighted empirical processes of dependent sequences, *The Annals of Probability Vol. 24, Number 4* (1996), 2098-2127.
- [42] Doukhan P., *Mixing : properties and examples*, Lecture Notes in Statistics 85, Springer Verlag (1994).
- [43] Doukhan P., P. Massart & E. Rio. Invariance principle for the empirical measure of a weakly dependent process. *Ann. Inst.H.Poincaré Probab. Statist.* 31-2, 393-427 (1995).
- [44] Chow.Y.Sand Teicher.H., *Probability Theory : Independence, Interchangeability, Martingales*, Springer Texts in Statistics (1997).

-
- [45] E. Liebscher Strong convergence of sums of strong mixing random variables with applications to density estimation. *Stochastic Processes and their Applications*. 65 (1996) 69-80.
- [46] M. Robinson, Nonparametric estimators for time séries, *J. Time Sériés Anal.* 4 No. 3 (1983) 185-207.
- [47] G.G. Roussas, Nonparametric estimation in mixing sequences of random variables, *J. Statist. Plann. Inference* 18 (1988) 135-149.
- [48] L.T. Tran, Kernal density estimation under dependence, *Statist. Probab. Lett.* 10 (1990) 193-201.
- [49] P.A. Nze and R. Rios, Density estimation in the l^∞ norm for mixing processes (in French), *C.R. Acad. Sci. Paris*, 320, Série I (1995) 1259-1262.
- [50] T.Y.Kim, Kernel density estimator for strong mixing processes, *Journal of Statistical Planning and Inference*, 333 (2005) 273-284.
- [51] G. Box and G. Jenkins, *Times series analysis : forecasting and control*, Holdenday (1976).
- [52] D.T. Pham and L.T. Tran. Some mixing properties of time series models, *Stochastic Process. Appl.*, 19 (1985), pp. 297-303.
- [53] D.W.K Andrews, Non-strong mixing processes, *J Appl. Prob*, 21 (1984), 930-934.
- [54] D. Bosq, *Nonparametric Statistics for Stochastic Processes*, Springer, (1998).
- [55] D. Bosq, (1996). *Nonparametric Statistics for Stochastic Processes. Estimation and Prediction. Lecture Notes in Statist.* 110. Springer, New York.
- [56] Yu.A Davydov, Mixing conditions for Markov chains. *Theor. Probab. Appl.*, (1973), 18, 2, 321-338.
- [57] E. Nummelin and P. Tuominen, Geometric ergodicity of Harris recurrent Markov chains with applications to renewal theory. *Stochastic Process (1982). Appl.* 12, 187-202.
- [58] J. D. Esary, F. Proschan & D. W Walkup. Association of Random Variables, with Applications. *Ann. Math. Statist.* 38 (1967), no. 5, 1466-1474.
- [59] T. E. Harris. A lower bound for the critical probability in a certaion percolation process. *Proc. Camb. Phil.Soc.* 59 (1960), 13-20.
- [60] C. M. Fortuin, P. W. Kasteleyn and J. Ginibre. Correlation inequalities on some partially ordered sets. *Comm. Math. Phys.* 22 (1971), no. 2, 89-103.
- [61] Yu.A Davydov, Invariance principle for stationary processes. *Theor. Probab. Appl.*, (1970), 15, 3, 498-509.

-
- [62] R. Durrett,. *Probability theory and examples. Cambridge series in statistical and probabilistic mathematics.*, (1991).
- [63] T. C. Mills,. *Time Series Techniques for Economists. Cambridge University Press*, (1990).
- [64] M.H. Hayes. *Statistical Digital Signal Processing and Modeling. New York : John Wiley & Sons*, (1996).
- [65] S.M. Kay, *Modern Spectral Estimation. Englewood Cliffs, NJ : Prentice Hall*, (1988).
- [66] P. Stoica and R. Moses. *Introduction to Spectral Analysis. Upper Saddle River, NJ : Prentice Hall*, (1997).
- [67] S.J. Orfanidis, *Introduction to Signal Processing. Upper Saddle River, NJ : Prentice Hall*, (1996).
- [68] D. B. Percival and A. T. Walden, *Spectral Analysis for Physical Applications. Cambridge University Press*.(1993).
- [69] S. M. Pandit and S. M. Wu, *Time Series and System Analysis with Applications. John Wiley & Sons*.(1983).
- [70] A. Belguerna and S. Benaissa. *The asymptotic convergence of the parameter estimator in the first order autoregressive model AR(1) under strong mixing conditions. International Journal of Mathematics and Statistics, Year 2014, Volume 15, Issue Number 3, 1-8.*
- [71] Bertrand : *Estimation du paramètre d'un processus autorégressif d'ordre 1. Loi limite de l'estimateur. 1988, Document de travail du Laboratoire de Proba. et Stat. de l'Université de Rouen.*
- [72] Billingsley, P. (1968) : *Convergence of Probability Measures. Wiley, New York.*
- [73] Brockwell P.J and Davis R.A. (1991) : *Time series : theory and methods. Springer Series in Statistics. Springer Verlag, New York, Second Edition.*
- [74] Brockwell P.J and Davis R.A. (1996) : *Introduction to Time series and Forecasting. New York : Springer.*
- [75] Chan N.H and Wei C.Z. (1987) : *Asymptotic inference for nearly nonstationary AR(1) processes. Annals of Statistics 15, 1050-1063.*
- [76] Cramer, H. (1946) : *Mathematical Methods of Statistics. Princeton University Press, Princeton.*
- [77] Davidson, R., AND J. G. Mackinnon (1993) : *Estimation and Inference in Econometrics. Oxford University Press, Oxford.*

-
- [78] Davis R and Resnick S. (1986) : *Limit theory for sample covariance and correlation functions of moving averages.* *Ann. Stat.*, 14, 533–558.
- [79] Gouriéroux, C., AND A. Monfort (1990) : *Séries Temporelles et Modèles Dynamiques.* *Economica, Paris.*
- [80] Karatzas, I. (1988) : "A Tutorial Introduction to Stochastic Analysis and its Applications," *Lectures notes, Office for Research and Development of the Hellenic Navy and Columbia University.*
- [81] Phillips, P. C. (1986) : "Understanding Spurious Regression in Econometrics," *Journal of Econometrics*, 33, 311-340.
- [82] Phillips, P. C. (1987) : "Time Series Regression with a Unit Root," *Econometrica*, 55, 277-301.
- [83] Silverman, B. W. (1986) *Density Estimation.* London : Chapman and Hall.
- [84] Slutsky E. (1937), *The Summation of Random Causes as the Source of Cyclical Processes,* *Econometrica*, 5, 105-146.
- [85] White, H. (1984) : *Asymptotic Theory for Econometricians.* Academic Press, Orlando.
- [86] Yule G.U. (1927), *On a Method of Investigating Periodicities in Disturbed Series with Special Reference to Wolfer's Sunspot Numbers*